# The organisation of social care in England: Markets, hierarchies and contract choices in residential care for older people.

Julien Forder

London School of Economics

Thesis submitted in consideration of PhD

April 2005

## Preface

I would like to thank my wife Joanne and my family for their support and their forbearance. This PhD was undertaken while I was working on the *Commissioning and Performance* project funded by the Department of Health. Some of the following was inspired by that project, but the work here is new and is the sole responsibility of the author. My gratitude is extended to my colleagues who worked with me on these projects. I would particularly like to thank my supervisor, Professor Martin Knapp, for giving me the time to do this, and for his valuable support along the way.

## Abstract

This thesis is concerned with strategic (economic) organisation, as applied to the long-term care system in England. This work adopts a *transaction cost* perspective. The main hypotheses are: first, that the transaction costs generated by (public sector) hierarchies in social care are *lower* than those generated in quasi-markets. Second, that production costs in hierarchies are *greater* than in markets. Third, that *contingent* contract use is associated with comparatively higher prices and mark-up rates, and greater net transaction costs. The motivation for this work is first to address perceived limitations of the theory in a comparative *public sector* application. Second, to inform the empirical and policy debate on social care reform.

Following an account of the historical policy and institutional context, a multi-period, comparative theoretical model was developed, building on the contract theory literature. It underpins a systematic empirical analysis of care home services – at local authority and care home level – for older people in 1998 and 1999. Various estimation techniques addressed the skewed nature of the data and the panel design.

The estimation results supported the theoretical hypotheses. Point estimates of marginal and average transaction costs were £6 and £21 per place per week respectively for hierarchies and £41 and £56 for placements under the market governance archetype, statistically significant differences. For production costs, a significant difference was found in the other direction: £89 for hierarchy and £55 for markets at the margin. Overall, the total (production + transaction) costs were not significantly different.

Contingent contract use was associated with higher prices relative to average variable costs of 8% of average price compared with non-contingent contracts. The analysis pointed to low profitability rates and that providers are not solely motivated by profit (only taking 55% of potential profit).

Policy implications were explored for both the markets-hierarchies and contracts analyses.

## Table of contents

## Figures

# Chapter 1. Introduction

## 1.1 Context

A significant proportion of all spending on publicly purchased social care goes not to providing services directly but to supporting the organisational processes designed to commission and deliver such care. These organisational processes and arrangements have a significant bearing on whether the right services get produced and delivered to the right people at the right time, and for the lowest cost. In other words, the way that the social care system is organised on both a micro and macro scale is important. There are a great number of possible organisational forms and arrangements that could be used. Choosing between them is a difficult task, with implications that can be measured according to a range of criteria. This work attempts to unravel these complexities and quantify the implications.

We have witnessed significant policy reform in the social care, health and public policy fields, both in England and elsewhere, that is centrally relevant to organisational questions (Wistow et al., 1996; Saltman, Figueras, and Sakellarides, 1998; Forder, 2002). Economic ideas have played a key role these reforms, although they are certainly not the only ideas that have influenced relevant policy development. Moreover, in the world of practical policy-making, even where economic ideas do serve to underpin policy, those ideas have been drawn from fairly superficial and selective sets of economic theory, an argument developed in the next chapter. Basic textbook models of demand, supply and the benefits of competition have frequently been the sole theoretical foundation. In contrast, some of the recent academic literature has been characterised by developments that take a considerably more in-depth and sophisticated view of the performance of alternative ways of organising and arranging economic activity. This literature is generally referred to as the economics of organisation (EO). It is a literature that provides a basis for *comparative* analysis of such organisational structures.

The main aim of this work is to assess the impact of different organisational arrangements used in the social care system in England, applying the tools and concepts of the economics of organisation. Relevant organisational considerations include the extent to which germane economic activity is undertaken in markets between independently owned and controlled stakeholders (social services departments, private providers, public sector providers, etc.) or whether it is conducted in bureaucracies, characterised by top-down management. Another key set of organisational characteristics relate to the nature of financial and other incentives that are in operation. For example, on what basis do those people that provide services get paid? Are they salaried or do they get paid a price for their services? Are the rates set in advance or retrospectively? Are incentives directly aligned with the costs of the services required or not? A third set of organisational arrangements concern how providers (and purchasers) are monitored to ensure standards. Is monitoring undertaken by government agencies, by purchasing authorities, by individuals, via self-regulation by professional providers, etc?

Whilst a very broad range of organisational dimensions potentially apply, many do align and can be gathered together to form a much smaller number of *governance archetypes*, a feature that this study exploits. In particular, this work assesses the choice between two sets of governance archetypes that are common in the literature (although often rather vaguely defined in that literature). The first is the choice between *markets* and *hierarchy* governance archetypes. In this case, the finer distinction is between the typical 'quasi-market' arrangements used in social care – involving local government funding and procurement of care with purchasing from privately owned, potentially competing providers – and traditional public bureaucratic or municipal arrangements. The second comparison is between types of contract used to conduct social care transactions, with a particular focus on the degree of *contract contingency*; that is, the degree to which contracts directly reflect characteristics and factors relevant to the transaction for the service.

Social care is a broad field. To limit the scope of this endeavour, the focus will be on older people's services, and in particular publicly funded residential care. The provision of care home services for older people takes the largest slice of all (public) social care expenditure – see chapter 4 for details. In particular, this sub-sector accounted for £4221m of gross spend in 2002/3 or 36% of gross local authority adult spend.

## 1.2 Structure of Phd

Having introduced the main aims and context this work, this chapter continues by laying some theoretical foundations that will apply throughout this work. It also describes the main sources of data. We turn in the next chapter to a review of the relevant theoretical and empirical literature, principally concerning economic organisational theory and the relatively modest breadth of this literature applied to social care. Having described the literature in some detail, chapter 2 provides an assessment of the fitness for purpose of the theory and empirical methodology in the literature. The chapter concludes that whilst there has been significant relevant work, a number of specific developments will be required to address our main research questions.

The conceptual building blocks of this approach, and of the associated empirical methodology, are laid out in chapter 3. The chapter describes the fundamental task of organising economic activity and then goes on to define transactions as the unit of analysis. It describes the agreements or contracts struck between individuals as they transact, and a broad framework for handling information and influence in transactions (the principal-agent framework). A key contribution of the chapter is to define and detail the concept of a 'governance structure' – that is the set of rules, protocols, agreements and regulations – that frame transactions and contracting. Governance structures are composed of a range of organisational dimensions, including those referred to above. The potential number of possible governance structures is high, but a much more limited number of governance *archetypes* can be used that have sufficient relevance in practice. Chapter 3 closes with an outline of the main theoretical propositions addressed in this work.

Chapter 4 discusses the historical, institutional and economic character of the social care system in England. This discussion is framed in the language and concepts of chapter 3. The aim is to translate those concepts into descriptions that are meaningful in social care. The analysis of its historical, institutional and economic character will point to any special features of social care in England that stand out relative to the general comparative governance concepts in the literature. A bespoke survey of commissioning arrangements (see below) provides rich descriptive information in chapter 4 about how social services departments are currently configured along the main governance dimensions identified earlier (i.e. about ownership, control, brokerage/care management arrangements, budget devolution, financial flows, levels of contract contingency and specification etc.). Along with (secondary) government data about activity and expenditure, the chapter paints a broad descriptive picture of social care organisation in England.

Chapter 5 develops theory for the analysis of market and hierarchy choices. Building on the literature outlined in chapter 2 – and in particular the contract theory of Grossman and Hart (1986), Hart (1995; 2003), Kreps (1990a; 1990c) and Milgrom and Roberts (1990; 1992) – a model is developed. The aim is to simultaneously model the impact of multiple contract types with explicit transaction costs and information asymmetry within a comparative market and hierarchies governance framework. Moreover, with reference to the conclusions of chapter 2, the model seeks to accommodate comparisons between public sector markets and hierarchies and also to incorporate a more appropriate treatment of social context and stakeholder motivation. The model is used to develop a set of hypotheses for empirical testing. The key result – based in particular on predicted behaviours with regard to investment, production/implementation, contracting and also opportunistic profiteering – is that transactions costs will be lower and production costs higher in public sector hierarchies compared to (quasi-) markets.

Chapter 6 tests these and related hypotheses with regard to the residential care for older people using data from all local authorities in England, including costs, service utilisation and input prices, needs-related factors and process indicators. Three sets of estimations are undertaken. The primary analysis is of the covariates of transaction costs and is undertaken to address hypothesis that hierarchies have lower transaction costs. Thereafter a model of total costs is fitted and is used to assess the overall impact of governance choices. Finally, an analysis of production costs is undertaken mainly to assess the precision of the governance variable in its impact on production costs. The main finding was that hierarchical, in-house local authority provision incurred transaction costs of about a third of those from equivalent market provision. Furthermore, adjusted production costs were found to be lower in markets than in hierarchies, although the difference was greater than that for transaction costs. Taken together, the *total* costs in markets were found to be slightly, but not significantly lower than in hierarchies.

Chapter 7 uses the same model foundations outlined below and adapts the model used in chapter 5 to focus on specific hypotheses concerning contract choice. The model in chapter 7 has parallels with that developed by Forder (1997a), but is greatly extended to address information, risk and cost adjustment properties of different contracts, as they affect both profit making organisations and also 'non-profit' providers. The key hypothesis is that contingent contracts will result in higher prices than non-contingent contracts. This effect is also found for non-profits although the price difference is reduced. In addition to these transactions benefits implications, transaction costs were also hypothesised to be higher for contingent contracts.

Chapter 8 describes the empirical specification used to test the above hypothesis and the results of the analysis. The empirical work utilises data from a survey of 600 care homes and 12000 residents. A residual demand function approach (see Baker and Bresnahan, 1985) is used to derive potential and actual price-cost margins for providers operating with different contracts. The results indicate that mark-up rates (surpluses and prices) are higher for providers with contingent contracts than those with non-contingent contracts, controlling for a range of confounding factors. The results also provide some indication of the motivation of providers, showing that providers are not solely motivated by profit. The results also produced some estimates of the economies of scale that exist in the care home market. The estimations suggest that transaction costs are higher for contingent contracts.

Chapter 9 draws out the main conclusions of this research. It rehearses the contributions to the literature that have been made. It also flags some policy relevance of the key results.

## 1.3 Theoretical background

### 1.3.1 Model foundations

The organisation of the social care system in England is complex and complicated. To make some progress, a somewhat stylised theoretical model is used. Whilst this approach inevitably sacrifices some realism, it does make the problem at hand more manageable. It allows us to make predictions that in turn can be tested. Chapter two reviews some of the literature that covers the relevant methodological arguments (see also Hodgson, 1988, for a good overall account). Clearly, there is a balance to be struck between realism of assumptions and tractability. The general modelling framework outlined below is the basis for the theoretical chapters 5 and 7 (and these chapters provide more detail/refinements as required for their purposes).

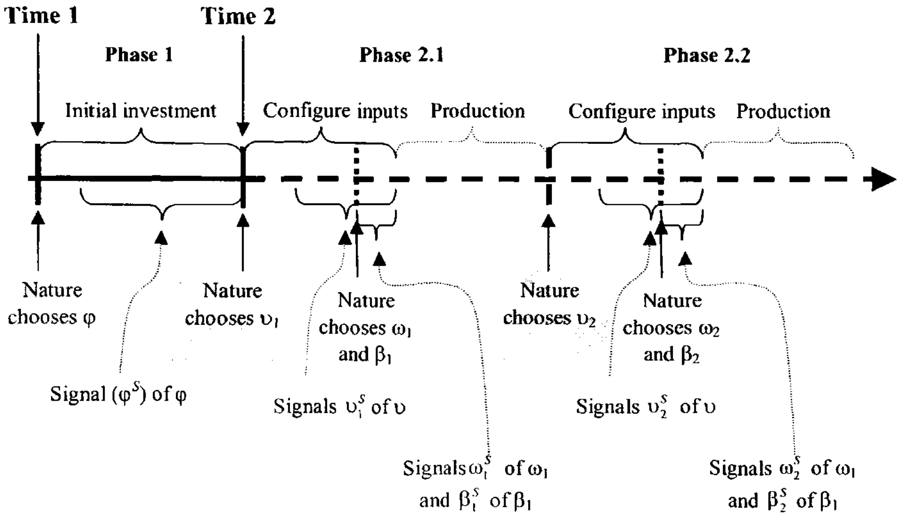
There are several broad processes in providing social care services. A first is investment in 'production' technology and more generally the process of bringing together the various capital, and skilled and unskilled labour inputs. Second, contracting and negotiation, which can be sequential to investment but can also run concurrently (i.e. we do not rule out the possibility of negotiations beginning even before the investment starts). Essentially, contracting is about agreeing the terms for the production and ultimate delivery of services. It is a flexible concept and applies just as relevantly to a work plan agreed between a manager/employer and employee as it does to a legal agreement between buyer and seller in a market. Third is the production of social care services, where 'production' takes a broad definition and involves the readying, deployment and use, etc., of services by people with care needs. At some time during this process reimbursement will take place as agreed.

Because there are a very wide range of relevant factors that potentially influence how these processes are conducted and the outcomes they produce – such as the needs characteristics of service users, prevailing labour and capital market conditions, government policy, regulation – it is important to reflect this complexity and uncertainty in the model. The usual method is to assume that the above activities take place within one of a great many possible 'states-of-the-world', each of which has a particular configuration of relevant circumstances. Furthermore, the exact nature of the state in which stakeholders find themselves is likely to become apparent only as time goes on, and relevant activities are undertaken. In other words, as the future unfolds it reveals relevant information.

The various elements of the modelling can be drawn together in a structured way. This structure details the timing and occurrence of relevant events, and is summarised in Figure 1-1. We divide the whole care process into two main time phases: investment starting at time 1 and provision/implementation or 'effort' starting at time 2, with contracting occurring concurrently. Furthermore, the second phase – the effort phase –

can be treated in aggregate or, as in the figure, can have two or more sub-periods explicitly reflecting repeated production following investment.

**Figure 1-1. Stylised timeline**




Uncertainty is framed by four parameters φ, ω, υ and β. The uncertainty regarding these parameters can be resolved at various times during the production process. They are determined 'by nature' for each $k = 1,\ldots, s$ states of the world.

At time 1, nature chooses φ, which represents contextual, external factors that affect investment e.g. local property market conditions. Investment then occurs to set up production with the relationship between investment and final product mediated by the parameter φ. Investment is at level $y^a$, and might be, for example, the construction of a care home.

At time 2, υ can be measured. This parameter reflects the needs characteristics of the user that can be measured (subject to error) by a pre-care assessment. Using the initial investment, care packages of inputs are put together and configured on the basis of υ. This configuration defines the type of output, $q(y^a(\varphi), \upsilon, \beta)$. The β parameter reflects the costs of caring for a person, given the prevailing technology and their risk characteristics, and it drives costs and outcomes once a person is in a care setting. It depends on the user's needs profile and also how their particular needs profile fits with the nature of the service provided in the home. It is therefore a refinement of initial assessed characteristics $\upsilon$. In other words, stakeholders that know $\beta$ also have information equivalent to that in $\upsilon$. To give an example, a person that presents to social services undergoes as assessment of need. A range of criteria applies and the person's needs and circumstances are evaluated against those criteria. Since this assessment is prior to care it can only relate to the person's pre-care circumstances and needs. The parameter $\upsilon$ concerns information about the relevant population at this stage. Once people who meet the criteria, based on $\upsilon$, are placed it is possible that their care needs will change. In addition, that person will have been in contact with the system for longer. Evaluation from this point will likely yield better, care-specific, information.

Before production (after time 2) nature also chooses a parameter, $\omega$, that captures factors that will affect productivity, but are beyond the control of the provider. It reflects, in particular, the type of factor inputs a provider uses and how these are combined, which has a bearing on the kinds of product types the provider is able to produce. When combined with $\upsilon$, $\omega$ also gives a good indication of $\beta$. During the ensuing production phase providers expend 'effort' $y^e$ to supply the care service. The variable $D$ measures the extent of production output by each provider (of their chosen type of service $q$).

Figure 1-1 indicates at what point nature determines key parameters. Stakeholders are able to measure these parameters thereafter but only at a (potentially substantial) cost. Moreover, this cost will differ between purchasers and providers for each parameter. For example, the cost for purchasers to measure $\beta$ is much higher than for providers. Stakeholders that do not measure a parameter are assumed, nonetheless, to get a noisy and tardy signal of the parameter, which is denoted by an $S$ superscript in the figure. This signal is insufficiently precise and anyway too late for contracting purposes. Stakeholders can, however, use it to gauge retrospectively the validity of reports made by others. When stakeholders propose a value in a report, a 'hat' embellishment is used: e.g. $\hat{\beta}, \hat{\upsilon}, \hat{\varphi}$. The recipients of the report assess its accuracy and accept the report with a probability inversely related to how far the reported value deviates from the signalled value. Assume that the stakeholder making the report expects it to be accepted with probability $r^j(\theta^j)$, $j$ = φ, υ, β, ω where θ is the difference between the reported and signalled value e.g. $\theta^\beta = \hat{\beta} - \beta^S = \hat{\beta} - \beta$. We assume that the signal is a noisy but unbiased indication of the actual value of the parameter. If the recipient of a report actually measures the parameter in question then any deviation (beyond some small reasonable degree of discrepancy) is sufficient for the report to be rejected. Overall then, $r = 1$ when $\theta = 0$. Also, $r_\theta < 0$ and $r_{\theta\theta} < 0$. The full specification of the costs of measuring these parameters, contracting costs and other are described in chapter 5.

The number of purchasers and providers operating in the system will depend on prevailing organisation arrangements; the number of products $i = 1$ to $N$ can vary with $N \geq 1$. At period 2 demand for product $i$ in state $k$ is, with full information:

$$(1.1) \quad D_{ki} = D_{ki}(p_{k1},...,p_{kN},\beta_k,\varphi_k,\upsilon_k)$$

where $p_{k1}$ to $p_{kN}$ are the prices charged by the provider of the $i$th product in the market. In practice, the full information case is unlikely to occur. Purchasers, in particular, will usually find it prohibitive to measure β (see chapter 5). They may also not measure φ. In this case demand is:

$$(1.2) \quad D_{ki} = D_{ki}(p_{k1},...,p_{kN},\hat{\beta}_k,\hat{\varphi}_k,\upsilon_k)$$

Otherwise the demand function has the normal properties: $\partial D_1/\partial p_1 < 0$ and $\partial D_1/\partial p_i > 0$ for $i = 2 ... N$.

The uncertainty that exists about the value of key parameters means that stakeholders with the better information – those making reports – can potentially misrepresent that information to less well informed parties. In accordance with standard modelling of informational problems (e.g. Rasmusen, 1992), any misrepresentation made by the agent will be detected with probability $1 - r$, as defined above. What happens if misrepresentation is detected is important. Generally, it is assumed that the potential contract fails to proceed and the agent reverts to reservation utility, generally set at zero (and exogenous in any case). It follows that expected period 2 demand in state $k$ is:

(1.3) $x_{ki} = r^{\beta}(\hat{\beta}_k) r^{\varphi}(\hat{\varphi}_k) D_{ki}(\hat{\beta}_k, \hat{\varphi}_k)$

This assumes that demand is zero with probability $\left(1 - r^{\beta}(\hat{\beta}_k) r^{\varphi}(\hat{\varphi}_k)\right)$.

So what are the respective payoff functions? Providers are assumed to maximise a utility function as follows:

(1.4)
$$U_i = \gamma_1 [v(\pi_{1i} - y - T) + \breve{\psi}(\rho_{1i})] + \gamma_2 [v(\pi_{2i} - y - T) + \breve{\psi}(\rho_{2i})] + \\ \ldots + \gamma_S [v(\pi_{Ni} - y - T) + \breve{\psi}(\rho_{Si})]$$

where $\pi_{ki}$ is provider $i$'s expected (second period) profit in state-of-the-world $k$:

(1.5) $\pi_{ki} = p_{ki}(x_{ki}) x_{ki} - C_{ki}(x_{ki})$

where $C$ are costs. Furthermore, in (1.4) the term $y$ reflects the total cost of investment and production effort. Transaction costs of various forms – which are detailed in chapter 5 – are captured by $T$.

Providers derive utility from profits according to the function $v(.)$. Marginal utility of income is assumed to be positive, $v' > 0$. We also assume that $v'' < 0$, $v''' = 0$ and $v' \geq |v''|$. The term $\psi$ is introduced to allow some degree of non-profit motivation. Providers are of two types, profit-*maximisers* and *satisficers*. The latter type of provider aims to strike a balance between profit making and being able to lower prices and so enable greater access to the service for users. For our purposes we assume that satisficers suffer disutility in proportion to (the present value of) prices where those are higher than a reasonable level – essentially if price is higher than the level that sustains normal profit, providers appreciate that access is more restricted than it could (viably) be and suffer disutility as a result. Should the purchaser wish to proceed with the contract then providers suffer disutility $\psi(\rho)$, where $\rho$ is the expected value of prices $p$ in to the future. For profit maximisers, $\psi = 0$ and $\psi' = 0$, and for satisficers, $\psi < 0$, $\psi' < 0$ and $\psi'' = 0$ (where $\rho$ is greater than the 'normal' level $\breve{\rho}$). Since providers have some expectation that the contract may not proceed, the disutility term enters (1.4) at its expected value: $\breve{\psi}(\rho)$. The most straightforward functional relationship in this case is $\breve{\psi}(\rho) = r\psi(\rho)$, i.e. $\breve{\psi}(\rho) = \psi(\rho)$ when $r = 1$, although see the theoretical chapters (5 and 7) for further details. We do, however, assume that any disutility is relatively modest relative to the marginal value attached to profit. Specifically:

*Assumption 1-1.* That: $v' + \breve{\psi}' > 0, \forall\pi$ and $v'(0) + \breve{\psi}' < v'(\infty)$.

Purchasers are assumed to maximise a utility function of the form:

$$(1.6) \quad Z = \gamma_1 z(V(x_1) - p(x_1)x_1) + .. + \gamma_S z(V(x_S) - p(x_S)x_S)$$

where $V$ is the purchasers total valuation of the output $D$ and is sum of marginal valuation:

$$(1.7) \quad V = \int_0^D R(y^a, x; \upsilon, \varphi, \beta)dx\,.$$

## 1.4 Empirical analyses

This work draws on three main sources of data: two specifically designed surveys – the first on providers and provision, the second on commissioning – and the routine data collections by the government about social care expenditure and activity.

### 1.4.1 Survey of residential care provision

A main source of data is a uniquely large and detailed national, cross-sectional survey of residential care and nursing homes for older people conducted in the autumn of 1996. It provides comprehensive information for 12,000 residents and over 600 care homes in 21 local authority areas in England. The principal aims of the survey were to provide a baseline description of homes and their users, and to explore the relationship between the costs or price of care and the dependency characteristics of residents. A comprehensive report of the survey provides details of methodology, sample frame and results (Netten et al., 1998).

The survey covered residential homes for older people managed by local authority social services departments and independent sector residential, nursing and dual registered homes for older people. The survey included residential and dual registered homes for older people with mental illness, but nursing homes which catered for elderly people with mental illness were not included if they were recorded simply as for people with mental illness in the database used for selecting the sample. Small homes, that is, those with fewer than four places, were not included in the survey.

Within the 21 local authorities in the survey, separate lists of all homes in the area were compiled for local authority homes, private residential and dual registered homes, voluntary residential and dual registered homes and registered nursing homes. Samples were selected randomly from the list with probability proportional to home size with size being defined as the number of places recorded on the sampling lists. Since the number of homes in London boroughs tends to be small, the number of private residential and dual registered homes and the number of registered nursing homes selected in London were each doubled. Within the selected homes, individual information was requested for a random sample of up to 20 residents per home (or all residents for smaller homes).

The fieldwork procedure involved an initial approach by letter to the selected homes, which was followed by a personal interview with the home manager. Resident information questionnaires were left to be collected latter, as were self-completion questionnaires for staff (for up to 20 relevant members of staff). The information collected in the personal interview included background information about the home, information about the type of care provided, the physical features of the home, staffing, contractual arrangements and charging arrangements. The residents questionnaire covered the characteristics of residents including origin of admission, funding sources, and dependency characteristics. The questionnaire for staff incorporated the Sheltered Care Environment Scale (SCES), developed by Moos and Lemke (1992). This scale is designed to capture the 'atmosphere' in the home, in terms of key characteristics: the levels of cohesion and conflict; the degree to which residents are encouraged to be independent or to disclose their feelings; the organisation of the daily routine; the influence residents have on the rules of the home; and the physical comfort of the home.

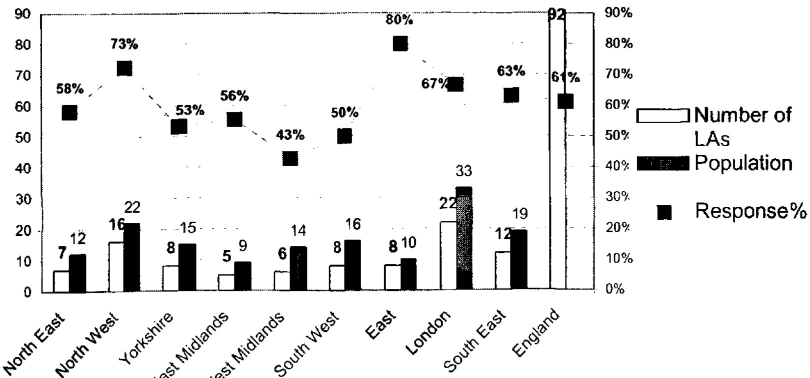
### 1.4.2 Survey of commissioning

A 2001 survey with a sample frame of all 150 local authorities in England was also a major source of data regarding commissioning arrangements (see MEOC Team, 2001). This survey utilised a postal questionnaire design, with letter and telephone follow-up. The questionnaire was structured and predominately involved closed-end categorical responses, some with detailing sub-questions. The questionnaire design drew on experience with two similar, but smaller sample surveys conducted in the mid and late 1990s.

The survey covered the following areas of commissioning: the extent and nature of purchaser – provider separation in local authorities; in-house and external purchasing process differentiation; budget holding; price setting (both for external and internal providers); typical price levels; relative purchasing power; perceived competitiveness; accreditation and review arrangements; contract types, and contract specification.

Figure 1-2 shows the overall response rate after this multiple follow-up process. Some 92 out of the 150 SSDs in England responded, a rate of 61 per cent of the population. There was some regional variation, ranging from 43 per cent in the West Midlands to 80 per cent for the East region. Analysis of potential regional effects did not find statistically significant differences for the main variables of interest – see section 4.5.3 of chapter 4.

**Figure 1-2. Commissioning survey, response rates from sample frame of 150 LAs**



### 1.4.3 National data

The national PSS data collections by the Department of Health were heavily utilised. These are validated sources available publicly on the Department of Health website and offer information about activity and expenditure for each local authority in England – each of the 150 "councils with social care responsibilities". Activity data are drawn from the collection: *Community Care Statistics – residential personal social services for adults, England,* which covers the number and types of care homes registered in England (i.e. capacity). Data on the number of places that local authorities fund is available from: *Community Care Statistics – supported residents (adults), England.* These data are currently available at: http://www.publications.doh.gov.uk/public/work_social_care.htm#catadult. Expenditure data are drawn from the PSS EX1 and RO3 returns that form the *Personal social services current expenditure in England* collection. These data are currently available at: http://www.publications.doh.gov.uk/public/work_expenditure.htm. Details of each of these data sources are given in subsequent chapters.

# Chapter 2. Economic theory, empirical findings and comparative governance: a literature review

## 2.1 Introduction

This chapter charts the organisational economics literature as a precursor to its use in addressing comparative organisational questions. The chapter describes the progress of this literature and considers its application to the social care questions of this research thesis. It is worth being clear at this stage about the need for the theory to be explicit about the institutional and organisational context in which economic activity takes place. Moreover, if we are to use theory in this work it must support overt comparisons between different institutional, organisational or 'governance' frameworks. Much of the mainstream economics literature has concerned the economic activity within market contexts, and even then with a very stylised treatment of market forms of organisation that downplay their institutional character. There are, however, a number of new schools of economic theory that are better suited, theories that go beyond neo-classical theory – the so called economics of organisation. This is a relatively new area of theory and as such is somewhat under-developed, without a consistently agreed core. Nonetheless, there is significant new work that is adding to the corpus of theory.

This chapter is structured as follows. The theoretical literature is described in the section 2. Section 3 offers a critical assessment of these theories in regard to their application to evaluating governance in social care. The fourth section considers the finding of this research in social care. The last section has conclusions.

## 2.2 Baseline theories

### 2.2.1 Neo-classical

Standard neo-classical economic theory is concerned primarily with exchange and production within markets. The high paradigm version is *general equilibrium theory* (Arrow and Hahn, 1971; Varian, 1978). Essentially, the general equilibrium model is the multi-period, multiple commodity version of the economic textbook perfectly competitive market model. Its relevance to questions of governance extends only as far as considering when this stylised market might fail to arrange efficient *complete contracts*. The standard market failures are imperfect market structure, economies of scale, information problems and externalities (Bator, 1957, 1958; Atkinson and Stiglitz, 1980).

Neo-classical economics has a fundamental set of assumptions, including the externalisation or exogeniety from the model of institutional effects, individual's tastes and preferences, social relationships (the *social environment*), and the technology of production.[1] This position is summarised in Figure 2-1. Another key assumption is that of rational maximising behaviour, meaning that: a decision maker can order 'states of the world'; preferences between these states of the world are transitive; and that decisions are made to maximise the utility obtained between different states of the world. Hence, the individual is assumed to possess exception information and computational powers.

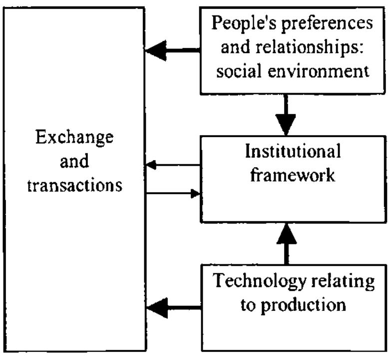



**Figure 2-1. Basic system diagram**

These assumptions add up to a very limited and abstract treatment of institutional frameworks. The market process is often abstracted to the theoretical construct of the Walrasian auctioneer (for a governance perspective see Arrow and Hahn, 1971; Sawyer, 1993) i.e. Adam Smith's *invisible* hand. The outcome is a *complete contingent claims contract*, which specifies all stakeholders' actions (production and exchange) and payments for *all* contingencies, both current and future. These assumptions are not

[1] These can be described by what Musgrave (1981) calls *domain* assumptions.

supposed to reflect reality, being a convenience that allows attention to be focused on the allocation of goods and services.

The idea of non-market organisation extends as far as a consideration of (private) firms, and even these are conceived as 'black boxes' in this paradigm. The organisation is taken to act as an indivisible decision-making unit, behaving in an analogous fashion to an individual entrepreneur. The objectives of this 'entrepreneur' and the organisation are treated as synonymous since the former is assumed to exercise full and effective control.

### 2.2.2 Institutional/behavioural

Within the behavioural tradition the interaction between social relationships and institutions is seen as paramount in explaining the behaviour of individuals. Organisations are not treated as 'black boxes'. Rather they are seen as a coalition of individuals each with their own goals (Cyert and March, 1963). Interaction between these individuals is treated as endogenous to a dynamic decision-making process that involves continual, implicit or explicit, bargaining and renegotiating. These decision-making processes are rooted in Herbert Simon's concept of 'bounded rationality' (Simon, 1955). Although a difficult concept to define precisely, it broadly means that individuals cannot engage in informed optimisation, being limited in both information and computational skills. Rather, their behaviour can be regarded as some non-optimising decision-making process, often termed 'satisficing'.

The behavioural approach provides a rich descriptive picture of the behavioural patterns that are likely to emerge within organisations such as private firms. But there are limitations; this corpus of theory under-emphasises external relationships. So whilst competition and the external regulatory environment will impact on internal behaviour, the response to actors within the organisation is limited by their resolution of internal conflict. Behavioural models have a focus away from the question of why non-market organisations exist. The approach is not truly comparative with regard to organisation. The original work also suffered the problem of not being very rigorous; rigor was traded for a relaxation of some of the more 'distasteful' assumptions found in many neo-classical models. More recent developments in evolutionary theories, for example Nelson and Winter (1982), do address some of these shortcomings e.g. with theoretical concepts of short-run production *routine* and the longer-run process of *search* to find better methods to characterise the firm.

### 2.2.3 New institutional economics: TCE

Encompassing the assumptions of rationality of the behavioural approach, New Institution Economics takes a further step from neo-classical economics by relaxing assumptions about the institutional framework in which decision are made. Building on the pioneering insights of Ronald Coase, institutional choices are made to economise on the (transactions) costs of organising production (Coase, 1937). Coase explained the existence of the capitalist firm as a voluntary response to the high costs of using the price (i.e. market) mechanism. This transaction cost-economising is the basis of Oliver Williamson's more general theory of alternative institutional forms (Williamson, 1975, 1979, 1985a, 1986; Williamson, 1994). New institutional economics (NIE) locates its analysis in *transactions* between individuals. In this way, the theory seeks to explain the institutional patterns that prevail by explicitly considering how alternative forms of economic organisation come into existence. Moreover, it recognises the importance of social relationships, and endogenous technology. It is closer to what might be called a system view. This is described in Figure 2-2 (see Hodgson, 1988).

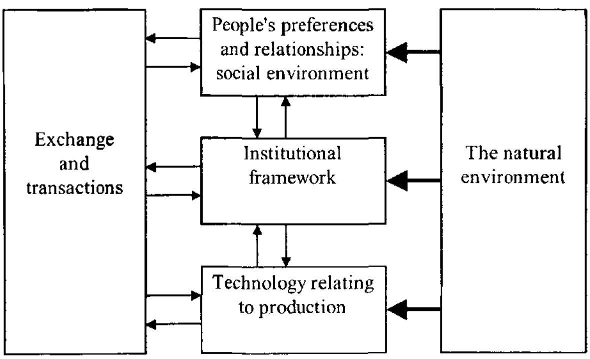



**Figure 2-2. Endogenous system**

The assumptive base of transaction cost economics is what Williamson (Op., cit.) calls human factors: *bounded rationality* and *opportunism (or human nature).* Following (Simon, 1955), individuals are bounded rational because they do not possess sufficient computational abilities to process relevant information, or find such computation too costly. Opportunism is a form of human behaviour in which (some) people act in self-interested ways, and are prepared to misrepresent relevant information if it promotes their own interests.

These assumptions – particularly bounded rationality – lead to a central tenet of transactions costs economics, that long-term contingent claims contracts will always be incomplete. To begin with, stakeholders cannot anticipate all feasible contingencies that might arise in the future, and so cannot make provision for them. Also, calculating the appropriate responses (for example, reimbursement structures that account for cost variations; pump-priming money for investments etc.) may be beyond the cognitive abilities of those people involved. Even if contingencies can be anticipated, actually specifying meaningful, clear and unambiguous terms that define both the contingency itself and how all parties might respond, particularly so that it is verifiable and enforceable legally (Hart, 1985), is at least very costly, and most likely impossible for all foreseen contingencies.

Consequently, there is a need for some process between the involved parties that enables them to arrive at an acceptable course of action to deal with non-contracted contingencies as they occur. Some form of adaptive bargaining and re-negotiation will be required. This agreed course is effectively an extension or refinement of the original contract, but it generally need only relate to the 'contingency' that has actually arisen, rather than all possible contingencies. A number of options present themselves (see Williamson, 1985, chapter 3). They can be primarily characterised by where the authority to make adaptations lies between parties to a transaction (more of this below). In *hierarchical* arrangements one party defers most authority to the other party to make decisions about adaptation to uncertain events. The firm is a good example of a pure hierarchy where an employee accepts to undertake actions specified by his or her employer.

There are serious questions as to why a person would give up their control to take orders from others, but unified control does mean that extensive prior agreements can be avoided. In *bilateral* arrangements parties to a transaction have no formal agreement (beyond the external legal framework) as to how uncertainty should be resolved. Instead they make adaptations by internal agreement, perhaps guided by convention or historical behaviour.

They agree new clauses in largely the same way as they did when agreeing the initial contract. A third category is the *trilateral* arrangement. Here a third party is granted powers of authority or arbitration to resolve unforeseen contingencies. The decision of the arbitrator – whose powers are specified in the contract – is usually enforceable according to provisions in contract law. The latter two options are generally conflated into one, since in practice resolutions are not entirely bilateral or trilateral, but are some mix of the two. In his earlier work, Williamson (1975) distinguishes between hierarchies and *markets*, with bilateral or trilateral arrangements as a subsequent refinement of market arrangements.

A key contribution of the institutional economics approach is the idea that transactions are conducted within *governance structures*. More specifically, Williamson defines a governance structure as "the institutional matrix within which the integrity of a transaction is decided" (1994, p. 102 or 1979 p. 233). It builds on the concept of an *institutional arrangement*, an idea with longer heritage, being "an arrangement between economic units that governs the ways in which these units can cooperate and/or compete" ((Davis and North, 1971), p 5-6.). More generally, North (1990) defines institutions as "the humanly devised constraints that shape human interaction" (p. 3). These constraints take "the form of rules and regulations; a set of procedures to detect deviations from the rules and regulations; and, finally, a set of moral ethical behavioural norms which ... constrain the way in which the rules and regulations are specified and enforcement is carried out" (p. 233). Furthermore, *formal* institutional constraints include "political (and judicial) rules [and] economic rules..." (p. 47) where the latter "define property rights, that is the bundle of rights over the use and the income to be derived from property and the ability to alienate an asset or a resource." (p. 47) (see also Coleman, 1990b; 1990a, on the role of rights in his theory of social action). Property rights, conferred in law, give the owner of assets involved in a transaction control about how they are used, particularly when non-contracted contingencies arise (Swedberg, 1994).

The other key contribution is the idea that choices between governance arrangements – e.g. between markets and hierarchies – have ramifications for how transactions are conducted, and specifically, the size of production and *transaction* cost they generate. It follows that choices between these governance arrangements are made to minimise the sum of production and transaction costs. The explicit treatment of transaction costs is the real departure in this literature from that before. Annex 2-1 summarises the main features of Williamson's theory.

What then determines the size of the transactions costs? Williamson describes three factors. First, *asset specificity*, the extent to which resources used in a specific transaction have a higher value in that use rather than another. It means that assets – physical or human – will become committed to particular uses with consequent adverse effects on competition. As well as for physical or technological reasons, asset specificity can accrue through use or familiarity; competition between a large number of competitors can be 'transformed' into limited number competition as assets in use gain a comparative advantage.

Asset specificity is an important cause of *hold-up* problems. Hold-up describes a situation when parties renege on prior agreements (particularly non-contractual promises) and force re-negotiation with the intention of securing more favourable terms. Specific investments tie a party into a bilateral relationship and undermine the credibility of their threat to walk away from the transaction when the less restricted partner proposes a renegotiation. There is an efficiency problem when, in anticipation of potential re-negotiation, under-investment occurs because the investor expects to lose some of the benefits of specific investments. Where specific investments would yield considerable production cost savings some form of hierarchical governance mechanism would be appropriate. Under unified control, the problem of hold-up becomes irrelevant and investment benefits can be gained.

The second factor is *uncertainty,* which is broadly defined to include a lack of information about contingencies that can, in principle, be anticipated, those that cannot, and those for which it is too costly to collect accurate information. It includes *asymmetric information* situations where one party to a transaction has more information than the other(s). Clearly uncertainty goes hand in hand with bounded rationality. Indeed, it is both the complexity and lack of information that limits rational decision-making.

The third factor is the *frequency* of a transaction. Frequently repeated transactions allow the cultivation of on-going relationships between stakeholders. These relationships will have an impact, in turn, on how future transactions are conducted.

These factors affect the size of transactions costs (Coase, 1937). Uncertainty and bounded rationality combine to increase the cost of long-term contracts in markets that would require a relatively full specification of complex future contingencies. Such transactions would be less costly where uncertainty could be absorbed in a sequential manner, i.e. with hierarchical arrangements. In addition, asset specificity and hence restricted competition combine with opportunism – especially with relatively infrequent transactions – to push up the net transaction costs of market based contracting (for an empirical analysis of specific assets in the fuel industry, see Joskow, 1987). Again, hierarchical organisation could circumvent these hold-up problems.

### 2.2.4 Contract theory

Contract theory – or perhaps more precisely, *in*complete contracts theory – extends the neo-classical model by treating the governance framework within which decisions are made as endogenous and accepts that both the governance structure and the institutional environment affect the economic behaviour of stakeholders. Contract theory also relaxes the neo-classical assumption of rationality, employing instead a (sometimes uneasy) blend of bounded rationality and hyper-rationality assumptions (see below).

Contract theory is the vehicle by which governance ideas from institutional economics have been formalised mathematically (see Hart, 1995, for a fuller account of the issues).[2] Under this broad epithet there are (at least) three theoretical developments or components that are especially central to this overall approach. First is the Grossman and Hart model of property rights in the context of incomplete contracts (Grossman and Hart, 1986). Grossman and Hart's model attends specifically to the problem of hold-up. It directly indicates efficiency advantages of hierarchical governance forms. Second is Milgrom and Robert's (1990) bargaining costs theory. This approach has its roots more firmly in principal-agent theory of neo-classical economics and is more attentive to the value of

[2] David Kreps (1996) contends that contract theory has developed from an attempt at a mathematical rendering of transaction cost economics. He sees an inevitability to this given the fundamental role of mathematical modelling in mainstream microeconomics ((Simon and Blume, 1994)). Those he argues with a mathematical inclination would no doubt argue that, because transaction cost economics is not specified formally, it is difficult to precisely determine transactions costs especially when the relevant human and environment factors act simultaneously (for example, (McGuinness, 1992)). Kreps (1996) argues that whilst mathematics does reduce problems to manageable proportions, dangers exist in missing the nuances of Williamson's work and the essence of the behaviouralist/institutionalist approaches (see below).

contracts (especially short-term contracts) and the reasons why these contracts might break down. The third is Kreps-Simon on reputation (Kreps, 1990b, 1990a; Kreps, 1990c, 1996; Simon, 1951). The Kreps-Simon model tackles the problem of motivating parties to transactions characterised by imperfect information.

#### 2.2.4.1 Property rights and incomplete contracts

The seminar contribution to the literature is the concept of the incomplete contract (Grossman and Hart, 1986; Hart, 2003). In particular, because it is hard to foresee and contract about the uncertain future, contracts cannot embody all relevant courses of agreed action for all possible contingencies. Moreover, when non-contracted contingencies do arise, their resolution depends significantly on who holds property rights i.e. ownership of the relevant asset. In particular, the *owner* of an asset has *residual rights of control* over how that asset is used.

There are two basic assumptions in the Grossman and Hart model. The first is that the supplier or provider can undertake an investment to reduce production costs that nonetheless ties them into a bilateral relationship with a less restricted partner. Moreover, the subsequent value of the product to the buyer depends on the prevailing demand circumstances, which are known only after the investment. The second assumption is that at the time of the investment the purchaser and provider cannot write a contract that commits them to a sale price specified for each possible contingency. They must wait until actual demand and other relevant circumstances are revealed and then bargain a price. As before, bounded rationality is invoked to justify this assumption; the required contract at the investment stage would be too complex to write.

Grossman and Hart (1986) consider the hold-up problem as outlined above. When a non-contracted contingency occurs after an investment, parties to the transaction have to negotiate a contract extension to cover this contingency. Absent property rights, the investing party expects at the time of investment for some of the benefits of their investment to be appropriated in negotiation over non-contracted contingencies, and so reduce their investment. However, ownership improves bargaining power and therefore investment incentives. In particular, if one party owns the investment assets as well as the production stage, that is, the transaction is conducted in a unified or hierarchical governance structure, then the hold-up problem becomes irrelevant; the owner has full rights to control and receive all (or most of) the returns of the asset. See Annex 2-2 for a (more) formal model. Here an example can illustrate the main points.

Suppose a purchaser wishes to buy residential care services. A provider is faced with the decision about the degree of investment in specialist equipment and home conversion. The benefits of the investment – and so the price the purchaser is willing to pay for placement – then depend on the needs of residents, demand conditions, local labour market dynamics, funding policy, regulatory environment and so forth. This complex situation effectively rules out the purchaser and provider from agreeing a complete contract in advance. Suppose that every £1 of investment lowers cost by £1.50 (or improved outcomes by £1.50). However, because the final price is set according to the difference between the purchaser's maximum valuation and the cost of the product, the provider can only expect to see a return of £0.75 (a half share). He therefore does not invest. Both parties therefore lose the return of £0.50 on every £1 of investment that might otherwise have taken place. Furthermore, if the cost of buying the provider's production equipment is less than this amount, it would be efficient for the purchaser to also become the owner of the means of production – a hierarchy.

#### 2.2.4.2 Bargaining cost theories and short-term contracts

In the context of incomplete contracts, governance structures emerge to guide new adaptations to and re-negotiation of existing contracts. Milgrom and Roberts (1990) argue that it is the bargaining costs of such adaptation that can cause market (governance) failure. In the absence of such costs, short-term contracts *will* be efficient. The pivotal feature of their model is that short-term contracts can efficiently allocate rights to a return from investment between stakeholders. As a result, Milgrom and Roberts show that were bargaining costs at zero, a series of short-terms contracts can, given a number of conditions, produce the same outcomes as a complete contract (e.g. no hold-up). This result is demonstrated in Annex 2-3 using the formal model developed for the Grossman-Hart theory (in Annex 2-2).

We can continue the above social care example to also show this result. In this case, at the point of investment the purchaser and provider agree what they expect the benefits of the investment to be in the future (of course they may turn out to be different because the world is complex, but that does not matter if the investment has already taken place). Suppose for simplicity that the future net benefits are shared equally. Then at the time of investment the stakeholders write a (short-term) contract – conditional only on the current investment level – where the purchaser pays for £0.50 of each £1 of investment. This means that the provider only has to spend £0.50 to invest with an expected return of half of £1.50 or £0.75. When the actual placement characteristics are known the purchaser and provider write another contract to agree the final price. Hence the investment goes ahead efficiently under a bilateral or market governance arrangement.

Milgrom and Roberts argue that the costs of agreeing and writing short-term contracts constitute *all* the transaction costs of market governance. An assumption made here is that binding short-term contracts can be written over the initial investments. This assumption is a crucial difference between the Milgrom-Roberts and Grossman-Hart models. When contracts can be written concerning initial investments, then there are no residual control issues and ownership becomes irrelevant. The only aspect of our example that has changed is that the stakeholders can write a binding short-run contract on the investment. Only the costs of this activity would impede reaching an efficient agreement.

Milgrom and Roberts' theory proceeds from the basic Coasian idea that when the costs of organisational decision-making are zero, negotiation always leads to an efficient outcome (Coase, 1960). They, nonetheless acknowledge a growing bargaining literature that does not (fully) support this claim (see surveys by Lyons and Varoufakis, 1989; Kennen and Wilson, 1993). In particular, when there are no obvious points of symmetry (Nash, 1953) around which to split a surplus, problems of hold-up and delay can occur as parties try to jockey to appropriate larger shares. In extreme cases where there are multiple players, bargaining may fail completely to provide a solution.

In addition, the theory described above embodies a number of more restrictive information assumptions. What happens when these are relaxed? The basic model has been extended to show that short-term contracts can also produce first best outcomes, absent short-term bargaining costs, with risk-averse stakeholders and moral hazard (or *ex post*) asymmetric information problems (Fudenberg, Holmstrom, and Milgrom, 1990).[3] Information problems relating to adverse selection are problematic and do render contracts inefficient. When one party to a transaction has *ex ante* private information about some factor that is relevant to the transaction, a first-best efficient contract cannot be designed. An example of this sort of information advantage is where a provider knows their underlying cost structure, but the purchaser does not.

Incentive compatibility (for truthful revelation of private information) is only achieved at the cost of some informational rent, a result from the mechanism design (adverse selection) literature (Baron and Myerson, 1982; Baron, 1989; Laffont and Tirole, 1993; Forder, 1997a). In other words, the provider can be induced to give information about costs to the purchaser, but only at a price. Specifically, more cost efficient providers will be allowed to under-invest (to lower costs) because they have an initial cost advantage over other providers and can use this to cover poor investment/productivity. For example, a residential care provider might have a less intensive (and so cheaper) staff training programme if it already had very low cost physical assets.

Following this account, bargaining costs include: the opportunity cost of bargaining (time, writing contracts...), costs of monitoring and enforcing the agreement (contract) and any (efficiency) "losses from failure to reach the most efficient agreement possible in the most efficient manner" (Milgrom and Roberts, 1990, p 65). The last of the three components includes (a) coordination of bargaining failures and (b) information rents that accrue to adverse selection problems.

#### 2.2.4.3 Reputation

When negotiating over non-contracted contingencies, stakeholders face a trade-off between the transactions costs associated with the process of adaptation and the risk of being exploited. These transaction costs include collecting information, monitoring, determining and then bargaining in respect to the best course of action, etc. Parties to a

[3] Other required conditions are that intertemporal transfers are possible and a number of technical limitations on common knowledge and the utility function.

transaction may decide to forgo or limit these activities. For example, individuals may be willing to defer authority and decision-making to a superior in a hierarchy (and hence reduce the costs they incur in dealing with uncertainty) if they believe the risk of exploitation by the hierarchical superior to be small. Likewise, a purchaser of services using a bilateral market contract can reduce transaction costs by reducing quality monitoring if he does not expect the provider to exploitative.

Mostly however, uncertainty is rife and so the *opportunity* for exploitation is high (anecdotally, this is a good description of most social care transactions). Why then do parties cede control rights and put themselves in potentially exploitative situations? Kreps (1996) believes that Simon's (1951) paper on the employment relationship is the seminal reference in this case. The central proposition is that an individual will cede some control to another because they are protected by the threat of mutually harmful exit by the employee should that person become convinced of exploitation. Put another way, the employer will refrain from exploitation in order to safeguard his or her reputation.

Kreps (1990a) argues that reputation provides a glue that permits mutually beneficial transactions to take place, where otherwise they would be too costly. Indeed, Kreps (1990a; 1990c; 1996) formalises these ideas using the folk-theorem from game theory (Fundenberg and Tirole, 1992). Essentially, this theorem shows that long term abstemious behaviour netting a modest payoff over many repeat transactions (for which a player attracts a good reputation and therefore secures repeat transactions) can be more attractive than short-run exploitation (which is likely to result in contract termination). In that case monitoring by the potentially exploited party can be reduced to the point where the expected short-run gains accrued before detection of exploitative behaviour, are just less than the long-run gains from behaviour that safeguards reputation.

To illustrate how this theorem applies to governance choices, consider two parties A (a manager) and B (a home-help worker) who have agreed an (incomplete) contract. Suppose that as the future unfolds they arrive at a contingency that is not specified in their contract. Moreover, A has relevant information about the event (e.g. a change in the financial regime), and the authority to take decisions, and B does not. Then, in a dynamic (repeated) scenario, A may accrue a reputation for being an honest and flexible manager such that B accepts what A tells him as the truth and is happy to concede authority.

In this case A need not be altruistic or moral or somehow constrained to be honest and act responsibly. Indeed, it is in A's interest to tell the truth and not exploit the situation because he conjectures that B will 'punish' him by terminating the contract (i.e. quitting) if A lies. So long as the value of a responsible use of authority outweigh the value of exploitation (and subsequent costly repercussions), a reputation equilibrium can hold.[4]

Here then an implicit *convention* is being followed: A does not exploit the situation for short-run gains and allows B some flexibility in fulfilling his contractual obligations. In return B can focus on the task at hand, work productively, without needing to establish (costly) safeguards on A, nor to duplicate work in collecting information or making calculations relevant to determining appropriate actions. In effect, a mutually beneficial *gift exchange* has taken place (Miller, 1992).

Whilst we consider the general shortcomings of contract theory below, there are two important conditions for the reputation model to work. First, it must be the case that each party can verify the actions of the other. Second, it must be the case that each party recognises and understands the actions of the other parties and knows how to respond in order to protect their reputation. Thus, each player must know what convention they will be playing and what the appropriate actions are in complying with this convention. If one party does not know what the appropriate 'comply' action is, then the reputation equilibrium may fall apart.

---

[4] Reputation appears to be able to overcome many forms of potential transaction failure (See Rasmusen, 1992, for a description of the infinitely repeated reputation game that he calls the 'grim strategy'). Early applications of this idea to product quality is Klein (1981)). However, a conceptual problem often dubbed the chainstore paradox (Selten, 1978, 1975) does complicate reputation models with *finite* repetition of transactions. Briefly, if the game has a definite endpoint, at that time a player's reputation is irrelevant and so he will exploit the situation if able. In anticipation, the other party will quit the game one transaction early, which means that the penultimate transaction is actually the last transaction. The game then unravels and the cooperative outcome is lost. The introduction of a small degree of uncertainty about the employer's type (i.e. perhaps the employer really is altruistic and therefore will not exploit) can overcome this problem (Milgrom and Roberts, 1982; Kreps and Wilson, 1982).

## 2.3 Critical application to comparative governance issues in social care - conceptual issues

### 2.3.1 Theoretical foundations for analysis of social care

In this section we consider the application of these broad theoretical approaches to social care with two principal and inter-linked aims. First, to review what the literature has provided by way of attempts to solve the governance problem. Second, to form some conclusions as to what the best theories are for the job, and thereby to point to required developments of the theory for our purposes.

### 2.3.2 Neo-classical

Neo-classical theory does not directly address the *comparative* governance question. What it does explain in detail is where markets might fail. Moreover, that markets can fail can be taken as an implicit case for alternative forms of governance. Traditional welfare economics has generally assumed this alternative to be 'government' (Atkinson and Stiglitz, 1980). The non-profit sector has also been considered as an alternative form of governance within neo-classical economics (Weisbrod, 1975, 1988).

'Standard' theoretical causes of market failure – imperfect competition, economies of scale, information asymmetry etc. – are well known (Bator, 1958; Barr, 1993) and need little rehearsing here, except, perhaps to point out the re-interpretation of the perfect competition condition as instead the need for perfect contestability (Baumol, Panzar, and Willig, 1982).

There have been a number of analyses of social care markets that draw on (neo-classical) industrial economics approaches (e.g. Forder, Knapp, and Wistow, 1996; Le Grand and Bartlett, 1993). A distinction is made between *structural* imperfections and *information* imperfections (Knapp et al., 1994; Forder, Knapp, and Wistow, 1996). The former concerns imperfect market contestability and competitiveness and the latter covers uncertainty and asymmetric information. In relation to asymmetric and incomplete information between purchaser and provider, there are many examples to draw on in social care (see section 2.5.1.2 below). Barriers to entry and exit are less significant. Regulation imposes some costs, but the relevance of sunk costs (Dixit, 1980) in a labour intensive industry is generally low (see section 2.5.1.1).

According to the neo-classical model these types of imperfections lead to market failure and imply high market governance costs. But how secure are these inferences? As we noted above, there are a number of areas of concern in using neo-classical theory as a basis for comparative governance evaluation. To begin with, neo-classical theory does not provide an account of why particular forms of governance are chosen; the 'required conditions' of this theory – i.e. textbook market failures – are joint necessary conditions for the existence of a textbook perfectly competitive market, and a first best outcome. But beyond this, the theory is largely silent, even as regards the consequence of one of the necessary conditions failing to hold. It tells us that either a complete contract can or cannot be written (which is of little practical relevance), but not the consequences of *some* degree of failure.

Furthermore, individuals' tastes, preferences and social relationships (embedded in a *social environment*) are also exogenous. As a result, a consideration of the feedback between social relationships and individual's preferences and values, on the one hand, with institutions or rules of the game, on the other, is precluded from the analysis.

These are not criticisms *per se* but rather a recognition that neo-classical micro-economics was not designed for comparative organisation purposes. Moreover, despite other criticisms (e.g. about rational maximisation assumptions – see Evers, 1993), the neo-classical approach, with its 'simplifying' assumptions, has been defended on the basis that it is meaningless to talk about the realism of assumptions because theories, being abstractions, cannot exhibit, nor are designed to exhibit complete realism (Friedman, 1953).
The question of whether a theory is realistic 'enough', it is argued, can only be settled by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories" (Friedman, 1953, p41; but see also Hodgson, 1988).

### 2.3.3 Transactions cost economics

#### 2.3.3.1 Applications to social care

Casual empiricism suggests that transaction cost economics (TCE) has significant relevance in social care. Most local authorities currently retain an in-house service that has many features consistent with *hierarchical* governance from the TCE literature (e.g. unified public ownership, top-down management, salaried employees etc.). Furthermore, as detailed in the next two chapters, councils also 'contract out' many of their services to private providers. These *market* transactions bear the relevant hallmarks: separately owned purchaser and provider, voluntary bilateral contracting for service, residual rights of control for providers, a residual claim by providers on financial surpluses etc.

Bartlett (1991) asserts that transaction cost economics represents a significant improvement upon the neo-classical approach in application to health care. In mainstream applications of neo-classical theory, transactions are relatively straightforward and have few of the conditions that cause high transactions (market governance) costs. In other words, even where contracts are incomplete, transactions can be undertaken efficiently with a sequence of short-term contracts in market settings (see Williamson, 1985, fig 2-1 and Kreps, 1990b, fig 20.1). However, the complexities of health and social care transactions combined with assumed bounded rationality suggest a role for alternative governance and hence the use of alternative theory (Bartlett, 1991, p53). Bartlett then goes on to review the basic transaction costs framework and considers its application for the study of the reforms and more specifically contracts choice in the NHS internal market.

Using a transaction cost economics approach Ferguson and Keen (1996) consider the transaction cost implications of strategies to improve information flow in the NHS (by using information and communication technologies – ICTs). They begin by interpreting transaction costs as similar to a per-unit tax on service provision (transactions). This 'tax' causes a sub-optimal level of output and therefore implies a deadweight efficiency loss. Specifically, these deadweight losses are *bargaining efficiency losses*. Ferguson and Keen themselves define transaction costs as the costs of acquiring information (search costs), the costs of bargaining and decision-making (negotiating costs) and the costs of enforcement and monitoring. They argue that information technology might reduce these types of transaction costs.

Two aspects of the transaction cost economics approach that Lunt et al., (1996) argue make it relevant for an analysis of community care are its concepts of *atmosphere* and *asset specificity*. The former is a concept that addresses the importance of attitudinal and ethical considerations in transactional relations. Lunt et al. (1996, p 376) note that Williamson's concept of atmosphere is used to capture the effects of the moral or value basis of transactions. Reference is made to altruistic motivations and reciprocity (e.g. see Titmuss, 1970). The social care relevance of atmosphere most clearly stems from the presence of voluntary sector providers in community care. As will be reviewed below, there is evidence that voluntary sector providers in social care have motivations, values and expectations that deviate substantially from those assumed in standard market theory (Kendall and Knapp, 1996; Forder, 2000). This hypothesis seems particularly apposite regarding voluntary organisations that have primarily campaigning and advocacy role (Smith and Lipsky, 1993; 6 and Forder, 1996). In addition to the volutary sector, there is also evidence that ostensibly for-profit, private sector providers exhibit motivations that depart from pure profit maximisation (Forder, 2000; Forder et al., 2004; Kendall et al., 2003). This suggests a blurring of the boundaries between the voluntary and private sectors.

Whilst there are many opportunities for specialisation in social care, from finance officer to care assistant, these are not transaction specific. There is a pool of people with relevant skills to draw on, and although some specific training would be required the costs are relatively low, as are the costs of people moving to similar jobs (there are 150 local authorities to choose between). Nonetheless, there will be some examples of highly specialised training for specific purposes where the costs of training for that individual purpose are large (Lunt, Manion, and Smith, 1996). In addition, physical assets tend to be more specific e.g. specially designed buildings (to meet the needs of disabled people). Also 'social capital' in the form of trust or reputation (Fukuyama, 1995) tends to be specific to a particular setting.

#### 2.3.3.2 Broad criticisms

The transaction cost economics/new institutional economics literature, especially the early writing, is generally not presented in the same formal/mathematical style as mainstream economics. In part of course this approach reflects its departure from the methodological tradition of neo-classical economics, particularly with its use of theoretical concepts that do not easily lend themselves to a formal treatment. But this approach is open to challenges concerning its internal consistency. It is also difficult to develop precise predictions for testing.

There are also more specific criticisms. One of the main arguments deployed in explaining hierarchical forms of organisation in TCE is that assets relevant to a transaction are highly specific. However, the contract theory approach (e.g. Milgrom and Roberts, 1990) shows that with appropriately crafted short-term contracts, asset specificity need not lead to inefficient outcomes (e.g. hold-up) in markets. The argument therefore is not whether transactions have specific assets, but whether enforceable short-term contracts can be written. And because *short-term* contracts are sufficient – with no need to agree complete contingent contracts – such contracts could be a practical proposition. The transaction cost of market activity is then the cost of deploying appropriate short-term contracts.

Another issue rests with the under-development of the transaction cost of hierarchical governance. Whilst Williamson is, for example, clear about the importance of recognising the limitations of feasible alternatives, and that alternative governance structures have their own costs and benefits, his account of hierarchical governance costs is somewhat limited. For example, the treatment of effort and the power of internal incentives could benefit from development. Moreover, Williamson makes reference to 'authority' in hierarchies that seems to go beyond the simple ceding of control to managers. Reputation and trust theory provides insights, but this area still feels incomplete – see section 2.4.1 below.

Another plank of TCE is the relevance of the social context of transactions. Context and social convention can influence how parties act when undertaking transactions. Granovetter (1985) criticises transaction cost economics on this point, arguing that Williamson does not sufficiently account for the social construction of behaviour and preferences: Williamson's 'atmosphere' goes some way but needs further development. Indeed, (Lunt, Manion, and Smith, 1996) recognise the importance of contributions by economic sociology in its contribution to identifying key influences on economic exchange. However, the authors argue that sociological approaches lack the coherence of an overall model, and should be seen as complementary to economic theories rather than substitutes (p. 380).

### 2.3.4 Quasi-markets

Although perhaps not a full theory as such, a significant, empirically grounded approach that has been applied in social care is described in the *quasi-markets* literature (Le Grand, 1991, 1992; Le Grand and Bartlett, 1993; Forder, Knapp, and Wistow, 1996). This approach suggests five 'conditions for success' – market structure, information, transaction costs and uncertainty, motivation and cream-skimming – that characterise the net benefits of quasi-markets. In this regard it has a neo-classical flavour, but beyond methodology, draws also on concepts from transaction cost economics. This quasi-market approach is explicitly concerned with public sector quasi-markets (rather than private market places). In having an essentially neo-classical methodology it shares the weakness of not really being comparative (see above). Failure to achieve the success conditions *implies* that alternative feasible organisational structures may have comparative advantages, but their costs and benefits are not considered. This literature does acknowledge the *second best* problem (Lipsey and Lancaster, 1956) – whereby if a success condition is violated, second best efficiency need not be achieved when the other conditions *are* met – but does not suggest a solution (which are generally very complicated – see Spulber, 1989).

Quasi-market analysis suggests that the nearer a contract is to being complete – and thereby having larger *ex ante* transaction costs – the less *ex post* transaction costs will be. Contract theory has a different perspective: not all feasible future contingencies need to be written into a (short-term) contract *ex ante* for the sequence of short-term contracts to be efficient *ex post*. Indeed, according to Milgrom and Roberts (1990, p 68) the agreement need specify only the immediate actions the parties will take and how they will be compensated. Therefore far fewer prevailing factors need to be explicitly accounted for in the contract, which greatly reduces *ex ante* bargaining costs (and so *ex ante* transaction costs). Some of the 'conditions for success' do not actually have to hold in this case; appropriately crafted short-term contracts could still produce good outcomes, despite apparent failings of some of the conditions.

Le Grand and Bartlett (1993) list transaction costs as a separate condition for success and identify *ex ante* and *ex post* transaction costs. Compare this to the definition of transaction costs in contract theory, where efficiency losses and therefore transaction costs are *caused* by violations of quasi-markets success conditions. Some part of total transaction costs is exogenous, but the other part is *endogenous*. Overall, either a broad definition of transaction costs should be adopted in which case its dependence on the other success conditions must be explicit. Or, a narrow definition is used, but with transaction costs cast as one element only of a comparative governance analysis.

### 2.3.5 Contract theory

Contract theory departs from neo-classical accounts with its prohibition on the writing of complete contracts (Hart, 2003). Contract theory is particularly promising because it is explicitly comparative and offers a formal account in the tradition of micro-economics. Nonetheless, with regard to social care governance questions, the bulk of incomplete contracts theory has been about the private capitalist firm as the alternative to market transactions. Only very recently has this theory developed to consider public/government hierarchies as alternatives to markets – or more precisely to quasi-markets where public purchasers buy in markets (Hart, 2003; Besley and Ghatak, 2001). Even then, these developments tend to focus only on specific problems e.g. provision of public (i.e. non-rival, non-excludable) goods (Besley and Ghatak, 2001) or on comparing types of public-private partnership with private providers and/or investors and only funding from government sources (Hart, 2003). One of the main theoretical contributions of the present work will be to develop the theory around the question of public hierarchical provision versus quasi-market provision (see chapter 5 in particular).

The empirical relevance of incomplete contracts theory – in advancing a combination of uncertainty, bounded rationality and transaction costs as the cause of incompleteness – is high. By contrast, although information imperfections in neo-classical theory – i.e. in the *information economics* branch of that literature – can lead to inefficiency, replicating some of the predictions of incomplete contract theory, this theory also indicates that rational actors can often write efficient incentive-compatible contracts. The problem is that these very sophisticated contracts are very rarely seen in practice. So, whilst the (complete contracts) information economics literature can provide useful insight and has provided a number of UK social and health care related works (e.g. Levaggi, 1996; Forder, 1997a), on balance the incomplete contracts approach has the greater promise.

## 2.4 Remaining problems and missing pieces of theories of comparative governance

There is a good foundation of theory available to address the main questions in this work. Nonetheless gaps and shortcomings do remain. Before considering two in more detail, it is worth re-iterating a main shortcoming of the above theory: its focus on *private* sector markets and firms. Public sector (quasi) markets and hierarchies have significant parallels but are clearly not completely analogous. This is an issue to be picked up in what follows of this work.

### 2.4.1 Rationality

For Williamson human rationality is of prime concern in understanding choices between methods of organising economic activity. In particular, he sees people as often unable to make the sophisticated and information-demanding calculations that are usually required to make markets efficient. Herbert Simon's seminal contribution of bounded rationality has a great deal of intuitive appeal. However, its current weakness (at least from the mainstream economics point of view) is its lack of precise definition for (mathematical) modelling purposes (but see Rubinstein, 1998). Transaction cost economics and contract theory invokes bounded rationality for the purpose of precluding complete contingent claims contracts, without which there would be no need for any form of governance of transactions.

It is important to note at this point that, according to the (hyper-rational) property rights school, some approximation to complete contracts need not rule out hierarchical organisations (firms) (Alchian and Demsetz, 1972; Jensen and Meckling, 1976). In particular, the firm is seen "as a nexus for a set of contractual relationships among individuals" (Jensen and Meckling, 1976m p. 310). According to this orthodoxy, there is no concept of 'authority' that goes beyond that provided under contract. This position is totally at odds with the economic sociology view, particularly the Weberian notion of authority and power (see a translation of Weber by Wittich and Roth, 1978). According to Weber, behaviour is affected in a way that is distinct from the constraints of economic 'market' power (as given by the contract) and is described as deriving by virtue of authority i.e. power to command and duty to obey (see also Hamilton and Feenstra, 1995). For Jensen and Meckling, the firm is nothing more than a legal fiction that serves as the nexus of contracts. Buchanan (1986) proceeds to criticise the transaction cost approach on this basis.

Lyons (1996) summarises the critique of the hyper-rational approach taking a practical, empirical perspective. Not only is the hyper-rational assumption somewhat distasteful in itself, when used it rules out empirically relevant phenomenon like conflict, hold-out and indeed, not instant resolution of contract negotiation, even when they are subject to asymmetric information between parties (Lyons and Varoufakis, 1989). There is also now a considerable body of experimental economics that finds widespread evidence of non-(hyper) rational behaviour (Roth and Schoumaker, 1983; Kagel and Roth, 1994).

All the comparative governance contract theory models also employ the bounded rationality construct to drive predictions of non-market governance. However, Kreps (1996) has highlighted a modelling weakness in its use. In transaction cost economics, bounded rationality works in three ways to prohibit complete contracts. First, that it is impossible to anticipate all the likely complicating factors that might affect a transaction. Second, that calculating the appropriate responses may be beyond the cognitive abilities of those people involved. Third, actually writing contract contingencies in a legally binding manner is, at the very least extremely hard and costly, and at worst it is impossible.

All three forms are (anecdotally) relevant in social care. Consider the example of a local authority purchaser setting up an agreement with a for-profit organisation to provide a specialist rehabilitation service for people with mental health problems. The example is pertinent to each of the three elements of bounded rationality. First, anticipating the care needs (and so costs of service) of all potential clients will be very difficult. Hence some contingencies will be unforeseen. Second, the contract needs to balance the incentives to provide good quality care, whilst not putting too much (net income) risk on the provider, and also to secure against any hold-up problems that might accrue to specific assets (equipment, staff training) in the transaction. Therefore an efficient contract may be very hard to calculate and negotiate. Finally, all clauses need to be written in an enforceable manner. This requirement in particular means objectively defined outcome and quality measures, something that has proved elusive in social care research (Knapp, 1984; Challis, 1992; Forder, Knapp, and Wistow, 1996).

Kreps notes that incomplete contract theory cites bounded rationality as a reason why complete contracts cannot be written, but also assumes that people are sufficiently rational to anticipate what decisions to take for the range of expected future contingencies. Their current behaviour is based on those expectations. These assumptions do not sit very comfortably with the above forms of bounded rationality, especially the first two. If all three forms of bounded rationality held precisely then writing short-term contracts at the time investment was occurring, in anticipation of the distribution of future outcomes, would be very difficult. And without confidence in this contract, hold-up problems are likely. Nonetheless, the actual short-term contract requirement is that both parties at the investment stage are happy that all reasonable contingencies are addressed, especially if the investment is somewhat 'lumpy' (i.e. not smoothly differentiable) as it often the case.[5]

More generally speaking, the boundedly rational decisions of individuals at any given time need not always amount overall to inefficient outcomes. If people can differentiate good outcomes from bad outcomes when they happen, then heuristic or rules of thumb behaviour can with repetition approximate the outcome of hyper-rational decision-making (Rubinstein, 1998). 'Natural selection' arguments are a good example. People who make the 'right' choices are more likely to 'survive' for the very reason that they are efficient. Tractable models of near rational behaviour should then be good at predicting behaviour at the limit.

### 2.4.2 Embeddedness and motivation

Granovetter (1985) levels an important criticism at economic theories of organisation. He argues that social relations in markets are more important, and those within hierarchies are less important, than suggested by economic theory of organisations (see also Dow, 1987; Perrow, 1990). Notwithstanding, Kreps's (1990a) work on reputation and ceded control in hierarchies, much of the new institutional economics literature supposes some intrinsic degree of 'authority' within hierarchies that motivates

[5] Then the investment decision becomes a gamble and full investment can still go ahead if parties believe that some potential contingencies have not be fully figured into calculations, so long as the risk of these contingencies is small.

employees. By contrast, the classic view of markets is of self-motivated actors pursuing exchange opportunities.

Putterman (1986) comments that we should not only look at the boundary between market and hierarchy (to see where this lies), but also to see both market and hierarchy as woven into the cloth of the wider economy. Others argue that Williamson places too much emphasis on the law as the institutional context of both governance structure and individual transaction (Hamilton and Feenstra, 1995). Also, by making hierarchy synonymous with authority in Weberian sense (see above), "the boundaries of economic organisation are determined by the reach of authoritative power and are not arbitrarily equated with the firm" (p. 62). In other words, it is inappropriate to see only activity within hierarchies as affected by (exogenous) social context; indeed, the distinction is too strongly drawn.

An essential point underlying these commentaries is that individuals are guided by general social values and norms (to some extent) in how they tackle (complex) transactions undertaken in any governance structure. Indeed, the importance of these experiences increase in proportion to the complexity and potential level of risk of a transaction: the less information an individual has about the specific transaction (particularly the type of player who he/she is dealing with) the more actions are likely to be tempered by broad social principals. Miller (1992) recognises this in the context of employment relations when he says that " 'rational choice' in such a setting may move an employee to make a 'gift' of costly effort simply because the ultimate effects of deviation from the norm may be large, uncertain and negative" (p. 206).

Grannovetter describes how transactions are *embedded* in conventions that exist in an individual's social network (see also Hannan and Freeman, 1984; Taylor and Hoggett, 1994a). In this way it is logically possible that a conflict can exist between, on the one hand, short-run gains that would be reaped by an 'instrumentally rational' person and, on the other hand, (instrumentally abstemious) behaviour of not flouting wider social conventions. Taylor-Gooby (1997) notes that the longer-run ramification of adhering to social conventions – that is maintaining a trust relationship – often yields higher efficiency: trust economises on transaction costs, particularly monitoring governance activities (Sako, 1992; Fukuyama, 1995).

There is also a related view that the institutional context is closely bound up with individuals' motivations – that the latter are endogenous to some degree. To caricature Le Grand (1997) somewhat, if policy treats people as instrumental (as being 'knaves' as Le Grand puts it) then we risk transforming trusting and altruistic people ('knights') into acting in instrumental or 'knavish' fashion. This effect works through the influence of policy on the social environment (see also Frey, 1997a, 1997b; Kendall, 2001).

Social environment and a departure from self-interested behaviour appear to be particularly relevant in social care. Not only are professional values of particular importance, but also caring values (Forder, 2000). Many voluntary organisations – and most notably smaller, local organisations – operating in social care would seem to defy instrumental rationality (Kendall, 2001; Taylor and Hoggett, 1994b). In health care, there is clear recognition of the organisation and the professionals within it, and between the organisation and its *context*. It follows from an application of organisational theory (McNulty and Ferlie, 2002). Health care organisations in England operate in a context of wider political and administrative management culture, which impinges on behaviour of the organisation, and can heavily influence change of the environment in which it works, for example forcing organisational change, mergers and re-structuring (Fulop et al., 2005).

The above arguments are persuasive but are by definition hard to model. A balance needs to be struck between being able to derive the essential results of economics of organisation theory – i.e. comparative governance propositions – and recognising how embeddedness and motivation will nuance these propositions. Kreps' (Op. Cit.) work on reputation in hierarchies is a very useful step forward in this respect. To a certain extent this approach also addresses the endogeneity of motivations, and if not of fundamental motivation then at least endogeneity of what a person expects to achieve, and their willing to trust others. An example is the 'grim strategy' in game theory whereby trust is maintained as long as it is reciprocated, but if not then an individual takes on a far more cynical or 'grim' position (Fundenberg and Tirole, 1992). Furthermore, a more flexible specification of motivation allowing non-profit objectives is also a component of the developing set of theory. This leaves more fundamental notions of (non-instrumental) trust to address. Can this be incorporated or will attempts to do so lead to arbitrary or even tautologous predictions?

## 2.5 Critical application to social care II: methodology issues

Two broad methodological approaches to the empirical application of the above theory are discussed in this section. First, is the *failure-performance* approach, which covers analyses that have sought to evaluate prevailing governance structures by assessing whether the theoretically determined 'required conditions' hold. For example, the quasi-markets programme – see section 2.3.4.

The second approach looks at outputs or at least process indicators directly. For example, studies that look at process indicators such as the flexibility of domiciliary care services (e.g. are they provided at times that people want?) when these services are provided in markets by independent sector organisations as compared to the previous hierarchical arrangements (Lewis, Bernstock, and Bovell, 1995; Lewis and Glennerster, 1996). Others look at the attainment of public policy goals directly, by eliciting the views of key stakeholders i.e. purchasers and providers (Wistow et al., 1996, chapter 7).

The second approach has the advantage that it (generally) needs to make fewer assumptions in order to link its process measures with outcome and efficiency conclusions. The failure-performance approach needs to infer how imperfections affect behaviour and how behaviour in turn has implications for efficiency. Ideally, analyses would determine how costs and *final outcomes* change with governance alternatives. Final outcomes refer to the fundamental benefits that are conferred, such as improvements in people's well-being, utility, quality of life etc. In practice, however, these outcomes are very difficult to measure. In the main, the empirical literature has employed the first approach. There are a few examples of the second. Netten and colleagues have used conjoint analysis to develop a older person's utility scale (Netten et al., 2002), but this has not yet be used in comparative studies. A number of studies have used intermediate outcomes, rather than final outcomes. For example, Forder (2000) considers the difference between the profitability and propensity to seek profits of providers of mental health services organised in different forms of market governance. In particular, non-profit providers operating in more network like arrangements made less profit, and sought less profit than other providers operating in regular quasi-markets.

### 2.5.1 Inferring governance costs: imperfections approach

The *failure-performance* approach includes the literature that has commented on the prevailing degree of imperfection (regarding the 'required conditions') in social care. A distinction can be made between *structural* imperfections and *information* imperfections (Knapp et al., 1994; Forder, Knapp, and Wistow, 1996). Literature relevant to 'human factors' i.e. rationality and motivations in social care is also reviewed.

#### 2.5.1.1 Structural imperfections: competitiveness/contestability

Taylor and Hoggett (1994a) argue that the strategies of local authorities in purchasing community care could decrease diversity of supply and perhaps market contestability. In particular they have concerns that small voluntary sector organisations will find the cost of market entry very high compared to their expected income. An example given by Taylor and Hoggett is insurance costs. Another example is the very high contracting cost organisations must bear once they enter the market (Gronbjerg, 1990). Case studies have identified high start-up costs, again especially difficult for smaller, specialist organisations, but suggest that the market is relatively competitive for the mainstream services (Hoyes and Means, 1993). Hoyes and Means also note that market failures can occur through overbearing use of monopsony power by local authorities as they push prices too low (see alsoForder et al., 2004).

A number of papers employing econometric techniques have explored the relationship between market structure and price. To use observations of market structure (i.e. numbers and types of providers) to comment on possible inefficiency, assumptions have to be made about the relationship between market structure and behaviour as well as between behaviour and efficiency. Econometric analyses can be used to estimate the former relationship rather than make assumptions about it. Forder and Netten (2000b) analyse data on residential and nursing home care for mainly elderly people. They find a statistically significant, but modest relationship between price and market structure, suggesting that competitiveness is relatively high. There have also been a number of studies of prices, demand and competitiveness in the US nursing home sector (Nyman, 1994; Nyman, 1989, 1985). Competitiveness appears to be lower in the US markets; however, entry into those markets in particular is more tightly regulated than in the UK. Studies of health care in England suggest evidence of a competition effect (Propper, 1996). There is also evidence that competition can reduce the quality of care in UK hospitals (Propper, Burgess, and Green, 2004).

#### 2.5.1.2 Information imperfections

The potential for information imperfections in social care appears high. Forder (1997a) considers the relationship between information problems (specifically, moral hazard problems relating to misrepresentation of client service cost characteristics) and the incentives embodied in current social care contracts. Data from a sample of residential care homes for older people are consistent with information efficiency shortfalls generated by providers operating under certain types of contract (see also Donaldson and Gerard, 1989).

Hoyes and Means (1993) draw on case study investigations; they comment that local authorities appeared to have poor and overly centralised information systems. Contract clauses often did not make reference to relevant factors of the transaction, particularly about quality. We might conclude that information search and monitoring governance activities are low and therefore that information imperfection remains significant. There are also more direct examples of poor information (Knapp et al., 1994). They found deficiencies in purchasers' knowledge of providers' prices and types of services.

### 2.5.2 Rationality and motivation

Kendall (2001) offers a number of different theoretical perspectives on motivation and provides evidence of motivations from a sample of residential care providers. In interviews, respondents more often reported professional and 'caring' (altruistic) motivations than profit or income related motivations. Forder (2000) shows the importance of motivations for efficiency and hence net governance costs. Data on providers of services for people with mental health problems casts serious doubt on the profit maximisation assumption. A comprehensive investigation of the motivations of home care providers in England comes to the same view (Matosevic et al., 2001; Kendall et al., 2003).

## 2.6 Conclusions: Theory and empirical techniques for evaluating social care governance

This literature review began with a description of the baseline economic theory contenders for an analysis of comparative governance in social care: neo-classical economics, transaction cost economics, and contract theory, the latter two constituting economics of organisation theories. The aim of this chapter was to consider the applicability and usefulness of the different economic theories in addressing the comparative governance and efficiency questions of this work (see chapter 1).

Neo-classical theory is to be rejected on these grounds. Its central weakness in this regard is that it treats governance and institutions (including formal and especially informal, social institutions) as exogenously determined. Moreover, where implications for governance can be inferred indirectly, for example, relating to the conditions whereby markets fail (to generate first-best efficient complete contracts), predictions have been poor. Many commentators have questioned the core assumptions of the neo-classical model, especially hyper-rationality and exogenous preferences.

*Economics of organisation* theory treats the organisational or 'governance' structures in which economic activity or 'transactions' are undertaken as endogenous. Governance matters in these models because complete contingent contracts of neo-classical general equilibrium theory are ruled out (generally by the bounded rationality of individuals and the prohibitive transaction costs that would be involved). The analysis concentrates on the nature of the processes and activities required to complete a transaction, rather than on the terms of production and exchange relations *per se* (as in the neo-classical paradigm). The choice of governance structure affects the costs of undertaking an efficient transaction (e.g. of collecting relevant information, determining and writing contracts, and monitoring for compliance...). Moreover, this relationship between cost and governance choice is taken according to certain attributes of a transaction (the initial information distribution and level of uncertainty, specificity of investments and so forth).

The central tenet of the theory is that transactions, which differ in their attributes, are aligned with governance structures in some discriminating manner. In fact, the key criterion in this literature for discriminating between governance structures is an efficiency one: the minimisation of *net* transaction costs. This approach therefore has significant potential in addressing our research question.

Two broad approaches in the economics of organisation are distinguished. The first is *transaction costs economics* (TCE). Although having long antecedents, TCE has introduced a new rigour of (economic) thinking about the existence and function of forms of economic organisation. The chapter also reports some recent developments, which although heavily rooted in TCE literature, go further in a number of ways; in particular, contract or rather incomplete contract theory, which uses formal (mathematical) modelling. In doing so it also addresses some shortcomings over the definition and cause of transaction costs in TCE.

This theory provides a sound foundation for this work. Nonetheless, it too has weaknesses to be addressed in the following chapters. First and foremost, contract theory has very largely to date been concerned with comparisons between privately owned firms as hierarchies and the (private) market. Our focus, however, is publicly funded activity: quasi-markets and public hierarchies/ bureaucracies.

Second is the issue of the treatment of stakeholder motivation and the influence of social context on transactions, drawing on concepts from economic sociology. There is a need to strike a balance between being able to derive the essential results of the theory – i.e. comparative governance propositions – and recognising how social context and motivation will nuance these propositions. David Kreps' work on reputation in modelling *trust*, particularly in hierarchies, is a very useful step forward in this respect and will be used below. To a certain extent this approach also addresses the endogeneity of motivations, and if not of fundamental motivation, then at least the endogeneity of what a person expects to achieve and their willingness to trust others. The theory in subsequent chapters will also adopt a more flexible specification of motivation, explicitly allowing non-profit objectives. This leaves more fundamental notions of (non-instrumental) trust to address. Can this be incorporated or will attempts to do so lead to arbitrary or even tautologous predictions?

Third, contract theory is somewhat selective in its treatment of bounded rationality. Complete contracts are ruled out by this assumption, but individuals are still able to form rather comprehensive *expectations* about the future. In what follows, we will take the approach of Milgrom and Roberts (1990), allowing a certain degree of rationality initially but modelling bounded rationality through the build-up of transactions costs. For example, stakeholders do not write complete contracts because the transaction costs of doing so are prohibitively high rather than due to a simple preclusion that comes from bounded rationality.

Turning to (empirical) methodology, two approaches were identified in the literature. The first – the failure-performance approach – involves observing the attributes of transactions within the prevailing governance structure to infer behaviour and therefore efficiency. The second approach attempts to directly measure behaviour and infer efficiency on that basis. Although the first approach has the advantage of a much easier measurement problem, it has the distinct disadvantage of having to make an additional layer of assumptions. These assumptions are required to infer behaviour from the observation of transaction attributes. Overall, the second approach appears more promising, particularly when under-pinned by a developed contract theory model.

## Annex 2-1. Transaction cost economics

Williamson (1993) summaries the crucial feature of transaction cost economics:

(i) The transaction is the basic unit of analysis;

(ii) Economic actors can undertake farsighted contracting but stops well short of hyper-rational extremes by conceding that all complex contracts are incomplete;

(iii) The critical dimensions of the transaction as mentioned above are frequency, uncertainty and asset specificity;

(iv) Williamson identifies 4 structures: market, hybrid, private bureau and public bureau;

(v) Each generic mode of governance displays discrete structural differences of both cost and outcome (benefit) according to nature of the transaction (item iii);

(vi) There are costs of *selective intervention*. These costs mean that interventions between semi-autonomous parts of a hierarchy do not always occur when there is an opportunity for expected net gain. Because hierarchies are unable to intervene selectively, they are unable to replicate market (high powered) incentives (Williamson, 1985b, chapter 6; Milgrom and Roberts, 1990).

(vii) Each generic mode of governance is supported by a distinctive form of contract law;

(viii) Williamson argues that transactions that differ in their attributes (item iii) are aligned with governance structures, which differ in their costs and benefits (item iv), in a 'discriminating – mainly transaction cost economising – way';

(ix) As stated above, transactions and governance structures are also embedded in a social environment (North, 1990; Miller, 1993). Changes in the nature of that e.g. in norms or conventions change the costs and benefits of particular governance structures;

(x) Williamson stresses that transaction cost economics is an exercise in comparative institutional analysis, concerning *feasible* alternatives.

## Annex 2-2. The Grossman-Hart model.

Consider a general framework that has two stakeholders and two periods. Stakeholders make some relevant action in each period that affects their final payoff. Conventionally, the first period's actions by each party, $x_1$ and $x_2$, are some form of investment decision. Actions in the second period, $y_1$ and $y_2$, are implementation actions. At the outset before any actions are undertaken, stakeholders can agree an allocation of control rights over the second period implementation decisions, $y_1$ and $y_2$. These control rights are assigned contractually. If the implementation decisions refer to physical assets then a ceding of control rights can be achieved by one party selling the assets to the other party.

Suppose the cost of action $x_i$ is $x_i$, $i$ = stakeholder 1, stakeholder 2 and the same for $y_i$. The total benefit that accrues is $S(x_1, x_2, y_1, y_2; C)$, being dependent on actions taken and the circumstances $C$ that prevails in time 1. Because the number of possible circumstances is so high, Grossman and Hart assume a contract on $x_1$ and $x_2$ in the first period cannot be written. Only in time two when the actual $C$ is known, can stakeholders contract. Because stakeholders cannot contract in the first period, individual expected benefit will depend on the distribution of period 2 benefits between stakeholders that occurs after action $x_i$. Grossman and Hart suppose that benefits are shared equally in some un-specified negotiation process. Hence, at time 1, each player can expect total payoffs of: $\pi_i = U_i + EV_i$ . Here $U_i$ is period 1 payoff, which equals $-x_i$ the cost of the investment. Also, $EV_i$ is period 2 expected payoff, which will be half the total benefits less the cost of the period 2 action, i.e. $EV_i = \frac{1}{2}S(x_1, x_2, y_1, y_2) - y_i$ . Consequently, each players' expected payoff is: $\pi_i = -x_i + \frac{1}{2}S(x_1, x_2, y_1, y_2) - y_i$ . In period 1 therefore, each player will invest to maximise expected payoffs i.e. $\partial\pi_i/\partial x_i = -1 + \frac{1}{2}S_{x_i} = 0$ or $S_{x_i} = \frac{1}{2}$.

Grossman and Hart compare this bilateral ownership outcome with a hierarchical or unified ownership case. Then total costs and benefits are as before, so the owner maximises: $\pi = S(x_1, x_2, y_1, y_2) - x_1 - x_2 - y_1 - y_2$ and so investment in this case is $\partial\pi_i/\partial x_i = -1 + S_{x_i} = 0$ or $S_{x_i} = 1$. Since it is assumed that benefits $S$ are strictly increasing in $x_i$ with marginal benefits greater than marginal costs, this analysis implies that investment and so net benefits are greater under unified ownership.

## Annex 2-3. The Milgrom-Roberts model

In contrast to the model in Annex 2-2, the two players can contract in the first period over investment decisions, $x_1$ and $x_2$. It is specifically because bargaining costs are zero that stakeholders can do so. The players will work out what efficient agreement would be reached for any circumstances $C$ in the second period. These expectations are then the basis on which the first period contract is determined, that is investments made and costs of the investments shared. Each player can calculate how much their and the other player's expected benefit is compared to each player's investment cost. The total costs of investment in period one can then be shared in proportions consistent with a half share of the total net benefit. Each player's expected payoff is $\pi_i = U_i(x_1, x_2) + EV_i = U_i(x_1, x_2) + E[W_i(S(x_1, x_2, y_1, y_2; C), y_1, y_2)|C]$ where $W_i$ is the proportion of the second period surplus that accrues to party $i$ (and depends on the realisation of $C$). In period 1, a contract will be written that is conditional on $x_1$ and $x_2$ so that the total payoff is shared, i.e. $\pi_i = \frac{1}{2}[U_1 + E[W_1(S, y_1, y_2)|C] + U_2 + E[W_2(S, y_1, y_2)|C]]$. To illustrate this point suppose that the common belief was that each player's investment (cost) $-x_i$ would yield the same marginal (expected) benefits $S_{x_1} = S_{x_2}$. Then $\pi_i = \frac{1}{2}[S(x_1, x_2, y_1, y_2) - x_1 - x_2 - y_1 - y_2]$ and so optimal investment is given by $\partial\pi_i/\partial x_i = -\frac{1}{2} + \frac{1}{2}S_{x_i} = 0$ or $S_{x_i} = 1$. In this case each stakeholder would pay for half the investment. Note that this contract will not specify terms for the second period and therefore does not need to be contingent on all possible values of $C$.

# Chapter 3. Analytical concepts and theoretical foundations

## 3.1 Introduction

This chapter considers how economic activity is organised. It develops concepts and frameworks that will be used to describe how the social care system (in England) is organised, drawing on the literature discussed in chapter 2. The organisation of a social care system concerns issues such as the ownership and control of key functions (i.e. of funding, assessment, care planning, purchasing, providing, monitoring etc.), how resources are allocated between stakeholders as they undertake these functions, how the system is regulated, and so forth. There are a great many options for how systems can be organised along these lines. For example, are purchasers and providers separately owned and controlled, are market forces and competition introduced, or are top-down bureaucratic methods used (Saltman, Busse, and Mossialos, 2002).

This and subsequent chapters focus on the implications of organisational choices in social care for efficiency, in particular, the impact on net transaction costs of organisational choices. Since choices about the organisation of *public* services such as social care need not be solely motivated by maximisation of efficiency, the analysis is not as such a positive analysis; it will not predict what organisation form (or more specifically, governance structure) should exist given relevant prevailing conditions. However, from a normative perspective concerning efficiency – or value for money, best value, or however it is labelled – the analysis will suggest whether the most efficient governance structure is currently being used. It is worth noting that in the case of publicly funded services, it is politicians and mainly central government politicians that decide the general organisational structures to be used in social care. And they may reference criteria such as equity, political expediency, social rights etc. as well as value for money/efficiency, in making their choices. In the case of private economic activity, it is consumers, entrepreneurs, regulators and so on that determine what organisational structure prevails (Coase, 1937).

The developing set of economics of organisation (EO) theory has much to offer in addressing comparative organisational questions as noted in the last chapter. It explicitly recognises that the way systems are organised has a strong bearing on how they perform, and that this form of organisation is not simply given, but instead chosen. Second, building on this recognition, the theory is continuing to refine its definition of forms of organisation. Third, the question of who makes choices about organisation, and to what ends, are central.

This chapter begins to develop the economics of organisation – especially the approach adopted in Milgrom and Roberts (1990; 1992) – for an application to social care organisational alternatives. Essentially it builds up in detail the concepts used in EO theory, applies them to social care, and then uses the result to assess the comparative efficiency properties of actual organisation arrangements used in social care. As to the structure of this chapter, after this introduction the chapter proceeds by defining efficiency in a way that is relevant to assessing organisational alternatives. It then turns to the question of what the basic elements or building blocks are of an 'organisational form'. The fourth section then casts this thinking in terms relevant to the social care system in England. The fifth section outlines theory that helps to determine the efficiency characteristics of organisational choices. This analysis is used to infer the comparative efficiency properties of different organisational arrangements and in this way underpins the empirical analysis conducted in subsequent chapters.

## 3.2 The problem of economic organisation

Organisation is the central concept in this chapter. Economic organisations are "created entities within and through which people interact to reach individual and collective goals" (Milgrom and Roberts, 1992, p. 19). But why do they exist? A fundamental observation about the economic world is that people can produce more, and realise economic gains, if they specialise in activities to produce goods and services, transacting with one another to acquire needed components and also final products and services. Whilst these gains to specialisation can be massive, they can only arise if people's actions and decisions are *coordinated* and people are appropriately *motivated.*

### 3.2.1 Coodination and motivation

The complexity of production of many services, not least social care services, gives rise not only to benefits of specialisation, but indeed, the absolute need for it. Going back to Adam Smith, it is clear that the volume, value and diversity of goods and services that can be consumed by individuals is very many times greater than if all individuals were entirely self sufficient (Blaug, 1986). The provision of social care has individuals devoting themselves to particular elements of the many stages of production. Of central importance is that this specialisation requires organisation.

The fundamental problem of organisation is twofold. First, is the coordination problem, that is, people must know what tasks need to be undertaken, how they should be accomplished and who should do what. For example, what kinds of social care services should be provided; how should they be produced and delivered to users; and who should do the commissioning and producing? The problem of organisation also involves a motivational problem, which is to ensure that the various individuals involved in these processes willingly do their parts. They should both accurately report information to implement and operate the right plan and also act as they are supposed to in order to carry out the plan.

## 3.3 Organisational forms – building blocks

### 3.3.1 Transactions and contracts

Specialisation leads to organisations that embody a myriad of interactions, negotiations, exchanges, and so forth, between people. It is these interactions or *transactions* that are at the core of EO theory, being the largest unit of economic activity that cannot be subdivided and performed by several different people (Milgrom and Roberts, 1992). The transaction is the process whereby individuals plan and implement activities to produce services, and agree terms of the exchanges of resources. The agreements struck between individuals as they coordinate within the organisation are *contracts*. These contracts are far broader than the formal legal agreements of the corporate world. Indeed, contracts can be informal, verbal, not enforceable or even verifiable by a third party. They do however specify each party's actions and rewards for each of a range of circumstances or contingencies. Contracting can be a continual process with new agreements being reached as new contingencies arise.

### 3.3.2 Principals and agents

Another theoretical building block of organisation economics is the classification of people or parties involved in transactions. Because transactions can be characterised by an imbalance of information and by contestability problems, there is likely to be a dependency relationship between the parties involved. In particular, one party to the transaction often has either more information and/or better bargaining power than the other party. In particular, the *principal* is a party who wishes to secure provision of some good or service but does not have the necessary specialised knowledge, skills or assets. The principal employs an *agent* to undertake this task and in the process delegates some control to that party (Grossman and Hart, 1983).

The problem for the principal in securing some service from the agent is either not knowing the true value of the benefits of the service, or being forced to accept those 'benefits' the agent wishes to supply. Either way the information imbalance or contestability problems make it difficult for the principal to motivate the agent to act, to a reasonable degree, in the interests of the principal. Although, the transaction will be mutually beneficial, (or why would it happen?) there may be an imbalance as to how the spoils are shared. In fact, the problem may be so significant than in the process of splitting the metaphorical pie, some of the pie can be wasted. These concepts are (gross) simplifications of the real world. But, when cast in these terms, theory can take us a long way – with predictive success – without being bogged down in spurious detail (Joskow, 1987, 1988; Shelanski and Klein, 1995; Baker and Hubbard, 2001, 2003).

In what follows we make the distinction between *purchasers* and *providers* in social care. This distinction is detailed in the next chapter within the historical context and current social care system in England. At this stage it is worth outlining the essential features using the terms of the principal-agent framework. Local authorities have a duty to provide social care for eligible users. To secure appropriate services the local authority acts as service commissioning or purchaser (the latter term has a slightly more narrow meaning) on behalf of the individual with care needs (even if the authority is commissioning its own in-house providers). This task involves an assessment of needs, a determination of an appropriate package of care (or more generally, care planning), and a determination of the person's and any carer's eligibility for local authority support with the funding and commissioning of those services. It also involves the (corporate) financing and 'buying' of eligible services. Local authorities can purchase from independent sector providers – by contracting out – or they 'purchase' from in-house providers.

Purchasing does not pre-suppose a framework for organising social care transactions. So purchasing can involve instructions to in-house providers to allocate services to individuals. The more traditional concept of purchasing is relevant when local authorities buy services from external providers. Nonetheless, whether it is the authority's contracting unit or other middle management grouping, in markets, hierarchies or other, local authorities are purchasers of public funded social care. In the above terms, the local authority purchaser is the *principal* in social care transactions. References to 'the purchaser' mean the *operational* manager of the purchasing function – he or she need not be the overall manager or chief operating officer of the social services department. Moreover, we are entitled to expect that the principal's motivations, preferences and constraints, will in some part reflect those of the most senior management and also of the elected members of the council.[6]

Providers are those organisations that supply contracted services. Independent sector providers supply under contract to the local authority purchaser. In-house providers are those organisational divisions within the local authority that specialise in the production of care. In large bureaucracies that conduct both purchasing and providing the distinction is often somewhat blurred e.g. senior managers can be responsible for both functions. Nonetheless, conceptually speaking, the operation of a social care system involves purchasing and providing function whatever governance arrangements are used. References to 'the provider' in most cases mean the main manager or decision-making within the providing unit i.e. the person responsible for operation of the production process. The provider is the *agent* in social care transactions. We need to take a moment to consider the role of (informal) carers. The above account should in no way be taken as underrating the importance of informal care. Nonetheless, informal care by definition involves transactions with a familial, household or close friends grouping, and whilst very interesting, is outside the remit of this work.

### 3.3.3 Governance arrangements

Individuals working in the social care system expend time, effort and resources in determining and discharging contracts in addition to direct production activities. There

[6] This potentially opens another set of (hierarchical) principal and agent relationships. However, these do not involve the direct transaction of services and will be more strategic in nature e.g. defining the principles under which the purchaser operates. This is interesting but beyond the scope of this work.

is a clear conceptual distinction between an individual's activities of production of services – e.g. caring for residents in a care home – and negotiating the care contract with the local authority and/or the resident. The ways these governance activities are carried out depends on the governance structure, which, as outlined in the last chapter, is the matrix of rules, regulations, protocols and conventions that pertain to the transaction (North, 1990; Williamson, 1979; Williamson, 1994).

Whilst the idea of governance structure is relatively simply, actually defining such a structure is much less so. The literature as outlined in chapter 2 has attempted to draw out relevant dimensions. These include: ownership, control and agency (brokerage and devolution) (Jensen and Meckling, 1976; Grossman and Hart, 1983, 1986; Coleman, 1990a; Williamson, 1975, 1985a); contract form and reimbursement incentives (Forder, 1997b; Hart, 1995; Fudenberg, Holmstrom, and Milgrom, 1990; Laffont and Tirole, 1993; MacNeil, 1985; Lyons and Mehta, 1997); regulation (Spulber, 1989; Stigler, 1971; Vickers and Yarrow, 1988); and social environment (Hamilton and Feenstra, 1995; Granovetter, 1995; Hannan and Freeman, 1984). These dimensions/concepts can self-evidently be applied in social care, although a detailed classification is left to the next chapter. Institutional nuances particular to social care in England are important as outlined in the next two chapters. Nonetheless, the main theoretical propositions of this work can be laid out 'barebones' style, as below, and then fleshed out later.

#### 3.3.3.1 Definition

Two of the above dimensions of governance structure are particularly important and so warrant elaboration. The first dimension concerns the degree of integration of purchasing and providing roles, and the ownership of the associated infrastructure. It is very useful to distinguish between ownership and control (Coleman, 1994). The former concerns, in particular, who owns the apparatus and assets of the purchasing function and who the providing function in social care. Put another way, are the principal (in this case the purchaser of social care services) and the agent (in this case the provider of care e.g. a care home) part of the same organisation or are they in separately owned organisations? As to the latter, ownership need not dictate the distribution between stakeholders of *control* over the various functions. An organisation with unified ownership may, for example, internally separate purchasing and providing. Or the owner of one set of assets may voluntarily pass or cede control to the owner of another set. In fact, the location of control is perhaps the key factor in explaining strategic performance, as is clear in incomplete contract theory.

The second dimension is *contract design*, which has a fundamental bearing on the operation of economic systems in general (Laffont and Tirole, 1993) and social care systems are no exception (Frank and Gaynor, 1991; Ma, 1994; Propper, 1995; Forder, 1997a). Oliver Williamson distinguishes types of incentives as either high- or low-powered (Williamson, 1985a). High-powered incentives are defined with reference to a stakeholder who "either by agreement or under the prevailing definition of property rights, appropriates a net revenue stream, the gross receipts and/or costs of which stream are influenced by the efforts expended by the economic agent." (Williamson 1985, p132). In other words, incentives are high powered when an individual can keep all the profits resulting from their efforts. Low-powered incentives feature some dilution of the relationship between profits/surpluses and efforts. Salaries are examples of low-powered incentives – individuals receive income that is only indirectly related to their efforts. As Williamson notes, the power of incentives depends on whether providers have control over their own actions and efforts and have the right to appropriate net income, either as a result of ownership or because this right was ceded contractually. It is a dimension that is closely aligned with ownership and control as described above.[7]

A closely related feature of contracts is whether agreed reimbursement is *contingent* on, i.e. tied to, the exogenous drivers of production cost. When stakeholders are salaried employees do not bear production costs and so their wage does not fluctuate (directly) with cost-relevant variation. When stakeholders are residual claimants however, the degree of contingency is relevant. If reimbursement is largely non-contingent, then providers are likely to experience significant variation in their net income.

#### 3.3.3.2 Governance archetypes

Even restricting our attention to the above dimensions creates a multitude of possible governance structures. However, in practice choices along particular dimensions tend to

[7] There is a significant literature on reimbursement incentives in contracts, including the optimal design of incentive power (Laffont and Tirole, 1993; Forder, 1997a). One practical development is the 'soft capitation' contract that employs a mix of high and low powered incentives (Frank and McGuire, 2000).

be correlated and can be grouped to reduce the number of alternative governance structures to just two main types (also see below on this point).

*Markets* have separately owned and controlled means of provision and purchasing. Contracts are determined in voluntary (bilateral) exchanges and contract adaptations are negotiated and resolved by both parties. Payment incentives are often high powered since ownership usually confers the right to appropriate residuals. It is however possible for lower or mixed reimbursement incentives to be embodied if parties cede some of the rights to income in the contract.

*Hierarchies* are characterised by decision-making authority (regarding adaptations to the contract) that is vested with one party (the hierarchical superior e.g. managers) being ceded to them by the other party (the subordinate e.g. employees), who accept the instructions of managers. Contracting is undertaken in a unified way and hierarchical subordinates are usually paid on a salary or equivalent low-powered incentive basis. Hierarchies commonly feature unified ownership, although that is not always the case since human assets can be hierarchically arranged but not owned by another (precluding slavery!). Employees can always leave if they wish.

These and other relevant features are summarised in Table 3-1 below. The table also includes a third combination, loosing termed 'network' governance. It is included to illustrate how many of the commonly attested features of networks in the literature can be re-produced by combining features of markets and hierarchies. For example, ownership is dispersed as in markets but control is often mostly ceded by one party voluntarily to the other party (the provider). Grant payment is very common. This form of payment mixes incentives regarding individual transactions. Although providers receive a lump-sum award and can keep the residual, this is usually spread over many transactions so allowing cost cross-subsidisation. Moreover, there are often circumstances that allow retrospective adjustments to the payment.

Some commentators distinguish 'networks' as a substantive separate category of governance structure (Powell, 1990; Rhodes, 1997, p 52). The argument often concerns the role of *trust* in transactions. Whereas hierarchies are characterised by top-down authority and markets by arms-length contracts, networks are characterised by trust and co-operation. Others comments reject this distinction, emphasising that markets and hierarchies are also embedded in social networks and that the role of trust in these structures is not systematically different (Granovetter, 1985). The relational contracts literature indicates that firms operating in (competitive) markets often engage in co-operative behaviour for mutual benefit (Dore, 1983; Sako, 1992; Kitson, Michie, and Sheehan Quinn, 2001). The formulation of the market governance archetype is centrally constructed around the idea of *bilateral* (rather than hierarchical) relationships and therefore it encompasses many of these concepts. Networks can be seen as either essentially bilateral/market or as hybrids between market and hierarchy. For these reasons, in what follows, only the markets and hierarchies archetypes are considered (which also makes the problem much more manageable!).

**Table 3-1. Governance options**

| **Dimension** | **Hierarchical** | **'Network'** | **Market (bilateral)** |
|---|---|---|---|
| ***Ownership-control-agency*** | | | |
| **Ownership** | Integrated | Dispersed | Dispersed |
| **Control** | Unified | Separate/relational | Separate/de-centralised |
| **Brokerage** | Purchasing agent | Agent | Individual |
| **Devolution** | Strategic (central) | Tactical | Tactical (local) |
| ***Contract design*** | | | |
| **Incentive type** | Low-powered (salaried) | mid-powered (grant) | High-powered (provider keeps profits) |
| **Specification** | Minimal, informal | Minimal, informal | Detailed, formal |
| **Length (duration)** | Short, frequent | Short, frequent | Long, infrequent |
| **Timing** | Retrospective | Retrospective | Prospective |
| **Contingency (linkage between payment and cost)** | High, costs and reimbursement linked | Intermediate | Low, fixed prices |
| ***Monitoring*** | | | |
| **Contract-specific** | Low – informal arrangements | Low | High – monitoring for compliance w.r.t. specifications |
| **Supply-side regulation** | Low – self-regulation | Low | High – Regular inspection |
| ***Social environment*** | | | |
| **Alignment of motivations by social convention** | High | High | Low |

## 3.4 Comparative governance

The economics of organisation theory offers a set of tools for addressing the comparative efficiency of different governance structures. This task is achieved in a number of ways.

First, the theoretical analysis provides a basis for defining governance choices. This assessment involves at one level simply interpreting organisational changes using the theoretical building blocks developed above. The theory is then used to understand why and how the organisation elements – which together form governance structures – impact on transaction outcomes.

Second, theory then frames the normative or comparative evaluative analysis of this work. In principle, different (sets of) governance arrangements can be compared against a number of criteria (e.g. equity implications, their public choice implications etc.), but their efficiency implications are the subject of this work. The degree to which coordination and motivation can be achieved determines transaction benefits, which in turn has a bearing on efficiency. Broadly speaking, when comparing alternatives, *transaction benefits* are the production costs of the service that is produced adjusted for the value/quality of the production outcomes. Transaction benefits need to be set against transaction costs. Standard evaluation approaches can be adopted; in particular cost benefit analysis (Mishan, 1976).

Transaction benefits are measured as: $V - C$, where $V$ is the total value of the final of the product and $C$ is the production cost. Transaction costs are $G$ and so net transaction benefits are:

$$(3\text{-}1) \quad (V - C) - G$$

The costs – production and transaction – can be measured in monetary terms in a relatively straightforward fashion. The outcomes $V$, however, incorporates the buyer's utility valuations of the product, and measuring this valuation in monetary terms is fraught with difficulties as the willingness to pay literature will testify (Donaldson, 1990). Instead, it is possible to take a shortcut by making certain assumptions. The approach adopted is to extrapolate production costs for each alternative given that the same outcome was achieved. Then the actual valuation of $V$ is irrelevant and net transaction benefits are given by adjusted production costs i.e. $\overline{C}(\overline{V})$ (see Knapp, 1984).

The relevant comparison is therefore:

$$(3\text{-}2) \quad \bar{V} - \bar{C}_1(\bar{V}) - G_1 - [\bar{V} - \bar{C}_2(\bar{V}) - G_2] = (\bar{C}_2(\bar{V}) + G_2) - (\bar{C}_1(\bar{V}) + G_1)$$

where the subscripts denote the alternatives being compared. If this function is positive then alternative 1 is the preferred choice. The $V$ term drops out of (3-2) as noted. The difficulty with this approach is the need to find proxies for $V$ that fully account for the implicit relationship between $C$ and $V$. These quality proxies may be hard to find, although their specification is much easier that having to place a value on $V$ directly. This issue is explored in the relevant empirical chapters below.

Governance structure choices do not directly impact on benefits and costs; rather these choices affect the actions of stakeholders as they work within these structures:

$$(3\text{-}3) \quad V^K(g_1^K,\ldots,g_n^K) - C^K(g_1^K,\ldots,g_n^K) - G^K(g_1^K,\ldots,g_n^K)$$

where $K$ is prevailing set of governance choices. The relationship between activities $g$ and $V$, $C$ and $G$ will differ for each $K$. The functions, $V^K(.)$, $C^K(.)$ and $G^K(.)$, for example, will reflect how well the motivation and co-ordination problems are solved in each governance set $K$. EO theory specifies these functions, and so in turn underpins the development of propositions for empirical testing and frames the empirical analysis. The key theoretical proposition is that a relationship exists between governance structure choices and outcomes, mediated by the features of the transaction and the principal-agent configurations involved.

In summary, our analysis of the comparative governance efficiency requires us first to define theoretically the conceptual building blocks. This task is used to develop archetypes or discrete sets of governance options. Then to use the theory to develop conjectures as to how people will behave within these different archetypes/sets. The second part is empirical; to identify these archetypes, measure transaction costs and adjusted production costs, and then to test the relationships between archetype/set indicators and these costs.

## 3.5 Main propositions

Two of the most relevant dimensions of governance choice are ownership/control and contract type. The former is central to distinguishing between market and hierarchy archetypes. With regard to contract type, the focus will be on the degree of contract contingency – the degree to which reimbursement rates reflect the specific costs of production – see section 3.3.3.1. Within market governance, we do in practice see a significant variation in contract contingency. Whilst the power of reimbursement incentives (see page 65) is also relevant in theory, in practice it tends to be highly correlated with the choice of market or hierarchy.

In care home markets, providers are almost exclusively residual claimants of one form or another, whether payment is contingent or not. In hierarchies, providers tend to have much lower powered reimbursement. Provider staff mostly receive a wage and although they often have a pre-determined budget, under-spends are retained at the centre, not by the provider unit. Over-spends are generally met by central reserves although persistent over-spends will tend to have ramifications for the employment status of the provider manager. These reasons mean that we focus only on contingency choices in markets and so avoid problems associated with the correlation with market or hierarchy choice.

In what follows, the main arguments and proposition with respect to these two choices are outlined. This constitutes an introduction to the analysis detailed in chapters 5 to 8 – where the ideas below are fully fleshed out – and aims to provide an overview.

### 3.5.1 Markets and hierarchies?

Hierarchies are characterised by top-down management decision-making. Planning the organisation's activities, such as what should be produced and how, is undertaken by the principal – see section 3.3.2 above – and these decisions then govern actual production. The principal is assumed to be a manager with key operational authority. We need to be clear, nonetheless, that this manager(s) will be working within a broader corporate and political context since we are referring to public sector commissioning. As outlined in chapter 1, we have taken purchasers as wishing to maximise the value of services provided (net of costs) – see (1.6). This is likely to be the consistent motivation of officers and members alike, particularly under cabinet local government arrangements, which feature a social care lead cabinet member. In practice, nonetheless, the function in (1.6) might not fully reflect the preferences of members even after they have sought to direct officers. For our purposes, as long as this influence activity is relatively stable and so (1.6) remains unchanged, we can leave this process to one side.

The planning function is undertaken in a unified way by managers and ceded by employees, who accept the plans of managers. By contrast, in (quasi-) markets the separation of purchasers from providers leaves each with a set of overlapping decisions, an overlap that can lead to some duplication of planning effort. Primarily to protect their usually different interests both purchasers and providers must each gather relevant intelligence (e.g. about users' need characteristics) in order to contract. Stylising the process, production requires an investment task and an implementation or 'effort' task, both of which can be very complicated (see chapter 1). Without a contract, effort, in particular, is unlikely to be of a level and type that is of mutual interest to the parties to the transaction.

In hierarchies, managers make decisions and instructions are passed to employees by fiat, restricting the costs of developing a 'contract'. Nonetheless, posturing, haggling and disagreement *can* occur between divisions of a hierarchical organisation, and are described as influence activities (see last chapter and Milgrom and Roberts, 1992). In markets parties must make explicit, objective contracts over a far wider range of contingencies. This contract determination is very expensive, comparatively speaking.

Plans will need adaptation when non-contracted contingencies arise. In hierarchies current activities can be adapted in a unilateral way, whilst in markets a further round of *joint* intelligence gathering and re-negotiation is required. A disadvantage with markets is the cost of these management activities; the contract is more costly to generate in markets than in hierarchies (Williamson, 1975, 1985a).

An integral part of modern complex organisations is the communication of information between relevant parties. It is a vital part of contract determination. The price mechanism can be a very efficient method of communicating relevant information (Milgrom and Roberts, 1992). Bureaucratic means of communication tend to operate on a quantity and quota basis and are less efficient.[8] In this respect markets might have a transaction cost advantage.

#### 3.5.1.1 Co-ordination

There are a number of options for parties to contract over *investment decisions* in markets. Essentially, the buyer and provider can choose to agree (cost-sharing) contracts in relation to investment decisions or, as is more common, the provider unilaterally makes investment decisions and bears the costs. The latter has problems, however, because although the provider bears the costs, they end up sharing the benefits with the buyer. The prospect of lower (marginal) value from investment may lead providers to under-invest. This problem is a form of *hold-up* as described in sections 2.2.3 and 2.2.4.1 of chapter 2. A sequence of short-run, cost sharing or *incentive* contracts could be agreed on investment, improving investment levels (section 2.2.4.2) but also generating greater transaction costs.

Hierarchies also have potential investment problems. Although there is no hold-up – because investment decisions are made by managers who are also the 'buyers' – employees/providers need to trust that they are not being exploited by managers (in respect of accepting reasonable instructions from managers). Even fears about exploitation may lead employees to interpret investment-related orders in a conservative fashion. Trust, reputation and social network effects as discussed above (especially section 2.4.2) all bear on the (expectation of the) likelihood of this exploitation. In addition, many commentators have noted politically determined limitations on investment funding in *public* sector hierarchies (*viz*. PPP and PFI policies).

Managers themselves may also be relatively under-motivated and so take low risk investment decisions. Public hierarchy managers are not exposed to the same performance incentives as providers in markets; much of this turns on the reward and failure regimes between hierarchies and markets respectively. Anecdotally, public organisations have tended to get 'bailed-out' if performance is poor up to a point. Thereafter, the risk of losing their jobs presumably offers a more robust source of

---

[8] One corollary is that price mechanisms tend to be somewhat *brittle*, that is when incorrect information is transmitted the effects are often more serious than in a quota system.

motivation for managers. Overall, it is difficult to say *a priori* whether investment will be higher in markets or in hierarchies. Much will depend on the asset specificity of investment, since it is the underlying cause of hold-up problems, as noted in the last chapter. Complexity is also a prime factor because it makes contracts particularly difficult and expensive to write.

Turning to the implementation or *effort* decision, the theory is clear on differences between markets and hierarchies. Incentives for providers to supply effort are high in markets; providers are residual claimants and so benefit directly from reduced costs. Effort is therefore *induced* in markets relative to the product contract, where this contract maintains quality. Without the product contract effort is likely to be misdirected relative to what the purchaser wants from the product as providers simply choose the least cost production strategy. The costs, therefore, of securing effort in markets are the costs of writing a well-specified product contract. These costs will be high and also a large component of the total cost will be fixed: to get even basic applicable effort a complex product specification is needed. But once the contract is specified then marginal bargaining costs for further *directed* effort are far less. The outcome for markets will be high (directed) effort – i.e. high transaction benefits – but also high transaction costs. Complexity and frequency of transactions play important parts because they increase the costs of writing product contracts in markets. Competition is also a prime factor because it potentially reduces negotiation costs (that stem from bargaining over terms in bilateral monopoly situations).

Unilateral efforts to cut costs (or improve quality) are not directly rewarded if providers are salaried, as is usually the case in hierarchies. Instead, managers offer instructions and effort is forthcoming to the extent that employees/providers adhere to those instructions. How much effort will depend on: how extensive are the instructions from managers, how reasonable employees find the instructions (i.e. are they being exploited?), how closely compliance is monitored and what sanctions there are for failure to comply. Overall, this implies low fixed, but high and rising marginal transaction costs. Optimal effort, therefore, is likely to be lower than in markets. But, this also implies lower total transaction costs associated with effort.

Another relevant co-ordination problem concerns the distribution of *risk* between purchaser and provider organisations in markets. The delivery of care services to people with a great variety of characteristics is fraught with uncertainties. The ability of providers to cope with the resultant income risk is inversely related to the size of organisations – big organisations can pool and spread risks more easily (Hey and Lambert, 1987). It follows that big organisations should optimally take on the majority of this risk. Public purchasers e.g. local authority social services departments are large, but often place most of the risk on providers by imposing fixed prices on the market (see next chapter and Wistow et al., 1996). Hierarchies have fewer such problems because risk can be spread and pooled throughout the organisation (by internal transfers). In markets inappropriate allocation of risk does occur and results in under-supply and excess prices.

#### 3.5.1.2 Motivation

Turning to motivation problems, providers often have better information about their production processes and costs than purchasers or managers. Purchasers/commissioners may not be able to base their instructions (their contract terms) on relevant, but hard to measure factors such as for example, provider productivity or production effort. 'Shirking' behaviour can then disguised amid the usual ups and downs in output that result from external conditions. It is a problem in hierarchies with their lack of incentives for effort (low-powered incentives). The problem is an extension of the cost-cutting co-ordination problem described above. The co-ordination problem of hierarchies hinged on the comprehensiveness of instructions to providers to improve productivity; the motivation problem rests with managers not being able to determine whether their instructions are being followed (or perhaps the inability of providers to specify their instructions in a way that is verifiable).

Shirking on effort is not a problem in markets because providers, who are residual claimants, make effort decisions – there is nothing to be gained by misrepresenting effort to purchasers. However, shirking on *quality* can be a problem. Even if quality can be written in contracts in a satisfactory manner, purchasers may not now be able to verify compliance (Dranove and Satterthwaite, 2000; Klein and Leffler, 1981; Propper, Burgess, and Green, 2004).

Poor information about the characteristics of service users and how those characteristics affect service costs can lead to problems in markets. This information problem largely depends on how providers are paid, specifically the degree of contingency of contracts. *Cost-exaggeration* can occur where reimbursement is related to some indicator of service cost (e.g. resident characteristics) and involves providers having incentives to overstate these cost-related indicators (see below and Forder, 1997a). Alternatively, when reimbursement is not contingent, providers may be inclined to over-select low-cost people, claiming to purchasers an average mix of users' needs. This *cream-skimming* behaviour stems from the availability of alternative providers, the choice of which accommodates selectivity. With hierarchies, where no alternative supplier is available, the problem is irrelevant. Since providers are salaried – i.e. reimbursement is non-contingent – there is also no incentive for cost-exaggeration in hierarchies.

All information problems will depend closely on the degree of complexity of the transaction. The more complex, the more opportunity there is for private information to be exploited. Frequency of transaction is also very relevant. Frequent interactions allow for reputation effects to develop, potentially mitigating adverse information problems (see last chapter).

#### 3.5.1.3 Transaction benefits

Adjusted production costs – i.e. transaction *benefits* – (3-2) are a function:

$$(3\text{-}4) \quad \overline{C}^K\left(y^{aK}, y^{eK}, \beta^K, \omega^K, \eta\right)$$

where $y^{aK}$ is investment, $y^{eK}$ is effort, $\beta$ is indicative of mis-representation, $\omega$ of shirking and $\eta$ is allocation of risk. Suppose that the first two terms have negative differentials, the latter two, positive differentials. We can then anticipate the propositions that are made in chapter 5. Based on the above discussion about likely investment, effort, risk etc., we cannot unambiguously say whether $\overline{C}^M < \overline{C}^H$ although this seems likely. In particular, the effort effect should overwhelm the other effects in social care. A range of mediating factors, explored in chapter 5, will affect the exact nature of this relationship; the competitiveness of markets is particularly important.

#### 3.5.1.4 Transaction costs

Summarising the above, there are four main sources of transaction cost: investment contracts, securing effort (by either direct contract or product contract), monitoring for information asymmetries and addressing risk:

$$(3\text{-}5) \quad G^K\left(y^{aK}, y^{eK}, \left\{\beta^K, \omega^K\right\}, \eta\right)$$

Types of transaction cost include measurement, bargaining and monitoring. Again, there is *a priori* ambiguity about the likely size of transaction costs under markets and hierarchies. However, the relative costs of determining and using product contracts in markets to induce effort, rather than using 'instructions' in hierarchies, is argued to be the telling source of difference as discussed above. Hence the proposition is: $G^M > G^H$.

The two propositions are detailed in chapter 5 and then tested in chapter 6.

### 3.5.2 Contract contingency

There are significant interdependencies between governance dimensions (see Table 3-1) With regard to contract design for example incentive type is strongly correlated with ownership. However, contract contingency varies *within* the markets archetype and so this correlation problem should be avoided. We could also consider contract timing but contracts are almost exclusively prospective in markets. Contract specification does vary significantly in social care 'market' arrangements, but specification is difficult to measure and categorise empirically. The most relevant theory to develop is therefore regarding contingency.

#### 3.5.2.1 Transaction benefits

The literature offers a solid body of theory on provider reimbursement in a procurement or 'out-sourcing' context (Laffont and Tirole, 1993). Different information-related opportunism, risk adjustment and the impact of cost-shocks are the main implications of choices about contract contingency. The theory is developed in chapter 7 and here we rehearse only the main arguments. Suppose that in purchasing care from a provider, the local authority purchaser either agrees a contract price that broadly reflects the cost of care for a person with given needs characteristics/case-mix – denoted $\beta$ – or one that does not. The degree of contingency of the former contingent contract is important. To benchmark, the degree of contingency can be measured relative to the impact of β on actual marginal cost: $p_{\hat{\beta}} = \frac{1}{h} c_{\beta}$, where $h$ is the degree of contingency parameter. The problem is that the purchaser does not known $c_{\beta}$ with certainty, which we can model more simply as the purchaser having only a noisy estimate of β. The purchaser assesses the user and so has an estimate of the referred case-mix $\beta^{I}$, but does not know β precisely, nor the size of $c(\beta^{I})$. This situation potentially leads to opportunistic behaviour on the part of so-inclined providers. Under a contingent contract – with a sufficient degree of contingency – they might overstate reports of β to secure a higher price (i.e. cost-exaggeration), so that reported case-mix is higher than actual case-mix: $\hat{\beta} > \beta = \beta^{I}$. Under a non-contingent contract, the provider might deliberately select users with lower than average case-mix (i.e. cream-skimmig) but claim that they are not being selective i.e. $\hat{\beta} = \beta^{I} > \beta$. In both cases $\hat{\beta} > \beta$, and since payment is dependent on $\hat{\beta}$ and costs on β, such behaviour generates an information rent. There are limitations to this behaviour; the greater the difference: $\theta = \hat{\beta} - \beta$, the more likely this opportunism is to being detected and suitable punitive action taken by the provider (e.g. terminating the contract and switching to a new provider). Also, in practice the cream-skimming strategy may be restricted because providers face limited demand – they will not be able to attract a sufficient number of low cost users and would rather take a high cost user than risk a 'void' i.e. an unfilled place. Cream-skimming without *free-replacement* will be more limited than cost-exaggeration.

The main problem with non-contingent contracts is that they foist case-mix risk onto providers. This can be a particular problem when providers are insufficiently large to spread risk. The adverse consequence is either overpaying for all users (except those with the very highest need) or risking poor outcomes for users. The latter may manifest in the longer-run with 'unlucky' but perfectly good providers becoming unsustainable, with some form of quality shirking for high dependency users, or more generally a two-tier system where it is difficult to place high cost users.

The final issue relates to systematic (unforeseen) cost shocks that affect all providers. Over time, the average dependency of users in care homes has been increasing (Forder and Netten, 2000a; Darton, Netten, and Forder, 2003). If prices are measured at a time when population dependency is higher than originally anticipated, the average price charged by providers with contingent contracts will be higher than those with non-contingent contracts.

Overall, in the short-run the production cost to the purchaser (i.e. price) is likely to be higher with contingent contracts, assuming that risk adversity is not overwhelmingly large. In turn, this implies that transaction benefits are higher with non-contingent contracts. The timeframe in question is however important. Providers with non-contingent contracts exposed to increasing risk and rising dependency can for a while absorb these cost pressures, but not indefinitely. If the use of non-contingent contracts means that providers leave and purchasers are very slow to respond in increasing prices, then the long-run transaction benefits of contingent contracts could be higher, despite cost exaggeration, and where prices for non-contingent contracts are actually lower. Furthermore, providers with contingent contracts are likely to have higher dependency users (because they do not cream-skim), a valued outcome for purchasers. Contingent contracts therefore have lower cost to outcome efficiency (i.e. high costs), but also outcomes of higher value to purchasers.

#### 3.5.2.2 Transaction costs

In terms of transaction costs, we would anticipate that contingent contracts are more expensive to determine. It is clearly simpler (in the short-term at least) for purchasers to operate with a single fixed (i.e. non-contingent) price. In chapters 7 – 9 we develop specific hypotheses along these lines and test them empirically.

## 3.6 Conclusion

Given the complexities of social care systems and so the substantial opportunties for specialisation, policy-makers are faced with a great many choices regarding appropriate arrangements for governance. These choices range from strategic issues about ownership and control of the apparatus of purchasing and provision, through choices about contracts and incentive structures, to ways of monitoring and regulating to ensure standards. Governance arrangements *facilitate* the production of care. Good governance gets the right services produced and delivered to the right people at the right time, and at the best cost. But governance activities also divert resources and so have a cost.

This chapter has drawn on economics of organisation (EO) theory to lay out the theoretical concepts with which to tackle comparative efficiency questions relating to the organisation of social care systems. The chapter introduces the main building blocks of such an analysis, including the transaction as unit of analysis, the definition of contracts, the principal-agent framework and the notion of governance structures and governance archetypes. These concepts frame the empirical strategy of this work: to ascertain comparative transaction benefits and costs of the different organisational arrangements we see in the social care system. In particular, the chapter defines and distinguishes market and hierarchical governance structure archetypes as they apply in a social care context. It also identifies the contract design dimension, which covers reimbursement incentives, price contingency, contract duration and so forth.

As a precursor to the chapters that follow this one, EO theory was used to provide an overview of the main propositions to the tested. These relate, first, to the choice between market and hierarchical governance and second to the choice between contingent and non-contingent contracts. As to the former, the basic contention is that whilst production will be more efficient in markets, the costs of writing product contracts in particular, will mean high transaction costs. In hierarchical forms of organisation, employees tend to cede decisions to managers – which has risks for productivity – but results in lower transaction costs. Markets have greater relative net benefits when there is potential for high competition, investments do not tie providers to specific purchasers, and complexity and uncertainty are relatively low.

As to the latter choice, contingent contracts are clearly more (transaction) costly to determine that non-contingent versions. Their relative transaction benefits hinge on the value of a better handling of risk set against the greater potential for cost exaggeration they embody.

# Chapter 4. Governance: an organisational economics interpretation

## 4.1 Introduction

In the previous chapter we outlined concepts relating to governance arrangements and choices. In this chapter we discuss the historical, institutional and economic character of the social care system in England. This discussion is framed in the language and concepts of chapter 3. The aim is to translate those concepts into descriptions that are meaningful in social care. In turn, in subsequent chapters the implications of an economics of organisation analysis can be couched in terms that are pertinent to social care policy. In addition, the analysis of its historical, institutional and economic character will suggest features of social care in England that are special or particular with regard to the general comparative governance concepts in chapter 3. For example, the public funding, political influence at a strategic level, and the significant brokerage function in social care at a micro-level will affect the inferences we are able to draw from a general deployment of the theory. Finally, this chapter will feed into the identification of relevant empirical indicators in social care that can be used to test the propositions developed in this thesis.

The chapter is structured as follows. After this introduction, in section 2 a brief (recent) history of social care in England is provided. The main intention is to describe key historical events which led us to the social care system today, and in particular referring to the previous chapter, how we arrived at the current set of governance arrangements. Inevitably this task will entail some positive analysis as to why certain choices were made. Indeed, it becomes clear that whilst economic efficiency has been a priority it is certainly not the only – or even the main – objective that policy makers have been pursuing in shaping the organisation of social care. This conclusion is particularly important because, even if decision-makers were rational, it would imply that the majority of organisation choices in social care might not be the most efficient.

In section 3, our attention narrows onto the organisational and economic character of social care. In particular, we present a description of the key stakeholders, sources of funding, service and other activity, regulation and so forth. Section 4 provides the empirical description. In particular, we look at essentially distinguishing characteristics of markets versus hierarchies, namely, ownership and control. Then we turn to a description of reimbursement, contracts, and contract contingency in social care in England.

## 4.2 History and policy context

The social care system in England inherited by the Labour government in 1997 reflects the legacy of at least two major sets of reforms/re-organisations. The first was the local government re-organisation of 1970 and the establishment of unified social services departments (SSDs) on 1 April 1971, both a result of the Seebohm Report in 1968 and the 1970 Local authority Act. The second was the series of reports and subsequent legislation in the late 1980s and early 1990s, including in particular, the Audit Commission report of 1986, the Griffiths report of 1988 and culminating in the NHS and Community Care Act 1990.

### 4.2.1 Seebohm

During the period from 1948 – when health and national assistance functions were taken from local authorities – to 1971 when social services departments were established, local authorities operated services in what many considered to be a fragmented and overlapping fashion (Timmins, 2001). The composition of statutory services we know currently as the personal social services did not exist, even conceptually, prior to the mid 1960s. A diverse range of other non-government organisations also provided services. Indeed, many of those services that became statutory responsibilities were originally provided by voluntary organisations (Kendall and Knapp, 1996). From their inception, SSDs combined the previous local authority children's departments, the welfare department and, in particular for mental health services, functions of the health department. They were to be headed by directors of social services who held statutory responsibilities on behalf of the Secretary of State. The set of service responsibilities included residential care, home care, meals, day centres, aids and adaptations, social work/counselling, and child protection. These *unified* SSDs were vested with complete responsibility for the needs of their areas, covering not just statutory provision, but also needs going beyond it. SSDs would address not only the specific needs of individuals and families but would also work at the level of the local community.

Social services were therefore to become 'universalistic' if not universal – anyone with need could access the system – and generic in terms of their professional outlook. The Seebohm report was a grand and comprehensive vision, but it came apart when funding tightened significantly in 1975. Other tensions were also to arise. Indeed, in keeping with its Poor Law origins of the 16th Century and essentially unaltered by the 1948 National Assistance Act, Social Services Departments presided over the provision of residential social services, in particular, that were orientated to the poor, having a welfare focus, and being means-tested.[9] Nonetheless, the Seebohm report did mark a significant leap forward for social care, in terms of its intention to comprehensively address the needs of the poor, the deprived and the distressed and in its re-organisation of social services.

Despite means-testing of people going into local authority homes, the significant increase in scope, availability and demand for social services, coupled with the tight controls over public expenditure, significantly stretched council resources. Pressure built up for alternative funding, particularly funds from central government. As a result, in 1979 an important change was made in the regulations governing social security so that payments – supplementary benefit for those on low incomes – became widespread in meeting the costs of residential care.

Supplementary benefit was eligible income to be assessed against the means-test for local authority homes. For those on supplementary benefit, their income would be brought up to the equivalent of the retirement pension and they would be charged this amount minus a modest personal allowance, and the local authority would make up the difference in the cost of care. For example, in 1991/92 the basic retirement pension was £52.00 per week and the personal allowance was £10.40. The resident would have to

[9] The Poor Laws provided general assistance to the poor in Britain. From the 16th century parishes were responsible for providing for their poor and levied a local rate to do this. The National Assistance Act 1948 separated responsibility for social security which became a Government responsibility and welfare – e.g. residential care - which remained with local councils.

pay the balance (£41.60) from their income, with the council making up the remaining total cost.

Supplementary benefit was also available to meet the costs of people moving into independent sector homes, a policy that has great significance for the mix of governance arrangements we see in social care today. Whilst the availability of this benefit was means-tested – for example in 1983, anyone with less than £3,000 in capital was eligible to apply for supplementary benefit to meet the costs of independent residential or nursing home care – no assessment of need was required in order to qualify for benefits. Moreover, the supplementary benefit system was part of the non-cash limited social security system. Consequently, with local authorities' own budgets cash-limited and despite the availability of some supplementary benefit for LA homes, the availability of these payments fuelled a major expansion of the independent residential and nursing home care sector. The effect on the numbers of placements in the independent sector is detailed below. But in terms of spending, in 1979 supplementary benefit to people in independent care homes cost £12m (12,000 claimants). By 1992, this cost had risen to over £2,500m (Laing, 1993). This arrangement for publicly supported residential care was essentially a non-needs-assessed voucher system. In other words, a system of market governance with individual commissioners, supported where relevant (according to the income test) by a voucher.

Before 1983, the amount of the supplementary benefit varied according to the price of the independent sector home chosen. In a modest attempt at rationalisation, a number of national price ceilings were imposed in 1983. These ceilings soon became the *de facto* price of state-supported independent residential care. A different ceiling price was set for London and non-London authorities and for each of the main clients groups (older people, people with mental health problems, learning difficulties etc.). For the older people client group (only) a price distinction was also made between 'standard' and especially frail people i.e. the price had a limited degree of contingency with respect to resident dependency.

### 4.2.2 Community care reforms

The start of the second major set of reforms, the effects of which reach us today, are marked by the House of Commons Social Services Select Committee report *Community*

*Care* in 1985, and shortly afterwards by the 1986 Audit Commission report, *Making a Reality of Community Care*. These two reports were highly critical of the status quo. There were many criticisms[10]:

- That the supplementary benefit/income support payments for independent sector residential care created perverse incentives to over use residential care in favour alternatives such as domiciliary care
- Organisational fragmentation and poor managerial responsibility
- Poor workforce planning
- Poor management, organisation and financing of the transfer of resources from the NHS to community care, including a lack of bridging money.
- A central government grant system that penalised local authorities in attempting to support additional spending using local tax (rates) revenue.
- At worst, a claim that social services were leaving many vulnerable people 'without care and at serious personal risk'

One of the key recommendations of the Audit Commission report was that the community care system needed urgent review. The then Secretary of State for Social Services, Norman Fowler, commissioned Sir Roy Griffiths, by that time the Prime Minister's special advisor on health care management, to conduct the review that was to be published in 1988. It concentrated on: the perverse incentive of the supplementary benefit arrangements – funding residential, but not domiciliary care – along with the lack of assessment of need; poor management and the supply-led nature of services.

There were three main recommendations that have a strong bearing on the shape of the system we have today. First, addressing the question of who should run the system. The need to consolidate the system was clear, but a number of choices were available. Broadly, the system could be run by: the NHS, by local authorities or by a new body. Griffiths choose to keep social care services with local authorities, reflecting a belief that the 'social model', not the medical model, was the most relevant. A main element of this consolidation was that the budget for (new) supplementary benefits/income support claimants should be transferred to local authorities, as a partially ring-fenced grant, but ultimately to form part of the council's overall (cash-limited) budget.

---

[10] See Wistow et al. (1994) and Timmins (2001)

Second, Griffiths recommended a system focused on the needs and preferences of users, rather than of the system. This would involve, principally, user assessment, care management, care planning and review. The intention was to put the users much more at the centre of the process, to allow systematic assessment and care planning around those needs, in a far more coherent and holistic fashion. Where people with more complex needs were involved, a care manager would act as commissioner to determine a package of care. These recommendations were adopted and care managers are now an everyday feature of social services. However, their working has in practice departed somewhat from the original conception. Today's care managers do act as commissioners for users, but they have large caseloads and more often act in the spirit of 'gatekeepers', rationing against eligibility criteria set by the local authority, rather than brokers acting for users. Nonetheless, care managers are, in this role, commissioners and not providers of services. They determine a person's care requirements and the budgetary implications, and establish *ex ante* demand for those services, independently of supply considerations.[11]

An emphasis on commissioning leads to the third set of recommendations. With its first set of recommendations, the Griffiths report was effectively a vote of confidence in the ability of local authorities to manage the system. But Griffiths was not, at the same time, giving a vote of confidence to the traditional bureaucratic/hierarchical model of local authority provision of services. Instead, in reflecting a theme that was developing across the Thatcherite welfare state reform agenda, the report argued for the establishment of a *mixed economy of care* and the concept of the *enabling* authority. This concept of enabling is particularly relevant to the present analysis. There are a number of interpretations of 'enabling' as applied to social care. A relatively early conception, essentially stemming from Seebohm, was of social services departments mobilising care resources from the community and voluntary sector, and also from informal carers. However, the concept used by Griffiths and adopted centrally in the White Paper *Caring for People* that was to follow, was concerned with the processes used to deliver services

[11] When this arrangement also involves separate, voluntary bilateral relationships between commissioning care managers (or at least the middle managers with oversight of individual care managers) and providers, then together this characterises (quasi-) market governance. As an alternative, after initial assessment, the care manager directly refers the client to the provider, e.g. the home care service, and it is the provider who works out the details of the care plan. This arrangement falls in the hierarchy category.

rather than the scope and development of care in the community. In particular, enabling as used at the time was synonymous with the use of market mechanisms to delivery care even when the purchaser, funder and decider of required services was a public body (Wistow et al., 1994; Lewis and Glennerster, 1996; Timmins, 2001; Means, Morbey, and Smith, 2002).

This definition of enabling, as noted, resonated with the broad brush of *new public management* that pervaded health, education as well as social services. A crystallising force was the Secretary of State for the Environment, Nicholas Ridley. His publication of *The Local Right: enabling not providing* (Ridley, 1988), argued that efficiency and effectiveness benefits would be forthcoming for councils who contract out service provision to potentially competing providers. New public management (NPM) retains universal public funding and ownership, and therefore is distinct from welfare reforms involving wholesale privatisation (with the State in a residual, funder-of-last resort, role). NPM does, nonetheless, embrace the main idea of privatisation, which is the erosion of previous monopoly or near-monopoly provision, but with the public sector in a funding and enabling role. Examples in health care run from the adoption of private sector management techniques to full scale re-organisation and merger activity (Fulop et al., 2005).

NPM was consistent with thinking from the US most readily exemplified by the work of Osborne and Gaebler (1992) and their well-know designation of the State adopting a "steering" role – i.e. arranging, designing and purchasing services – rather than "rowing" i.e. providing. By having, on the one hand, to manage such 'contractualisation' and, on the other hand, to compete with independent providers for contracted out services, the public sector would, it was argued, be forced to become more market-oriented and more enterprising. It would, in the words of Osborne and Gaebler, make governments more "entrepreneurial". These concepts resonate very closely with the governance concepts of chapter two. Movement from a rowing to steering role is consistent with a change from public hierarchy to (quasi-) *market*.

Historically speaking, 'markets' have been a feature of the social care system to a much greater extent than has been the case for other public services. This situation reflects to a large degree the role of voluntary organisations in the social services (Kendall and

Knapp, 1996). Indeed, the early 'social services' were mainly voluntary sector organised and funded. Only later did local authorities assume a more dominant position. Although it is a stretch to describe the early voluntary sector arrangements as market governance as we know it now – indeed those arrangements might be more likened to *private* hierarchies – they did entrench the concept of non-government, voluntary, and diversely owned provision. Having a large and viable non-government provider sector during a period when funding responsibilities were shifted to public sources, accommodated a substantial mixing of governance arrangements. And unlike other public services such as health care, the public sector assumption of responsibility for social care was not complete.

The White Paper, *Caring for People*, which set out the main policy groundwork to be legislated subsequently as the 1990 NHS and Community Care Act, endorsed the vision laid out by Griffiths. In particular, it included "the separation, to at least some degree, of purchasing and providing functions within social services departments; the development and support of increased levels of activity by private and voluntary providers; and the regulation of provider agencies in all sectors ... through procedures of service specification and contracting" (Wistow et al., 1994, p. 21). Also endorsed were the dual functions of needs assessment and care management.[12] The Act was to be implemented from 1 April 1993, although the funding transition from central social security budgets to local councils was to be phased over 3 years.

Although the Department of Health provided guidance as to a range of options for a purchaser – provider separation, no particular model was imposed. Moreover, the Government was explicit about not forcing a rapid change of pace. In practice, as will be discussed in more detail below, local councils made varying interpretations of how far to take this separation, some of which reflected local political preferences.

[12] Subsequent policy and practice guidance clarified, for implementation, the concept of a purchaser-provider separation ((Department of Health, 1990, 1991)). This guidance also laid out a number of additional policy components for social services departments to operate in the enabling role. First, the *devolution of budgets* to staff responsible for assessment and design of care packages for individuals in need (a shift away from historical service-led budget allocations). Second, the use of *service specifications* to describe the context, objectives, quality, inputs, process, outputs, outcomes and monitoring of services. Third, the use of contracts to indicate prices, terms and conditions and specifications of the service to be provided. The service specification was to be a statement about how a social services department wished a particular service to be delivered. It constituted the starting point for the determination of the actual contract specification, which would be finalised through negotiation with the provider. Another component in the guidance was the monitoring of contracts for compliance.

Nonetheless, the degree to which councils were adopting internal market-like arrangements pre-dated this decision, at least for residential care. After all, many councils had, pragmatically or otherwise, seen the benefits of shifting cost burdens onto the social security system for residential care, although this of course meant the use of the independent – and mainly private, for-profit – sector. Some councils had also been contracting out using their own funding prior to the widespread use of social security funding. By the time that councils took full responsibility for commissioning publicly funded residential care after 1993, the ratio of public to private provision varied substantially across the country. Some councils went further after 1993 and became, if anything, only residual providers of residential care. Others retained a majority of in-house provision, and in the years to follow the 1993 implementation date had not shifted substantially from the traditional public bureaucracy model.

### 4.2.3 The New Labour approach

By the time Labour came to power in 1997, the local authority-commissioned market was well established. The 1998 White Paper, *Modernising Social Services*, stressed individual independence (including greater user influence), more effective prevention and rehabilitation, and a broadening of the supply-side regulation of services. The Labour government were not wedded to public sector provision, with the then Secretary of State, Alan Milburn stating: 'it is no longer who provides the social care that matters. It is the quality of care that counts'. But neither did that mean an unfettered embracing of markets. Instead, consistent with the developing agenda on public sector reform, the virtues of a 'third way' were extolled (Blair, 1998; Giddens, 1998). With reference to chapter 3, the third way can be likened to a network governance structure emphasising long-term partnership arrangements between commissioners built on trust (Osborne, 1997; Rhodes, 1995). Adopting a "third way" of running the NHS was described, at the time, by the Prime Minister as "a turning point for the NHS. It replaces the internal market with integrated care". It was said to constitute an explicit rejection both of the "old centralised command and control systems of the 1970s" and of the "divisive internal markets systems of the 1990s". This, however, does not represent the wholesale replacement of one governing structure by another (as it certainly turned out in the NHS).

Regarding social care, the Department of Health published an Agreement, *Building Capacity and Partnership in Care* in 2001, which laid out the aim to provide "a framework for future working relationships between providers and commissioners locally, geared to delivering the services that people need and expect" (p. 5). In particular, "to promote the establishment of close and harmonious working relationships, good communication, and to foster constructive co-operation between all parties involved in providing care and support services for adults." (p. 5/6), and "to establish a way of working that: ... promotes mutual trust" (p. 6).

As argued in the last chapter, differentiating 'networks' from markets is problematic. Markets of almost any type require at least minimum amounts of trust. It is therefore difficult to see whether this new policy direction produced a systematically different set of behaviours. Whilst it is the case that the majority of large care home providers did sign up to the Agreement, macro-level behaviour at least did not appear to change materially. Indeed, concerns about the level of prices in the market, and about the impact of regulations following from the Care Standards Act 2000, were still significant among providers (Laing, 2004). Market analysts Laing and Buisson indicate that, although slowing, the rate of home closures was still significant after 2001 (Laing and Buisson, 2003).

## 4.3 Where does this leave us?

These policy developments have left what might be described as a somewhat eclectic system. It still has significant origins in the Poor Law and its successor, the 1948 National Assistance Act, in that local authorities have a responsibility to provide accommodation and services to people who are "substantially and permanently handicapped by illness, injury or congenital deformity" subject to a means-test. The reforms of the 1970s consolidated the social services responsibility of councils and constituted an attempt to make a universalistic service i.e. potentially open to anyone to use (but still means-tested).

As regards *residential care* for older people specifically, the establishment of Social Services Departments in itself did not much change the nature or level of provision. Departments continued to provide residential care under part III of the 1948 Act. The loosening of social security rules to allow funding of residential care placements did represent a significant change however. For a start, the (publicly funded) demand side now included not only local authorities (and the NHS) but also individuals effectively purchasing their own care and drawing on social security funding. Furthermore, on the supply side, these arrangements really presented an opportunity for expansion of private sector provision.

Whilst local authorities have regained the position of being practically the sole demanders of publicly funded care – the legacy of the 1993 changes – the role for independent sector as suppliers is now fully established. Indeed, independent sector care home places constitute the largest share. In effect, therefore, many councils from 1993 inherited voucher forms of quasi-markets (Le Grand, 1991). With their own provision significantly run down and with a responsive independent sector, council purchasers had no (economic) choice but to contract out a significant amount of care to the independent sector. It has meant that we have running in parallel an in-house (hierarchical) system and a (quasi-) market system, both of which, as yet, are far greater than being 'residual' special cases. Moreover, both are funded from the same source – local authority budgets – which removes a key 'degree of freedom' from any comparison of governance arrangements. So whilst there was a flourishing private market in the 1990s, this was substantially funded by an entirely separate social security system, which operated on different principles – a different means test, no real assessment and test for need, a multitude of buyers – greatly complicating any comparison. For the purposes of this thesis, the comparative governance structure analysis outlined in the last chapter will be between council-funded contracting out – the 'market' governance structure – and council-funded in-house provision – the 'hierarchy' governance structure.

A further fortuitous characteristic of the present eclectic system is the wide variation between the 150 local authorities with social services responsibilities in terms of the balance between quasi-market and in-house arrangements. As noted above, although there is perhaps some correlation between the political flavour of councils and their choices regarding this balance, there are also some significant legacy affects. Moreover, the decision criteria – although beyond the scope of this thesis – are far from dominated by ideology. For example, Hampshire county council have just embarked on a substantial programme to increase their in-house provision.

Where markets are in use, one of the consequences of the NPM approach has been an increased awareness of the role and consequences of choices about contractual relationships between councils and the independent sector market. The Department of Social Security (DSS) voucher market, involving relationships between individuals and their choice of home, operated with what amounts to fixed price spot contracts, otherwise known as call-off contracts (see below for more detail). Although the system is now different, this 'voucher' system had, and still continues to have, an important legacy. Many authorities continue to purchase care on a call-off basis with a limited tariff of prices (as is outlined in more detail below). In some cases, for older people's placements, councils simply took the standard DSS rate as the price they would pay. Some councils retained the DSS 'enhanced' rate distinction, applying it specifically in cases where, in addition to physical dependency, residents also had significant mental health problems (generally beyond baseline levels of depression and/or dementia). Few councils chose to operate with a variable price determined, not in advance for a categor of resident, but set according to the needs of the particular person to be placed (Wistow et al., 1996).

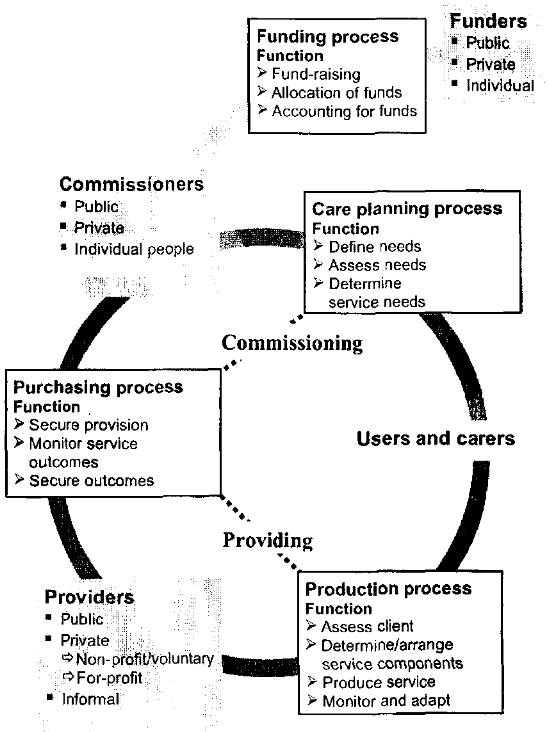
In general then, after 1993 contract contingency was limited in scope and categorical in nature rather than continuous. Moreover, there were few categories – often just two – and then quite widely spaced. As a result only homes that were prepared to cater for a relatively wide range of needs of residents actually experienced any contingency at all (Netten et al., 1998). Homes that catered for frail older people without significant cognitive impairment often worked with a single non-contingent price.

After 1993 the spot or call-off contract was the normal model, and largely continues to be today (more detail can be found below). Put another way, block contracts – where capacity is purchased in advance and paid for whether in use or not at a given time – were rare (Netten et al., 1998). The exception was in relation to council homes that 'floated off' in the immediate period before 1993. Very briefly, in 1991 and 1992 a window of opportunity existed whereby residents in a council home that transferred int the independent sector would become eligible for additional social security support to help meet costs.[13] Councils that floated off homes – often into 'trusts' with the local

[13] See Wistow et al., 1994 chapter 6 for a full and excellent account.

management fully retained – could benefit from a sizeable cost shunt to non social service budgets. Subsequently, and contentiously from the point of view of the wider independent sector, these trusts often attracted block contracts from the council. Furthermore, in the years to follow, more commissioners and providers came to see the general benefits of block contract arrangements. Nonetheless, the proportion of total independent sector places purchased under block contract remains small as described below.

**Figure 4-1. Social care stakeholders and processes**




## 4.4 Principals and agents in social care

Having briefly outlined the main historical developments in the care market, we can now look at how the system is currently configured in more detail. Figure 4-1 identifies four key stakeholders and shows the processes that link them. It is a very stylised picture but does frame some of the descriptive analysis reported below.

A typical example of the social care process would be of an older person who has had a short-term crisis – perhaps a fall – and decides, with others, that they can no longer manage at home. The local council Social Services Department is funded by central and local taxation to provide services or support for people who cannot afford services themselves. The council will act as the main commissioner. First, a social worker/care manager undertakes an assessment and develops a care plan with the person and their family. Second, the council procures a placement in the care home of the individual's choice (broadly speaking). In this example, the provider is likely to be an independent sector care home.

### 4.4.1 Users, need and demand

Formal services are provided to *users and carers*. In theory, social services in England are 'universalistic' as noted above in that anyone can approach their council for help with social care needs, although most people will not need such help in their lifetime (or necessarily qualify for council financial support). Councils then have a statutory duty to assess the individual's needs and circumstances. Then, subject to financial and other constraints and in accordance with a means-test for user contributions, to make available services to meet the individual's needs (or to make a direct payment). Table 4-1, reproduced from table 2 of Comas-Herrera et al. (2003), describes the prevalence of functional dependency in the UK. It reports the percentage of people that are unable to manage *activities of daily living* (ADLs), and instrumental ADLs (IADLs). The former include activities such as bathing/showering, washing, dressing, feeding, and toileting. The latter include shopping, laundry, cleaning, cooking etc.

Population dependency levels are good indicators of need. In particular, people with difficulties relating to two or more ADLs would benefit from social care, including residential care. The demand for services is affected by many more factors including, importantly, available funding. The data show a positive relationship between age and greater functional dependency. For example over a quarter of all men and two-fifths of women aged 85 or more experience difficulties with 2 or more ADLs. There is currently an important debate – beyond the current remit – about how need changes with the ageing of the population, and whether morbidity will become 'compressed'.

**Table 4-1. Estimated percentage of the older population of the UK with different levels of functional dependency, 2000.**

| | None | IADL | 1 ADL | 2+ADL |
|---|---|---|---|---|
| Males | | | | |
| 65-69 | 85 | 4 | 4 | 7 |
| 70-74 | 83 | 4 | 6 | 7 |
| 75-79 | 73 | 9 | 7 | 11 |
| 80-84 | 55 | 13 | 16 | 17 |
| 85+ | 45 | 12 | 15 | 28 |
| Females | | | | |
| 65-69 | 82 | 4 | 5 | 9 |
| 70-74 | 75 | 6 | 8 | 11 |
| 75-79 | 62 | 6 | 15 | 17 |
| 80-84 | 45 | 12 | 17 | 26 |
| 85+ | 25 | 17 | 15 | 43 |
| All | 67 | 8 | 10 | 15 |

### 4.4.2 Commissioning and funding

The main *commissioners* in the care market are councils, individuals themselves (and their families), the NHS and to a minor degree charitable and corporate organisations. There are 150 councils with social services responsibilities (CSSRs), ranging in size from small London Boroughs to large Shire counties, and they are the main commissioners. Many councils have a distinct Social Services Department (SSDs), although some combine housing and social services and others have delegated some social care responsibilities to their partner PCT (or care trust). Some are single tier like the London Boroughs and unitary authorities, some are two-tier having both county and district administrative levels. Some SSDs organise themselves on a client group basis, some on a geographical basis. In any case children's services are now separate. Although the exact arrangements vary, a subset of council members have a specific responsibility for social services and oversee the work of the director of social services (under cabinet style arrangements there is usually a single lead member for social services). Again local arrangements vary – and more detail is provided below in section 4.6.1 – but most authorities operate with two broad types of commissioner. First, a contracts unit that secures services via contracts at a strategic level. Second, care managers who work directly with service users calling-off services as required.

Commissioners principally undertake two functions – the assessment and care-planning function and the purchasing function. They are also involved with funding, at least in the allocation of funds. *Assessment and care-planning* is the process of determining and understanding an individual's needs and the risks they create, and translating these into a plan about what services could be used to ameliorate the risks. Councils are required in law to undertake an assessment of needs. Currently the framework laid out in *Fair Access to Care* (Department of Health, 2002) provides a range of domains of dependency and risk that can be used to determine care needs. Once need is established commissioners can work with the person to decide what services would best address those needs. Care planning takes financial considerations into account, both the assessed personal resources of the care recipient and also the relevant budgets of the commissioning organisation (various legal test cases have established that councils are able to take account of their financial situation when determining service eligibility – see *LA social services letter* LASSL(1997)13). Once a care plan is established commissioners need to 'purchase' the required services.

*Purchasing* will differ according to the governance arrangements that are in operation, but essentially involves a specification of requirements, a tender for providers (in a market arrangement), agreeing a contract, monitoring for delivery and compliance, invoicing and payment, and finally any revisions to arrangements. It is not concerned with the technology and operation of care production, only the characteristics of services, the implications for users and the costs of those services (Wistow et al., 1996). The specification of requirements – in effect a determination of demand – involves definition and mapping of need, and decisions about what constitutes eligible needs and therefore the relationship between need characteristics and service demand. A stock-take of current resource availability both in terms of current capacity/service cost and budgetary constraints is also part of the purchasing task. Together this information can be used to determine the extent and type of services to be purchased, or procured (Audit Commission, 1997a).

*Funders* may or may not be distinct from commissioners. Where public support is available, a layer of public administration is required to determine funding levels in total and to allocate these funds to commissioners. For care that is privately funded – e.g. in entirety or as user charges – the individual is the funder.

#### 4.4.2.1 Public funds

As of 2004/5 Government in England funds about £10.5 billion of adult social care from the public purse. This funding level is determined by Department of Health and HM Treasury *Spending Reviews*, comprising analysis of baseline pressures and new policy requirements. Spending reviews work on a 2/3-year cycle. The last was in 2004 and covered spending for 2005/06, 06/07 and 07/08, although the last year of review period can be modified by the subsequent spending review (due in 2006).

Almost all public spending on social care is now made via local councils (with a small amount of long-term care spending by the NHS – see below). Most is allocated in the Revenue Support Grant to councils with a proportion also made up from council tax.[14]

#### 4.4.2.2 Charges

Councils also raise money from charges to people using services. This will raise total adult social care spending to over £12.5 billion by the end of the 2004/5. In the case of care homes, for people with very significant medical needs, NHS continuing care is available. Recipients of NHS continuing care pay no fees (although they do lose or have reduced main state benefits – e.g. pension – after a short period). Otherwise, for people with social care needs there are a number of options that apply in the long-term. People with assets in excess of £20,000 are required to pay the full price of care themselves.[15] They are effectively private purchasers. At a minimum they will be paying £300 per week, probably much more.

[14] During the spending review process the Government works out the total required spending for each service the council provides (not just social care) and adds this up to give total council spending. A calculation is also made as to how much as council is expected to raise through local council tax and through business rates. Government grant – the Revenue Support Grant – to councils is the difference between total council spending and expected revenue through local taxation. The RSG presently accounts for about two-thirds of total spending. However, it is then up to councils, not central government, as to exactly how to divide their funds between services. A small proportion – only about 2% of total social care spending – is in the form of specific 'ring-fenced' grants that must be spent as dictated by the central government. As it happens, the total amount of public money on adult social care that councils have been spending recently is very nearly the same as the government's suggested figure. But in principle, if central government changed the amount it contributes on paper, councils need not actually make the same change in practice (although this would clearly create local resource pressures).

[15] This and associated rates are for 2004/5 but are normally uprated on an annual basis.

In all other cases where people have capital of less than £20,000 the council will make a contribution and purchase on behalf of users. Indeed in these cases the council can limit the total price it will pay for care.[16]

Essentially therefore, support for care homes is a safety net system. People pay for their own care entirely until those costs impoverish them. Thereafter the State pays the excess of cost over the person's income. The average gross cost of an older person looked after in residential and nursing home care to the English councils in 2002/3 was £360 per week. The average client contribution was £125 per week of this total. It is worth noting that many older people rely on the state pension system to provide them with income, some of which is then used to pay charges. Until recently (Oct 2003), people moving into independent sector homes could claim Residential Allowance (RA) of about £60 a week to use towards their charges. The RA was not available for people moving into the local authorities own homes.[17]

Local authorities – as commissioners and funders, not providers – have their overall performance assessed by the Commission for Social Care Inspection (CSCI). There are currently six broad domains in the assessment criteria, addressing council processes and outcomes for users and their families (see Social Services Inspectorate, 2004). Councils are star-rated as a result of the assessment process, and a number of improvement incentives and sanctions are then applied according to the rating a council receives.

### 4.4.3 Provision of services

*Providers* are organisations that produce the agreed services for the user. Providers can of course be distinguished according to the services they provide, for example, home care or

---

[16] Where people have capital of less than £20,000 both their capital and income are taken into account. If capital is less than £12,250 this is ignored as is any income it produces. If capital is between £12,250 and £20,000 the actual income is ignored, but for each £250 of capital an income – called "tariff income" – of £1 per week is assumed. The tariff income is added to the rest of the person's income in order to calculate their care charge. Because total income including tariff income can exceed the value of all actual income a person has coming in, the idea is that the person will have to divest of some capital to secure the extra income required. Eventually, capital will fall to the lower threshold (£12,500) and tariff income will become zero. Individuals with assets below £20,000 are required to contribute *all* their income (plus any tariff income) as fees less an allowance for person income, currently £18.10 per week. The local authority and the NHS pay the rest. The NHS will pay a fixed amount in three bands towards the cost of nursing care needs that residents may have. In the short-term – up to 3 months – the value of a person's home is disregarded in the means test (only their current income and other assets are taken into account), making it easier for people to keep their old home for a while should a person wish to return to the community.

[17] Generally, after a short period, people claiming Attendance Allowance lose their eligibility on entering a care home.

residential care. Indeed *care homes* fall into two groups – those that provide nursing care (by registered nurses) i.e. what were called nursing homes and those without nursing care (providing only personal care) i.e. those that were called residential care homes. They can also be distinguished according to how they are operated and by whom. There are four main provider sectors. First, the public sector that includes mainly local authorities but also others such as the NHS. Second, the voluntary sector which comprises formal organisations independent of government that, although they may earn profits, are bound by a non distribution constraint, which means that they cannot distribute profits to any owners or shareholders. Many voluntary organisations – and most in the social care sectors – have charitable status, conferring certain tax advantages (Knapp and Kendall, 2001; Kendall, 2003; Kendall and Knapp, 1996). Third, the private sector, which is not constrained in is distribution of profits. In recent years, the private sector has become a much more important provider of social services, especially in residential care (see below). Finally, there is the informal sector, principally composed of individual carers (family members and others). Generally, for informal care no contingent payment is made.[18]

With regard to care homes, standards of care are laid out in the provisions of the 1984 Residential Care Act and subsequently in the Care Standards Act 2000. Standards relating to staffing inputs are probably the most important, although there is a range of domains covering individuals' choice, health, personal care, daily activities, complaints, environment and management. Homes need to be registered to receive public funding. New homes must complete a registration process. Once established homes are inspected by CSCI, currently on a bi-annual basis, with respect to the above domains.

#### 4.4.3.1 Providers

There is a considerable diversity of providers in terms of home size, ownership and legal structure. Table 4-2 reports the average number of places per home, by home type. Nursing homes are bigger on average than residential care homes. The PSSRU survey of care homes – a large and comprehensive survey of the industry – provides figures for average home size and range of sizes as was the case in 1996 – see Table 4-3 (Netten et

[18] Our focus here is on formal care services. Nonetheless it would be remiss not to mention at least the contribution of informal care. Informal care remains a very important substitute for formal services (Pickard et al., 2000; Hirst, 2001).

al., 1998). The legal structure of independent sector providers spans commercial for-profit, limited companies, through partnership, sole-proprietors, small community non-profits to National Charitable organisations. Laing and Buisson provide information about *major providers* – any London Stock Exchange company or other organisation with 3 or more homes. Major providers market share of the for-profit home sector was 34.5% in April 2003. Whilst this has grown recently in terms of market share, major provider holdings of bed numbers has not changed in the recent period (Laing and Buisson, 2003). Overall, small businesses continue to dominate the sector. Major not-for-profit providers command a much higher market share at around 65%. Unlike for-profit major providers who focus on nursing homes, not-for-profit major providers tend to operate more in the residential home market.

**Table 4-2. Average places per home, older people, England, 2001, all homes**

| Home type | Homes | Average places per home | Year |
|---|---|---|---|
| Nursing[1] | 2414 | 39 | 2003 |
| Independent residential – excluding small homes[2] | 7008 | 23 | 2001 |
| LA residential | 1116 | 35 | 2001 |

Notes: (1) Laing and Buisson, 2003
(2) Department of Health Statistical bulletin,

**Table 4-3. Size of home by home type (PSSRU survey, 1996)**

| Number of places | Local authority | Private residential | Voluntary residential | Dual registered homes | Nursing homes | All homes |
|---|---|---|---|---|---|---|
| Mean | 35.0 | 20.0 | 30.3 | 39.2 | 37.7 | 27.5 |
| Minimum | 12 | 6 | 8 | 9 | 9 | 6 |
| Maximum | 66 | 56 | 100 | 77 | 180 | 180 |
| N | 167 | 148 | 122 | 76 | 159 | 672 |

## 4.5 Expenditure and activity

We now turn to the question of what this system spends and produces.

### 4.5.1 Expenditure

Local authorities are the main sources of social care expenditure. Local authority social care spending has been increasing significantly in real terms, having more than doubled in the last 10 years (Table 4-4). Total public spending has not increased as rapidly however because in the 3 years from 1993, much of the increase in local authority spending was through an inheriting responsibility for people who would have previously been funded using social security payments. The corresponding DSS budget was transferred to councils in a staged process – the special transitional grant was made over a period of 3 years. The social security system retained responsibility for those people who were receiving the income support payment before 1993. However, by 2002/3 very few of these 'preserved rights' cases remained and those who did were transferred to the local authority system. The latest figures in the table give a good indication of total public spending on social care, although they are a slight under-estimate of the grand total because there are two further, albeit modest, sources of public support: NHS funding and private spending that is supported by universal social security payments such as attendance allowance (i.e. not private spending backed by private means). Exact figures on the social security supported expenditure are not available but we can nonetheless conclude that total public spending on all social care is more than 1.4 per cent of GDP.

**Table 4-4. Personal social services gross expenditure - England**

| Year | Cash terms | 2000/01 Prices | % GDP |
|---|---|---|---|
| | £ billions | £ billions | |
| 1986-87 | 3.1 | 5.4 | 0.7% |
| 1987-88 | 3.4 | 5.7 | 0.7% |
| 1988-89 | 3.8 | 5.8 | 0.7% |
| 1989-90 | 4.2 | 6.1 | 0.8% |
| 1990-91 | 4.7 | 6.4 | 0.8% |
| 1991-92 | 5.1 | 6.5 | 0.8% |
| 1992-93 | 5.5 | 6.6 | 0.9% |
| 1993-94 | 6.3 | 7.3 | 0.9% |
| 1994-95 | 7.5 | 8.7 | 1.0% |
| 1995-96 | 8.4 | 9.4 | 1.1% |
| 1996-97 | 9.3 | 10.2 | 1.1% |
| 1997-98 | 10.0 | 10.7 | 1.2% |
| 1998-99 | 10.8 | 11.3 | 1.2% |
| 1999-2000 | 12.0 | 12.3 | 1.3% |
| 2000-2001 | 12.8 | 12.8 | 1.3% |
| 2001-2002 | 13.6 | 13.2 | 1.3% |
| 2002-2003 | 15.2 | 14.3 | 1.4% |

*Source: DH Gross PSS expenditure*

In this study, we focus on services for older people, which constitute the largest client group. Gross spending on older people's services by councils was £6860m in 2002/3, which is some 45% of the total spend (or just under 60% of total adult spending i.e. excluding services for children). Table 4-5 gives a break down of spending, distinguishing council and private spending for older people. Estimates suggest that purely private spending – where the council makes no contribution – are in the order of £1575m for residential care or nearly two-fifths of the (supported) PSS gross spend.

**Table 4-5. Expenditure on social services for older people – England 2002/3**

| £millions | Total | Non-residential | Residential | Assessment |
|---|---|---|---|---|
| PSS gross | 6860 | 1990 | 4250 | 620 |
| PSS net | 5070 | 1740 | 2710 | 620 |
| User fee | 1790 | 250 | 1540 | |
| Private | 2000 | 425 | 1575 | |
| TOTAL | 8860 | 2415 | 5825 | 620 |

Notes: PSS expenditure data are from DH Statistical Bulletin 2004/02.
Private expenditure data are unpublished PSSRU estimates (with thanks to Raphael Wittenberg).
Private and user fee costs are partly met from disability benefits.
NHS costs (including nursing care in nursing homes for privately funded residents) is excluded.

Laing and Buission provide estimates of the proportion of care home residents by source of funding – see Table 4-6. Their estimates are for the UK rather than England but should apply (with perhaps a slightly higher proportion of self payers in England than for the UK as a whole). The NHS pays for about 5 per cent of residents in independent care homes, nearly all of them in nursing homes. Laing and Buisson suggest that the NHS currently funds just over 50,000 long-term care places in the UK, mostly for the care of long-term geriatric and psycho-geriatric patients. Some 20,000 places are purchased from the independent sector with the rest provided in-house (Laing and Buisson, 2003). Unlike local authority funded care, this NHS 'continuing care' does not involve any contribution from patients (even to cover 'hotel' costs), although patients do lose some or all of their state benefits such as state retirement pensions and attendance allowance. There is a fine – and controversial – line between eligibility for fully-funded NHS continuing care and means-tested nursing care – essentially hinging on whether the resident's primary need is for healthcare.

Income support preserved rights cases are included in the table but entries for 2003 are zero because, as noted above, the remaining few preserved rights cases were transferred to local authority funding. As to self-payers, exact numbers are not routinely available. Generally, they are determined by subtracting state-funded residents from the total. The Department of Health did carry out two censuses of self-payers in nursing homes, finding 42,700 in June/July 2001 and 41580 in Dec 2002. These figures when grossed up to the UK level match the Laing and Bussion figures in Table 4-6 closely.

**Table 4-6. Sources of finance for residents in the private and voluntary nursing and residential care homes for elderly and physically disabled people, UK 2003.**

| | Nursing homes | | Residential homes | | Both | |
|---|---|---|---|---|---|---|
| | **000s** | **%** | **000s** | **%** | **000s** | **%** |
| Income Support (preserved rights) | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| Local authority | 97 | 56.7 | 151 | 70.2 | 248 | 64.2 |
| NHS | 20 | 11.7 | 0 | 0.0 | 20 | 5.2 |
| Self-pay | 54 | 31.6 | 64 | 29.8 | 118 | 30.6 |
| Total | 171 | 100.0 | 215 | 100.0 | 386 | 100.0 |

*Source. Laing and Bussion (2003)*

Table 4-7 shows that in real terms PSS spending on older people has increased significantly since 1994/5, being over half as much again by 2002/3. However, the rate of increase was not quite as great as for total social care spend. Beyond any special transitional grant legacy effects, from 1997/8 to 2002/3 real spending has increased by 24% for older people.

**Table 4-7. Older people's spend 94/5 to 02/3**

| **Year** | **Gross PSS expenditure - older people** | |
|---|---|---|
| | **Market prices** | **Real (2000/1 prices)** |
| 1994..95 | 3567 | 4128 |
| 1995..96 | 4070 | 4563 |
| 1996..97 | 4575 | 4983 |
| 1997..98 | 4912 | 5203 |
| 1998..99 | 5216 | 5405 |
| 1999..2000 | 5644 | 5771 |
| 2000..2001 | 5899 | 5899 |
| 2001..02 | 6173 | 5982 |
| 2002..03 | 6860 | 6454 |

*Source. Personal Social Services expenditure and unit costs: Department of Health*

Table 4-8 provides details of expenditure on services to older people supported by the local authority. Residential and nursing home care command the lion's share of expenditure for older people. The share of resources going to residential care has remained largely constant in recent years. The proportion used to fund nursing home care rose significantly after 1993 but largely due local authorities taking on responsibility for funding nursing homes. In real terms, since 1997/8 gross expenditure on care homes has increase by just under 30% to 2002/3.

**Table 4-8. LA expenditure on services for older people – England, £000s and %**

| | 1994/5 | | 1997/8 | | 2002/3 | |
|---|---|---|---|---|---|---|
| Assessment and care management | 355,386 | 10% | 428,519 | 9% | 622,471 | 9% |
| Nursing home placements | 405,290 | 11% | 1,026,491 | 21% | 1,479,751 | 22% |
| Residential care home placements | 1,455,888 | 41% | 1,877,345 | 38% | 2,741,712 | 40% |
| Supported and other accommodation | 38,290 | 1% | 37,908 | 1% | 24,522 | <1% |
| Direct payments | .. | | .. | | 10,698 | <1% |
| Home care | 947,123 | 27% | 1,119,895 | 23% | 1,395,093 | 20% |
| Day care | 155,527 | 4% | 185,899 | 4% | 286,905 | 4% |
| Equipment and adaptations | .. | | .. | | 67,254 | 1% |
| Meals | .. | | .. | | 95,469 | 1% |
| Other non..residential costs | 209,319 | 6% | 235,491 | 5% | .. | |
| Other services to older people | .. | | .. | | 136,598 | 2% |
| TOTAL OLDER PEOPLE | 3,566,822 | 100% | 4,911,549 | 100% | 6,860,472 | 100% |

*Source: Government RO3 and PSS EX1 Gross PSS expenditure data*

The focus of the empirical work is residential and nursing care homes for older people. It concerns the £4221m spent in 2002/3 (or 36% of gross local authority adult spend).

### 4.5.2 Activity

There have been significant changes both in the total number of residential and nursing home places and in the nature of which organisations provide those services. Table 4-9 shows the numbers of places available over the last two decades and more. It refers to the total number of registered places by sector and represents total capacity, not the number of places actually purchased by various commissioners (see below for the number of such 'publicly supported' places).

**Table 4-9. Residential and nursing home places for older people, including older people with mental health problems, and younger physically disabled people: places by sector in the England, 1976–2001**

| | Local authority | | | Voluntary | | | Private | | | Dual registered | | | Nursing homes | Total | Total 65 over |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | Older | YD | Total | Older | YD | Total | Older | YD | Total | Older | YD | | | |
| 3/1976 | 110796 | | | 32789 | | | 26412 | | | | | | | | |
| 3/1977 | 112631 | | | 33022 | | | 28126 | | | | | | | | |
| 3/1978 | 113574 | | | 33292 | | | 30073 | | | | | | | | |
| 3/1979 | 113592 | | | 33912 | | | 31998 | | | | | | | | |
| 3/1980 | 114103 | | | 34957 | | | 35764 | | | | | | | | |
| 3/1981 | 114921 | | | 36881 | | | 39253 | | | | | | | | |
| 3/1982 | 115493 | | | 36743 | | | 44346 | | | | | | 18197 | | |
| 3/1983 | 115913 | | | 37613 | | | 51760 | | | | | | 22554 | | |
| 3/1984 | 116430 | | | 38242 | | | 63072 | | | | | | 27346 | | |
| 3/1985 | 116080 | | | 37466 | | | 80041 | | | | | | 33869 | | |
| 3/1986 | 115609 | | | 36000 | | | 92605 | | | | | | 41570 | | |
| 3/1987 | 114189 | | | 34374 | | | 106108 | | | | | | 53920 | | |
| 3/1988 | 112422 | | | 34402 | | | 116668 | | | | | | 57007 | | |
| 3/1989 | 109194 | | | 34166 | | | 135369 | | | | | | 73601 | | |
| 3/1990 | 105380 | | | 34960 | | | 145457 | | | | | | 89616 | | |
| 3/1991 | 97853 | | | 36685 | | | 155315 | | | | | | 108979 | | |
| 3/1992 | 86676 | | | 40608 | | | 158990 | | | | | | 123957 | | |
| 3/1993 | 77012 | | | 34539 | | | 146015 | | | 11194 | | | 144325 | 413085 | |
| 3/1994 | 68899 | 67,401 | 1,498 | 35619 | 31,269 | 4,350 | 143274 | 142,460 | 814 | 13788 | 12,413 | 1,375 | 148459 | 410039 | 402002 |
| 3/1995 | 63810 | 62,443 | 1,367 | 35266 | 31,435 | 3,831 | 142109 | 141,372 | 737 | 15049 | 13,462 | 1,587 | 155412 | 411646 | 404124 |
| 3/1996 | 61425 | 60,155 | 1,270 | 35270 | 31,511 | 3,759 | 141738 | 141,013 | 725 | 18354 | 16,603 | 1,751 | 150527 | 407314 | 399809 |
| 3/1997 | 55659 | 54,680 | 979 | 37186 | 32,460 | 4,726 | 145275 | 144,293 | 982 | 24636 | 23,544 | 1,092 | 154984 | 417740 | 409961 |
| 3/1998 | 53804 | 52,734 | 1,070 | 37123 | 33,012 | 4,111 | 148913 | 148,011 | 902 | 30587 | 29,135 | 1,452 | 162341 | 432768 | 425233 |
| 3/1999 | 50176 | 49,112 | 1,064 | 34981 | 31,055 | 3,926 | 150179 | 148,997 | 1,182 | 31843 | 30,708 | 1,135 | 159629 | 426808 | 419501 |
| 3/2000 | 46475 | 45,210 | 1,265 | 36760 | 32,518 | 4,242 | 147392 | 146,125 | 1,267 | 36332 | 34,973 | 1,359 | 152663 | 419622 | 411489 |
| 3/2001 | 42723 | 41,832 | 891 | 37390 | 33,566 | 3,824 | 145493 | 144,625 | 868 | 37015 | 35,769 | 1,246 | 147529 | 410150 | 403321 |

*Source: DH community care statistics 2001*. Notes (1) Interpolated from total nursing home places.

During the early part of the time series reported, the national data combined places for older people with places for younger physically disabled people, although the latter constitute only a very small share (e.g. about 2% in 1994). Data on nursing homes has generally been available by type of nursing home rather than client group. Nonetheless, some attribution of bed/place has been made to older people (by Department of Health statisticians) and this data is given in the table. The data reported are those collected by the Department of Health. After 2001, collection of data on registered homes – rather than those publicly purchased – was passed to CSCI. The data collection by CSCI has not been on exactly the same basis as before, and was patchy during the transition period. Nonetheless, the latest totals produced by CSCI are very much in line with the data reported in the table (i.e. 371328 places for older people in 2003). The market analysts Laing and Buission undertake an annual survey of care home places in the UK. When scaled down to the England population total residential places are slightly lower (by 13,400 or 6%).

**Figure 4-2. Places in care homes for older people – by provider type, 1970-2001**

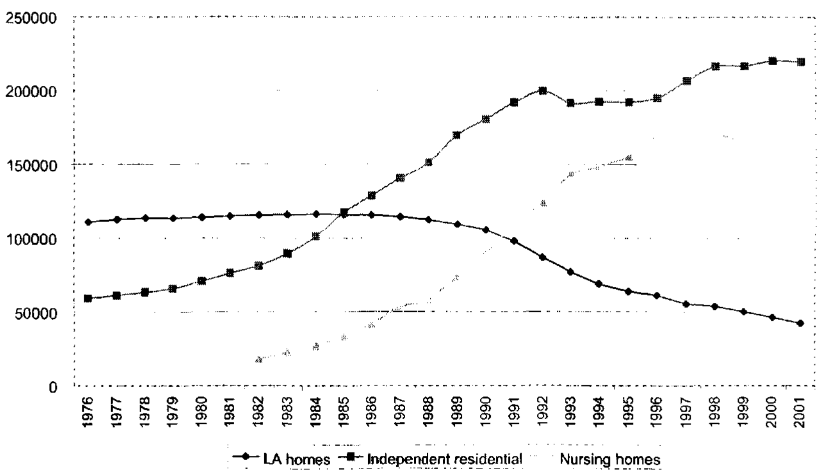
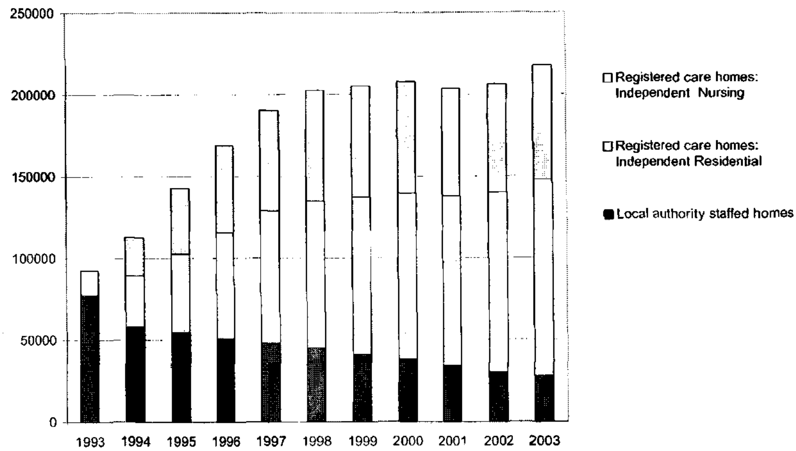



Figure 4-2 shows the places in care homes over time. The key comparison is local authority residential and independent sector residential places. As regards the former, it is clear that from the late 1980s and early 1990s – at the time of the Griffith Report and subsequent legislation – the number of places declined from being relatively constant. The dip in residential care places is due to the removal from the time series of younger disabled adults. Figure 4-3 shows how the number of residents that are supported by local authority has changed since local authorities took on responsibilities for new residents who otherwise would have been funded by the social security system.

**Figure 4-3. Local authority supported residents, people over 65 (all client group**




The numbers are also provided in Table 4-10. The data show a very significant incre post 1993. Much of this growth represents the transfer of care responsibility from th social security system to local government, although after the three year transition period supported placements continue to grow.

**Table 4-10. Local authority supported care home places, over 65s, England**

| | All home types | | Local authority | | Independent | | | |
|---|---|---|---|---|---|---|---|---|
| | Supported | Total Places | Supported | Total Places | Residential Supported | Nursing Supported | Total supported | T p |
| 1993 | 92,710 | | 78,000 | | 14,710 | 0 | 14,710 | |
| 1994 | 112,939 | 402,002 | 58,297 | 67,401 | 31,530 | 23,112 | 54642 | 3 |
| 1995 | 142,994 | 404,124 | 54,773 | 62,443 | 48,347 | 39,874 | 88221 | 3 |
| 1996 | 168,962 | 399,809 | 51,028 | 60,155 | 65,089 | 52,845 | 117934 | 3 |
| 1997 | 190,145 | 409,961 | 48,276 | 54,680 | 81,082 | 60,787 | 141869 | 3 |
| 1998 | 202,725 | 425,233 | 45,350 | 52,734 | 89,925 | 67,450 | 157375 | 3 |
| 1999 | 205,325 | 419,501 | 41,255 | 49,112 | 96,520 | 67,550 | 164070 | 3 |
| 2000 | 207,920 | 411,489 | 38,555 | 45,210 | 101,795 | 67,570 | 169365 | 3 |
| 2001 | 204,070 | 403,321 | 34,380 | 41,832 | 104,210 | 65,480 | 169690 | 3 |
| 2002 | 206,195 | | 30,245 | | 110,120 | 65,830 | 175950 | |
| 2003 | 217,640 | | 28,130 | | 120,235 | 69,275 | 189510 | |

The use of local authority homes – in-house providers – declined sharply in the couple of years to 1994 but thereafter has declined gradually. There is some sign that this trend is now flattening out.

## 4.6 Governance arrangements

Services are generally easy to identify and distinguish. Governance, being a far more strategic and wide-ranging concept, is less so. As discussed in the last chapter there are a range of distinguishing factors to illustrate governance choices – see table 3-1 in chapter 3.

A number of empirical indicators are used to give these theoretical concepts some practical meaning. The indicators are drawn from four data sets (see also chapter 1 section 1.6). The first dataset is secondary data on activity and expenditure, as described above, which is available from routine Department of Health collections. Second, the survey of residential care data that provides information on commissioning arrangements from a provider's perspective. The third source is the PSSRU commissioning survey. The fourth is the data from the small LA sample broad mixed economy study. These data can be used to describe organisational structures and contract reimbursement incentives.

### 4.6.1 Organisational structure

As noted in the last chapter it is useful to distinguish, conceptually, between the distribution of *ownership* of the purchasing and providing function on the one hand, and the distribution of *control* over those processes on the other. The former distribution is most easily measured by the relative size of independent sector supply. In addition, where the residential care sector is characterised by extensive public provision, it is likely to imply unified ownership. Some councils have almost no in-house residential care for older people and clearly rely totally on market arrangements. Others have a mix, and the average numbers are given in the above tables and charts.

Ownership need not dictate the distribution between stakeholders of control over the various functions. In theory, at least, a public funding body could, for example, cede purchasing functions such as user assessment to independently owned providers, creating an organisation that could at least partially be described as a privately owned hierarchical organisation. In practice, local authorities are required by law to assess and so make arrangements to meet assessed needs (subject to eligibility etc.) and so remain public sector purchasers in governance terms. Effectively therefore only public hierarchies are possible where the public sector organisation jointly 'purchases' and provides.

The other, more relevant, possibility is the (public sector) 'internal' market. In this variant there is public ownership of both the means/resources of commissioning and also provision. However, control is dispersed within the organisation. Provider units are managerially separate and do not account to purchasers. Moreover, purchasers have some freedom to choose different (public sector) providers, perhaps in-house providers from other local authorities. This model has strong market-like features (although does retain some hierarchical elements e.g. providers often have relatively low powered incentives as they are only partial residual claimants). It is, nonetheless, sufficiently different from the (pure) hierarchical model to have different behaviour implications. We need therefore to see how relevant this model is in an empirical sense.

In multi-tiered organisations (i.e. social services departments) where an agent or 'broker' works on behalf of the service users, choices are available as to the location of that purchasing decision-making. Often purchasing functions are spread across different management tiers in the organisation, but the key issue is the location of *operational* control and purchasing budgets. Authority to sign-off on budgets may be held centrally or devolved to local agents, such as care managers (Davies, 1992; Wistow et al., 1996). The degree of *devolution* and the location of budgets helps us to understand whether an internal market is in operation, or whether a more traditional hierarchical model is in place. For example, as we discuss below, in many cases budgets can be devolved directly to in-house *provider* managers. Clearly, this arrangement implies that the process for commissioning care from the independent sector is systematically different from the process to secure in-house provision since the latter are directly allocated a budget, and the former have services purchased under contract. Or put differently, it rules out the processes from buying from in-house and independent sector being exactly the same.

The PSSRU commissioning survey posed a number of questions that allow inferences about these arrangements. Some 90 councils (of 150) responded to the questionnaire. We first asked about the location of the purchaser provider separation within councils. The majority of respondents (42 or 47%) began the separation of functions at the third organisational tier i.e. below Assistant Director level. Thirty-one councils indicated the second tier, with the rest (17 or 19%) at fourth tier or below. Whilst the particular organisational arrangements in LAs differed considerably, this information is suggestive of whether control is, on the whole, de-centralised or unified. Indeed, where the social services department is separated at AD level we might expect greater de-centralisation than a separation at lower management tiers.

We also asked whether the process for arranging care from independent sector providers that councils operated was different from the process of arranging similar care from in-house providers. There are clearly a number of levels to which this question can apply. Where respondents state there is a difference this response is sufficient for us to reject there being an internal market. A null response, on the other hand, need not indicate the existence of an internal market because respondents could interpret this question to be about just assessment and care-planning, not including contracting as well. For these reasons it is likely that Table 4-11 underestimates the number of authorities with different processes (i.e. a third).

**Table 4-11. Is the process for arranging care from independent sector providers different from the process of arranging similar care from in-house providers?**

| | Number | Percent |
|---|---|---|
| No difference in arrangements between independent and in-house providers | 54 | 60% |
| Yes; securing in-house and external services involves different processes | 27 | 30% |
| Not applicable/don't know | 9 | 10% |
| Total | 90 | 100% |

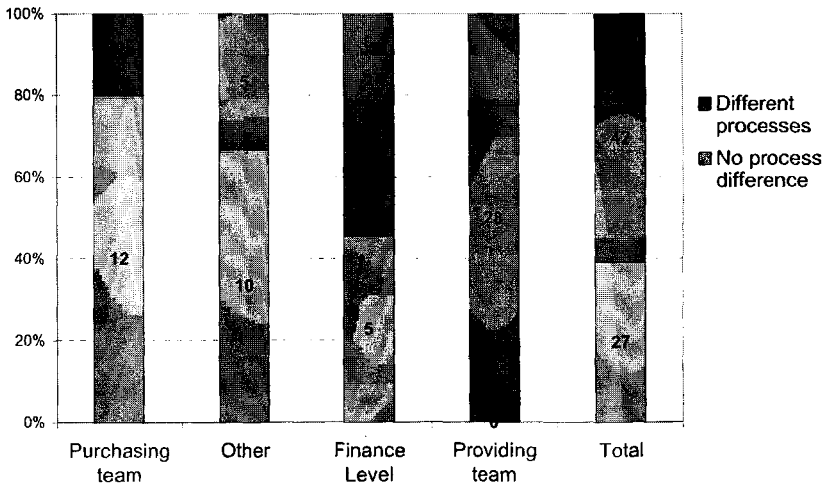
Because the question about processes may underestimate the number of councils that have hierarchical arrangements for the purchase of in-house services, we can also look at where respondents stated they had located budgets for securing in-house services. Table 4-12 reports this information from the survey. We can see that around two-fifths of councils allocate budgets directly to providers, which rules out their operating an internal market. Taking the results from these two questions together, the data suggest at least 60% of councils have hierarchical arrangements for in-house services – see Figure 4-4.

**Table 4-12. Typical location of budgets for securing residential care**

| Care manager | External | | Internal/in-house | |
|---|---|---|---|---|
| LA or SSD finance level | 11 | 14% | 12 | 17% |
| Purchasing team manager | 49 | 61% | 16 | 23% |
| Care manager | 0 | 0% | 0 | 0% |
| Providing team manager | 1 | 1% | 28 | 39% |
| Other | 19 | 24% | 15 | 21% |
| Total | 80 | 100% | 71 | 100% |

*Source: Commissioning survey*

**Figure 4-4. Proportion of LAs with no difference in the processes for securing in-house from external provision – by location of budgets for securing *in-house* provision**




### 4.6.1.1 The relationship between control and ownership

In practice, decisions about ownership and distribution of control are very unlikely to be made independently. The unit production costs of in-house providers have been consistently higher than those in the independent sector. If councils are willing to operate as though in a market with their in-house providers, the next step to actually outsourcing is rather small, and likely to be comparatively more efficient in that case.

We therefore hypothesise that the proportion of supported places retained in-house by local authorities is positively related to an *absence* of an internal market (see Table 3-1). This hypothesis can be tested by comparing, for each local authority, the proportion of total placements made in-house (rather than in the independent sector) with the indicators of control distribution described above. Multiple regression was used to estimate the simultaneous relationship between this measure and the indicators of control distribution. Since decisions about out-sourcing are also influenced by a range of local authority specific factors, such as political control, historical supply etc, a 'local authority type' control factor was also used. Local authorities are categorised as one of five types: shire, metropolitan, unitary, inner London and outer London. Data on control was available for 71 local authorities from the PSSRU commissioning survey. The clearest indicator of unified control was used – that is, where *providers* directly hold budgets i.e. the 28 of 71 councils described in Table 4-12. The dependent variable – an indicator of (integrated) ownership – is the ratio of supply by in-house providers to total residential care home supply, with high values indicating high integration. Supply is measured as the number (actually, incidence) of (supported) placements by provider type. See Table 4-13 for descriptive statistics.

**Table 4-13. Ownership and control variables – descriptive statistics**

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Prop. res care LA | 0.285 | 0.154 | 0.004 | 0.670 |
| Unified control: In-house budgets held by provider manager | 0.386 | 0.490 | 0 | 1 |

Linear regression was used to estimate the relationships. Although testing could not reject the normality of the dependent variable (Shapiro-Wilk W test for normal data, $p$ = 0.29), in view of the relatively small sample size (71 cases) the estimation was also bootstrapped (1000 repetitions). Since these ownership and control indicators of hierarchical governance are *a priori* jointly determined, an instrumental variables regression was also undertaken with indicators being instrumented by the proportion of places supplied in-house lagged one year. A Basmann-Sargan test could not reject the null hypothesis of correctly specified instruments in the IV regression ($p$ = 0.504). OLS was also conducted for comparison.

Table 4-14 reports the results of the estimation. The *unified control* variable is significantly correlated with the ownership ratio, in both the linear model and in the models allowing for simultaneity. Bootstrapped 95% confidence intervals also indicate that the relationship is significantly different from zero. Beyond that, variation in the ownership ratio is significantly associated with the political control of the council, with labour councils more likely to have centralised ownership. A number of local authority type dummies and a property prices variable were included to account for council prosperity and costs.

**Table 4-14. The relationship between 'ownership' and 'control' indicators of governance choice ($n$ = 71) – dependent variable: centralised ownership**

| | 2SLS[1] | | | | 3SLS | | | | OLS[2] | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Bootstrap 95% CI | | | | Bootstrap 95% CI | | | |
| | Coeff | t-stat | lower | upper | Coeff | t-stat | lower | upper | Coeff | t-stat |
| Unified control | 0.47 | 3 | 0.16 | 0.78 | 0.49 | 3.56 | 0.19 | 0.80 | 0.10 | 2.59 |
| Lab seats (% total) | 0.27 | 1.81 | | | 0.28 | 1.95 | | | 0.18 | 1.88 |
| London LA | -0.24 | -2.44 | | | -0.25 | -2.58 | | | -0.16 | -2.45 |
| Property prices (log) | 0.34 | 3.12 | | | 0.35 | 3.46 | | | 0.15 | 2.61 |
| Coastal LA | -0.24 | -1.64 | | | -0.24 | -1.76 | | | -0.16 | -1.73 |
| Const | -3.91 | -2.92 | | | -4.03 | -3.27 | | | -1.54 | -2.22 |

Notes
[1] Basman-Sargan test for instrument specification ($Chi^2$, 8, 0.95) of 15.51
[2] Reset specification test: $p > 0.66$ and for heteroscedasticity, $p > 0.64$

Unified control corresponds to an increase in the proportion of placements made in-house, other things equal, of nearly 50 per cent for both the instrument variable models. The magnitude of the relationship supports our use of ownership as a proxy of an indicator of hierarchical governance.

### 4.6.2 Reimbursement contracts

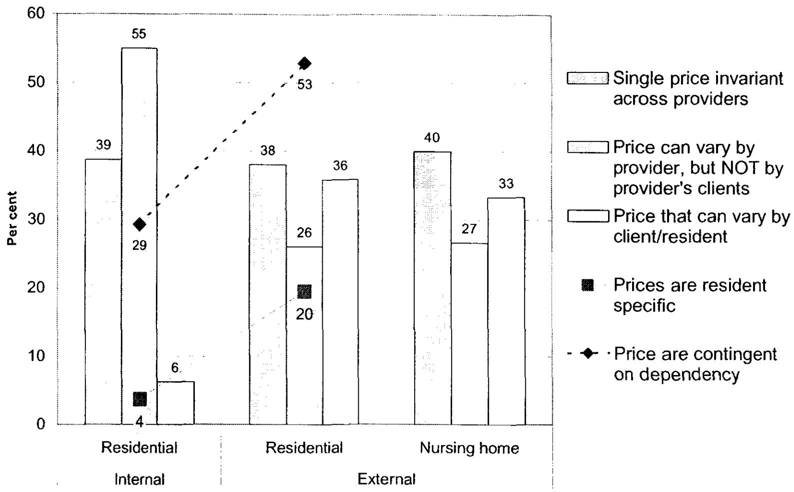
Table 3-1 in the last chapter described a number of governance dimensions concerning contracts such as timing, length, specification etc. Of arguably the most empirical relevance is contract contingency. To establish the use of contingent contracts in the residential care market, we have two sources of data. First, local authority reporting of the types of contracts they offer – and specifically their responses to the commissioning survey. Second, providers' views as to the types of contracts they have. There is some potential for disagreement between these sources, not only because the samples are different, but also because there are many options as to how contracts can be contingent, e.g. by client group, by case-mix, by severity etc., and only a subset need apply to any particular provider. In chapter 8 we use the latter since we are interested in how providers respond to different contingencies, and this will depend on their own perceptions. Partly because provider's perceptions are covered later, and because that data is much more specific, here we will cover the results from the commissioning survey in more detail.

The survey asked a series of questions about exactly how LAs set their prices for purchase of care from independent (and also in-house) providers. The survey had a response from 92 local authorities on these issues. The first question was: how are prices determined for the purchase of (publicly funded, long-term) care from independent sector (private and voluntary) providers? There were three options:

| | |
|---|---|
| OPTION 1 | a single price that is *invariant* across different (publicly-funded) providers in the authority (and so is also invariant with respect to clients/residents) |
| OPTION 2 | a price that can vary by provider, but does NOT vary by (publicly-funded) clients/residents served by the provider |
| OPTION 3 | a price that can vary by (publicly-funded) client/resident (and so can vary by provider) |

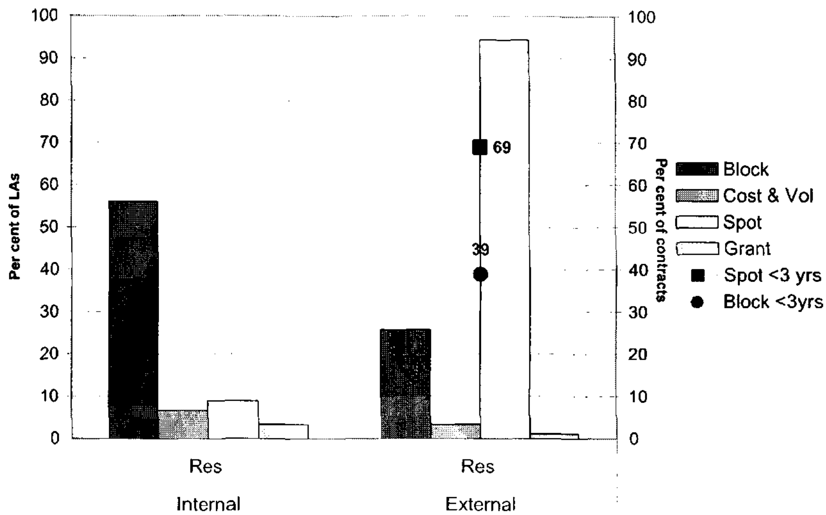
Figure 4-5 describes the responses. For residential care from external providers the most common option was the first. However, the distribution across the three options was fairly uniform. The figure also shows the responses to a number of follow-up questions. Those respondents that selected options 2 or 3, were asked whether the price typically reflects residents' dependency levels? In the case of option 2 it would be the provider level price that varies with client dependency, implying either some average dependency level effect, or more likely the client group specialisation of the home, particularly relevant for those homes that specialise in caring for older people with dementia or other cognitive impairment. For those selecting option 3, prices may clearly vary on a per client basis. Altogether, just over half (53%) of the 92 respondents worked with prices that did reflect dependency (within the older people's client group). We also asked those selecting option 3, whether their prices were determined at the time of placement, being specific to the particular resident (i.e. not set in advance). This question tries to distinguish between tariff based pricing systems and those apparently more rare cases where prices are truly individual resident specific. For independent sector residential care 20% of councils reported this arrangement, which might nonetheless be a little on the high side. The corresponding level of contingency for internal arrangements with in-house providers is far lower as shown in figure.

**Figure 4-5. Pricing options and dependency contingency: contracts with independent sector (external) and in-house care homes – percentages of LAs within the sample, 2001 ($n$ = 92)**




The survey also asked about the *quantity* contingency of contracts i.e. whether average reimbursement per unit depends on the total quantity purchased. Block contracts are quantity contingent in this way because payment rates are tied to the total quantity purchased. Spot or call-off contracts make little or no such allowance – payment rate is constant regardless of the total amount purchased. Cost and volume contracts represent something of a hybrid of the two types. Figure 4-6 shows that spot contracts constitute the very significant majority. Moreover, although a fifth of councils also report using block contracts, it is likely that at least some of these contracts are for independent trusts and specific charitable providers (see page 91). For mainstream average sized independent sector homes the likelihood of block contracts being used is small.

**Figure 4-6. Types of contact used to secure residential care, and contract durations - % of authorities and contracts, 2001**




In chapter 8 we analyse contract contingency from a provider's perspective. We use the Survey of Residential Care, which provided a response from over 600 homes in England. Among a wide range of other questions (see Netten et al., 1998) the survey asked about pricing arrangements with the local authority purchaser. In particular, homes were asked whether they operated with a price that was "pre-set by local authorities dependent on the type of resident (e.g. level of dependency)". Table 4-15 describes the responses. Overall, 38% of homes regarded price as being affected by client dependency.

**Table 4-15. Perceived user dependency contingency of price – independent sector homes, % of homes by type of home**

| | Sector | % of homes | N | SD |
|---|---|---|---|---|
| Residential | Private | 46.4% | 112 | 0.50 |
| | Voluntary | 35.9% | 78 | 0.48 |
| Nursing home | | 31.0% | 113 | 0.46 |
| All types | | 38.0% | 303 | 0.49 |

*Source: PSSRU survey of residential care*

We can make some comparison with the Commissioning Survey funding, but with a number of caveats. First, the commissioning survey was undertaken about 5 years after the homes survey and we might expect the underlying trend to be an increasing use of contingent contract arrangements. Secondly, homes' perceptions are based on their own specific 'market' of types of residents, whereas the local authority covers the whole local spectrum. Third, homes in the home survey were drawn from a sub-sample of local authorities. Nonetheless, there was a high consistency between the commissioning survey results and the two-fifths of homes reporting contingency in the provider survey. It specifically depends on whether we take as the most relevant comparator from the commissioning survey, the 35% of LAs that reported option 3 (i.e. client level contingency) or the overall level of dependency contingency (which includes inter- but not intra-home contingency), which just over 50% of LAs reported.

### 4.6.3 Response rates

As noted in chapter 1, the commissioning survey data showed some regional variation in response rates. To test for potential problems as a result, we tested for significant differences between the 9 administrative regions of England for the main variables of interest. Fisher's exact test was used and the probabilities are as follows.

- whether LAs had an internal market ($p = 0.345$)
- whether there was a different commissioning process for in-house compared with independent providers ($p = 0.865$)
- key budget holder – in-house residential ($p = 0.011$)
- key budget holder – external residential ($p = 0.031$)
- external pricing options ($p = 0.214$)
- Internal pricing options ($p = 0.152$)

In most cases no significant difference was found with the exception of key budget holding. This has potential implications for our 'internal markets' variable – see above. However, the internal market variable is based on budget holding information, and it was not significant. Therefore, along with closer inspection of the data, we infer that the significance of the Fisher Exact test in this case was due not to regional differences in budget holding by providers, but due to differences in budget holding at purchasing team level rather than at a more strategic level. This finding does not undermine any conclusions about control and ownership relationships.

## 4.7 Conclusions

This chapter outlines some of the main historical antecedents to the current social care system. Three are most relevant. First, the local government re-organisation of 1970 and the establishment of unified Social Services Departments (SSDs) on 1 April 1971 as a result of the Seebohm Report in 1968 and the 1970 Local authority Act. Second, the series of reports and legislation of the late 1980s and early 1990s, including in particular, the Audit Commission report of 1986, the Griffiths report of 1988 and culminating in the NHS and Community Care Act 1990. Third, the 1997 White Paper and New Labour policy that prompted a shift away from the unfettered market, but was not a return to municipal bureaucracies of the 1970s.

These policy developments have left what might be described as a somewhat eclectic system. Although the reforms of the 1970s consolidated the social services responsibility of councils and constituted an attempt to make a 'universalistic' service, it remained means-tested and rooted in the provisions of the 1948 National Assistance Act. Social services were not viewed as a universal service because the majority of people could go through their lives and not use social care. The loosening of social security rules created, for a time, a more diverse demand-side but now the vast majority of state-supported care is LA-purchased. Nonetheless, these arrangements did result in a far larger and more diverse supply side. The legacy therefore is a present system with wide variation between the 150 local authorities with social services responsibilities in terms of the balance between quasi-market and in-house arrangements. Furthermore, a consequence of the new public management policies, and thereafter New Labour's third way has been new approaches and variety in the contractual relationships between councils and the independent sector providers.

This chapter has described the variety of organisational and contractual arrangements empirically by looking at changes in expenditure and activity over time, and also by describing a number of cross-sectional indicators of ownership and control, and reimbursement arrangements. If pushed to generalise, the residential care system in England is typically composed of market governance arrangements, with even some indication of internal markets in some areas. Nonetheless, hierarchical arrangements do remain a significant minority arrangement, perhaps covering a fifth of publicly funded residential care for older people.

# Chapter 5. Markets and hierarchies

The institutional arrangements that govern the undertaking of transactions are many and varied. As described in chapter 4, social care is no exception, involving very complicated transactional arrangements by most standards. Nonetheless, as indicated in chapter 3, many of the relevant dimensions are aligned around two archetype governance structures, markets and hierarchies. This chapter considers the economic advantages and disadvantages of this choice in efficiency terms – a normative judgement as also described in chapter 3. As such this chapter underpins the development of relevant empirical hypothesis regarding the likely transaction and production costs of markets and hierarchies in social care.

This chapter attempts to make more formal the arguments made in chapter 3 about the comparative net benefits of markets and hierarchies, using the concepts developed in that chapter and underpinned by the model outlined in chapter 1. The claims advanced in chapter 3 – that in comparison to markets, hierarchies have lower transaction costs and higher production costs – are analysed. The aim is to detail the problem formally, explore the internal consistency of the arguments and so develop testable propositions. The analysis below throws up many nuances and particularities that significantly colour our consideration of the choice between market and hierarchical governance.

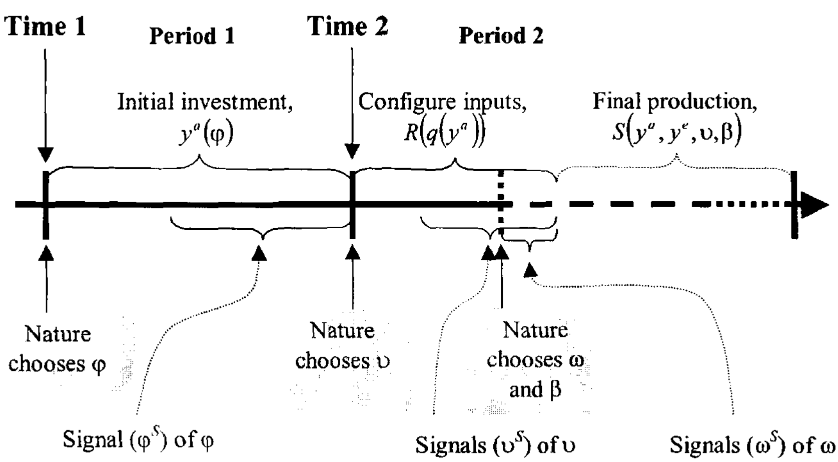
Chapter 3 described the particular features that distinguish market and hierarchical governance structures. It emphasised the need, when comparing markets and hierarchies to comprehensively compare their benefits *net* of their costs, including net transaction costs. As outlined in that chapter, there are transaction costs and benefits in terms of both the *co-ordination* and *motivation* problems, where in practice, economic activity is not perfectly co-ordinated or motivated.

The structure of this chapter is as follows. First, it develops the model described in chapter 1 to characterise market and hierarchy governance structures. Second, it draws out the net transaction benefits of each. Third, it considers a range of mediating features, and in setting these in a social care context, infers the likely effects on transaction and production costs of choices between social care markets and hierarchies. Fourth, it considers how governance tasks will be carried out in each governance structure. Fifth, the net transaction and production costs are assessed on this basis. Finally, empirical propositions are advance and discussed.

## 5.1 A model

As noted in chapter 1, the theoretical model has two time periods. In this chapter we need only consider an aggregate second 'provision' period. This slightly simplified structure is given in Figure 1-1.

**Figure 5-1. Stylised timeline**




The nature of uncertainty is the same as in chapter 1 i.e. uncertainty exists regarding the state of world, which is summarised by the parameters $\varphi$, $\omega$, $\upsilon$ and $\beta$, and resolved at various times during the production process. To briefly recap, at time 1 nature chooses $\varphi$, which we can think of as contextual, external factors that affect investment. Investment occurs to generate inputs into the final product with the relationship between investment, $y^a$, and the final product mediated by the parameter $\varphi$. Investment is assumed to have a proportional cost, $y^a$.

At time 2 $\upsilon$ can be measured and shortly thereafter $\beta$. The former parameter reflects, for example, the user's needs characteristics that can be measured (subject to error) by a pre-care assessment. The latter term $\beta$ reflects the costs of caring for a person (for given outcome). It depends on the user's needs profile and also how their particular needs profile fits with the nature of the service provided. It is therefore a refinement of initial assessed characteristics $\upsilon$ (i.e. knowing $\beta$ is sufficient to also know the value of $\upsilon$) – see section 1.3.1 on page 15 and thereafter.

Ideally investment should be geared to $\upsilon$ to allow the best 'care technology' to be in place. However, since this parameter is not chosen by nature at the point of investment, only a general investment 'type' is possible (at some given expectation of $\upsilon$). Using the initial investment, care packages of inputs are put together and configured. In turn, this configuration defines the type of output, $q(y^a(\varphi), \upsilon, \beta)$. The benefit function $R$ measures how well these care packages fit the characteristics of users, given the initial investment, $y^a$. We assume that configuration involves a fixed cost, which without loss of generality, can be set at zero. After the care packages are put together they are implemented for users in the home, with type as above, and to an extent $D$ i.e. the number of care packages, or simply, 'output'. The nature of the 'product' will depend on how well residents actually respond to care package, and this is in part indicated by the parameter $\beta$. So a home might provide places that are aimed at people with, for example, more severe needs. However, the final configuration of the service will depend on the characteristics of the people actually placed. So if more people than expected have behavioural problems, or nursing care needs, or whatever, the actual service provided will adapt to meet those needs (e.g. a changed staff mix, even changes to the physical environment).

The *costs* of producing the care package depends on a level of effort, $y^e$, made by providers, as well as the characteristics of the care package and the provider type, $\omega$. The cost of effort to the provider is $e(y^e)$ and its effectiveness ("productivity") depends on the parameter $\omega$.

This chapter is concerned with alternative ways to organise social care transactions. The 'transaction' is therefore the focus, and as a result, our attention is on the relationship between the purchaser and provider undertaking the transaction. We therefore refer to 'the' (single) purchaser and provider.

Transactions produce an exchange surplus to the stakeholders involved:

$$(5.1) \quad TS = V(y^a, D; \upsilon, \varphi, \beta) - C(y^a, y^e, D; \upsilon, \varphi, \beta, \omega)$$

where the level of investment is $y^a$, and output levels are $D$. The function $V$ is the total value of the final product to the purchaser, and total surplus ($TS$) is this value less the total costs of its production, $C$. The latter are affected by the degree of cost cutting effort employed by the provider.

To be more explicit about the functional form of (5.1), we can define total value as the integral of output $x$ from 0 up to level $D$:

$$(5\text{-}2) \quad V = \int_0^D R(y^a, q, x; \upsilon, \varphi, \beta)dx$$

where $R_D < 0$. Costs are assumed to take the form:

$$(5\text{-}3) \quad C(D) = c(D)D$$

to allow, to some extent, for economies of scale, and in particular, we assume, $C_D > 0$. We will also assume that average costs fall relatively slowly such that $R_D < c_D < 0, \forall D$, and that $C_{DD} < 0$ and $C_{DDD} = 0$. We re-write surplus as:

$$(5\text{-}4) \quad S = \int_0^D R(y^a, q, x; \upsilon, \varphi, \beta)dx - C(y^a, y^e, D; \upsilon, \varphi, \omega, \beta)$$

where the optimal, $x = D^*$ is the output level such that $R(D^*) = C_D$, and so where $S_D = R - C_D = 0$. The above is simple (cardinal) statement of net benefit. People derive value from services provided. The amount of value will depend on their preferences (although we need not be concerned with the exact impact of preferences), but will be offset by the (opportunity) costs involved.

Purchasers are budget limited and so the number of care packages can increase when costs fall. Function (5-4) can also be written as follows, being explicit about service type $q$:

$$(5\text{-}5)\quad S = \int_0^D R\left(q\left(y^a(\varphi), y^e, \upsilon\right)\right)dx - c\left(q\left(y^a(\varphi), \upsilon\right), y^e, \right)D$$

### 5.1.1 Effort and investment tasks

The 'investment' and 'effort' actions here are conceived to be complex and multi-dimensional. They are ongoing *projects* rather than one-offs and therefore require a string of decisions and specifications. As described in chapter 3, the opportunity to specialise leads to a distinction, in model terms, between purchasers and providers. In this framework, production is characterised by *effort*, and is made (only) by providers, primarily involving labour/human assets. In theoretical terms, providers are *agents*. In practice, there are also labour inputs into purchasing tasks, and these tasks have costs. However, these tasks are closely concerned with management, being a statutory duty of authorities, and can be treated as undertaken by the *principal*. Furthermore, all councils organise these purchasing activities under the same broad governance arrangements (i.e. hierarchies), whether or not they use markets or hierarchies for care production. For these reasons we do not need to construct principal-agent relationships in purchasing. We therefore sidestep issues of organisation of labour and effort in purchasing.[19]

Prior to discussing governance structure issues, we need to be clear how governance tasks could be arranged in relation to production effort and investment. As is common in the literature, we suppose that it is impossible to write a contract directly on effort, although a 'contract' specifying instructions as to how efforts should be directed is possible. If a purchaser wished to develop such a contract they would have to negotiate with the provider. Instructions would have to be made in a way that was clear and, to be effective, minimised the chance of misinterpretation by providers – deliberate or

[19] In cost terms, purchasing activities are also much smaller than in magnitude.

otherwise. With relevance to these ideas, we divide costs into two types. First, in making appropriate measurement of relevant circumstances. And second, in undertaking (unilateral) planning to draw up instructions. Third, in 'contracting' with providers to convey these instructions. Often the latter two are conflated into one 'contracting' or 'bargaining' cost category (e.g. see chapter 2), but here the distinction is useful. In particular, if the provider unilaterally undertakes effort then they need to measure and plan, but not contract. If the purchaser wishes to convey instructions, however, they would have to measure, plan *and contract.*

As regards investment, we are primarily referring to the process of putting in place physical assets. Again a stream of decisions and specification are required in this task. In this case, either purchasers or providers can unilaterally decide investment levels and contracting need not be required. Measurement tasks would be required by the investing party as well as the need to plan, but there is no contacting activity, so no contracting cost. Alternatively, purchaser and provider could undertake investment together. Since they are organisationally distinct, some form of coordination would be required – they would, in short, have to agree an investment contract. Both would have to measure and plan, and both would have to agree a verifiable contract in this case. In theory, the investment process does involve some labour input and hence a potential need to contract in relation to effort. However, since this would apply whatever the governance arrangement – unilateral or contracted investment – we can ignore it from the analysis below.

In the main, these individual governance tasks have a fixed and variable cost element. To limit the complexity of our problem, we will assume that measurement tasks are (mainly) fixed. The others are variable: they increase with the size of the effort or investment task, albeit to a diminishing degree. In other words, to write and monitor a contract for any amount of investment or effort involves a significant cost. Beyond the initial (large cost) level, marginal increases in costs diminish rapidly (for example, as in Figure 5-2).

**Figure 5-2. Contracting costs**

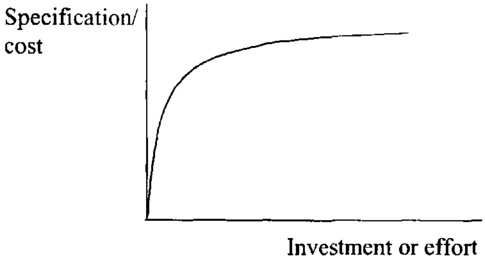



Table 5-1 summaries the various governance costs outlined above in relation to effort and investment tasks. It lists the options for how effort or investment could be organised. Below we will describe how these various costs – $T$ for measurement, $W$ for planning and $B$ for contracting/bargaining – can differ for each option according to the governance structure – market or hierarchy – that is in operation.

**Table 5-1. Governance costs**

| | Provider | | | Purchaser | | |
|---|---|---|---|---|---|---|
| | Measure | Plan | Contract | Measure | Plan | Contract |
| ***Effort*** | | | | | | |
| Provider undertakes unilaterally | $T^{y^eK}$ | $W^{y^e}$ | 0 | 0 | 0 | 0 |
| Purchaser instructs (contracts with) the provider | 0 | 0 | $B^{y^eK} \geq 0$ | $T^{y^eL}$ | $W^{y^e}$ | $B^{y^eL}$ |
| ***Investment*** | | | | | | |
| Provider undertakes unilaterally | $T^{y^nK}$ | $W^{y^n}$ | 0 | 0 | 0 | 0 |
| Purchaser undertakes unilaterally | 0 | 0 | 0 | $T^{y^nL}$ | $W^{y^n}$ | 0 |
| Purchaser and provider contract | $T^{y^nK}$ | $W^{y^n}$ | $B^{y^nK}$ | $T^{y^nL}$ | $W^{y^n}$ | $B^{y^nL}$ |

The two options regarding effort both involve an (equal) planning cost. Because this cancels out in comparison, we simply set it to zero below. Treating measurement costs as fixed, it then means that the marginal *transaction* cost of a provider unilaterally undertaking effort is zero. If the purchaser instructs the provider, however, the marginal transaction cost is $\frac{\partial \left[ B^{y^eK} + B^{y^eL} \right]}{\partial y^e}$. Planning costs for investment also cancel out and so are dropped. Again this simplifies the analysis by making unilateral actions – here investment – have a zero marginal cost. When parties contract on investment, marginal costs are non-zero.

### 5.1.2 Product contracts

Contracts can also be written on product characteristics (type and output) – called the *product* contract, denoted the $q$-contract. The costs of determining product contracts are assumed to be very similar in profile to those of effort or investment contracting i.e. initially high but then diminishing. The rationale for this assumption is detailed below. At this stage, it is worth being explicit that we are dealing with a complex product and a contract that can be potentially verified by a third party. Not all contracts need to be verifiable of course (again, this is discussed below). But when referring below to a product contract, it means a verifiable one. Clearly, the cost of determining a contract will depend on what that contract is designed to achieve. The more detail that is included, the less is the chance for mis-interpretation later, particularly as the state of the world reveals itself. We suppose that to set up negotiation, to commit to an agreement, to draft a legal (or legalistic) document etc., means high initial costs, even for a minimally specified contract. In addition, it is likely that some 'critical mass' of specification will be required if a product contract has any chance of being robust against potential challenge later.[20]

### 5.1.3 Non-contracting activities

As discussed, contracts need not be agreed between stakeholders for investment and effort to be made; the stakeholder actually undertaking these actions may decide to do so unilaterally. Of any total investment or effort, some proportion may be undertaken by contract and some, the residual, may be non-contracted. In notation, respectively, we have:

(5-6) $y^j = y^{jC} + y^{jN}, j = a, e$

---

[20] The detail does not need to be in each individual contract. In social care, broadly speaking, the public sector as overall commissioner has put in place a whole raft of inspection and regulation against a set of standards – as laid out in the Care Standards Act 2000 and subsequent regulation. We could certainly interpret these as part of a product contract. Individual commissions may not have to bear these costs, but taken as a whole the public sector in commissioning services from the care market, certainly does. In the past, moreover, the equivalent regulation did not apply to in-house providers.

where the $j$ superscript refers to either investment ($a$) or effort ($e$). In terms of notation, we need to be clear that any activity that is indirectly induced by a contract is regarded, nonetheless, as contracted (otherwise, no effort would be contracted effort).

Product specifications can be written in a (potentially verifiable) contract, $q_k^C$, left subject to a non-contracted agreement, denoted $q_k^N$, or not agreed at all. Because contracted and non-contracted agreements should not overlap, we can assume that:

(5-7) $q_k = q_k^C + q_k^N$

Contracted product specifications can vary. For example, contracts could have precise details as to the required product, i.e. high $q_k^C$, or could be very loosely or informally defined i.e. low $q_k^C$. When relevant specifications are left non-contracted or not agreed at all, producers have residual control.

This conception of product contracts is at a high level of abstraction. It also implicitly assumes a framework of contract law and conventions exists in the wider economy to underpin the use of product contracts. The aim is to capture the idea of a general choice between a highly specified but expensive contract, or a loosely specified, cheap agreement, both of which have been used in social care markets in England (see chapters 3 and 4, and Wistow et al., 1996). We also want to capture the distinction between a legalistic, verifiable contract and a 'relational', informal contract (Sako, 1992).

### 5.1.4 Transaction costs

To be precise about these transaction costs, we use notation that is explicit about the governance structure in use. We define $B^{ij}$ as the bargaining cost in governance structure $i = M, H$ (market or hierarchy) and to stakeholder $j = K, L$ (provider or purchaser). We assume that:

$$(5\text{-}8) \quad B^{ij} = B_k^{y^{a}ij}\left(y^{aC}\right) + B_k^{y^{e}ij}\left(y^{eC}\right) + \sum_{k=1}^{s} B_k^{qij}$$

that is, total bargaining costs are the sum of any bargaining on investment ($y^{aC}$) and effort ($y^{eC}$), and also product type $q$, for each relevant state of the world. Stakeholders only realise one investment and effort project, so although bargaining costs of these activities might differ by state, there is no need to sum them across states. Bargaining costs increase with the level of the *contracted* action, that is, $\partial B_k^{y^{a}ij}\left(y^{aC}\right)/\partial y^{aC} > 0$ and $\partial B_k^{y^{e}ij}\left(y^{eC}\right)/\partial y^{eC} > 0$. Unilaterally undertaken actions, as noted above, require no bargaining costs.

In addition we can also be more explicit about who bears (fixed) measurement costs.[21] In order to ascertain the nature as described by the parameters, φ, υ and ω, of the current state $k$, stakeholders must pay measurement cost $T^{fij}$, where, $f$ is the parameter to be measured, $f$ = φ, υ, ω, in governance structure $i = M, H$ (market or hierarchy) and to stakeholder $j = K, L$ (provider or purchaser). To contract on investment requires information on φ. To contract on $q$ (and $D$) requires φ and υ, as do instructions regarding effort.

#### 5.1.4.1 Notation

Regarding notation with respect to the parameters, $I$ = φ, υ, ω, β, a signal is denoted $I^S$. To re-cap, a signal is information a person receives without properly measuring so it is late and imperfect. A report from a party that has measured $I$ to the other party that has not, is written $\hat{I}$.

In what follows we distinguish between behaviour that is first-best and that which is optimal given transaction costs or other local constraints. In particular, $y^{e*F}$ denotes first best effort (in this case), $\tilde{y}^e$ denotes the optimal level, generally for markets and $y^{e*}$

[21] This distinction is made because people can of course measure the state of the world without having to contract. In theory measurement could be treated as variable, with a variable rather than fixed cost. This added complication does not however add to the comparison of markets versus hierarchies – assuming that any stakeholder optimises a measurement decision, the result of a lower marginal measurement cost under one governance structure compared to another is just a different total measurement cost. This is how it is included in the model.

denotes alternative optimal level – for hierarchies – where comparisons between optimal levels are warranted.

Stakeholders are initially assumed to be risk neutral. The implications of this assumption are explored latter. Nonetheless, in that we are dealing with organisations rather than individuals, risk neutrality might be argued to be appropriate.

## 5.2 Optimal governance

Governance choices involve a consideration of both organisational structure and contractual arrangements. As was discussed in previous chapters, a key dimension of organisational structure is not so much who owns the means of purchasing and providing, but rather who *controls* them. In particular, the defining characteristic of a hierarchical transaction is that one party (the "subordinate") is willing to cede considerable control of actions that directly affect them to the other party ("super-ordinates"). Market or bilateral transactions are characterised by a sharing of this control, where decisions are subject to mutual agreement.

The types of contractual arrangements used between parties tends to be related to the choice of organisational structure – see 3.3.3.2 especially of chapter 3. Nonetheless, whilst this chapter considers explicit choices between market and hierarchy, it does consider a range of relevant contractual arrangements, particularly for market structures.

In hierarchies, stakeholders cede control over the specification of actions and price setting to hierarchical super-ordinates (managers). The super-ordinate then collects information about $\varphi$, $\upsilon$ and possibly $\omega$, and determines contract specifications when they are needed. In this way, transaction costs are low: only super-ordinates have measurement costs. Bargaining costs are minimal because bargaining is one-sided and single clause (that is relating to only one state of the world, the current one). The subordinate will have to accept reports about relevant circumstances from managers, although they can validate these reports to some extent retrospectively when, through experience, they gain signals of state-of-the-world parameters.

In markets, stakeholders negotiate over all product characteristics and prices $\{q, D, p\}$; each has a stake in these decisions and a share of control rights over agreements about them. In other words, control is dispersed. Stakeholders may also contract over investment and effort actions, although this depends on the transaction costs of doing so. In order to exercise control shares, stakeholders need information on circumstances pertinent to all actions and prices. All stakeholders must have some information about circumstances $\varphi$ and $\upsilon$. If a stakeholder were to rely totally on reports of these circumstances, they would be ceding control over associated actions, and transaction would effectively be hierarchical.

This chapter is concerned with social care services transactions between purchasers and providers. To re-cap, in hierarchies, the 'purchaser' is the super-ordinate instructing the provider – as subordinate – about the actions required to produce social care. We are concerned with the interface between care planning and production, and it is transactions across this interface that are hierarchically organised. Super-ordinates need not map directly onto traditional management organisational forms. Indeed, here the key function of super-ordinates is that they collect information about $\varphi$ and $\upsilon$, and so they need not be 'senior managers' as the terms is commonly used. More likely, middle managers that might secure this information from other, non-provider subordinates (e.g. care managers), are super-ordinates.

### 5.2.1 Market

Each stakeholder collects information $\upsilon$ and $\varphi$. The purchaser and provider negotiate over the product contract (in $q$, $D$ and $P$), but the provider makes decisions about effort, $y^e$. In many cases, the provider also makes decisions about investment, $y^a$, although there are situations where purchasers and providers might directly contract on $y^a$. Generally, therefore, in addition to measurement costs, the main transaction costs concern the bargaining costs of determining the product contract. These costs will be large.

The general expected utility function for providers is (1.4) in chapter 1. Here we expand some of the terms in that function (detailing transaction costs $T$ and $B$, and also overall costs to include provider effort as while as other production costs) so that provider expected utility in markets will be:

$$(5.9)\ U=\sum_{k=1}^{S}\gamma_k\left[-y_k^a-T^{\varphi MK}+\left(p_k-c_k\left(D_k\right)\right)D_k r^{MK}-e\left(y_k^e\right)r^{MK}-T^{\upsilon MK}-B^{MK}+\psi\left(p_k\right)r^{MK}\right]$$

The superscripts on $T_k$ refer respectively to the circumstance for which information is collected, $\upsilon$ or $\varphi$; whether it is collected under market, $M$ or hierarchy, $H$; and whether collection is by the purchaser $L$ or the provider $K$. Thus, for example, $T_k^{\varphi MK}$ is the cost of collecting information about circumstance $\varphi$ under market governance by the provider. The term $B$ is the cost of contracting, and depends on the type of contract used – see next section. In most cases it will be sizeable.

The utility function includes a term, $r^{MK}$, which is the probability that, in this case the provider, will have their report of relevant information accepted (and therefore have the transaction proceed) – see section 1.4.1 in chapter 1. It allows for the provider to mis-report information if they wish, but with an attendant risk of losing the contract. This probability will depend on how much information the purchaser collects, and therefore what basis for comparison they have with the provider's report. In particular, $r^{MK}\left(\beta,\hat{\beta},\omega,\hat{\omega}\right)<1$ if the report differs from the actual value (where we have assumed that the purchaser's signal is unbiased): either, $\hat{\beta}\neq\beta$ or $\hat{\omega}\neq\omega$. If the misrepresenting party, in this case the provider, is caught then we assume that payments are forfeit: the purchaser expresses zero demand ($D = 0$), otherwise demand is $D_k$ – see chapter 1. Therefore, in the above utility function, with probability $r^{MK}$, the provider's output is $D$ with unit revenue $p$. Expected demand is $x_k=r^{MK}D$.

The term $\psi$ reflects disutility from excessive pricing, which limits access to the service. Changes in prices, beyond some baseline price, increase disutility: $\psi'<0$. Nonetheless, increases in profit per unit output, and so price, always yield net marginal increases in utility, other things equal, that is: $\psi'(\infty)>-1$ (assumption 1-1 – see chapter 1).

To sum up, provider utility is expected profit (price, $p$, minus production cost $c$ for each unit expected output $x$) minus investment cost $y^a$, minus expected effort cost $e(y^e)$, minus contracting/bargaining costs, $B$, minus measurement costs $T$ and finally adjusted for disutility $\psi$.

Overall then in state *k*, provider utility is

$$
(5.10)\quad \begin{aligned} U &= -y_k^a - T^{\varphi MK} - T^{\upsilon MK} + U^{II} \\ &= -y_k^a - T^{\varphi MK} - T^{\upsilon MK} + (p_k - c_k)x_k - r^{MK} e(y_k^e) - B^{MK} + r^{MK}\psi(p_k) \end{aligned}
$$

where $U^{II}$ is second period utility.

For the purchaser utility is:

$$
(5.11)\quad Z = \sum_{k=1}^{S} \gamma_k \left( r^{MK} \left( \int_0^{D_k} R_k(x)dx - p_k D_k \right) + \left(1 - r^{MK}\right)z - T_k^{\upsilon ML} - T_k^{\varphi ML} - B^{ML} \right)
$$

The purchaser's utility depends on whether the current provider is discharged (for misrepresentation). If the provider is not discharged then output is *D*; otherwise, it is assumed that the purchaser can find a replacement provider, and secures utility *z*. As a shorthand we can write $Z^0 = \left(1 - r^{MK}\right)z$.

#### 5.2.1.1 Contract types

Consider first *adaptive contracts*, where stakeholders wait until all output relevant parameters of the state of the world can be measured. In the present model, this means $\varphi$ and $\upsilon$, and so contracts are agreed at time 2. In other words, the provider makes a unilateral decision about investment, without having agreed a contract. When they are determined, these 'CA' contracts take the form of a service specification *q*, *D*, and payment for provision. Since most relevant information is known at time 2, the market appraisal (giving $\varphi$) and the (pre-care) assessment (giving $\upsilon$) have already been undertaken, service specification is straightforward; the contract only has to cover the current situation, not all possible circumstances. Bargaining costs are thus low compared with having to make specifications for all possible contingencies. Payment involves a division of period 2 surplus. We assume, in particular, that stakeholders write contracts based on the expected utilities (in state *k*) at the end of period 2, which are respectively:

$$U^{II} = (p_k - c_k)x_k + r^{MK}\psi(\rho_k) - r^{MK}e(y_k^e) - B^{MK} \tag{5-12}$$

for the provider, and

$$Z^{II} = r^{MK}\int_0^{D_k} R_k(x)dx - r^{MK}p_kD_k + Z^0 - B^{ML} = -\int_{p=R(rD_k)}^{\infty} x_k(p)dp + Z^0 - B^{ML} \tag{5-13}$$

There are many ways that the purchaser and provider could agree how to share the transaction surplus.[22] In markets, we will assume a Nash bargaining model ((Nash, 1953)). Bargaining over price occurs to allocate the exchange surplus between stakeholders. To simplify the problem, assume (innocuously) that $\psi(\rho_k) = D\Psi(\rho_k)$ i.e. that disutility of excessive prices is felt per unit output.[23] The standard result is the sharing of surplus is according to a share parameter $\mu \in (0,1)$. Surplus, in this case, is the sum of second period utility i.e. $U^{II} + Z^{II}$ from (5-12) and (5-13). Using (5-5) we have:

$$U^{II} = \mu r^{MK}S + \mu\left(r^{MK}\psi - r^{MK}e(y^e) + Z^0 - B^{MK} - B^{ML}\right) \tag{5-14}$$

and

$$Z^{II} = (1-\mu)r^{MK}S + (1-\mu)\left(r^{MK}\psi - r^{MK}e(y^e) + Z^0 - B^{MK} - B^{ML}\right) \tag{5-15}$$

In other words the provider gets a share $\mu$ of the surplus with the purchaser gets the rest. A full derivation is in Annex 5-1. State $k$ total utility, from (5.9), is $U_k = U^{II} - y_k^a - T^{\varphi MK} - T^{\upsilon MK}$, or substituting the above:

$$U_k = \mu r^{MK}S + \mu\left(r^{MK}\psi - r^{MK}e(y^e) + Z^0 - B^{MK} - B^{ML}\right) - y_k^a - T^{\varphi MK} - T^{\upsilon MK} \tag{5-16}$$

This is just adding the investment and measurement costs that are borne at time 1. This means that total purchaser utility is:

$$Z_k = (1-\mu)r^{MK}S + (1-\mu)\left(r^{MK}\psi - r^{MK}e(y^e) + Z^0 - B^{MK} - B^{ML}\right) - T_k^{\upsilon ML} - T_k^{\varphi ML} \tag{5-17}$$

---

[22] It is important that they share the surplus efficiently, but the material result is that they do share, and not so much by how much exactly.

[23] Assumption 1-1 means: $\psi'(\rho_k) = D\Psi'(\rho_k) > -1$ (see chapter 1).

The second option is the *full contingent contract* that involves determining a series of specifications and associated reimbursements for all potential states of the world. This type of contract is really only a theoretical possibility. As is universally assumed in the literature, the transaction costs of actually producing a full contingent-claims contract are prohibitive (see chapter 2). We will set it aside at this point, although for completeness, this form of contract is detailed in Annex 5-2.

A third alternative is the *cost sharing* or incentive contract. It is based closely on the contract type discussed informally in section 2.2.4.2 and Annex 2-3 where individuals do write a contract to cover investment costs at the time of investment because they can anticipate and agree on what benefits each will accrue from that investment and so share the its costs accordingly.

This 'CS' contract has two parts. At time 1, stakeholders contract on investment agreeing to share investment costs in accordance with their expectation of net utility accrued for each possible $\upsilon$, given the investment in the first period (and circumstance $\varphi$). At time 2, once the actual value of $\upsilon$ is known, parties contract for product specification and prices given period 1 investment.

The first period problem is to choose the shares, $\phi^C = 1 - \phi^R$ for provider and purchaser respectively, of investment cost $y^a$, over utilities at time 1, which are

$$(5\text{-}18)\quad U^{I2} = -\phi^C y^a + E\left[U^{II}\left(y^a\right)\middle|\upsilon\right] - T^{\varphi MK} - T^{\upsilon MK}$$

$$(5\text{-}19)\quad Z^{I2} = -\phi^R y^a + E\left[Z^{II}\left(y^a\right)\middle|\upsilon\right] - T^{\varphi ML} - T^{\upsilon ML}$$

and $\phi^C + \phi^R = 1$ (the superscripts mean state 'I' and version '2' of the function to distinguish the CA contract case). Expected utility in period 2 is given by (5-14) and (5-15) respectively to give:

$$(5\text{-}20)\quad U^{I2} = -\phi^C y^a + \mu\Pi\left(y^a\left(\phi^C\right)\right) - T^{\varphi MK} - T^{\upsilon MK}$$

and

(5-21) $Z^{I2} = -\phi^R y^a + (1-\mu)\Pi\left(y^a\left(\phi^C\right)\right) - T^{\varphi ML} - T^{\upsilon ML}$

where $\Pi = r^{MK}S + r^{MK}\psi - r^{MK}e\left(y^e\right) + Z^0 - B^{ML} - B^{MK}$.[24] Differentiating these functions respectively with respect to the share gives:

(5-22) $\dfrac{\partial U^{I2}}{\partial \phi^C} = -\phi^C y^a_{\phi^C} - y^a + \mu\Pi_{y^a} y^a_{\phi^C}$

and

(5-23) $\dfrac{\partial Z^{I2}}{\partial \phi^R} = -\phi^R y^a_{\phi^R} - y^a + (1-\mu)\Pi_{y^a} y^a_{\phi^R}$

which simply indicates how the purchasers or providers utility would change if their share changed. Now, $y^a_{\phi^R} = y^a_{\phi^C}\dfrac{\partial \phi^C}{\partial \phi^R} = -y^a_{\phi^C}$ and $\phi^R = 1-\phi^C$, so:

(5-24) $\dfrac{\partial Z^{I2}}{\partial \phi^C} = \dfrac{\partial Z^{I2}}{\partial \phi^R}\dfrac{\partial \phi^R}{\partial \phi^C} = -y^a_{\phi^C} + \phi^C y^a_{\phi^C} + y^a + \Pi_{y^a} y^a_{\phi^C} - \mu\Pi_{y^a} y^a_{\phi^C}$

Suppose we are comparing this situation with the CA contract situation (where $\phi^C = 1$). There are two options. First, the purchaser will happily take a share of the investment cost so long as $\dfrac{\partial Z^{I2}}{\partial \phi^C} \le 0$. However, when $\dfrac{\partial Z^{I2}}{\partial \phi^C} = 0$, the provider is likely to see $\dfrac{\partial U^{I2}}{\partial \phi^C} < 0$ i.e. the provider still has gains to make. The second option is therefore for the provider to offer further compensation to the purchaser in return for a higher cost share. This could continue until,

(5-25) $\dfrac{\partial U^{I2}}{\partial \phi^C} + \dfrac{\partial Z^{I2}}{\partial \phi^C} = 0$

---

[24] Just to clarify the notation, strictly some of the bargaining costs in the $B$'s are incurred in the first period and so $\Pi$ includes not only expected second period profit, but also first period bargaining costs.

At a share of $\tilde{\phi}^C$, any small change in $\phi^C$ implies a gain to one party that is completely offset by the loss to the other: $\frac{\partial U^{I2}}{\partial \phi^C}\left(\tilde{\phi}^C\right)+\frac{\partial Z^{I2}}{\partial \phi^C}\left(\tilde{\phi}^C\right)=0$. Put another way, at values $\phi^C \neq \tilde{\phi}^C$, then we would have $\frac{\partial U^{I2}}{\partial \phi^C} \neq \frac{\partial Z^{I2}}{\partial \phi^C}$ and it would be possible for one party to compensate the other party for their loss as a result of the change in $\phi^C$ and still gain. In other words (noting the transaction costs), the CS contract allows purchaser and provider to jointly maximise their utilities.

### 5.2.2 Hierarchy

Hierarchies are characterised by top-down management as decision-making authority is vested in managers and ceded by employees, who accept the plans of managers. Planning the organisation's activities, such as what should be produced and how, is undertaken by mangers and these decisions then govern actual production. Generally, subordinates retain only decisions about their own effort.

In terms of the model, managers (purchasers) collect information $\varphi$ and $\upsilon$, and determine $D$, $q$ and $p$. Indeed, that 'providers' do not collect information $\varphi$ and $\upsilon$ is what meaningfully differentiates hierarchies from markets. To not collect this information and rely on reports from other parties, is to be effectively ceding control over decisions to that party, being unable to verify with any certainty the actual conditions that pertain. A central characteristic of hierarchical arrangements therefore is the lack of duplication of governance tasks, particularly, measurement and bargaining, as both parties collect relevant information and contract accordingly (Kreps, 1996).

Managers in hierarchies make decisions about investment, $y^a$, and bear production costs, $c$. They also provide instructions to producers that have a direct bearing on effort $y^e$. These instructions concern activities that employees undertake that both produce the desired quality and quantity of service and determine the costs of production. However, as transaction costs are positive, there are always some residual, un-contracted aspects of effort. Hence, the producer/employee has some control over $y^e$. Hierarchies differ from market-based transactions in that an explicit product contract is not (usually) agreed. Since control is ceded to the super-ordinate who plans the nature of the required product, $q$, there is no bargaining required to determine it, and therefore no duplication of effort as in a market. Instead, a 'contract' in the form of instructions to subordinates – conceived here as directions regarding efforts – serves as the means to secure output. These instructions are by fiat and often adaptive and so, whilst subject to influence, are likely to have much lower transaction costs than formal product contracts.[25]

Utility for the purchaser is:

(5-26)

$$Z = \sum_{k=1}^{S} \gamma_k \left[ r^{HL}\left(\varphi_k^S, \hat{\varphi}_k, \upsilon_k^S, \hat{\upsilon}_k\right)\left(\int_0^D R_k(x)dx - c_k D_k - \left(p_k - c_k\right)D_k\right) - T_k^{\upsilon HL} - y_k^a - T_k^{\varphi HL} - B^{HL}\right] + \sum_{k=1}^{S} \gamma_k \left[\left(1 - r^{HL}\left(\varphi_k^S, \hat{\varphi}_k, \upsilon_k^S, \hat{\upsilon}_k\right)\right)Z^M\right]$$

or writing $z^W = \int_0^D R_k(x)dx - c_k D_k - \left(p_k - c_k\right)D_k$, it is more succinctly,

$$(5.27)\quad Z = \sum_{k=1}^{S} \gamma_k \left[r^{HL} z^W - T_k^{\upsilon HL} - y_k^a - T_k^{\varphi HL} - B^{HL} + \left(1 - r^{HL}\right)Z^M\right]$$

The main difference here compared to the market case is that now the 'purchaser' – i.e. the hierarchical super-ordinate – directly bears investment and production costs. In period 1, the purchaser pays cost $T_k^{\varphi HL}$ to determine state $\varphi$ and cost $y_k^a$ to make an investment of that level. In period 2, the purchaser informs the provider i.e. the sub-ordinate of production requirements. If the sub-ordinate complies then the payoff for the super-ordinate is $\int_0^D R(x)dx$ less the non-wage costs of production $cD$ and the 'wage rent', which is $p_k^w D_k = \left(p_k - c_k\right)D_k$.

Hierarchical purchasers direct the provider's efforts. Although the state of the world is known – so that instructions only need to cover current circumstances – like investment, production effort is complicated and multi-dimensional, and specification and

[25] In relation to the literature (i.e. Kreps, 1990a) we are dealing with ceding of control as in the employer-employee relationship, and not with the case where the purchaser/principal cedes control to the provider (although the same principles hold of reducing transactions costs against the chance of being exploited).

monitoring of provider actions will be a costly task. The term $B^{HL}$ represents these costs. We assume that this cost varies according to how thorough the instructions are regarding how the provider is to deploy efforts. To model this feature, suppose that total effort has two components, effort that is contracted, $y^{eC}$, and effort that is not, $y^{eN}$, so that:

$$(5\text{-}28)\quad y^{e} = y^{eC} + y^{eN}$$

Note that transaction costs are increasing in contracted effort: $B^{HL}_{y^{eC}} > 0$.

5.2.2.1 Public sector hierarchies and publicly funded services

The probability that the sub-ordinate will comply with instructions depends on how truthful the super-ordinate is about relevant prevailing circumstances (i.e. about φ and υ). Since the provider/sub-ordinate does not collect information about prevailing circumstances at this stage, instead receives only a (noisy) signal $\varphi^S$ and $\upsilon^S$, there is an opportunity for the purchaser to misrepresent its value and thereby reduce the subordinates payment, and hence their utility given circumstances φ and υ. In practical terms this assumes that the purchaser undertakes the assessment and then reports the results to the provider. The provider will gain information later during the provision of care, but at this stage relies on the purchaser's report (along with the noisy signal).[26] The super-ordinate's choices regarding such 'exploitation' turn on what happens when the provider detects misrepresentation, the utility consequences of which are $Z^M$.

Providers that detect misrepresentation have two options. First, they can simply quit. Second, they can stay on but scale back effort i.e. to shirk on effort by not complying fully with manager's effort instructions. Shirking involves subordinates misrepresenting the value of ω, in order to hide reduced effort. Super-ordinates could in theory measure this private information ω and therefore automatically detect shirking. However, doing so involves substantial transaction costs $T^{\omega HL}$. Also, even if they do measure, super-ordinates may still prefer not to sack 'shirkers' because to attempt to do so when the

[26] There are of course principal-agent problems as between the 'management' super-ordinate and the staff actually conducting the assessment. These staff are employees but not providing employees and this distinction is important in a comparison with providers in markets.

purchaser is misrepresenting $\varphi$ and/or $\upsilon$, is likely to damage its reputation. When fighting dismissal or attempting to gain some compensation it is very much in the interest of ex-employees to broadcast their grievances regarding the exploitation by the super-ordinate. There may also be other safeguard to protect the employee e.g. in employment law (see in particular Williamson, 1994) and/or from unions. The consequence is that even when measuring $\omega$, some employees could shirk and keep their jobs.

If shirkers are dismissed this occurs before production but the adverse consequences to the super-ordinate's reputation will mean to hire replacement employees will cost more than wages paid when there is no misrepresentation of $\varphi$ and/or $\upsilon$ (see next section for details of how subordinates are replaced). The same situation applies when providers decide to quit themselves (the first option above). If shirking occurs, sub optimal effort will mean lower utility for the super-ordinate, or if re-hiring (or renegotiation of employees remuneration packages) occurs this will involve a re-contracting cost/higher wages.

Overall then these assumptions are to convey that managers face a potential cost if they try to exploit their employees *too much*. These assumptions are really a question of degree, but it seems reasonable to assume that at some point, exploitation goes too far and the repercussions affect the organisation. We make these assumptions because any resultant shirking or re-hiring cost will be inefficient relative to the first best situation, and of course in comparison with markets.

In model terms, the utility the super-ordinate can expect if their misrepresentation of $\varphi$ and/or $\upsilon$ is detected will be less than that level they would have accrued had no misrepresentation taken place. We denote this difference as $m^S(\theta^H)$ for the shirking case or $m^R(\theta^H)$, for the renegotiation/replacement case where $m^k(\theta^H) > 0, \forall\theta^H > 0, k = S, R$ noting that the size of super-ordinate's misrepresentation is $\theta^H = \theta^H(\varphi_k^S(\varphi_k), \hat{\varphi}_k, \upsilon_k^S(\upsilon_k), \hat{\upsilon}_k)$, with $\theta^H = 0$ being no misrepresentation.[27] In the limit

[27] Within hierarchies common ownership and employment, and continuous, adaptive and often relatively informal and inter-personal interaction between super-ordinate and subordinate will facilitate the

case, the costs of *detected* misrepresentation may be so large as to make the hierarchy untenable so that the funder switches purchasing to market-based providers (see chapter 3). Then super-ordinate utility will be at their reservation, $Z^0$. Whatever the final outcome, it is also reasonable to expect that the cost $m$ will also be non-negatively related to $\theta^H$: $m_\theta \geq 0$.

If purchasers do not directly measure $\omega$ – because, say, measurement is too expensive – they must rely on a noisy signal. This gives providers a (further) opportunity to shirk on effort, claiming in their report $\hat{\omega}$, that $\omega$ is more adverse than the purchaser believes. We will stylise this process somewhat for convenience. At the beginning of the production period, providers make their report $\hat{\omega}$. This is accepted with probability of $r^{HK}(\omega^S, \hat{\omega}) \in [0,1]$, in which case the current provider goes on to produce output $D$ with instructions $y^{eC}$, based on circumstances $\hat{\omega}$. Otherwise, the provider is dismissed, and a new sub-ordinate is hired, who in turn makes a report $\hat{\omega}$.

Providers that quit because they have detected purchaser misrepresentation receive utility $U^M = \tilde{U}(\theta^H = 0)$, the utility of a market based provider given conditions $\varphi$ and $\upsilon$. Providers that are released because their shirking was detected find other employment to receive utility $U^0$; moreover because they get a 'bad reference' since they were sacked rather than resign, $U^0 < U^M$. In this case, when providers detect misrepresentation by the super-ordinate, $U^0$ is higher than when no detection occurs because the credibility of the purchaser's bad reference suffers in the former case. Providers that are not released receive wages $w_k$, payable in arrears (which can be written $w_k = p_k^w D_k$, the unit wage rent $p_k^w$ by output). Their utility is $u^W = (w - e(y_k^e) + \psi\{\rho_k\})$, wages less the disutility of both effort and excessive prices. Wages are the subordinates share of the transaction surplus, and so do not change underlying costs. Changes in wages, in other words, do not change optimal output.

The provider's utility is:

---

development of social relationships between stakeholders. As such super-ordinates may suffer direct disutility from the deliberate exploitation of fellow workers.

$$(5.29)\quad U=\sum_{k=1}^{S}\gamma_k\left[r^{HL}\left(r^{HK}\left(w_k-e\left(y_k^e\right)+\psi\{\rho_k\}\right)+\left(1-r^{HK}\right)U^0\right)+\left(1-r^{HL}\right)U^M\right]$$

or, writing $u^W=\left(w-e\left(y_k^e\right)+\psi\{\rho_k\}\right)$, and considering just state $k$,

$$(5\text{-}30)\quad U_k=r^{HL}\left(r^{HK}u^W+\left(1-r^{HK}\right)U^0\right)+\left(1-r^{HL}\right)U^M$$

To clarify somewhat, if there are no information shortfalls, the provider's utility is $U_k=u^W=\left(w-e\left(y_k^e\right)+\psi\{\rho_k\}\right)$ i.e. wages less effort costs and any disutility the provider feels about service user's access to services. Using (5-4), surplus can be distinguished by payment, i.e., $\int_0^D R(x)dx-pD+(p-c)D=S$. Substituting wage rent into (5-26) gives,

$$(5.31)\quad Z=\sum_{k=1}^{S}\gamma_k\left[r^{HL}\left(S_k-w_k\right)-y_k^a-T_k^{\upsilon HL}-T_k^{\varphi HL}-B^{HL}+\left(1-r^{HL}\right)Z^M\right]$$

#### 5.2.2.2 Ceded control and misrepresentation

In hierarchies in period 1, the super-ordinate (i.e. the manager and in this case the purchaser) measures $\varphi$, chooses investment $y^a$, and makes a report $\hat{\varphi}$ to the provider. To be able to hire subordinates – given a good reputation – the super-ordinate needs to at least equal the provider's opportunity utility. Assuming some form of market does exist, the utility a provider could expect if they moved to providing in that market serves as a good benchmark. This benchmark applies given subordinates perceptions of the values of $\varphi$ and $\upsilon$, opportunity utility is a function of super-ordinate reports about the state-of-the-world or equivalently misrepresentation, $\theta^H$, that is: $\tilde{U}\left(\theta^H\right)$. This is an *ex ante* determined value and is assumed to be taken as constant during the ensuing deliberations of hierarchical stakeholders.

This level sets the wages paid in arrears to the provider:[28]

(5-32) $u^W = \tilde{U}(\theta^H)$

and so,

(5-33) $w \equiv \tilde{U}(\theta^H) + e(y_k^e) - \psi$

The maximisation problem where super-ordinates rely on a signal of ω is:

(5.34) $\max_w Z = r^{HL}(\theta^H)(S_k - w_k) - y_k^a - T_k^{\upsilon HL} - T_k^{\varphi HL} - B^{HL} + (1 - r^{HL}(\theta^H))Z^M$

subject to

$w \geq \tilde{U}(\theta^H) + e(y_k^e) - \psi$

Substituting for *w* gives:

(5-35) $\max_w Z = r^{HL}(\theta^H)(S_k - \tilde{U}(\theta^H) - e(y_k^e) + \psi) - y_k^a - T_k^{\upsilon HL} - T_k^{\varphi HL} - B^{HL} + (1 - r^{HL}(\theta^H))Z^M$

If providers accept the claims of purchasers about circumstances φ and υ, their utility is, from (5-32), $u^W = \tilde{U}(\theta^H)$. Super-ordinates have an incentive to claim that prevailing circumstances were deleterious so that providers should only expect a low payout and so low $\tilde{U}(\theta^H)$. Suppose that the chance of avoiding detection $r^{HL}$ is a decreasing function of $\theta^H$, such that $r_{\theta^H}^{HL} < 0$, for all $\theta^H$. Then the choice for the purchaser regarding $\theta^H$ can be found by differentiating *Z* with respect to $\theta^H$.

[28] Strictly, because providers might shirk, be detected and so dismissed, the amount the purchaser/super-ordinate expects to pay is: $r^{HK}u^W + (1 - r^{HK})r^{HK}u^W + (1 - r^{HK})^2 r^{HK}u^W + (1 - r^{HK})^3 r^{HK}u^W + \ldots$ or summing, $r^{HK}u^W\left[\frac{1}{1-(1-r^{HK})}\right] = u^W$. It is this function that is set to: $\tilde{U}(\theta^H)$.

The value of $Z$ will depend on whether subordinates detect misrepresentation, and if so, how they respond, as outlined above. To make these considerations clear, we can write the relevant functions i.e. $S(\theta,\breve{\theta})$, $e(\theta,\breve{\theta})$, $z^{W}(\theta,\breve{\theta})$ and $Z^{M}(\theta,\breve{\theta})$ being explicit about their dependence on, respectively, actual misrepresentation and misrepresentation detected by the subordinate, the latter being $\breve{\theta}$.

$$(5\text{-}36)\quad \max_{w} Z = r^{HL}(\theta^{H})z^{W}(\theta^{H},0) - y_{k}^{a} - T_{k}^{\upsilon HL} - T_{k}^{\varphi HL} - B^{HL} + (1 - r^{HL}(\theta^{H}))Z^{M}(\theta^{H},\theta^{H})$$

where, $z^{W}(\theta^{H},0) = S(y^{e}(\theta^{H},0)) - \tilde{U}(\theta^{H}) - e(y^{e}*(\theta^{H})) + \psi(\theta^{H}) \equiv S(y^{e}*) - w(\theta^{H},y^{e}*)$, noting that when providers do not detect misrepresentation they put in optimal instructed effort:[29] $S(y^{e}(\theta^{H},0)) = S(y^{e}*)$. In other words, $z^{W}(\theta^{H},0)$ is the super-ordinate's utility given that subordinates do not detect misrepresentation of $\varphi$ or $\upsilon$, and so super-ordinates can *expect* their effort instructions to be followed, where the required effort levels are optimal.[30] Thus, for each $\theta^{H}$ the subordinate is happy, prior to production, to accept $w(\theta^{H},y^{e}*)$, the wage decided by the super-ordinate.

Alternatively, if misrepresentation is detected after the super-ordinate decides payment $w(\theta^{H},y^{e}*)$, so that it is treated as fixed, then the subordinates can be expected to shirk. Thus, utility in this case is the no (detected) misrepresentation level less $m^{S}(\theta^{H})$ as outlined in section 5.2.2.1 i.e.,

$$(5\text{-}37)\quad \begin{aligned} Z^{M}(\theta^{H},\theta^{H}) &= z^{W} - m^{k}(\theta^{H}) \\ &= S(y^{e}) - w(\theta^{H},y^{e}*) - m^{k}(\theta^{H}) \end{aligned}$$

---

[29] Although optimal, $y^{e}*$, we show below that effort is not first best $y^{e*F}$

[30] Note that $\theta^{H}$ is the degree of misrepresentation – it does not affect the actual value of underlying parameters, $\varphi$ and $\upsilon$. Other things equal optimal values of effort would be based just on the actual value of these parameters. However, it may be the case that $y^{e*}$ choices are (potential) signals of $\theta^{H}$, but is only follows if the impact of effort on $S$ is mediated by $\varphi$ and $\upsilon$. If instead, $\varphi$ and $\upsilon$ simply shift the $S$ function, then $y^{e*}$ is not a function of $\varphi$ and $\upsilon$ and no is no signalling effect. It is an innocuous simplification to assume the latter case to hold. Even in the former case, a lot is assumed by way of knowledge of precise functional relationships by subordinates for them to appreciate the signalling effect.

Assuming that purchasers make optimal decisions about investment for all $\theta^H$, i.e. $Z_{y^a} = r^{HL} z^W_{y^a} - 1 - B_{y^a} + \left(1 - r^{HL}\left(\theta^H\right)\right) Z^M_{y^a} = 0$, then $Z_{y^a}\left(y^{a*}\right) y^a_\theta = 0$, and so,

$$(5\text{-}38)\quad \frac{\partial Z}{\partial \theta^H} = r^{HL}_{\theta^H} z^W\left(\theta^H, 0\right) + r^{HL} z^W_{\theta^H} + \left(1 - r^{HL}\right) Z^M_\theta - r^{HL}_{\theta^H} Z^M\left(\theta^H, \theta^H\right)$$

or using (5-37),

$$(5\text{-}39)\quad \begin{aligned} \frac{\partial Z}{\partial \theta^H} = {} & r^{HL}_{\theta^H}\left(S\left(y^e *\right) - w\left(\theta^H, y^e *\right)\right) - r^{HL} w_{\theta^H} - \left(1 - r^{HL}\right) w_{\theta^H} \\ & + \left(1 - r^{HL}\right)\left[S_{y^e}\left(y^e\right) y^e_\theta - m^S_\theta\right] - r^{HL}_{\theta^H} Z^M \end{aligned}$$

or

$$(5\text{-}40)\quad \frac{\partial Z}{\partial \theta^H} = r^{HL}_{\theta^H}\left(S\left(y^e *\right) - w\left(\theta^H, y^e *\right) - Z^M\right) - w_{\theta^H} + \left(1 - r^{HL}\right)\left[S_{y^e}\left(y^e\right) y^e_\theta - m^S_\theta\right]$$

Alternatively, if purchasers simply re-set wages to no misrepresentation levels if misrepresentation is detected, then the differential is[31]:

$$(5\text{-}41)\quad \frac{\partial Z}{\partial \theta^H} = r^{HL}_{\theta^H}\left(S\left(y^e *\right) - w\left(\theta^H, y^e *\right) - Z^M\right) - r^{HL} w_{\theta^H} + \left(1 - r^{HL}\right)\left[-m^R_\theta\right]$$

Thus, in either the no detection or detection case payment $w$ to subordinates is reduced in line with *claimed* circumstances $\varphi$ and $\upsilon$ that imply $\theta^H$. However, in the latter case, subordinates reduce effort in response, and so surplus $S$ is also reduced. Indeed, along with the expected replacement costs $m$, the fall in $S$ is the penalty the super-ordinate faces if misrepresentation is detected.

---

[31] In practice, there are significant costs to re-setting wages. First, working relationships with providers are already soured and increasing wages may be insufficient to rectify the damage. In other words, providers may still shirk even with increased wages. Second, restoring wages might be difficult (a) because it sends out clear signals of purchaser's exploitative behaviour to government, users and potential employees. And (b), to restore effort, wages would have to rise to a level higher than the no-misrepresentation level (so as to overcome the trust deficit), and this might not be feasible within budget constraints. Therefore, unless shirking has a very detrimental second order effect on $S$, these significant costs - meaning that $m^R > m^S$ – suggests that super-ordinates would prefer to let subordinates to shirk. We will suppose this to be the case in what follows.

What then is the sign of (5-40)? At $\theta^H = 0$, by definition, $r^{HL} = 1$ and $r^{HL}_{\theta^H} = 0$, and therefore $\left.\frac{\partial Z}{\partial \theta^H}\right|_{\theta^H=0} = -w_{\theta^H} > 0$. This follows because $w_{\theta^H} < 0$ (see (5-33)); super-ordinates lower payment to subordinates below the agreed tariff under the guise that circumstances, φ or υ are worse than then subordinate expects. Indeed, the benefits of misrepresentation are to reduce the share of exchange surplus that goes to the provider.

Overall at $\theta^H = 0$, there are marginal benefits to misrepresentation. In turn, there is a positive probability that either providers shirk or have wages above market rates in public sector hierarchies.

But how much will super-ordinates represent $\theta^H > 0$? As regards the first term of (5-40), when $\theta^H > 0$ we have:

$$
\begin{aligned}
&S\left(y^{e*}\left(\theta^H,0\right)\right) - w\left(\theta^H, y^{e*}\right) - Z^M\left(\theta^H,\theta^H\right) \\
(5\text{-}42)\quad &= S\left(y^{e*}\left(\theta^H,0\right)\right) - w\left(\theta^H, y^{e*}\right) - S\left(y^{e}\left(\theta^H,\theta^H\right)\right) + w\left(\theta^H, y^{e*}\right) + m^S\left(\theta^H\right) \\
&= S\left(y^{e*}\left(\theta^H,0\right)\right) - S\left(y^{e}\left(\theta^H,\theta^H\right)\right) + m^S\left(\theta^H\right) \\
&> 0
\end{aligned}
$$

When misrepresentation occurs providers shirk and so effort is not at its optimal level and net surplus is less than is optimal value. Also, $r^{HL}_{\theta^H} < 0$ by construction for $\theta^H > 0$ and so the first term in (5-40) is negative when $\theta^H > 0$ (and zero when $\theta^H = 0$). The second term remains positive. Replacement cost $m$ is positive, and non-negatively related to θ. Detected misrepresentation induces shirking so $y^e_{\theta^H} < 0$. The marginal effect of a fall in effort is to reduce surplus, $S_{y^e} > 0$, which means that the last term is negative. Overall, for small increases in $\theta^H$ above zero, the sign of (5-40) is ambiguous. For more significant increases, because of its second order effect $r^{HL}_{\theta^H\theta^H} < 0$, the first term will dominate and the differential $Z_\theta$ will reduce to zero. Also, as $r^{HL} \rightarrow 0$, we have $S_{y^e}\left(\theta^H, y^e\right)y^e_\theta < w_\theta < 0$. This follows because providers are likely to shirk to an extent such that effort savings at least corresponds to their loss in income (just prior to the production phase) when payment is set at the level given $\theta^H$ with expected *optimal* effort i.e. when provider utility is:

(5-43) $u^W = \left(w\left(\theta^H, e\left(y_k^e *\right)\right) - e\left(y_k^e\right) + \psi\right)$

To restore their utility with a change in effort, we have $u_0^W = w_\theta - e_{y^e} y_\theta^e = 0$.

Furthermore, because effort is then sub-optimal, marginal benefits of effort exceed marginal costs: $S_{y^e}\left(\theta^H, y^e\right) > e_{y^e}$ given the assumed second order effects. Hence, we have

(5-44) $S_{y^e}\left(\theta^H, y^e\right) y_\theta^e < w_\theta < 0$

since $y_\theta^e < 0$.

Brushing over the detail, the above indicates that hierarchical managers face a trade-off. They benefit from using their better information to paint the world as more adverse than it really is, and so get more out of providers for less. But they always accept there is a chance that this ploy will backfire. If it does the result is production that falls short of the most efficient i.e. reducing the surplus that the purchaser values. This begins to convey that inefficient behaviour can occur in hierarchies even before any consideration of transaction costs.

The problem for purchasers as modelled above is entirely instrumental. However, in addition to the penalties in *m* (which apply only if the stakeholder is *caught*), purchasers may intrinsically dislike exploiting providers (even modestly).[32] This effect may be

[32] Reputation effects will strengthen the negative marginal effect. Suppose the cost *m* represents not just chance of losing the current surplus, but also all potential future transactions and associated net surpluses. Unless discount rates are very high, or expected repeat transactions very low, the present value of the opportunity cost *m* stream will be many multiples higher than the single period *m*. This is a standard folk theorem result (see Fundenberg and Tirole, 1992; Kreps and Wilson, 1982). Furthermore, super-ordinates may intrinsically dislike opportunistic behaviour. Its discovery could damage social relationships, or simply cause a 'loss of face' for super-ordinates. In this model, the purchaser is an agent for users collectively and indeed, the population of users of care services would benefit from opportunistic behaviour by the super-ordinate, through some combination of quantity and quality of service. In this regard therefore, for the super-ordinate to suffer disutility from misrepresentation, they are introducing imperfections into the agency relationship with users. Nonetheless, organisation theory would indicate that social relationships within organisations can develop to create these effects (for a pertinent overview

modelled as akin to a transaction cost in the purchaser's utility function, $Z$, with this transaction cost an increasing function of $\theta$ e.g. $B^{m}(\theta^{H})$ where $B^{m}_{\theta}(\theta^{H}) > 0$. Then even at $\theta^{H} = 0$, (5-40) reduces to:

$$(5\text{-}45)\quad \left.\frac{\partial Z}{\partial \theta^{H}}\right|_{\theta^{H}=0} = -w_{\theta^{H}} + B^{m}_{\theta}$$

which need not be positive if the intrinsic dislike is strong.

### 5.2.3 Where no one undertakes governance tasks

Above, we have shown how governance structure alternatives are distinguished according to the allocation of governance tasks. The implicit assumption is that some level of governance tasks is worth undertaking. In particular, that it is worth someone measuring the state of the world and writing a product specification. We could imagine a provider unilaterally investing and producing some generic product based only on expectations of $\varphi$ and $\upsilon$, and then selling it in a simple transaction to the purchaser. The problem with this case is that the product is likely to be very ill-fitting with the actual state of the world, and thereby of low value to the purchaser, such that they may not wish to buy the product at any price that covers its cost. Moreover, both investment and effort levels will be below first best levels. Overcoming these problems could certainly justify the transaction costs, especially for complex and multi-dimensional products where the efficiency shortfalls will be large.

## 5.3 Transaction benefits and costs

### 5.3.1 Co-ordination

Hierarchies are characterised by poor incentives to cut costs and to be responsive to changes in demand. These efforts are not rewarded (at least not directly) if providers are salaried, as is usually the case in hierarchies. Likewise, altering supply (e.g. from residential to home-based care) generates few (direct) benefits for providers but all the

---

see Smelser and Swedberg, 1994). Also, note that whilst the purchaser's efforts are ultimately to satisfy the objective of best value for users, lines of accountability for purchasers are (at present) hierarchically arranged with central and local government as super-ordinates. Such a configuration is entirely consistent with purchasers wishing to maximise their surpluses, but can allow opportunities for purchaser's own preferences to work through.

costs. In markets the converse is true: efficient, targeted production is rewarded with higher profits, at least in theory. The size of the problem in hierarchies depends crucially on how closely providers are managed. If their instructions include explicit activities to cut costs or to adapt supply then the problem is minimised. But will *managers* have incentives to put in sufficient effort to make these detailed instructions? Also, managers can expect 'influence' from the provider division that will require efforts to appease. In any case in hierarchies individual initiative is likely to be under-rewarded, relative to markets, and therefore under-supplied.

Providers operating in markets would appear to have better incentives to make investments that improve net benefits. However, these providers may be much less eager to invest if those investments tie the provider to particular purchasers; that is, supply becomes dedicated to the purchaser – see chapter 2. The result would be the hold-up problem, described in chapter 2 i.e. under-investment. In hierarchical organisations the provider division operates under manager's instructions and so the level of investment is not (directly) a provider issue. Hierarchical providers may still not have incentives to make the investment at least cost, but under-investment for the above reasons is less likely. *Public* hierarchies may face external constraints on investment in the form of difficulties in raising the required capital.

Some of these ideas can be formalised using the models developed above. We can use as a baseline, the nature of coordination and motivation in a world without transaction costs. Absent transaction costs, all parties collect information $\varphi$, $\upsilon$ $\beta$ and $\omega$ once they are revealed by nature. Motivation problems disappear since misrepresentation is fully detected.

Using utility functions (5.9) and (5.11) and where $r^j = 1\ \forall j$ and $T^j = 0\ \forall j$ welfare can written:

$$(5\text{-}46)\quad \begin{aligned} W_k = U_k + Z_k &= (p_k - c_k)D_k - y_k^a - e(y_k^e) + \psi(\rho_k) + \int_0^D R(x)dx - pD \\ &= S - y_k^a - e(y_k^e) + \psi(\rho_k) \end{aligned}$$

We would have exactly the same result if we combined hierarchical purchaser and provider utilities (respectively, (5-26) and (5.29)) under these conditions. Differentiating this function with respect to investment and effort gives first best conditions:

(5-47) $$\frac{\partial S_k}{\partial y_k^a}\left(y_k^a *^F ; E[\upsilon,\beta,\omega]\right)-1=0$$

(where $E[.]$ is the expected value of the as yet not revealed parameters) and

(5-48) $$\frac{\partial S_k}{\partial y_k^e}\left(y_k^e *^F\right)-e_{y_k^e}=0$$

We now turn to considering markets and hierarchies where transaction costs are nonzero.

#### 5.3.1.1 Investment

In markets, the provider makes decisions about $y_k^a$ and $y_k^e$. We can consider different market arrangements according to contracting options chosen. First, *adaptive contracting* which has the provider making the investment decision based on expected utility for each $\upsilon$, and on the understanding that the purchaser will procure a share of the benefits of the investment. Taking expectations over $\upsilon$ of provider utility (5-16), gives $E[U_k|\upsilon]=E\left[\mu r^{MK}S+\mu\left(r^{MK}\psi-r^{MK}e(y^e)+Z^0-B^{MK}-B^{ML}\right)-y_k^a-T^{\varphi MK}-T^{\upsilon MK}|\upsilon\right]$. The provider alone makes investment decisions and therefore $B_k^{y^a ij}=0$ in (5-8). (At this stage the purchaser need not make a measurement of $\varphi$, but since this information will be needed in negotiating product contracts in the second period, they will need to do so at that time). Furthermore, purchaser and provider have to contract only on one state of the world in the second period, so bargaining costs (but not measurement) will be very low. In any case, second period bargaining costs are irrelevant to investment decisions. Therefore, differentiating with respect to $y_k^a$, gives the first order condition:

(5.49) $$r^{MK}\mu\frac{\partial S_k}{\partial y_k^a}\left(\tilde{y}_k^a\right)-1=0$$

which involves no marginal bargaining costs. Nonetheless, this level of investment is inefficient using (5-47) as a benchmark. Since we assume that $S_{y_k^a y_k^a} < 0$, this implies that $y_k^a *^F > \tilde{y}_k^a$ i.e. investment is below the first best level. Moreover because $S_{y_k^a} > 0$ for all $y$ then $S *^F > \tilde{S}$. The inefficiency of CA contracts with regard to investment is due to mainly to the problems of renegotiation. Parties in this case do not write binding contracts at time 1 as in the complete contracts case (in order to avoid the very high bargaining costs) and essentially wait until after the investment when $\upsilon$ is known to agree terms. By time 2 the provider's investment costs are sunk and do not enter into negotiations about future surplus division. Nevertheless, future surplus at time 2 is dependent on the nature of the first period investment. Therefore, the provider bears the full costs of the investment but can expect only a share of the benefits, in accordance with the agreed division of transaction surplus.

Since investment is made before $\upsilon$ is known and so on the basis of expectations about $\upsilon$, an alternative is for stakeholders to commit to a contract written at time 1 on expected product characteristics and associated prices, $E[\{q, x, p\}]$, a so called *generic* contract. Such a contract would certainly save transaction costs since it would be a single state contract. Two problems arise however. First, when the actual value of $\upsilon$ is known, it might be so different from the expected value that the purchaser would benefit from renegotiation of the contract for ideal specification at this time $\{q(\upsilon), x(\upsilon), p(\upsilon)\}$ even though they would lose the sharing of investment costs. In anticipation, providers would under-invest. In particular, suppose the vector $Q^e = E[\{q, x, p\}]$ represents the generic contract. Moreover, this generic product is optimal for $\upsilon^e$ (expected $\upsilon$). Then, expected surplus is:

$$(5\text{-}50)\quad E\left[S\left(y_k^a * \left(Q^e\right), Q^e\right)\right] = \sum_{k=1}^{s} \gamma_k S\left(y_k^a * \left(Q^e\right), Q^e; \upsilon_k\right) < S\left(y_k^a * \left(Q^e\right), Q^e; \upsilon^e\right)$$

This inequality follows because for all states $Q * (\upsilon_k) \neq Q^e$, except by chance a state where $\upsilon_k = \upsilon^e$. If the shortfall in surplus is large, then conceivably,

$$(5\text{-}51)\quad S\left(\tilde{y}_k^a, Q^*(\upsilon_k);\upsilon_k\right) > E\left[S\left(y_k^{a*}(Q^e), Q^e\right)\right]$$

in which case the provider would be better off making unilateral investments without a contracted product exchange terms. The latter is the CA contract outcome with investment level $\tilde{y}_k^a$. We are trading higher levels of investment against poorly tailored services. Given the diverse and user specific nature of social care, the likelihood of (5-51) is reasonably high.

Second, even if the first did not apply, it may be that verification of the product is context specific. For example, it might be very difficult to write a specification that does not make reference to the state of the world for which it was designed. Ascertaining compliance in a different state would then call for abstract judgements. Any of this may undermine the safeguards that inhibit the purchaser from renegotiating the contract.

Absent transaction costs, *complete contracts* would give first best investment levels. However, the transaction costs are so high as to make this type of contract a practical irrelevance – see Annex 5-3 for details.

The final contract type is the investment *cost sharing* or *incentive* contract. Such a contract involves far fewer transaction costs than complete contracts because, like CA contracts, in the second period only product specifications and prices based on prevailing $\upsilon$ are determined. Unlike CA contracts however, in the first period parties contract on initial investment. In particular, the purchaser and provider agree to share the initial investment cost and write a contract accordingly. The utility functions are (5-20) and (5-21), where $\phi^C$ and $\phi^R$ are the respective provider and purchaser shares, such that $\phi^C + \phi^R = 1$. Differentiation of (5-20) indicates the level of investment the provider wishes:

$$(5\text{-}52)\quad \frac{\partial U^{I2}}{\partial y^a} = -\phi^C + \mu\Pi_{y^a} = 0$$

Similarly the purchaser wishes investment to be at a level given by:

$$(5\text{-}53)\quad \frac{\partial Z^{I2}}{\partial y^a} = -\phi^R + (1-\mu)\Pi_{y^a} = -(1-\phi^C) + (1-\mu)\Pi_{y^a} = 0$$

At time 1 both parties anticipate the value of $\mu$, which indicates expected sharing of the transaction surplus at time 2. Given $\mu$, the desired optimal level of investment for each will depend on $\phi^C$ because this is the share of the marginal cost of the investment each will bear. At time 1 each party will need to come to an agreement about $\phi^C$. A failure to do so puts them in the situation of a CA contract. With zero transaction costs of determining the investment contract, doing so will be the best option for both. The net gains from improved investment can be shared. In particular, the provider can fully compensate the purchaser for sharing some of the original investment costs and still gain. The purchaser likewise knows that by sharing some of the investment cost, the provider will be induced to improve investment and therefore the purchaser will be sharing a larger surplus at time 2 (by an amount given by $\mu$).

In fact, as noted above, the parties could agree to go beyond this point, until the joint gains are exhausted (this was option 2 for the CS contract). From a CA contract baseline (where effectively $\phi^C = 1$), this mutually beneficial situation will continue so long as the absolute value of $\frac{\partial U^{I2}}{\partial \phi^C}$ exceeds $\frac{\partial Z^{I2}}{\partial \phi^C}$ because in that case, the purchaser could be fully compensated for any loss, should there be any.[33] Using (5-22), (5-24) and (5-52) in (5-25) gives:

$$(5\text{-}54)\quad \frac{\partial U^{I2}}{\partial \phi^C} + \frac{\partial Z^{I2}}{\partial \phi^C} = -y^a_{\phi^C}\left(1-\Pi_{y^a}\right) = -y^a_{\phi^C}\left(1-\frac{\phi^C}{\mu}\right)$$

[33] This requirement certainly holds at $\phi^C = 1$. Now, $\frac{\partial U^{I2}}{\partial \phi^C} = -y^a, \forall \phi^C$ making use of (5-52). Also, at $\phi^C = 1$, we have $\frac{\partial Z^{I2}}{\partial \phi^C}\left(\phi^C = 1\right) = y^a + y^a_{\phi^C}\left(\frac{1}{\mu} - 1\right) < y^a$, noting that $y^a > 0$. Initially, the share of the gains from improved investment will likely exceed the share of costs the purchaser bears.

It is clear therefore that we have the equality $\phi^C = \mu$ at the optimal sharing level i.e. for $\frac{\partial U^{12}}{\partial \phi^C} + \frac{\partial Z^{12}}{\partial \phi^C} = 0$. With sharing at this level, the purchaser also chooses investment $\tilde{y}^{aK}$. Substituting for $\phi^C$ in (5-53) gives:

$$\frac{\partial Z^{12}}{\partial y^a} = -(1-\mu)+(1-\mu)\Pi_{y^a} = -1+\Pi_{y^a} = \frac{\partial U^{12}}{\partial y^a}(\mu)=0.$$

Differentiating Π, we have $\Pi_{y^a} = r^{MK} S_{y^a} - B^{ML}_{y^a} - B^{MK}_{y^a}$. By assumption, $\Pi_{y^a y^a} = r^{MK} S_{y^a y^a} < 0$. Since $\Pi_{y^a} = \phi^C / \mu$, then $\Pi_{y^a \phi^C} > 0$ it follows that $y^a_{\phi^C} < 0$. So how does investment with this contract compare to the first best? At the chosen sharing level, the level of contracted investment, $y^{aC}$, is given by differentiation of first period utility:

$$\begin{aligned} \frac{\partial U^{12}}{\partial y^{aC}} = \frac{\partial Z^{12}}{\partial y^{aC}} &= \Pi_{y^{aC}} - 1 = r^{MK} S_{y^a} \frac{\partial y^a}{\partial y^{aC}} - B^{ML}_{y^{aC}} - B^{MK}_{y^{aC}} - 1 \\ &= r^{MK} S_{y^a} - B^{y^a ML}_{y^{aC}} - B^{y^a MK}_{y^{aC}} - 1 = 0 \end{aligned} \tag{5-55}$$

Here $\frac{\partial y^a}{\partial y^{aC}} = 1$ (differentiating (5-6)) because the provider has no incentive to make unilateral investments if a cost share can be agreed with the purchaser, as follows. Even when there are no motivation problems (i.e. $r^{MK} = 1$), contracted investment levels still fall short of first best levels as a result of the bargaining costs of the investment contract. At $\tilde{y}^{aC}$ where $\Pi_{y^{aC}}(\tilde{y}^{aC}) = 1$, it is the case that $r^{MK} S_{y^a}(\tilde{y}^{aC}) > 1$. Where transaction costs are very high it may be the case that $\mu r^{MK} S_{y^a}(y^a) > r^{MK} S_{y^a}(y^a) - B^{y^a ML}_{y^{aC}}(y^a) - B^{y^a MK}_{y^{aC}}(y^a)$ for all $y^a$, and therefore both parties are better off reverting to a CA contract with no pre-contracted investment. However, since marginal bargaining costs are high initially and then reduce quickly relative to surplus, either the provider unilaterally invests at the beginning (as a CA contract) or not at all under a CS contract. Total investment $\tilde{y}^a = \tilde{y}^{aC} + \tilde{y}^{aN}$ will remain below first best levels because, in any case, $r^{MK} S_{y^a}(\tilde{y}^a) > 1$.

In *hierarchies*, investment is under the control of the purchaser whose interests are represented by the maximisation problem above. In this case, with optimal decisions about $\theta^H$, we have purchaser utility, from (5-36), of:

$$(5\text{-}56)\quad Z = r^{HL}(\theta^H)(S(y^e*(\theta^H)) - w(\theta^H, y^e*)) - y_k^a - T_k^{\upsilon HL} - T_k^{\varphi HL} - B^{HL} + (1 - r^{HL}(\theta^H))Z^M$$

The first order condition is:

$$(5\text{-}57)\quad Z_{y^a} = r^{HL} z_{y^a}^W - 1 + (1 - r^{HL}(\theta^H))Z_{y^a}^M = 0$$

Because investment is undertaken (unilaterally) by the purchaser, there are no marginal bargaining costs (planning costs were set to zero to simplify comparisons with markets). Now $z_{y^a}^W = \dfrac{\partial S_k}{\partial y_k^a}$ in this function, but what of $Z_{y^a}^M$? When misrepresentation of $\theta^H > 0$ occurs and it is detected, $Z^M$ is given by (5-37). Thus, $Z_{y^a}^M = (S_{y^a} - m_{y^a})$. The value of $m_{y^a}$ depends on whether or not replacement has an adverse effect on the (marginal) benefits of investment. Clearly, if inefficiencies are so large that the hierarchy becomes untenable and the local authority switches funds to market-based providers then the benefits of investment will be affected. Suppose that the hierarchy becomes (completely) untenable with probability $\sigma^H$. Then, $m$ can be written:

$$(5\text{-}58)\quad m = (1 - \sigma^H)m^0(\theta^H) + \sigma^H(z^W(\theta^H, 0) - Z^0)$$

so that if after shirking in response to super-ordinate misrepresentation, subordinate are fired, but the effects on reputation then make the hierarchy untenable (which, *ex ante*, is ascribed a probability $\sigma^H$), the result is $m = z^W(\theta^H, 0) - Z^0$, and so $Z^M = Z^0$, the purchaser's reservation utility. Regarding investment, we have $m_{y^a} = \sigma^H S_{y^a}$, which makes (5-57),

$$(5\text{-}59)\quad Z_{y^a} = S_{y^a} - 1 - (1 - r^{HK})\sigma^H S_{y^a} = 0$$

If purchasers do not exploit provider/subordinates, that is, do not misrepresent prevailing circumstances then $r^{HL} = 1$ and (5-59) is equivalent to the first-best condition (even then . It is also first best when $\sigma^{H} = 0$ for any $r^{HL}$. Above in section 5.2.2.2, we have argued that misrepresentation does occur however.

Whilst above we have implicitly referred to generic hierarchies, final investment in *public* hierarchies will depend on the costs of capital and the influence of any higher layer of government. If local decision-makers face an unconstrained supply of capital funds then the above arguments hold without further qualification. However, public hierarchies are often constrained to access capital via centrally determined public debt conduits. Although this means that the costs of capital are relatively low (since government issued debt is very unlikely to default), the extent of borrowing is subject to centrally determined protocols. These protocols are unable to fully account for local contexts and tend to act to the lowest common denominator with a consequent conservative assessment of the need for capital locally. Moreover, such funding is 'on balance sheet' counting against PSBR and so carries a political imperative to keep it low. Under these arrangements whilst $S_{y^{a}}(\varphi, \upsilon)$ remains unchanged from a local perspective the marginal costs of investment may not just be a linear function (i.e. not just equal to $y^{a}$), instead increasing for big projects, in which case relative to the first order condition (5-59), we have $S_{y^{a}}(\varphi, \upsilon) > 1$, even when $r^{HL} = 1$.[34]

### 5.3.1.2 Effort

In markets there is no need to directly contract on effort however as it shown below; providers unilaterally decide effort levels and purchasers need only to form expectations of $\omega$ and $\beta$. In hierarchies, purchasers (as super-ordinates) induce effort decisions (because only they have measured $\varphi$ and $\upsilon$). The precision of purchasers' instructions

[34] These arguments hold if investment in funded by central public capital mechanisms. However, they need not be the only mechanisms available to public hierarchies. First, capital projects may be administered locally, perhaps via an issue of municipal bonds as is common in the US. This approach would alleviate centrally determined limits on borrowing. Second, investment projects may be out-sourced to the market so as to reduce cost inefficiencies. Third, some form of PFI or PPP arrangement could be used where the investment project is made in the market with the new assets leased to the public hierarchy. These alternatives would change the cost of investment for hierarchies relative to markets.

regarding effort is improved with knowledge of ω (hence requiring its measurement). Otherwise, the purchaser relies on the signal $\omega^S$. The degree of uncertainty associated with the latter allows providers to misrepresent, to some extent, effort and so costs, in order to improve their income (their surplus share). The exact form of such misrepresentation depends on the prevailing governance structure.

In *markets*, efforts are made in period 2, by which time φ and υ are known and price and output are finalised (indeed, purchasers may already have paid the provider). Providers decide the actual level of effort and this determines the costs and the quality/type of product being produced: $q = q(y^e;\omega)$, where $q_{y^e} > 0$. After the contract is determined and prices are fixed, providers do not intrinsically value product type, although they are concerned to produce quality sufficient to meet contractual obligations. Moreover, we will assume that $\frac{\partial e_{y^e}}{\partial q} > 0$ i.e. there is greater marginal effort cost in producing high quality/type services. In that producing quality/type requires effort, providers will select the product type that minimises the effort to cost relationship within the confines of the product specification in the contract. With no contract, providers would choose $q = 0$. Purchasers will want to induce effort levels by creating a product contract specification, subject to the bargaining and measurement costs of making such refinements. The purchaser values quality/type and, although they incur transaction costs in securing quality, at the margin the extra value is greater than the extra bargaining costs. This assumption is especially valid when the product is very complex and the range of feasible types/qualities is very broad, because then zero quality/type is practically useless; the purchaser would simply not pay for such a low quality of product. If bargaining costs were very high at the margin these products would not be produced. Therefore we assume: $V_q > B_q$ (which holds at the very least for quality/type up to the level $q \leq q^{*F}$).

In a zero transaction cost world, purchasers will write a contract that completely and unambiguously specifies the optimal product quality, $q^{*F}$. Providers will then select optimal effort, subject to producing $q^{*F}$. However, where marginal transaction costs accrue to agreeing a verifiable contract specification, purchasers can expect effort to fall below first best levels. Even if purchasers set prices at the first best level, providers will use the ambiguity of an incomplete contract specification to select the least effort-cost product type. In anticipation, purchasers optimally reduce prices to that consistent with $\tilde{y}^e < y^{e*F}$.

Utility for the purchaser at time 2 is then $Z^{II}$ from (5-15). Without provider misrepresentation of β i.e. with $r^{MK} = 1$, this is:

$$(5\text{-}60)\quad Z^{II} = (1-\mu)r^{MK}S + (1-\mu)\left(\psi - e\left(y^e\right) + Z^0 - B^{MK} - B^{ML}\right) \equiv U^{II}\frac{1}{1+\Psi'}$$

Differentiating this function by $y^{eC}$, the amount of effort induced by the product contract gives:

$$(5\text{-}61)\quad \begin{aligned} &\frac{\partial S_k}{\partial y^{eC}}\left(\tilde{y}^{eC}\right) - e_{y^{eC}} - \left(B_q^{q^C MK} + B_q^{q^C ML}\right)\frac{\partial q}{\partial y^{eC}} \\ &= V_q\frac{\partial q}{\partial y^{eC}} - C_{y^{eC}} - e_{y^{eC}} - \left(B_q^{q^C MK} + B_q^{q^C ML}\right)\frac{\partial q}{\partial y^{eC}} = 0 \end{aligned}$$

Purchaser and provider agree a product contract[35] that implies a required effort $\tilde{y}^{eC}$. This (contracted) effort level will be at the first-best level only when marginal bargaining costs are zero (see (5-48)). So when these costs are positive, contracted specification is: $\tilde{q}^C < q^{*F}$ and so $\tilde{y}^{eC} < y^{e*F}$. *After* contracts are determined at time 2, the provider's utility reduces from (5-14) to:

$$(5\text{-}62)\quad U^{III} = P\left(\tilde{y}^{eC}\right) - C\left(y^e\right) - e\left(y^e\right)$$

where product type $\tilde{q}^C$ and implied (contracted) effort $\tilde{y}^{eC}$ have already been determined. Payment, *P*, will be a function of shared surplus, production costs and bargaining costs. Differentiating gives:

---

[35] Note that $q$ in this case is the minimum product type/quality that can be achieved from writing contracts with bargaining costs, $B$

$$(5\text{-}63)\quad \begin{aligned}\frac{\partial U^{III}}{\partial y^e} &= -C_{y^e}(\tilde{q}^C) - e_{y^e}(\tilde{q}^C) = -c_{y^e}(\tilde{q}^C)D - e_{y^e}(\tilde{q}^C) \\ &= S_{y^e}(\tilde{q}^C) - V_q \frac{\partial q}{\partial y^e} - e_{y^e}(\tilde{q}^C) < S_{y^e}(\tilde{q}^C) - e_{y^e}(\tilde{q}^C)\end{aligned}$$

which means that the optimal $\tilde{y}^e \neq y^e{*}^F$. Providers do not supply effort unilaterally at $\tilde{q}^C, \tilde{y}^{eC}$. In the case of complex products:

$$(5\text{-}64)\, \frac{\partial U^{III}}{\partial y^e} = -C_{y^e}(\tilde{q}^C) - e_{y^e}(\tilde{q}^C) = \left(-V_q + B_q^{q^C M}\right)\frac{\partial q}{\partial y^{eC}} < 0$$

where $B_q^{q^C M} = \left(B_q^{q^C MK} + B_q^{q^C ML}\right) < V_q$ as assumed.[36] Overall, then, $\tilde{y}^e = \tilde{y}^{eC} < y^e{*}^F$.

Failing to achieve a first best with an incomplete contract is a commitment problem, which is common in the contracts literature (see chapter 2). Nonetheless, there are commitment mechanisms other than the enforceable contract. For example, the purchaser may trust that the provider will not act opportunistically, trust that the provider could value and wish to earn. In this case, contracts were have far less specification. Such *relational contracts* would have lower transaction costs[37] but trust ensures the provider's efforts are greater than the case absent trust (see chapter 2).

Overall, we assume that to induce any effort, some kind of formal contract specification is needed (at the very least to determine a price) and so $B(0) > 0$. To induce greater

[36] If (marginal) bargaining costs were very high, the differential in (5-64) could become positive, implying that providers were content to unilaterally supply effort. In any case, the purchasers are unlikely to forgo *any* sort of product contract. Without one, quality would then fall to zero: $\frac{\partial U^{III}}{\partial q} = -\frac{\partial e_{y^e}}{\partial q} < 0$, which would only happen if $B_q^{q^C M}(0) > V_q(0)$.

[37] Even with relational contracts some bargaining costs are incurred. First, there are the costs of establishing contracted specifications, however minimal these are. Second, although not laid out in an explicit contract, purchasers still have to communicate *some* preferences regarding the types of service they wish to be produced. One would imagine these costs to be very much lower than the costs of establishing detailed highly specified, legally binding product contracts, but they still exist so the first-best solution is not obtainable. Moreover, both parties need to undertake some form of measurement of the prevailing state of the world, i.e. υ, in order to place their preferences in context.

levels of effort, formal contracts can be successively augmented by less costly, more informal provisions, and so $\frac{\partial^2 \left(B^{qMK} + B^{qML}\right)}{\partial y^{e^2}} < 0$.

*Hierarchical* providers choose their efforts according to instructions from managers and receive utility (5.29), with a wage income $w$. Managers supply instructions about $y^{eC}$ and can monitor for compliance with probability $r^{HK}$. Decisions about instructions about $y^{eC}$ are made by the purchaser to maximise their utility $Z$ in period 2, after uncertainty about $\varphi$ and $\upsilon$ is resolved. Based on (5-56), purchaser utility is:

$$(5\text{-}65)\quad Z^{HH} = \left(S_k + \psi - e\left(y_k^e\right)\right) - \tilde{U} - B^{HL}$$

Here $B^{HL}$ are total transaction costs that the super-ordinate faces at time 2. They include, in particular, the costs of contracting to generate instructions.[38]

Differentiating gives:

$$(5\text{-}66)\quad \frac{\partial Z^{HH}}{\partial y^{eC}}\left(y^{eC*}\right) = S_{y^{eC}} - e_{y^{eC}} - B^{HL}_{y^{eC}} = 0$$

Instructions by managers, even if fully complied with, will not generate first best levels of effort if marginal contracting costs are greater than zero. Indeed, at $y^{eC*} = y^{e*F}$, $\frac{\partial Z^{HH}}{\partial y^{eC}}\left(y^{e*F}\right) = -B^{HL}_{y^{eC}} < 0$ using (5-48). Since, by definition $S_{y^e y^e} < -e_{y^e y^e} < 0$, we have $y^{eC*} < y^{e*F}$. What is the provider's utility at this stage? The value will depend on whether super-ordinates are opportunistic. Provider utility is given by (5-30). If misrepresentation is detected it is only relevant to the question of provider effort in the case where providers do not quit. Purchasers may be content to let providers shirk in these circumstances because their alternative is to re-negotiate the wage structure and/or

[38] They also could include social utility costs on super-ordinates in leveraging effort from subordinates with whom they have personal/social relationships. If the super-ordinate is fearful that imposing more 'work' on subordinates could damage social relationship they value, effort levels will be suffer relative to those levels considered appropriate absent such concerns.

sack and re-hire workers, and this course of action is very expensive. Above we defined the super-ordinate's utility as $Z^M$ if exploitation is detected – see (5-37). Moreover, if $Z^{M^R}$ is the utility where renegotiation is needed and $Z^{M^S}$ when providers shirk, then we may find that

$$(5\text{-}67)\ Z^{M^R} - Z^{M^S} = S\left(y^{eC}*\right) - w\left(0, y^{eC}*\right) - m^R\left(\theta^H\right) - \left[S\left(y^{eC}\right) - w\left(\theta^H, y^{eC}*\right) - m^S\left(\theta^H\right)\right] < 0$$

noting that by assumption $m^R\left(\theta^H\right) > m^S\left(\theta^H\right)$ and $w\left(0, y^{eC}*\right) > w\left(\theta^H, y^{eC}*\right)$. In general, purchasers will either allow shirking or will re-negotiate depending on the relative size of these factors. If the former is chosen by the purchaser, the maximum size of additional shirking is given when $Z^{M^R} - Z^{M^S} = 0$. At this point, effort under the shirking option will be less than $y^{eC}*$, i.e.,

$$(5\text{-}68)\ \tilde{y}^{eC}* \in \left\{y^{eC} : S\left(y^{eC}*\right) - w\left(0, y^{eC}*\right) - m^R\left(\theta^H\right) - \left[S\left(\tilde{y}^{eC}*\right) - w\left(\theta^H, y^{eC}*\right) - m^S\left(\theta^H\right)\right] = 0\right\}$$

In this case effort will be allowed to fall below that level implied by $\dfrac{\partial Z^{HH}}{\partial y^{eC}} = 0$ as given by (5-66) (even if the purchaser knows β and ω). The purchaser would anticipate the possibility of this shortfall in effort, and it would affect the size of $B^{HL}_{y^{eC}}$ in (5-66).

In relation to costs, in hierarchies we assume that instructions yield essentially proportional bargaining costs so that, $\dfrac{\partial^2 B^{HL}}{\partial y^{e^2}} \approx 0$, and that $B^{HL}(0) = 0$.

#### 5.3.1.3 Pricing and monopoly

The above model assumes two stakeholders to a transaction, but a transaction characterised by efficient bargaining given bargaining costs (cf. the effects of competition below). However, the literature is replete with models that accommodate inefficient bargaining (see Lyons and Varoufakis, 1989, for a good survey). The outcome is often deadweight efficiency losses, the classic example being the non-price discriminating monopolist. Nonetheless, the root cause is in many ways the existence of contracting costs, as in the investment problem, since contingent contracts can in theory be written to alleviate the coordination failures (given the contracting costs). In relation to the classic monopoly problem, stakeholders could agree contracts with non-uniform pricing ((Vickers, 1996)). The purchaser simply makes a lump-sum transfer to the provider equal to the excess profit the provider would reap from imposing an inefficient bargaining solution (i.e. from restricting supply), and then agrees terms consistent with the efficient solution. There are some timing issues and concerns about renegotiation but these can be solved. As before, the real impediment is the cost of producing a complete contract.

#### 5.3.1.4 Risk sharing

A final co-ordination problem concerns the distribution of risk between purchaser and provider organisations in markets. The delivery of social care services to people with a great variety of characteristics is fraught with uncertainties. This can have differential consequences in markets and hierarchies, depending largely on the size of organisations bearing the risk and stakeholder reimbursement arrangements (which affect the distribution of risk). Generally speaking, where organisations are risk averse, expected utility of a stream of profits is less than the utility of the expected value of that stream i.e. $Eu(\pi) < u(E(\pi))$. However, the size of the difference is negatively related to the number of transactions that can be expected within a period i.e. $Eu(\pi) \to u(E(\pi))$ as $n \to \infty$. Annex 5-4 demonstrates this result. Relative to large organisations that conduct many transactions, smaller organisations are likely to suffer lower expected utility other things equal. Hierarchies tend to be large, market providers relatively small. The latter might therefore require compensation, often in the form of higher prices per transaction.

### 5.3.2 Motivation

Information pertinent to a given transaction falls essentially into two classes. Relevant factors may be *external* to both purchaser and provider e.g. the assessed dependency characteristics of referred users. Information φ and υ are of this type and so $T^{jMK} \approx T^{jML}$, for $j$ = φ, υ. Other relevant factors concern the specific circumstances of either the purchaser or the provider, and is *private information* to use the term from information economics. For example, the implications of the care technology utilised by the providers. Information ω and β fall into this category. The former concerns the provider's productivity, which we assume is set by nature. The latter relates to the experienced characteristics of residents (and is to some extent under the control of providers, in that in some arrangements providers are able to select and reject particular users).

For private information, either β or ω, measurement cost to the purchaser is much higher than to the provider: $T^{jMK} < T^{jML}$, for $j = \omega, \beta$. In fact, the difference is often very sizeable, to the point where the net benefits to the purchaser of measuring these parameters is in question. Nonetheless, there are generally adverse consequences of relying only on signals, and these vary according to the governance structure in use. We explore the details in chapter 7, although we consider the problem in general here (see also Forder, 1997a). We first consider asymmetric information about β and then about ω.

#### 5.3.2.1 Misrepresentation of care (cost) type

In social care the cost and effectiveness of services depend closely on the specific characteristics of users of those services, and in particular, how they interact with the care technology of the provider (Netten et al., 1998). Private information about β can give rise to two problems *in markets*.

The first problem, *cost exaggeration*, involves providers exaggerating their report of β to purchasers. The provider claims that the costs of the care package of the required type for the resident i.e. β are higher than the purchaser's estimate. The second information problem is *cream-skimming*, whereby providers try to select clients who have expected care costs below the purchaser's estimate but claim that these clients have an actual cost in the region of the purchaser's estimate (Forder, 1997b). This is a form of *adverse selection* (Glennerster and Matsaganis, 1993).

In the model, the purchaser makes an assessment of the client – and so ascertains υ. Furthermore, this in effect provides a signal of β. However, uncertainty remains about the final costs of care since assessment tools cannot fully predict the costs of care on an individual basis. The signal of β allows purchasers to make a judgement on the validity of the provider's report. In particular, we assume that they accept the provider's report with a probability $r$, where, given the deviation, $\theta = \hat{\beta} - \beta = \hat{\beta} - \beta^S$, we assume that $r(0) = 1$, and that $r_\theta \leq 0$, $r_{\theta\theta} < 0$. In other words, the chances of misrepresentation being detected increase (exponentially) with the size of the distortion $\theta$.

The implications depend on how misrepresentation affects (net) provider income. As noted in chapter 4, contracts can be contingent on user characteristics (in as much as these affect the costs of care). The initial assessment – yielding parameter $\upsilon$ – can provide relevant information, although parameter $\beta$ is the better information in this regard. Contracts can be directly contingent on $\beta$ with prices set after an initial placement period in negotiation between the purchaser and provider. Alternatively, a price tariff according to dependency might be pre-determined. Often, the criteria for classifying users are very vague. In this case, even where prices are initially contingent on $\upsilon$, there is typically scope for revision after the initial (trial) period of a placement, whereby prices become contingent on the refined information in $\beta$. Both give opportunities for cost exaggeration. In other cases, if the estimate in $\upsilon$ turns out to be greater than actual dependency as measured by $\beta$, the provider may withhold this information from the purchaser.

Contracts can also be non-contingent in this regard, in which case cost exaggeration would not affect provider income. In this latter case, cream skimming would be expected – lowering the actual value of $\beta$ would reduce costs, given prices.[39] In either case, the general effect is that misrepresentation of $\theta$ has positive effects on the share of surplus accrued by the provider.

At time 2 at the point where prices and outputs are agreed, the provider's utility is, using (5-14),

$$(5\text{-}69)\quad U^{II} = \mu r^{MK} S(\hat{\beta}, \hat{\omega}; \beta, \omega) + \mu\left(r^{MK}\psi - r^{MK} e(y^e) + Z^0 - B^{MK} - B^{ML}\right)$$

[39] We explore the detail of these behaviours in chapter 7.

where $S$ can be written $S(\theta)$ a function of information difference $\theta$. If we differentiate (5-81) by $\theta$ to represent either of these forms of information exploitation we have:

$$(5\text{-}70)\quad \frac{\partial U^{II}}{\partial\theta}=\mu\left[S(\theta_k)+\psi-e\left(y^e\right)\right]\frac{\partial r^{MK}}{\partial\theta}+\mu r^{MK}\left(S_\theta+\psi_p p_\theta-e_{y^e}y^e_\theta\right)$$

The first term is negative since $r_\theta<0$ and for providers to be in business, surplus less disutility and effort is greater than zero. Overall, the second term measures the benefits of misrepresentation to the provider, where this distortion is not detected. $S_\theta$ measures the effect of an increase of $\theta$ on the perceived value of the output by the purchaser – which translates from the provider's perspective to an increase in prices – and/or a reduction in costs. A change in $\theta$ is likely to have repercussions for optimal effort – although this is not figured into the product contract because in this case the purchaser accepts the provider's report. Moreover, because the purchaser does not directly measure effort and since the provider determines actual effort we need only consider the case where $y^e_\theta=0$. The provider will only change effort if it is beneficial to do so; if a strategy of misrepresentation increases utility without an effort response, then if the chosen optimal effort response is also figured in, it cannot decrease the benefit of change. The second term of (5-70) then reduces to:

$$(5\text{-}71)\quad \mu r^{MK}\left(S_\theta\big|_{y^e}+\psi_p p_\theta\right)$$

Suppose we are considering changes in misrepresentation from $\theta$ to $\theta'$, where $\theta'>\theta$. With unchanging effort and so marginal costs,

$$(5\text{-}72)\quad \begin{aligned} S(\theta')\big|_{y^e} &= \int_0^{D(\theta')}R(\theta',x)dx-C(\theta') \\ &= \int_0^{D(\theta)}R(\theta,x)dx+\int_0^{D(\theta)}\left[R'(\theta')-R(\theta)\right]dx+\int_{D(\theta)}^{D(\theta')}R(\theta')dx \\ &\quad -\int_0^{D(\theta)}C_x(\theta,x)dx-\int_{D(\theta)}^{D(\theta')}C_x(\theta',x)dx \\ &= S(\theta)\big|_{y^e}+\int_0^{D(\theta)}\left[R'(\theta')-R(\theta)\right]dx+\int_{D(\theta)}^{D(\theta')}R(\theta')dx-\int_{D(\theta)}^{D(\theta')}C_x(\theta',x)dx>S(\theta)\big|_{y^e} \end{aligned}$$

The inequality holds because $R(\theta', x) > C_x, \forall x \in [D(\theta), D(\theta'))$, and at $x = D(\theta')$, we have $R(\theta', D(\theta')) = C_x$. So we can state that $S_\theta|_{y^e} > 0$ by virtue of an assumed smoothness of the benefits function. Furthermore, by assumption, $\psi'(\infty) > -1$, and therefore, $S_\theta|_{y^e} + \psi_p p_\theta|_{y^e} = 1 + \psi_p > 0$. This inequality follows because $S$ is divided between purchaser and provider by price so that, where $D_\theta \geq 0$, it follows given $y^e$ that $0 < p_\theta \leq S_\theta$. Therefore (5-71), which is the second term in (5-70), is positive.

Since the first term of (5-70) is negative whilst the second is positive, we cannot in general sign this differential. However, at the point where there is no misrepresentation, we have $r(0) = 1$. Moreover, although $r_{\theta\theta}(0) < 0$ and $r_\theta(\tilde{\theta}) < 0$, $\tilde{\theta} > 0$, at $\theta = 0$ we might expect that $r_\theta(0) = 0$, or is at least very small. In other words, the chances of a small degree of misrepresentation being detected are almost negligible. But of course as the size of this distortion increases so the chance of detection increases (and after some point, where misrepresentation is flagrant, detection is more or less certain).

Suppose that $r_\theta(0) = 0$, then we can write (5-70) as

$$(5\text{-}73) \quad \frac{\partial U^{II}}{\partial \theta} = \mu\left(S_\theta + \psi_p p_\theta - e_{y^e} y^e_\theta\right) > 0$$

which of course means that $\theta = 0$ is not a solution and that some exploitation of information is optimal. The benefits for the provider will come from the purchaser believing that surplus $S$ is greater than it actually is, given actual $\beta$, and consequently paying the provider more to make up the provider's share. Chapter 7 provides a detailed consideration of incentives, contract types and information problems, showing that contract contingency is particularly important.

In *hierarchies,* employees do not have wages that are contingent on reported $\beta$. Exaggerating this parameter would not have direct positive consequences (on income). Moreover, since providers have ceded control there are no opportunities to reject and replace referred clients. Consequently, providers cannot cream-skim clients to reduce effort (given effort instructions based on signalled β). Although there are no (direct) benefits to exaggerating the care costs of the prevailing cohort of users – since wages are set in advance, based on the average – there may be some indirect benefits of misreporting of β to compound uncertainty about ω, and so further enhance the benefits of shirking, which we consider next.

### 5.3.2.2 Productivity 'shirking'

As described above in section 5.3.1.2, in hierarchies even when purchasers/managers (and providers/subordinates) have perfect information about ω, coordination problems can arise because there are costs in developing and applying appropriate sets of instructions. However, the problem is compounded when purchasers do not measure ω. Not only will instructions be incomplete, but the provider might also be motivated to shirk on those instructions, $y^{eC}*$, that have been specified. A truthful report of ω is consistent with full compliance with instructions $y^{eC}*$. However, misreporting implies provider efforts away from this value. Therefore, we can conceive $r^{HK\tau}$ as measuring the deviation of $y^{eC}$ from the instructed value. Using (5-30), those providers remaining in the industry after they make a decision about whether the purchasers is being exploitative have utility (i.e. they have chosen not to opt for alternative utility $U^M$) of:

$$(5\text{-}74)\ U^{HII} = r^{HK\tau}\left(w\left(\theta^H, y^e*\right) - e\left(y^e\right) + \psi\right) + \left(1 - r^{HK\tau}\right)U^{0\tau}$$

where τ indicates whether or not the provider detected misrepresentation by the purchaser. Where such misrepresentation was detected $\tau = 1$ and $r^{HK1} = r^{HK}\left\{y^{eC} - \tilde{y}^{eC}*\right\}$, such that $r^{HK}\{0\} = 1$ and $r^{HK}_{y^e} > 0$. Where misrepresentation is not detected (or does not occur) $r^{HK0} = r^{HK}\left\{y^{eC} - y^{eC}*\right\}$. As indicated in section 5.3.1.2, $y^{eC}* > \tilde{y}^{eC}*$.

In the case of no detection of super-ordinarte misrepresentation, if shirking $y^{eC} < y^{eC*}$ is detected by the super-ordinate then the subordinate is dismissed and will receive utility $U^{00}$. This utility is that accrued in a market alternative (at $\theta^H = 0$), but with some penalty for having a 'bad reference'. It is in any case less than

$u^W(\theta^H,0)=w(\theta^H,y^{eC}*)-e(y^{eC}*)+\psi$. In the no shirking case, $y^{eC}=y^{eC*}$, at which point $r^{HK0}=1$. If we differentiate (5-74) with respect to $y^{eC}$ we find:

$$(5\text{-}75)\quad U^{HH}_{y^{eC}}=-r^{HK0}e_{y^{eC}}+r^{HK0}_{y^{eC}}\left(w_k-e(y^{eC})+\psi-U^{00}\right)$$

At $y^{eC}=y^{eC}*$, $r^{HK0}=1$ and so:

$$(5\text{-}76)\quad U^{HH}_{y^{eC}}=-e_{y^{eC}}+r^{HK0}_{y^{eC}}\left(w_k-e(y^{eC}*)+\psi-U^{00}\right)$$

The first term is negative, but the second is positive. In all likelihood, at $y^{eC}=y^{eC}*$, the marginal change in the probability of detection will be very small, making the differential negative. In that case providers will engage in a modest amount of shirking. However, for higher levels of shirking given the non-linear nature of the detection function, the sign of this function is likely to revert to positive.

The corresponding case is when super-ordinates misrepresent $\varphi$ and $\upsilon$ i.e. when $\tau=1$. In this case, providers are allowed to 'shirk' to effort $\tilde{y}^{eC}*$. Then, $r^{HK1}<1$ only when effort is below this allowed level, i.e. when $y^{eC}<\tilde{y}^{eC}*$. Differentiating at $\tilde{y}^{eC}*$ we have:

$$(5\text{-}77)\quad \begin{aligned} U^{HH}_{y^{eC}}(\tilde{y}^{eC}*) &= -r^{HK1}e_{y^{eC}}+r^{HK1}_{y^{eC}}\left(w_k-e(y^{eC})+\psi-U^{01}\right) \\ &= -e_{y^{eC}}+r^{HK1}_{y^{eC}}\left(w_k-e(y^{eC})+\psi-U^{01}\right) \end{aligned}$$

which has an ambiguous sign as before. Arguably we might find that $U^{01}$ is slightly higher than $U^{00}$ if only because a super-ordinate who is misrepresenting information has less creditability in supplying a 'bad reference'. Alternative providers might be willing to give the benefit of the doubt to dismissed subordinates in that case. Hence we might find slightly higher shirking from the agreed effort level when $\tau=1$.

The $r^{HK}_{y^{eC}}$ term reflects the sensitivity of purchaser monitoring of deviations in effort from instructed levels. If for very small deviations from optimal effort the chance of detection changes very little (from no chance of detection) then overall, $U_{y^{eC}}\left(y^{eC*}\right) < 0$ and $U_{y^{eC}}\left(\tilde{y}^{eC*}\right) < 0$, which implies that effort will be below the levels given in section 5.3.1.2 i.e. below $y^{eC*} < y^{e*F}$ and $\tilde{y}^{eC*} < y^{e*F}$. Together with the subordinate lacking the motivation to unilaterally supply effort, i.e. $y^{eN} = 0$, this implies that actual effort is below the first best level $y^{e} < y^{e*F}$.

In markets, misreporting of $\omega$ brings no benefits because the contract is already determined based on expected $\omega$. If $\omega$ was known by providers before contracts were determined then there would be a potential benefit in misleading the purchaser about productivity and so affect the terms of the agreement. This effect would be akin to the cream-skimming result discussed above.

## 5.4 Mediating features

### 5.4.1 Timing – frequency, duration and longevity

*Frequency and duration* are key features of transactions since they affect opportunities for *reputation* to have a bearing on behaviour. Reputation is a very powerful transaction cost economising mechanism: in repeated transactions stakeholders can have an incentive to maintain a good reputation (Kreps and Wilson, 1982; Milgrom and Roberts, 1982; Roth and Schoumaker, 1983; Fundenberg and Tirole, 1992). In particular, a person can accept the truthfulness of claims by other people if the good reputation of those others would be undermined if they cheat, lie or misrepresent, and if loss of reputation is harmful. If the transaction is repeated frequently enough and if a good relationship would continue indefinitely then long-term truthfulness is more beneficial than short-term opportunism.

This mechanism is central to the operation of hierarchies where subordinates (e.g. employees) are willing to cede control (and thereby risk exploitation) because the employer has a reputation to protect (Kreps, 1990a; Kreps, 1996). Reputation is needed to ensure that employees can expect a reasonable share of the surplus generated by the relationship. A similar argument applies to the use of minimum specification, adaptive 'relational' contracts since concerns about reputation commit stakeholder to act reasonably with reference to non-contracted contingencies.

The central benefit of reputation in these cases is the low unit measurement and bargaining costs they entail. Where control is ceded duplication of measurement is avoided. In addition, because one party is given authority voluntarily, protracting bargaining is unlikely. Reputation also reduces transaction costs in markets, for example, by reducing the need for monitoring to combat misrepresentation of private information (Klein and Leffler, 1981; Forder, Knapp, and Wistow, 1996).

### 5.4.2 Complexity, uncertainty and private information

Unit transaction costs of all three types – measurement, bargaining and monitoring – are positively related to the degree of *complexity* and *private* information relating to the transaction. In particular, where complexity is high, bargaining costs especially will be substantial. When information is asymmetrically distributed i.e. much is *ex ante* private, measurement costs will be high. The relative extent of private information and complexity is likely to have repercussions that depend on the type of governance structure that is used. From a governance point of view, hierarchical arrangements are often the better option when dealing with the more complex transactions. Hierarchies have lower measurement costs overall. Also, their top-down managerial arrangements mean that they can adapt at relatively low cost to new contingencies in contracts.

### 5.4.3 Competition and asset specificity

Levels of *competitiveness* are important in a number of regards. First, bargaining costs depend on the level of competitiveness. If there is only one provider then negotiation can become protracted. But add one more provider and haggling and bargaining can be cut short by the threat of the purchaser playing suppliers off against one and other, so reducing unit bargaining costs. In relation to the above model, we have sidestepped this issue by assuming efficient bargaining, but where it is an issue, it will add to transaction costs. Second, when competition is healthy, prices in markets are good mechanisms for transmitting information, which acts to reduce measurement costs (Milgrom and Roberts, 1992). Hierarchies use quantity adjustment mechanisms, which are less efficient.

Competition levels will also affect coordination and motivation. First, high levels of competition largely undermine rent-seeking behaviour since competition forces providers to act efficiently in order to survive (Tirole, 1988 ). For example, absent other costs, the under-investment problem is addressed since, even with CA contracts, the threat of the purchaser switching provider will force a higher level of investment $y^a$. Similarly, where a purchaser can easily switch providers, there is a lesser need to write safeguards in contracts, and therefore reduce its bargaining costs for all levels of effort and investment. If a provider does not supply an appropriate product then absent legally binding contracts and third party enforcement, the purchaser can threaten to switch the contract to an alternative provider. Second, competition can help address some shirking (and slacking) problems by allowing benchmarking of competitors, that is by allowing principals to compare agent/provider's observed behaviour – i.e. yardstick competition (Schleifer, 1985).

*Asset specificity* – where the use of certain assets is tied closely to a particular transaction – reduces potential competition with the implications as indicated above, specifically causing under investment problems. Asset specificity is a problem that largely applies to physical assets and therefore is absent in hierarchical governance structures where both production and purchasing (physical) assets are under unified ownership (Grossman and Hart, 1986). Networks would partially address these problems since control is often ceded to one party (even if ownership is still separate).

For these reasons healthy competition acts to reduce the comparative net transaction costs of hierarchies compared with markets. One counterfactual relates to the problem of cream-skimming, which is driven by the availability of alternative providers. Competition generates choice that accommodates selectivity. With hierarchies where no alternative supplier is available the problem is irrelevant.

### 5.4.4 Social context

*Social context* is also highly relevant to health care transactions as outlined in chapter 2, section 3.6.2. Grannovetter and others describe how actions that arise in specific transactions are *embedded* in conventions that exist in an individual's social sphere (Granovetter, 1985; Hannan and Freeman, 1984; Hamilton and Feenstra, 1995). These conventions may work against narrow economic considerations. For example, parties to transactions in societies that value personal honour may be less like to exploit their position – to cheat – than societies with more 'pragmatic' values (Sako, 1992; Hodgson, 1988; Granovetter, 1985). *Social capital* or more generally inherent *trust* can produce very similar effects as reputation (Kreps, 1996). Trust is essentially deterrence-based: i.e. "people do what they say they will do because they fear the consequence of doing otherwise" (Doney, Cannon, and Mullen, 1998, p. 605). Trust is sustained insofar as there is a visible sanction likely to be applied where trust is violated.

An alternative perspective sees trust as not so calculative, stemming instead from a more inherent alignment of motivations, as a 'presumed reliability' (Giddens, 1990). From this alternative perspective, the instrumental conception of trust (e.g. as reputation) reduces trust misleadingly to a matter of risk assessment, and misses the point. As Giddens puts it, trust in general is largely "blind trust" (1990, p. 33). Where trust is high for whatever reason net transactions costs will be considerably reduced. In markets with high trust relations, many of the formal mechanisms of arms-length contracting become redundant.

## 5.5 Net (optimal) effects and (variable) transaction costs

What are the likely transaction benefits and costs in social care between different governance archetypes? The comparison can be made in a number of pertinent dimensions.

### 5.5.1 Efficient contracts and market power

At a given level of investment and effort, and state-of-the-world, the transaction between purchaser and provider generates a certain surplus that can be allocated between parties. The final product cost to the purchaser will include the share of surplus that goes to providers and appears as profit. For comparative purposes between markets and hierarchies, we have assumed that, at the baseline, hierarchical providers can expect to achieve utility $\widetilde{U}$, all other things equal, which is the level of utility perceived to be achieved in markets given the same set of circumstances. Therefore, if the comparator is a competitive provider market then hierarchy providers will also perceive the fallback to be the representative utility of a provider in a similarly competitive market. Hence, distributional considerations in relation to market power are precluded for our purposes.[40]

What this assumption implies is that for the same levels of investment and effort, and state-of-the-world, the production cost of hierarchies and markets would be identical. In practice when markets are very competitive, and the comparator is a public hierarchy, political control might be expected to give hierarchical workers some protection and therefore, a slightly higher proportion of the surplus.

We have assumed efficient bargaining between purchaser and provider in the above analysis. As noted in section 5.3.1.3, in markets, there is a potential for classic monopoly deadweight loss. However, with relatively modest transaction costs this problem can be overcome e.g. non-uniform pricing. Also, as reviewed in section 2.2.4.2, there are examples of efficiency shortfalls concerning distributional issues (e.g. strikes and hold-outs). There is no particular reason to suppose that markets are more or less susceptible to these problems compared with hierarchies. In the former there is less unionisation, but in the later there is less bargaining (because employees cede control).

### 5.5.2 Investment

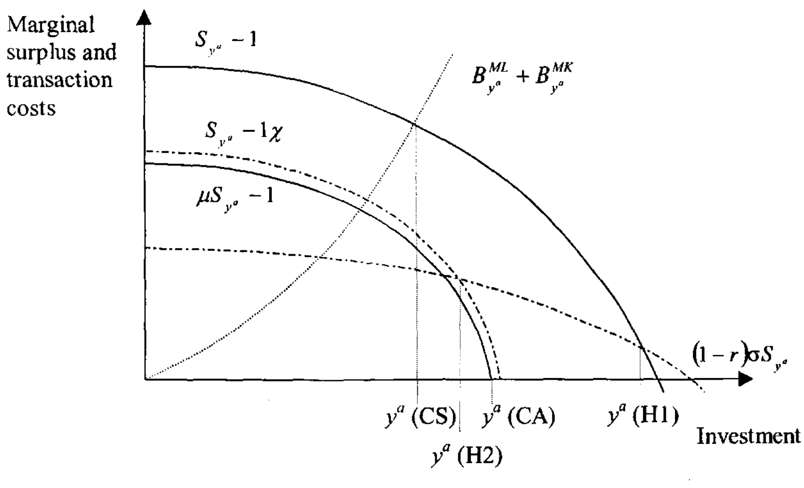
Adaptive contracts (CA contracts) do not involve any contract specification at the time of investment and so no bargaining costs are incurred at this stage. However, as indicated above, without a contract, providers can expect purchasers to appropriate a share of the investment surplus, and this induces under-investment. Total surplus is therefore reduced. Figure 5-3 shows investment decisions as undertaken in markets and in hierarchies with hypothetical cost and benefit functions for illustration. In the figure the investment level for CA contracts is, in this case, $y^a$ (CA). The alternative is for stakeholders to employ cost sharing or incentive contracts (CS contracts). Since these contracts involve directly contracting on the investment task, transaction costs are incurred. However, precluding transactions cost, these contracts secure first best investment levels. In the figure the implied investment level is $y^a$ (CS). In practice,

[40] In practice, we might assume that hierarchical providers under-estimate the competitiveness of alternative markets. This is purely an empirical proposition based on the observation that in practice residential care markets have been very competitive, perhaps more that local authority purchasers have expected. This proposition would imply that: $\mu^M < \bar{\mu}^M \Rightarrow \tilde{U}^H$ i.e. provider market power is lower than hierarchical stakeholders assume, which in turn affects their perception of a reasonable opportunity utility for hierarchical providers.

market stakeholders will choose between CA and CS contracts according to which generates the greatest net surplus, which is given in the figure as the area under the relevant benefit curve (up to the corresponding investment level) less the area under the associated transaction costs curve. We assume that the transaction costs of writing complete contracts (CC contracts) is greater than those for CS contracts. Therefore, given the benefits function is the same, stakeholders would always choose CS or CA contracts – as such CC contracts are not shown in the figure.

In hierarchical arrangements, investment decisions are internalised so that given investment costs, first best levels of investment can potentially be achieved. The limiting factor is if after investment hierarchies become untenable as a result of excessive (detected) exploitation of providers. Since some (small) degree of exploitation is optimal, investment will fall below first best levels. Therefore, if costs of making the investment were the same as in markets, hierarchical investment levels would be at $y^a$ (H1) in the figure, above the market level (and also generating more surplus).

**Figure 5-3. Investment decisions**




However, we can question this assumption of the same (marginal) investment costs. We have not assumed that any 'effort' or productivity is involved in investing in the model, although if we did, then for the reasons stated in the previous section, the marginal costs of investment are likely to be higher in hierarchies. This is an argument often used to support PFI programmes (Grout, 2003). If we accept that (marginal) investment cost is higher, then as shown in the figure, $\chi > 1$ and hierarchical investment levels would be at

$y^a$ (H2). Moreover there will be some value of $\chi$ such that net surplus from investment could be increased by out-sourcing investment to the (private) market sector.

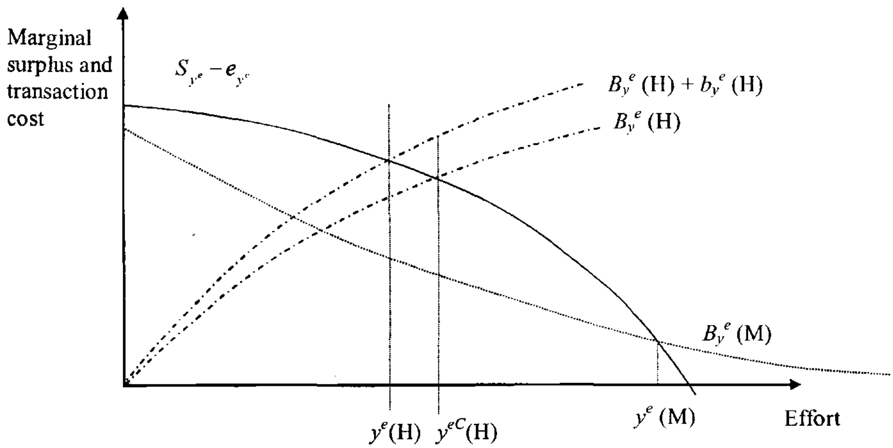
The investment cost includes not only the construction and training costs (direct costs), but also the costs of debt. Public sector projects can usually secure lower debt costs than private sector counterparts, and so when projects are outsourced if their transaction costs and direct costs were the same as for market projects, investment levels in public sector projects would be greater (Grout, 2003).

### 5.5.3 Effort

In hierarchies, when misrepresentation occurs, sub-ordinates might choose (and be allowed) to shirk below $y^{eC}*$. Actual net surplus would correspond to the lower level of effort, but the instructions given to sub-ordinates – for the purposes of generating effort $y^{eC}*$ and baseline from which they shirk – would relate to $y^{eC}*$ and transaction costs would be at this higher level. In effect per unit of effort, transaction costs would be increased by $b_y{}^e$ up to actual effort level $y^e$. In Figure 5-4 contracted effort, as given by (5-66), is $y^{eC}$ $(H)$ ($\equiv y^{eC}*$). However, with the possibility of 'allowed' shirking depending on $r^{HL}$ $(\theta^H)$ and also the relative size of $m^S$ and $m^R$, actual effort will fall below this level and so expected actual effort is $y^e$ $(H)$. At this level, $S_{y^e} - e_{y^e}$ need not equal $B_{y^e}(H) + b_{y^e}(H)$ since shirking behaviour reduces effort below the optimal. Transaction costs as measured by the area under $B_{y^e}(H)$ are still incurred up to effort $y^{eC}$ $(H)$, with the additional component from $y^e$ $(H)$ to $y^{eC}$ $(H)$ equivalent to the area $\int_0^{y^e(H)} b_{y^e}(H) dy^e$.[41] Net surplus will however be given by the area under $S_{y^e} - e_{y^e}$ between 0 and $y^e$ $(H)$. Net surplus less transaction costs is the area between $S_{y^e} - e_{y^e}$ and $B_{y^e}(H) + b_{y^e}(H)$ from 0 to $y^e$ $(H)$.

[41] Note that incurring these additional transaction costs per unit of effort, purchasers may wish to increase effort beyond $y^{eC}$ $(H)$ in anticipation of shirking. However, sub-ordinates will not increase effort further with given wages $w$.

**Figure 5-4. Effort decisions**




Effort is induced in markets rather than instructed as in hierarchies. Inducement comes from the product contract that involves a price for providing the service as specified and allows providers to keep the residual after costs as profits. Effort impacts on costs and on product specification $q$. Without the specification of required service characteristics effort may be misdirected. Indeed, where the specification allows some scope for interpretation (is not *complete*) after prices are fixed, providers will apply effort in producing the absolute basic product that has the lowest cost to effort relationship. *Directed* effort as relevant to $S$ requires a product specification, and indeed, the greater is its specificity, the more likely that optimal levels of directed effort are applied and the desired product quality produced. But greater specificity means greater transaction costs. Moreover, marginal transaction costs will be large with respect of low levels of directed effort, and diminishing thereafter; even to get basic applicable effort, for a reasonably complex service, a complex product specification is needed. Writing such a specification has a large fixed cost component.

Figure 5-4 shows these bargaining costs of effort in markets as being high initially and then diminishing. Two relevant features apply in this case. First, the marginal bargaining cost function is likely to diminish more rapidly than the marginal net surplus, at least initially, as bargaining is practically a fixed cost of inducing effort. Second, marginal net surplus initially exceeds this marginal transaction cost otherwise projects would be unfeasible overall. As a result, at the optimal effort, marginal bargaining costs are likely to be very close if not at zero, and indeed lower than marginal contracting cost for hierarchies. With the same marginal net surplus function as between markets and hierarchies therefore, effort is hypothesised to be at a higher level in markets than in hierarchies.

Despite marginal bargaining costs in markets being below those in hierarchies at optimal effort, because they are high initially in markets bargaining costs in total could certainly exceed those in hierarchies. For this reason it is difficult to say whether the total net utility (welfare) that is generated from effort is unambiguously greater in markets.

The problem of shirking is especially significant in public sector hierarchies where managers themselves are unlikely to benefit substantially from efforts to be creative in motivating employees. In particular, there will be much reliance on the threat of outsourcing of provision, but even then, many managers would retain their positions. Sanctions on social services directors will come from local electorates (weakly) and from central government performance assessment. Where SSDs with predominantly hierarchical arrangements are under-performing an option for directors is to increase the proportion of care purchased externally in the market. Nonetheless, these incentives are indirect and so shirking levels in public hierarchies could be expected to exceed even those of private hierarchies. The result in the model is not only low productivity but could also be products with inappropriate characteristics. Translated in practice, the result is high unit costs and services that are supply-led and non-responsive to local needs (Savas et al., 1998).

Overall, the empirical hypothesis is that effort is greater in markets and so production costs will lower (other things equal), but that transactions costs are lower in hierarchies.

### 5.5.4 Motivation

One of the key problems with using markets in social and health care is the inherent information complexity of these services that can give rise to problems whereby providers misrepresent key characteristics (including the cost and type of service provided). Above we saw that in theory some degree of misrepresentation is optimal in markets. Below it is shown that such misrepresentation increases production costs (because purchasers think they are getting a higher value service than is actually the case).

There is no specific transaction cost consequence as such, unless purchasers decide to monitor providers more than they would do otherwise. Monitoring is generally expensive and quite lumpy in its application, so unlikely to be undertaken unless purchasers otherwise expect significant misrepresentation to occur. Overall, in markets either production costs are pushed up or monitoring transaction costs are increased relative to the no misrepresentation case. The anecdotal evidence (see previous chapter) is that social services purchasers do not engage in much monitoring, but it is difficult to establish what the baseline is compared to say hierarchies.

Shirking occurs in hierarchies. As described above there are circumstances in which super-ordinates may 'allow' some shirking in order to retain staff. Beyond this level shirking can occur because information about effort is costly to collect. As with misrepresentation in markets, monitoring may be undertaken specifically to address this problem – i.e. to measure $\omega$ – however for the above reasons there must be some expectation of considerable shirking otherwise.

### 5.5.5 Risk

Risk management is generally facilitated by large size – so risks can be spread and pooled – as shown above. Hierarchical organisation of services then has the better risk management characteristics, at least in regard to the benefits of size. In theory, in markets the purchaser could design contracts that efficiently allocate risk to providers (i.e. retain most of the risk) but in practice this appears not to happen, and in any case it requires greater contracting transaction costs. We would anticipate that market providers charge a premium to carry this risk i.e. that overall production costs are higher than in the risk neutral case. In theory, hierarchies avoid these premiums, although in practice, this characterisation of public sector hierarchies may be a little strained. It is not inconceivable to imagine that risk averse provider managers, who are instructed to stay within budget, might keep a little back to cover unforeseen contingencies.

Nonetheless, overall hierarchies are in a better position to deal with risk and should have lower production costs, *ceteris paribus*, as a result.

## 5.6 Fixed transaction costs (measurement)

In addition to the variable transaction costs just considered, there are also sizeable fixed transaction costs resulting mainly from measurement activities. 'Purchasers' in both markets and hierarchies have to measure all relevant information, i.e. $T^{\varphi ML}$ and $T^{\varphi HL}$ regarding $\varphi$ and $T^{\upsilon ML}$ and $T^{\upsilon HL}$ regarding $\upsilon$. Providers in markets also measure this information, i.e. $T^{\varphi MK} > 0$ and $T^{\upsilon MK} > 0$ but those in hierarchies do not: $T^{\varphi HK} = 0$ and $T^{\upsilon HK} = 0$.

It is difficult to form *ex ante* judgements as to the relative sizes of $T^{jHK}$ compared to $T^{jMK}$ for each $j = \varphi, \upsilon$. However, it would seem reasonable to assume that because hierarchies do not have a purchaser-provider distinction as such, and that 'purchasers' i.e. managers are diffused throughout the organisation that the costs of measurement are slightly lower than in markets. This argument is particularly relevant to the $\upsilon$ parameter where 'provider-side' insights are especially helpful.

## 5.7 Net transaction costs

Based on the above sections, the following table summarises our hypothesis about transaction costs *faced by purchasers*.

**Table 5-2. Transaction costs (purchaser)**

| | | **Market** | **Hierarchy** |
|---|---|---|---|
| **Variable** | Investment | 0 (CA) or $\int_0^{y^a(CS)} B^{ML}_{y^a} dy^a > 0$ (CS) | 0 |
| | Effort | $\int_0^{y^e(M)} B^{ML}_{y^e} dy^e > 0$ | $0 < \int_0^{y^e(H)} \left(B^{HL}_{y^e} + b^{HL}_{y^e}\right) dy^e \left[ < \int_0^{y^e(M)} B^{ML}_{y^e} dy^e \right]$ |
| | Motivation | $\geq 0$ | 0 |
| | Risk | 0? | 0? |
| **Fixed** | | $T^{\varphi ML} + T^{\upsilon ML}$ | $T^{\varphi HL} + T^{\upsilon HL} \leq T^{\varphi ML} + T^{\upsilon ML}$ |
| **Total** | | $\boldsymbol{TC^M}$ | $\boldsymbol{TC^H < TC^M}$ |

Together therefore individual components lead to the first main empirical hypothesis:

*Hypothesis H1:*

*Total purchaser transaction costs in hierarchies are less than in markets:* $\boldsymbol{TC^H < TC^M}$

To re-cap, *provider* transaction costs also accrue in markets: $\int_0^{y^e(M)} B_{y^e}^{MK} dy^e > 0$ in agreeing the product contract in relation to effort. Also, $\int_0^{y^a(CS)} B_{y^a}^{MK} dy^a > 0$ under CS contracts from agreeing an investment contract. Because for both markets and hierarchies we have set (the assumed equal) planning costs to zero, there are no transaction costs for providers in hierarchies. Hence if *H1* holds for purchaser transaction costs it will also hold for *total* transaction costs, which we label hypothesis *H1a*.

## 5.8 Net production costs

The above comparative prediction of the level of effort and investment as between market and hierarchy implies comparative production costs as follows. Should the transaction go ahead final provider payment will be:

(5-78) $P = U^{II}(\mu) - \mu\psi + C + \vartheta y^a$

Period 2 utility less excess price disutility is the net monetary share to the provider. Payment is the sum of production cost, shared investment cost and provider net utility. In markets with CA contracts, payment to cover investment is made in the first period and so does not figure above i.e. $\vartheta = 0$. With cost sharing and CC contracts investment does figure in payment considerations with payment covering the provider's contribution to the costs i.e. $\vartheta = \mu$. Purchasers (super-ordinates) direct the level of investment in hierarchies, although in terms of accounting data these costs may show up in provider's unit costs, which are equivalent to payments. In any case, the associated level of market utility will be the basis for the opportunity utility calculation for hierarchical providers. As given by the constraint (5-32), $u^W = \tilde{U}^{II} \equiv U^{II}$, and so (5-78) applies to market and hierarchical providers.

#### 5.8.1.1 Effort

Provider effort is at the contacted level i.e. $\tilde{y}^e = \tilde{y}^{yC}$ as indicated by (5-63). Also, $U^{II} = (1 + \Psi')Z^{II}$ and so using (5-60), $U^{II} = \mu\left(r^{MK}S + \psi - e(y^e) + Z^0 - B^{MK} - B^{ML}\right)$ and therefore, $\frac{\partial U^{II}}{\partial y^e} = \mu\left(S_{y^e} - e_{y^e} - \left(B_q^{q^C MK} + B_q^{q^C ML}\right)\frac{\partial q}{\partial y^e}\right) = \mu\left(S_{y^e} - e_{y^e} - B_{y^e}\right)$. Hence differentiating (5-78) gives,

$$(5\text{-}79)\quad P_{y^e} = U^{II}_{y^e} + C_{y^e} = \mu\left(V_{y^e} - C_{y^e} - e_{y^e} - B_{y^e}\right) + C_{y^e} = \mu V_{y^e} - \mu\left(e_{y^e} + B_{y^e}\right) + C_{y^e}(1-\mu)$$

The sign of this differential depends on the relative size of the impact of effort on costs and on the purchaser's valuation of the product (via the effect on quality). When $\mu V_{y^e} < -C_{y^e}(1-\mu)$ then prices fall as effort increases: $P_{y^e} < 0$. In fact, since in relation to provider motivation, $\mu \leq ½$, the condition reduces to $V_{y^e} < -C_{y^e}$.[42] Even if this is not satisfied, the combination of marginal effort and bargaining costs may ensure that marginal price changes are downwards (although we have assumed that $V_{y^e} > B_{y^e}$).

It follows, *ceteris paribus*, that where effort levels markets are higher than those in hierarchies, total payment in markets will be lower.

#### 5.8.1.2 Investment

For investment, given that $U^{II} = \mu r^{MK}S + \mu\left(r^{MK}\psi - r^{MK}e(y^e) + Z^0 - B^{MK} - B^{ML}\right)$, the differential is (discounting provider misrepresentation):

$$(5\text{-}80)\quad P_{y^a} = U^{II}_{y^a} + C_{y^a} + \vartheta = \mu S_{y^a} + C_{y^a} + \vartheta - \mu B_{y^a} = \mu V_{y^a} + (1-\mu)C_{y^a} + \vartheta(1-\mu) - \mu B_{y^a}$$

We cannot sign this function *a priori* since it is likely that increased levels of investment will reduce costs. If investment is cost neutral in that it changes the product but not its costs, then $P_{y^a} > 0$ : under cost-sharing or CC contracts below optimal level

[42] It is possible for the sign to change if the marginal value of the product to the purchaser is large – because this (directed) effort yields substantial increases in quality.

of investment, $\mu S_{y^a} + \vartheta - \mu B_{y^a} = \mu(S_{y^a} + 1 - \mu B_{y^a}) > 0$ and under CA contracts, $\vartheta = 0$ and $S_{y^a} > 0$. In practice, investment is required to allow the production of product type $q$. Where products improve in quality so generally their production costs increase.

#### 5.8.1.3 Motivation failures

Turning to motivation failure, productivity shirking has the same effect as the 'poor incentives' coordination failure above, that is, reduces comparative effort levels and therefore implies higher comparative payments, other things being equal.

Regarding misrepresentation of care/cost type, β, the effect on payment is given by the differentiation of (5-78) with respect to $\theta^{\beta}$, which at the (first best) no misrepresentation level is (dropping the β superscript),

(5-81) $P_\theta = U_\theta^{II} - \mu\psi_\theta + C_0 \geq 0$

At the no misrepresentation level, $U_\theta^{II}$ is given by (5-73) and is greater than zero. Although it will increase provider utility, cream-skimming will not result in a change of prices (because price is set by the purchaser – the gain comes from providers selecting lower cost users – in (5-81) $U_\theta^{II} > 0$ is offset by $C_\theta = -U_\theta^{II} < 0$). If in the agreed contract, price is contingent on reported β to a sufficient degree, cost exaggeration will occur instead. Then the increase in provider utility as given by (5-73) will come from an increase in $P$. Essentially, with reference to (5-62), if the contract is such that $P_0 > -C_0$ (at $r^{MK} = 1$) then cost-exaggeration will result. Otherwise, providers will cream-skim. There is a possibility of some change in optimal effort, that could have a reducing effect on costs, but this is likely to be dominated by the demand effect. Overall, providers who are opportunistic in this way gain from a price increase as outlined above, an increase that is based on an exaggerated report of β – see also chapters 7 and 8.

#### 5.8.1.4 Market power

*Ceteris paribus*, $U^{II}$ will be a function of μ, the market power parameter. Above we argued that $u^W = \tilde{U}^{II} \equiv U^{II}$. However, it may be the case when market competition is particularly harsh that public hierarchical equivalent 'market power' is slightly higher than for markets i.e. $u^W \geq \tilde{U}^H \equiv U^H$. The implication is that for the same effort and investment, and in same state-of-the-world, market production cost may be slightly lower than in hierarchies.

An increase in competition/contestability – given monopsony purchasing – will reduce μ. A fall in μ will not affect effort in markets or hierarchies, although it may reduce investment under CA contracts i.e. $\frac{\partial y^a}{\partial \mu} \geq 0$. However, with the above cost function assumptions, $P_{y^a}(\mu) > 0$ and provider's utility being positively related to its net surplus share, i.e. $U^H_\mu(y^a) > 0$, so:

(5-82) $$P_\mu = U^H_\mu(y^a) + U^H_{y^a}\frac{\partial y^a}{\partial \mu} - \psi + C_{y^a}\frac{\partial y^a}{\partial \mu} + \vartheta\frac{\partial y^a}{\partial \mu} = U^H_\mu(y^a) - \psi + P_{y^a}\frac{\partial y^a}{\partial \mu} > 0$$

## 5.9 Overall net production costs

Production costs are a function of the levels of investment, effort, misrepresentation and market power in the above model. The baseline position is that if all relevant variables were the same for market and hierarchy, respective utilities would be equal. The following table summarises the above section:

| | | ***Markets cf. Hierarchies*** | **Sign** | **Market PCs** |
|---|---|---|---|---|
| **Co-ordination** | **Effort** | $\tilde{y}^{eC}(M) > y^{eC}*(H) \geq \tilde{y}^{eC}*(H)$ given $\hat{\omega} = \omega$ for both $M, H$ | $P_{y^e} < 0$ | Lower |
| | **Investment** | $\tilde{y}^a(M) <> y^a*(H)$ | $P_{y^a}$ | ? |
| **Motivation** | **Misrepre-sentation β** | $\theta^{\beta M} \equiv (\hat{\beta}^M - \beta^M) > (\hat{\beta}^H - \beta^H) \equiv \theta^{\beta H} = 0$ | $P_{\theta^M} \geq 0$ | Higher |
| | **Shirking ω** | $\hat{\omega}^H \neq \hat{\omega}^M = \omega$ implying $\tilde{y}^e(M) = \tilde{y}^{eC}(M) > y^{eC}*(H) > y^e*(H)$ | $P_{y^{eC}} \equiv P_{y^e} < 0$ | Lower |
| | **Market power μ** | $\mu^M < \bar{\mu}^M \Rightarrow \tilde{U}^H$ | $P_\mu > 0$ | Lower |

The results for production costs are ambiguous *a priori*. It would not be unreasonable, however, to expect the effort considerations to be dominant, especially as the motivation effects could well cancel out. Moreover, if anything, anecdotal evidence suggests that investment in markets is slightly better, again resulting in lower costs. Overall, we posit the following hypothesis:

*Hypothesis H2:*

*Total production costs in hierarchies are greater than in markets:* $PC^H > PC^M$

## 5.10 Discussion

### 5.10.1 Empirical hypothesis

Summarising the above discussion, the investment, effort and motivation failure arguments made above suggest that (variable) transactions costs will be lower and production costs higher in public sector hierarchies compared to markets. The fixed transaction costs arguments reinforce this hypothesis since fixed transaction costs (mainly measurement costs) in hierarchies are, in terms of the above theory, lower than in markets.

In addition to investment, effort and motivation effects, the above also noted that hierarchies and markets differ in how they deal with risk. The arguments related to risk tend to countervail our main hypothesis somewhat because we would expect market providers to charge risk premiums that push up production costs to purchasers. Only where efficient risk sharing contracts are used would we expect that production cost risk premiums are reduced to levels equivalent to those in hierarchies.[43] Risk effects will be significant in social care (see chapter 7) but the relative risk effects as between markets and hierarchies are very unlikely to overturn the sum of the other variable transaction and the fixed transaction cost differences.

With the caveats about risk, what drives the transactions costs differences is the degree to which control rights are ceded. High ceding in hierarchies will reduce transaction

[43] The problem is that in practice, it does not appear that these types of contracts are being used much.

costs but removes incentives that mitigate slacking or shirking – depending on the reimbursement of hierarchical subordinates. Generally, hierarchical provider productivity is only affected by the effort being induced through purchaser's instructions.

Reputation and other safeguards are the glue that holds together hierarchies and the ceding of control rights. If reputation was absent, hierarchical providers would expect to be exploited and receive low payment, and, in response would either quit or shirk heavily. Without a (at least reasonable) reputation, hiring back providers to replace those that quit would be at higher than normal market rates. The combination of shirking and/or high wage costs would make for very inefficient production. Such inefficiency might ultimately make the hierarchy untenable to government, and so a switch to market modes of operation would likely follow. In other words, where reputation effects can work, although net production costs may still be higher in hierarchies than in markets, the savings in transaction costs could be very large. A considerable saving can be made in hierarchies in avoiding the high costs of negotiating and writing product contracts and in the duplication of fixed (measurement) transaction costs.

Reputation effects can also work in markets to curb problems like cost and quality misrepresentation. The consequence is likely to be a reduction in production costs compared to situations where this type of misrepresentation does occur (unless significant monitoring has been undertaken in which case the consequence would be reduced transaction costs – this is not however consistent with the observations about monitoring). Reputation in markets would therefore reinforce the above hypothesis.

Complexity and competition are also relevant, as outlined above, in affecting the transaction to production cost equation. The two main hypotheses are explored empirically in the next chapter.

## Annex 5-1

Restricting ourselves to state-of-the world $k$ – as stakeholders are risk neutral – suppose that the purchaser's offer price is $p^R$ and the provider's offer price is $p^C$. Bargaining is efficient so that all expected surplus from the *current* transaction (between these two parties) is allocated,

(5-83) $r^{MK}\int_0^D R(x)dx - p^R x + (p^C - c)x = r^{MK}S$

which follows multiplying (5-4) by $r^{MK}$ and where output is $x$ up to $D$. The optimisation problem is therefore:

(5.84) $\max_p Z^{II}U^{II}$

s.t.

$r^{MK}S - r^{MK}\int_0^D R(x)dx + p^R x - (p^C - c)x = 0$

This constrained optimisation problem gives two first order conditions:

(5.85) $U^{II}Z^{II}_{p^R} + \lambda x = 0$

and

(5.86) $U^{II}_{p^C}Z^{II} - \lambda x = 0$

so that:

(5.87) $Z^{II} = U^{II}\dfrac{-Z^{II}_{p^R}}{U^{II}_{p^C}} = U^{II}\dfrac{-Z^{II}_{p^R}}{U^{II}_{p^C}}$

where, differentiating, $Z^{II}_{p^R} = -x < 0$ and $U^{II}_{p^C} = x(1+\Psi') > 0$. Changes in payment $p^R = p^C$ do not affect output levels since output is set to a level $D$ such that $R(D) = C_x$. The constraint can be re-written as $Z^{II} - Z^0 + B^{ML} + U^{II} - r^{MK}\psi + B^{MK} + e(y^e)r^{MK} = r^{MK}S$ and $Z^{II} = U^{II}\dfrac{1}{1+\Psi'}$ from (5.87), so:

(5-88) $U^{II} = \left(\dfrac{1+\Psi'}{2+\Psi'}\right)\left(r^{MK}S + r^{MK}\psi - r^{MK}e(y^e) + Z^0 - B^{MK} - B^{ML}\right)$

or (5-14) as above, writing $\left(\frac{1+\Psi'}{2+\Psi'}\right)=\mu$. In the profit maximising case, for instance, absent transaction costs and effort costs, $Z^{II}_{p^R}/U^{II}_{p^C}=-1$, or equivalently $\Psi'=0$, and so ${}^{\pi}U^{II}=\frac{1}{2}r^{MK}S$. Otherwise, when $\Psi'<0$, $\mu<½$. Substituting for $U^{II}$ using (5-88) into $Z^{II}=U^{II}\frac{1}{1+\Psi'}$ from (5.87), gives:

$$Z^{II}=(1-\mu)r^{MK}S+(1-\mu)\left(r^{MK}\psi-r^{MK}e(y^e)+Z^0-B^{MK}-B^{ML}\right)$$

For the CC contracts case, expected surplus now includes utility from period I: $(R-p^R)x+(p^C-c)x-y^a=S^I\equiv r^{MK}S-y^a$. The Nash bargaining outcome is as above, given from the maximisation $Z^IU^I$ subject to this expected surplus constraint, which can be expressed as $Z^{II}-Z^0+B^{ML}+U^I-r^{MK}\psi+B^{MK}+r^{MK}e(y^e)=S^I$.

## Annex 5-2

Full contingent contracts are negotiated at time 1 when φ is known. With these contracts, because the nature of the product $q$, $D$ and the associated payment terms for all possible states of the world are specified, then whatever state does actually arise (i.e. what υ), each party will have an agreed course of action/payment, tailored in advance for that state. Such a contract can be enforced at time 2. The exact nature of this enforcement depends on prevailing contract law, but in any case the contract is the basis for damages to be levied for non-compliance. The contract would specify $\{x^{I*}(\varphi,\upsilon),q^{I*}(y^a*(\varphi),\upsilon)\}$, $\forall\upsilon$ and payment $p^I*$, and also a set of damages clauses: $M^{KM}>0$ if $\{x,q\}\neq\{x^{I*},q^I*\}$ and $M^{LM}>0$ if $p(\varphi,\upsilon)\neq p^I*(\varphi,\upsilon)$. Here the superscript $I$ denotes that variables are agreed at time 1. At time 2, after the provider has made investment $y^a$, the purchaser will wish to renegotiate the contract (as in the above case where no time 1 contract is agreed in order to divide the transaction surplus net of the investment cost). Would the provider agree? If reparations for breech were expected to be sufficiently large, then the provider would not agree. If reparations were expected to be quite small, then reservation utility plus reparations may be less than the surplus from a renegotiated contract. This issue will generally turn on the expected attitude of the courts (see Hart, 1995). Since full contingent (CC) contracts are negotiated at time 1 they can account for the investment cost. Thus, we add this cost to (5-12) to get provider (partial) utility of:

$$(5\text{-}89)\quad U^{I} = (p_k - c_k)x_k - y^{a} + r^{MK}\psi(\rho_k) - r^{MK}e(y^{e}) - B^{MK}$$

Purchaser (partial) utility is unchanged: $Z^{I} = Z^{II}$. Furthermore, expected surplus is now: $(R - p^{R})x + (p^{C} - c)x - y^{a} = S^{I} \equiv r^{MK}S - y^{a}$ i.e. now including period $I$ utility. The Nash bargaining outcome is directly analogous with the above, except that all period utility is summed – see Annex 5-1:

$$(5\text{-}90)\quad U^{I} = \mu S^{I} + \mu\psi = \mu(r^{MK}S - y^{a}) + \mu(r^{MK}\psi - r^{MK}e(y^{e}) + Z^{0} - B^{MK} - B^{ML})$$

Purchaser utility is:

$$(5\text{-}91)\quad Z^{I} = (1-\mu)(r^{MK}S - y^{a}) + (1-\mu)(r^{MK}\psi - r^{MK}e(y^{e}) + Z^{0} - B^{MK} - B^{ML})$$

## Annex 5-3

Under complete contracts stakeholders agree at time 1 prices, $p$, and final outputs $\{D, q\}$ for each state of the world $\upsilon$, given known $\varphi$. Since $\upsilon$ is unknown at time 1, investment is at the level required to fulfil the expected contracted product specification and price. Specifically, the investment level under CC contracts is the value that maximises both expected utilities, (5-90) and (5-91), that is,

$E[U^{I}(\varphi)|\upsilon] = E[\mu(r^{MK}S - y^{a}) + \mu(r^{MK}\psi - r^{MK}e(y^{e}) + Z^{0} - B^{MK} - B^{ML})|\upsilon]$ and

$E[Z^{I}|\upsilon] = E[(1-\mu)(r^{MK}S(\upsilon^{e}) - y^{a}) + (1-\mu)(r^{MK}\psi - r^{MK}e(y^{e}) + Z^{0} - B^{MK} - B^{ML})|\upsilon]$.

Differentiating either of these functions with respect to the investment level induced by the product contract gives:

$$(5\text{-}92)\quad r^{MK}\frac{\partial S}{\partial y^{aC}}(\tilde{y}^{aCC}) - 1 - B^{MK}_{y^{aC}} - B^{ML}_{y^{aC}} = r^{MK}\frac{\partial S}{\partial y^{aC}}(\tilde{y}^{aCC}) - 1 - \sum_{k=1}^{s}\left(B^{qMK}_{q} + B^{qML}_{q}\right)\frac{\partial q_k}{\partial y^{aC}} = 0$$

The product type specified in the contract for each $\upsilon$ is that implied by this level of investment, i.e. $q = q(\tilde{y}^{aCC}, \upsilon, \varphi)$ and output $D$, at a price that shares the surplus generated by $\tilde{y}^{aCC}$. Should the provider produce $y^{aC} \neq \tilde{y}^{aCC}$ then

$q^C(y^{aC},\upsilon,\varphi) \neq q^C(\tilde{y}^{aCC},\upsilon,\varphi)$, which is a contract breech and subject to (prohibitively high) damages. Similarly, if the purchaser failed to pay the agreed $p(\tilde{y}^{aCC},\upsilon,\varphi)$, they would have to pay damages. Specification of product terms $\{q,D,p\}$ for each contingency is required to ensure an appropriate specification whatever the actual circumstances that prevail at the time of any challenges to the contract. The existence of bargaining costs prohibits the writing of fully specified product contracts and therefore the actual contract cannot induce providers to supply the first-best investment.[44] All parties are aware of this, and so the payments made reflect the anticipated, not the first best level of investment. It might be assumed that providers would unilaterally supply the additional investment beyond $\tilde{y}^{aCC}$. However, for this additional investment the provider would bear the full costs but expect to reap only a share of the benefits i.e. as in the CA contracts case, and so $y^a < y^{a*F}$ where transaction costs exist.

Bargaining costs are a serious constraint. It is difficult to get an idea of magnitude of abstract terms such as $B^{MK}$, but clearly, the need to calculate optimal service configurations and pricing for all conceivable values of $\upsilon$ (and potentially $\beta$ and $\omega$ and given one of many possible states implied by $\varphi$) is a demanding task. Many commentators argue the bounded rational nature of stakeholders and the consequent near infinite costs of this task, to effectively rule out CC contracts (see chapter 2). Nonetheless, describing this type of contract is useful because it helps make clear why CA contracts fail to deliver first best outcomes (even absent motivation problems).

## Annex 5-4

In order to investigate the affects of scale, we will consider organisations with net income drawn from some distribution with constant variance, $\sigma_\pi^2$. For each transaction $i$ providers receive a variable net income with a mean of $\bar{\pi}$. Managers (perhaps reflecting the preferences of stakeholders) are risk averse with utility $u = u(\Pi)$, such that, $u(0) = 0, u' > 0, u'' < 0, u''' = 0$. Performing a Taylor series expansion on $u$ about $u = 0$:

---

[44] Only complete contracts that have full specifications for each contingency will be assured of inducing the first best level of investment (given expectations at the time of actual investment). Where only a partial specification exists due to transaction costs, the investment will be too generic, as it were, under-supplied, with the provider expecting, *ex post*, for a lack of precision to work against him.

(5-93) $u(\Pi)=u(\bar{\pi})+u'(\bar{\pi})(-(\bar{\pi}-\pi))+\frac{1}{2}u''(\bar{\pi})(-(\bar{\pi}-\pi))^2=u(\bar{\pi})-u'(\bar{\pi})(\bar{\pi}-\pi)+\frac{1}{2}u''(\bar{\pi})((\bar{\pi}-\pi))^2$

where $\Pi=\bar{\pi}+(\pi-\bar{\pi})=\bar{\pi}-(\bar{\pi}-\pi)$. Taking expectations we have:

(5-94)
$$\begin{aligned} Eu(\Pi) &= u(\bar{\pi})-u'(\bar{\pi})(E[\bar{\pi}]-E[\pi])+\tfrac{1}{2}u''(\bar{\pi})(E[\bar{\pi}]-E[\pi])^2 \\ &= u(\bar{\pi})-u'(\bar{\pi})(E[\bar{\pi}]-E[\pi])+\tfrac{1}{2}u''(\bar{\pi})(E[\bar{\pi}]^2+E[\pi]^2-E[\bar{\pi}]E[\pi]) \end{aligned}$$

As $\bar{\pi}=\frac{1}{n}\sum_{i=1}^{n}\pi_i$ and $\bar{\pi}^2=\left(\frac{1}{n}\sum_{i=1}^{n}\pi_i\right)^2$, so we have $E[\bar{\pi}]=E\left[\frac{1}{n}\sum_{i=1}^{n}\pi_i\right]$ and $E[\bar{\pi}]^2=E[\bar{\pi}^2]=E\left[\left(\frac{1}{n}\sum_{i=1}^{n}\pi_i\right)^2\right]$. Also, $E[\pi]=E\left[\frac{1}{n}\sum_{i=1}^{n}\pi_i\right]$, which means,

(5-95)
$$Eu(\Pi)=u(E(\bar{\pi}))+\tfrac{1}{2}u''(\bar{\pi})\left(E\left[\left(\frac{1}{n}\sum_{i=1}^{n}\pi_i\right)^2\right]+E\left[\frac{1}{n}\sum_{i=1}^{n}\pi_i\right]^2-2E\left[\frac{1}{n}\sum_{i=1}^{n}\pi_i\right]E\left[\frac{1}{n}\sum_{i=1}^{n}\pi_i\right]\right)$$

or

(5-96) $$Eu(\Pi)=u(E(\bar{\pi}))+\tfrac{1}{2}u''(\bar{\pi})\left(E\left[\left(\frac{1}{n}\sum_{i=1}^{n}\pi_i\right)^2\right]-E\left[\frac{1}{n}\sum_{i=1}^{n}\pi_i\right]^2\right)$$

or

(5-97)
$$Eu(\Pi)=u(E(\bar{\pi}))+\tfrac{1}{2}u''(\bar{\pi})\frac{1}{n^2}\left(E\left[\left(\sum_{i=1}^{n}\pi_i\right)^2\right]-E\left[\sum_{i=1}^{n}\pi_i\right]^2\right)=u(\bar{\pi})+\tfrac{1}{2}u''(\bar{\pi})\frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n}\pi_i\right)$$

and so,

(5-98)
$$Eu(\Pi)-u(E(\bar{\pi}))=\tfrac{1}{2}\frac{u''(\bar{\pi})}{n^2}\operatorname{var}\left(\sum_{i=1}^{n}\pi_i\right)=\tfrac{1}{2}\frac{u''(\bar{\pi})}{n^2}\sum_{i=1}^{n}\operatorname{var}(\pi_i)=\tfrac{1}{2}\frac{u''(\bar{\pi})}{n}\operatorname{var}(\pi_i)=\tfrac{1}{2}u''(\bar{\pi})\frac{\sigma_\pi^2}{n}<0$$

which indicates the negative utility associated with risk. It is clear however that this disutility lessens in size as the number of transactions $n$ increases, so that as $n\to\infty$, so $Eu(\Pi)=Eu(\pi)\to u(\bar{\pi})$. The variance of net return, which gives rise to negative utility is diminishing as the number of samples (transactions) tends to its probability limit: mean costs for larger organisations tend to the income level.

# Chapter 6. Empirical analysis of governance archetypes

## 6.1 Introduction

This chapter seeks to address the two main hypothesis developed in chapter 5, namely that purchaser (and total) transaction costs in hierarchies are less than in markets: $TC^H < TC^M$ and that total production costs in hierarchies are greater than in markets: $PC^H > PC^M$. The empirical focus is on residential care for older people. The empirical investigation exploits the variation across England in the proportion of places that are organised under broadly (quasi-) market like and (public) hierarchy like arrangements – see chapters 3 and 4. This chapter proceeds in section 6.2 by describing the cost modelling approach, specifically the specification of the cost functions to be estimated. Section 6.3 describes the main data used in the analysis. It reports the derivation of the cost variables. The estimation techniques used are outlined in section 6.4, and the results of the analysis are given in section 6.5. Conclusions follow in section 6.6.

### 6.1.1 Governance arrangements

The main empirical task in this chapter is the attribution of costs – transaction and production – to governance alternatives i.e. to explore how transaction costs differ between governance arrangements. In chapter 3 a theoretical definition of governance arrangements was developed. Our practical measure used here relies on the distribution of *ownership* as its basis, whilst the theoretical definition includes issues relating to the distribution of *control*. However, the analysis in chapter 4 (section 4.6.1.1) showed first that there was a strong correlation between ownership and control. Specifically, synthetic regression using data on control and incentives, as well as ownership, at local authority level – but with a much smaller sample size – supported the use of ownership distribution as a summary statistic. Moreover, even 'internal markets' in social care are likely to have significant 'hierarchical' features since provider (home) managers are directly employed by the local authority and are accountable ultimately to the corporate centre.

Taking an output focus, the empirical indicator for 'governance archetype' is the relative number of care home placements made according to ownership distribution.

Specifically, a placement is made under hierarchical arrangements if ownership of the means of purchasing and provision are *integrated* i.e. both public sector. A placement is made under market arrangements if ownership is dispersed. Because we are concerned only with publicly funded, and so publicly purchased services, the latter is equivalent to the number of placements made to independently owned providers.

### 6.1.2 Costs

An inclusive definition of 'total costs' in the study is of total local authority *social services* expenditure less expenditure explicitly identified as allocated to non-older client groups. Total costs comprise both the costs of production of services and the transaction costs of facilitating production and allocation of services (see chapter 3). Identifying production costs by governance type and client group is straightforward since our governance definition is output based. However, the transaction cost element is not output based. Total costs data, and by subtraction, transaction cost data, are not categorised by output type. In this study statistical modelling is used to allocate these transaction costs. Transaction costs take a broad definition in this analysis being essentially total costs less the costs of production of services.

## 6.2 Cost modelling

The aim of the analysis is to test a number of hypotheses that concern the relationship between social care costs – total, production and transaction – and governance arrangements. A statistical model is used to estimate the association between variation in these costs across local authorities (at two points in time) and the proportion of total residential care output provided by local authorities (hierarchies), rather than purchased from the independent sector – i.e. our governance archetype indicator. The local authority Social Services Department (SSD) is the unit of analysis.

### 6.2.1 Cost-theoretic underpinnings

A cost model underpins the analysis. The production cost function for each service takes the standard form:

$$(6\text{-}1) \quad C_i^k = C_i^k\left(q_i^k, w_i, \tau_i\right)$$

where $q$ is service output, with the superscript $k = h, m$ indicating hierarchical and market output respectively. Residential costs are described by the vector $C = \{C_i^h, C_i^m\}$. The subscript $i = 1,.., N$, denotes the local authority SSD. The vector $w_i$ is the various inputs costs, $\tau_i$ is a vector of cost shift factors. The vector $\tau_i$ includes, in particular, client dependency characteristics that affect the costs of providing output of required standard. Transaction costs are:

$$(6\text{-}2) \quad T_i = T_i\left(q_i^1,..,q_i^k,..,q_i^s, w_i, \tau_i\right)$$

The governance choice by the local authority is assumed to be according to the following programme:

$$(6\text{-}3) \quad \max_{q_i} U_i = U_i\left(q_i, \theta\right)$$

Subject to production and transaction costs summing to total expenditure $E$:

$$(6\text{-}4) \quad E_i = \sum_k q^k C^k + T(q)$$

where $q$ is the vector of outputs and $\theta$ are preferences. This problem has the usual first order condition:

$$(6\text{-}5) \quad \frac{U_{q^1}}{\left(C_i^1 + q_i^1 C_{q^1}^1\right) + T_{q^1}} = ... = \frac{U_{q^k}}{\left(C_i^k + q_i^k C_{q^k}^k\right) + T_{q^k}} = ... = \frac{U_{q^s}}{\left(C_i^s + q_i^s C_{q^s}^s\right) + T_{q^s}}$$

which can be solved for optimal outputs $q_i^k = q_i^{k1}\left(q_i^{-k}, C_i, w, \tau, \theta\right)$, with $C_i$ the vector of production costs, and $q_i^{-k}$ is the vector of outputs other than output $k$. Substituting using (6-1) gives $q_i^k = q_i^{k2}\left(q_i^{-k}, w, \tau, \theta\right)$. We are interested in the three outputs: LA residential care, in-house residential care and domiciliary care, which will be denoted $l \subset k : k = h, m, nr$ with the remaining $s - 3$ outputs denoted by $j \subset k : j \not\subset l$, and the vector of $j$ outputs referred to as $q_i^{-j}$. The $s - 3$ output equations can be solved simultaneously for the $s - 3$ outputs $j$ (assuming a solution exists e.g. a linear equation system) to give the vector of other outputs as (partial) reduced-form equations:

$$(6\text{-}6) \quad q_i^{-j} = q_i^{-j}\left(q_i^h, q_i^m, q_i^{nr}, w, \tau, \theta\right)$$

These outputs can be substituted firstly into the transaction cost function (6-2):

$$(6\text{-}7) \quad Y_i^T = T_i^R\left(q_i^h, q_i^m, q_i^{nr}, q_i^{-j}\left(q_i^h, q_i^m, q_i^{nr}, w, \tau, \theta\right), w_i, \tau_i\right)$$

or

$$(6\text{-}8) \quad Y_i^T = T_i^R\left(q_i^h, q_i^m, q_i^{nr}, w_i, \tau_i, \theta\right)$$

to give the *residual* transaction cost function, $Y_i^T$. Similarly, residual total cost, and residual production cost functions are:

$$(6\text{-}9) \quad Y_i^l = Y_i^l\left(q_i^h, q_i^m, q_i^{nr}, w_i, \tau_i, \theta\right), \qquad l = E, C$$

i.e. where $Y_i^C$ is the (residual) production cost function and $Y_i^E$ is the (residual) total cost function. Depending on exact functional form, these residual cost functions can be re-specified as:

$$(6\text{-}10) \quad Y_i^l = Y_i^l\left(\Omega_i, q_i^T, q_i^{nr}, w_i, \tau_i, \theta\right), \quad l = E, C, T$$

where $\Omega_i = q_i^h / \left(q_i^h + q_i^m\right) = q_i^h / q_i^T$ is the governance indicator variable.

A Taylor series expansion (to second order) in logs by differences $\{\tilde{q}_i^h, \tilde{q}_i^m, \tilde{w}_i\}$ can be used to approximate this transaction cost function.

$$
\begin{aligned}
\tilde{Y}_i^T \approx\ & \tilde{Y}_i^T(0,0,0, w_i, \tau_i, \theta_i) + \frac{\partial \tilde{Y}_i^T}{\partial \tilde{q}_i^h}\tilde{q}_i^h + \frac{\partial \tilde{Y}_i^T}{\partial \tilde{q}_i^m}\tilde{q}_i^m + \frac{\partial \tilde{Y}_i^T}{\partial \tilde{w}_i}\tilde{w}_i \\
(6\text{-}11) \qquad & + \tfrac{1}{2}\left[\frac{\partial^2 \tilde{Y}_i^T}{\partial \tilde{q}_i^h \partial \tilde{q}_i^h}\left(\tilde{q}_i^h\right)^2 + \frac{\partial^2 \tilde{Y}_i^T}{\partial \tilde{q}_i^m \partial \tilde{q}_i^m}\left(\tilde{q}_i^m\right)^2 + \frac{\partial^2 \tilde{Y}_i^T}{\partial \tilde{w}_i \partial \tilde{w}_i}\left(\tilde{w}_i\right)^2\right] \\
& + \frac{\partial^2 \tilde{Y}_i^T}{\partial \tilde{q}_i^h \partial \tilde{q}_i^m}\left(\tilde{q}_i^h \tilde{q}_i^m\right) + \frac{\partial^2 \tilde{Y}_i^T}{\partial \tilde{q}_i^h \partial \tilde{w}_i}\left(\tilde{q}_i^h \tilde{w}_i\right) + \frac{\partial^2 \tilde{Y}_i^T}{\partial \tilde{q}_i^m \partial \tilde{w}_i}\left(\tilde{q}_i^m \tilde{w}_i\right)
\end{aligned}
$$

assuming that $\frac{\partial^2 \tilde{Y}_i^T}{\partial\tilde{\kappa}^k \partial\tilde{\kappa}^j} = \frac{\partial^2 \tilde{Y}_i^T}{\partial\tilde{\kappa}^j \partial\tilde{\kappa}^k}$ where $\tilde{\kappa} = \{\tilde{q}_i^h, \tilde{q}_i^m, \tilde{w}_i\}$ and the superscript refers to the *j*th element. Here $\tilde{\kappa} = \log\kappa$.

This approximation function has a number of attractive properties, but most importantly imposes no *a priori* restrictions on cross-elasticities between outputs, and inputs and outputs.

### 6.2.2 Empirical specification

Three sets of estimations are undertaken. The primary analysis is of the covariates of transaction costs and is undertaken to address the hypothesis that hierarchies have lower transaction costs. Thereafter a model of total costs is fitted and is used to assess the overall impact of governance choices. Finally, an analysis of production costs is undertaken mainly to assess the precision of the governance variable in its impact on production costs.

Two models are fitted for the transaction costs analysis. The first is the stochastic counterpart of (6-10), with power-transformed variables:

$$\text{(6-12)} \quad \log Y_i^T = \beta_1^T \tilde{\Omega}_i + \beta_2^T \tilde{q}_i^T + \beta_3^T \tilde{q}_i^{nr} + \beta_4^T \log w_i + \langle \beta_5, \tau_i \rangle$$

The Box-Cox metric: $\tilde{x} = (x^\lambda - 1)/\lambda$ is used to tackle the problem of zero values on raw-scales that prevents the use of logarithmic transformations. Nonetheless, the natural log is the limiting case of the Box-Cox metric: $\lim_{\lambda \to 0} \tilde{x} = \ln x$. This ratio model has total residential care output as a control factor. The coefficient $\beta_2^T$ therefore measures substitution effects with given total output.

The second approach to estimating transaction costs was to use the generalised translog multiproduct cost function (GTMCF) (Caves, Christensen, and Tretheway, 1980):

$$(6\text{-}13)\quad \log Y_i^T = \alpha + \beta_h \tilde{q}_i^h + \beta_m \tilde{q}_i^m + \beta_w \log w_i + \beta_{hh} (\tilde{q}_i^h)^2 + \beta_{mm} (\tilde{q}_i^m)^2 + \beta_{ww} (\log w_i)^2 + \beta_{hm} \tilde{q}_i^h \tilde{q}_i^m + \beta_{hw} \tilde{q}_i^h \tilde{w}_i + \beta_{mw} \tilde{q}_i^m \tilde{w}_i$$

where $\alpha = \alpha_0 + \beta_n^T \tilde{q}_i^{nr} + \langle \beta_\tau, \tau_i \rangle$

This is a flexible functional form that is linear in parameters, the empirical counterpart of (6-8), and so has no restrictions on cross-elasticities. However, the Box-Cox transformation is used instead of logs for variables that have zero values on the raw scale.

The total cost and production cost models are the complement to (6-13):

$$(6\text{-}14)\quad \log Y_i^l = \beta_1^l \tilde{\Omega}_i + \beta_2^l \tilde{q}_i^T + \beta_3^l \tilde{q}_i^{nr} + \beta_4^l \log w_i + \langle \beta_5^l \tau_i \rangle, \quad l = E, C$$

Total costs are by definition at the local authority level as they include transaction cost elements.[45]

### 6.2.3 Empirical hypothesis

Table 6-1 lists the signs of the relevant coefficients in the respective models that are consistent with the two main hypotheses developed in chapter 5. We would expect authorities with relatively high proportions of hierarchical provision (that is publicly owned provision) to be associated with relatively low transaction costs, given total output and other control factors. The reverse would be the case for production costs. As to total (transaction and production) costs, it is not *a priori* clear what the sign would be because it is the sum of two countervailing effects.

The analysis is focused on the local authority i.e. purchaser transaction costs. This is motivated mainly because it is local authorities that make decisions about governance

[45] Production costs are service-specific and these costs can be attributed *a priori* to individual production units, that is, care home types. Indeed, analysis conducted elsewhere at the home level show that local authority (hierarchy) homes have higher production costs than homes in the independent (market) sector, other things equal (Netten et al., 1998). Nonetheless, an LA level analysis was also performed to assess the precision of the governance variable in its impact on costs at this level, and to provide some basis for interpretation and/or calibration of the LA level results in correspondence with the home level results.

structures. In any case, as argued in section 5.7 of chapter 5, if the data support hypothesis *H1* then we have a great deal of confidence that they would also support *H1a* (that total transaction costs are higher in markets), although we should be explicit that we are not directly testing *H1a*. Any reference to transaction cost below is to purchaser transaction cost, unless specifically attributed to providers.

**Table 6-1. Expected signs**

| Cost | Model | Variable | Coefficient | Sign | Hypothesis (Chapter 5) |
|---|---|---|---|---|---|
| Transaction (log) | Ratio | Hierarchical/LA output proportion (Box Cox) | $\beta_1^T$ | -ve | H1 (see section 5.7) |
| Transaction (log) | Translog | Net LA output (Box Cox) | $\beta_h + 2\beta_{hh}\overline{\tilde{q}_i^h} + \beta_{hm}\overline{\tilde{q}_i^m} + \beta_{hw}\overline{\tilde{w}_i}$ | +ve | |
| Production | Ratio | Hierarchical/LA output proportion (Box Cox) | $\beta_1^C$ | +ve | H2 (see section 5.9) |
| Total | Ratio | Hierarchical/LA output proportion (Box Cox) | $\beta_1^E$ | ? | |

## 6.3 Residential care in English local authorities

The data are drawn from routinely collected Government data sources – see chapter 1. For further details see Table 6-24 in Annex 6-1.

### 6.3.1 The Costs of residential care

Expenditure data are used to construct three cost dependent variables. Categories available in local authority expenditure collections (the revenue outturn statistics, RO3 – see Table 6-24) are listed in Table 6-2. This categorisation can be aggregated to 7 expenditure variables, which in turn, can be reduced to our three cost variables:[46]

*Total cost = older people production costs (*P1 *to* P7*) + older people transaction costs (*T1 *to* T9*)*

---

[46] The expenditure data include revenue expenditure and capital charges (which includes debt redemption, interest, and capital expenditure charged to revenue account). As ever, the particular accounting convention that underpins these figures may not generate an accurate economic opportunity cost figure. A bespoke bottom-up economic costing is clearly not possible for all expenditure items relating to care of older people for all Social Services departments in England. For our *comparative* purposes anyway, there is no reason to expect accounting conventions to differ by provision type.

**Table 6-2. Expenditure – older people client groups**

| Expenditure category - older people | Transaction | Production |
|---|---|---|
| *Assessment & Commissioning costs* | | |
| Senior management ( eg : Assistant Director costs ) | T1 | |
| Care management / social work | T2 | |
| Management costs (SSMSS)[1] | T3 | |
| *Residential costs* | | |
| Residential care placements : own LA provision | | P1 |
| Residential care placements : commissioned placements | | P2 |
| Nursing placements | | P3 |
| Other residential services for older people | | P4 |
| Management costs (SSMSS): residential | T4 | |
| *Non-residential costs* | | |
| Home care / home help | | P5 |
| Day centres | | P6 |
| Other non-residential costs | | P7 |
| Management costs (SSMSS): non-residential | T5 | |
| *Central strategic* (apportioned to older people client group) | | |
| Strategic management, planning, etc. | T6 | |
| Registration and inspection | T7 | |
| Complaints procedures | T8 | |
| Management costs (SSMSS) | T9 | |

Notes [1] SSMSS: Social Services Management and Support Services

Strategic management costs (T6 + T7 + T8+ T8) are not allocated by client group in the expenditure data.[47] They include more general governance activity such as registration and inspection. Total costs include the production costs of services for older people – residential care, nursing home care, domiciliary care and other forms of residential care. In addition, the costs of care management, planning and assessment are also components.

Descriptive statistics concerning these cost measures for the pooled sample are reported in Table 6-3. Table 6-4 reports the data for the 1998/9 sample, whilst Table 6-5 has the results for the 1997/8 sample. As is usual for cost data each sample exhibits a rightward skew, which has implications for the estimations as discussed below.

[47] Total costs relate to the older people's client group, although in that the data on strategic management (which are transaction costs) are not related to output and client group, a small proportion of the transaction costs that in reality apply to non-older services are included in our measure of 'older people' transaction and total costs. This restriction should not be a problem unless our governance indicator defined by output relating to older people services is strongly negatively correlated to an equivalent measure for non-older people client groups such as mental health and learning disabilities, which is not suggested by the evidence. Absence such correlation the counting of some additional non-older people transaction cost merely introduces more.

One relevant issue is that it is possible that some transaction costs are bound up with the measure of independent sector production cost, taking the form, for example, of providers having to do their own assessments of clients. This would act to exacerbate the results below that markets have high transaction costs and low production costs.

**Table 6-3. Costs (£000s) – pooled sample**

| Cost | Mean | Std. Dev. | Min | Max | Skewness | Kurtosis | Median |
|---|---|---|---|---|---|---|---|
| Total | 34608.84 | 20796.41 | 2329.00 | 125309.00 | 1.59 | 5.93 | 30003.50 |
| Production | 11840.26 | 7721.54 | 518.55 | 43482.00 | 1.65 | 6.13 | 9967.00 |
| Transaction | 5363.24 | 3335.58 | 448.77 | 18279.78 | 1.82 | 6.78 | 4498.76 |

N = 230

**Table 6-4. Costs (£000s) – 1998/9 sample**

| Cost | Mean | Std. Dev. | Min | Max | Skewness | Kurtosis | Median |
|---|---|---|---|---|---|---|---|
| Total | 35783.52 | 21531.61 | 2359.00 | 125309.00 | 1.56 | 5.86 | 30788.00 |
| Production | 12326.63 | 8038.76 | 518.55 | 43482.00 | 1.66 | 6.29 | 10324.00 |
| Transaction | 5749.63 | 3539.47 | 479.18 | 18279.78 | 1.71 | 6.17 | 4875.16 |

N = 113

**Table 6-5. Costs (£000s) – 1997/8 sample**

| Cost | Mean | Std. Dev. | Min | Max | Skewness | Kurtosis | Median |
|---|---|---|---|---|---|---|---|
| Total | 33474.32 | 20088.06 | 2329.00 | 116902.00 | 1.61 | 5.93 | 29708.00 |
| Production | 11370.52 | 7406.49 | 531.75 | 40436.00 | 1.61 | 5.75 | 9221.00 |
| Transaction | 4990.05 | 3095.60 | 448.77 | 16823.78 | 1.93 | 7.43 | 4349.34 |

N = 117

### 6.3.2 The characteristics of residential care

Residential care is characterised for our purposes in terms of activity or output levels, the costs of inputs such as labour, and also a number of process indicators. Table 6-6 summarises the empirical indicators and lists the transformations used to address skewed data.

#### 6.3.2.1 Activity

The analysis aims to allocate costs to different governance arrangements – market and hierarchy. In practice almost all authorities use some combination of these two modes, and to measure their relative proportionate use in each authority we use the proportion of total output that is provided under each arrangement. The latter measure is our 'governance indicator', $\Omega_i$ and it is constructed using output, $q_i^h$ and $q_i^m$ – see section 6.2.1 (the merits of this definition are discussed in the conclusion).

**Table 6-6. Variables - definition**

| Category | Variable | Definition | Transformations |
|---|---|---|---|
| Output | LA Output | LA provider residential care output in resident-weeks | Box-Cox |
| | Ind Output | Independent provider residential care output in resident-weeks | Box-Cox |
| | Output ratio | Ratio of LA provider output to total residential care output | Box-Cox |
| | Output ratio (BC) × wage (log) | Ratio of LA provider output to total residential care output (Box-Cox) × wage (log) | Box-Cox, log |
| | Total output | Total residential care output: resident weeks | Log |
| | Domiciliary output | Older people helped to live at home per 1,000 population aged 65 and over | Log |
| | Domiciliary output | Older people helped to live at home | Linear |
| Input costs | Wage | Social care wages | Linear<br>Log |
| Population | Pop 65+ | Population aged 65 plus | Log |
| | Pop 65-75 | Population aged 65 to 75 | Log |
| | Pop ratio | Ratio of population aged 75 and over to population aged between 65 and 75 | Log<br>Linear |
| Dependency | SMR | Standardised mortality ratio (uk=100) | Linear |
| | Hosp% | Number of hospital admissions of people aged 75 and over with a diagnosis of hypothermia per older population | Linear |
| | Gender | Proportion of females to males in LA | Linear |
| Process factors | Statement | Percentage of people receiving a needs/service statement | Log<br>Linear |
| | Single room% | Percentage offered a single room | Linear<br>Dummy |
| Time | First period | Observation for 1997/8 (first time period) | Dummy |
| LA dummies | Met | Metropolitan LA | Dummy |
| | Shire | Shire LA | Dummy |
| | InnLon | Inner London LA | Dummy |
| | OutLon | Outer London LA | Dummy |

Activity information is reported in Table 6-7 for the whole sample, Table 6-8 for the 1998/9 cohort and in Table 6-9 for the 1997/8 cohort. The average ratio of local authority resident-weeks to independent sector resident-weeks in 1998/9 was a third, down slightly from its value in 1997/8. Throughout the analysis period 2 authorities had no in-house provision of residential care for older people and relied entirely on market means of securing service for their clients. The greatest proportion of in-house use was just over three quarters of all resident-weeks. Using total resident-weeks estimates, the average production cost per week is £280.

**Table 6-7. Activity – pooled sample (N = 230)**

| Variable | | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|---|
| *Residential care* | | | | | |
| Supported residents - LA homes | LA Supp | 354.17 | 291.26 | 0.00 | 1637.00 |
| Supp. residents - Indep. homes | Ind Supp | 813.00 | 529.42 | 39.00 | 2667.00 |
| Resident weeks – LA homes | LA Output | 15847.70 | 14083.66 | 0.00 | 77804.45 |
| Resident weeks – Indep. homes | Ind Output | 29117.22 | 20794.85 | 1296.38 | 110029.80 |
| Resident weeks – ratio (LA/all) | Output ratio | 0.34 | 0.16 | 0.00 | 0.76 |
| *Domiciliary care* | | | | | |
| Older people helped to live at home (no. in LA) | Dom. Output | 3822.96 | 2899.49 | 162.83 | 19597.62 |

**Table 6-8. Activity – 1998/9 sample (N = 113)**

| Variable | | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|---|
| *Residential care* | | | | | |
| Supported residents - LA homes | LA Supp | 346.32 | 281.03 | 0.00 | 1197.00 |
| Supp. residents - Indep. homes | Ind Supp | 848.41 | 563.54 | 39.00 | 2667.00 |
| Resident weeks - LA homes | LA Output | 15457.04 | 13482.66 | 0.00 | 58571.57 |
| Resident weeks – Indep. homes | Ind Output | 30389.03 | 22205.45 | 1296.38 | 110029.80 |
| Resident weeks – ratio (LA/all) | Output ratio | 0.33 | 0.15 | 0.00 | 0.64 |
| *Domiciliary care* | | | | | |
| Older people helped to live at home (no. in LA) | Dom. Output | 3828.54 | 2934.27 | 164.33 | 19536.91 |

**Table 6-9. Activity – 1997/8 sample (N = 117)**

| Variable | | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|---|
| *Residential care* | | | | | |
| Supported residents - LA homes | LA Supp | 361.75 | 301.82 | 0.00 | 1637.00 |
| Supp. residents - Indep. homes | Ind Supp | 778.81 | 494.27 | 40.00 | 2432.00 |
| Resident weeks - LA homes | LA Output | 16225.01 | 14688.88 | 0.00 | 77804.45 |
| Resident weeks – Indep. homes | Ind Output | 27888.90 | 19351.79 | 1329.38 | 94037.44 |
| Resident weeks – ratio (LA/all) | Output ratio | 0.36 | 0.17 | 0.00 | 0.76 |
| *Domiciliary care* | | | | | |
| Older people helped to live at home (no. in LA) | Dom. output | 3817.568 | 2878.104 | 162.825 | 19597.62 |

Domiciliary care activity is also listed in the tables, measured here as the number of older people in the local authority in receipt of domiciliary care services. This number is equivalent to about 8 per cent of the population over 65 and 18 per cent of the population over 75. The highest proportion of over 75's receiving this care was 41 per cent.

Activity is asymmetrically distributed with a rightward skew (skew = 1.65 and Kurtosis = 5.57 for LA resident-weeks and skew = 1.65 and Kurtosis = 5.72 for independent sector resident-weeks). A (natural) logarithmic power transformation reduces this asymmetry and was employed in the model for the total weeks variable. The governance variable $\Omega_i$ was also transformed. The existence of zero values for the ratio $\Omega_i$ rules out a log transform and instead a Box-Cox transformation was used – see section 6.2.2.

#### 6.3.2.2 Control factors

A wide range of other factors compound the relationship between the cost and governance – the $w_i$ and $\tau_i$ terms in (6-1) and following equations. The $w_i$ are inputs costs such as labour and capital costs. The vector $\tau_i$ are shift factors including, client dependency, population, authority types, and process/quality indicators. Table 6-10 below describes the empirical proxies used in the analysis..

### 6.3.3 Sample sizes and Missing values

At the start of 1998, England had 132 local authorities. During the course of the year 10 of these local authorities (Shire counties) were re-organised into 28 new local authorities, making 150 local authorities from 1999 (and currently). From these 10 Shire LAs, 19 new unitary authorities were created, as was one new Shire LA. Two of the original Shire LAs ceased to exist, with eight remaining in name, but with new boundaries.

In order to have a consistent panel dataset the 20 new authorities and the eight Shires with new boundaries were deleted. Accordingly the data set used 122 authorities over the two-year period, a total valid sample of 244. Eight cases (Brent 98/9, Bromley 98/9, Stockport 98/9, Wandsworth 98/9, and Somerset and Tameside in both periods) were dropped where a positive expenditure was recorded in maintaining people in homes but the activity statistics recorded zero places. A further 2 cases (the Isles of Scilly in both periods) were dropped because information on domiciliary care was missing. Four cases (two local authorities in both periods: Southwark and North Lincolnshire) were dropped due to missing data on the number of LA supported residents. Hence the regression sample was 230 cases and 117 local authorities (95 % of the valid sample).

**Table 6-10. Descriptive statistics – pooled sample (N=230)**

| Variable name | Factor | Mean | Std.Dev. | Min | Max |
|---|---|---|---|---|---|
| ***Input cost*** | | | | | |
| Wage | Input cost | 5.157E+01 | 1.422E+01 | 3.483E+01 | 8.810E+01 |
| Dwell price | Average dwelling prices, all dwellings (1998) | 8.676E+04 | 2.842E+04 | 5.705E+04 | 1.278E+05 |
| ***Population*** | | | | | |
| Popall | Population - all | 3.247E+02 | 2.181E+02 | 5.000E+00 | 1.231E+03 |
| Pop 65-75 | Population aged 65 to 75 | 2.727E+04 | 1.981E+04 | 5.260E+02 | 1.050E+05 |
| Pop 65+ | Population aged 65 and over | 5.081E+04 | 3.717E+04 | 9.750E+02 | 1.928E+05 |
| Pop ratio | Ratio of population aged 75 and over to population aged between 65 and 75 | 8.619E-01 | 1.083E-01 | 5.755E-01 | 1.199E+00 |
| Pop density | Population per Km squared | 2.628E+03 | 2.735E+03 | 6.100E+01 | 1.367E+04 |
| ***Dependency*** | | | | | |
| SMR | Standardised mortality rate (UK=100) | 9.974E+01 | 1.020E+01 | 6.500E+01 | 1.240E+02 |
| Gender | Proportion of females to males in LA | 1.031E+00 | 4.300E-02 | 6.667E-01 | 1.122E+00 |
| Hosp% | Number of hospital admissions of people aged 75 and over with a diagnosis of hypothermia | 1.058E-02 | 3.570E-03 | 2.261E-04 | 2.003E-02 |
| ***Process Indicators*** | | | | | |
| Statement | Percentage of people receiving a needs/service statement | 7.219E+01 | 2.227E+01 | 5.000E+00 | 1.000E+02 |
| Single rm% | Percentage offered a single room | 9.360E+01 | 1.043E+01 | 4.700E+01 | 1.000E+02 |
| ***LA types*** | | | | | |
| Met | Metropolitan LA | 3.000E-01 | 4.593E-01 | 0.000E+00 | 1.000E+00 |
| Shire | Shire LA | 2.087E-01 | 4.073E-01 | 0.000E+00 | 1.000E+00 |
| InnLon | Inner London LA | 1.000E-01 | 3.007E-01 | 0.000E+00 | 1.000E+00 |
| OutLon | Outer London LA | 1.652E-01 | 3.722E-01 | 0.000E+00 | 1.000E+00 |
| Unitary | Unitary LA | 2.261E-01 | 4.192E-01 | 0.000E+00 | 1.000E+00 |

## 6.4 Estimation techniques

In brief, two estimators were used. The first is the population-averaged general estimating equations (GEE) model, which is analogous to the generalised linear model (GLM) and was used to address both skewed data issues and the panel dataset. The second is the (one-way) GLS random-effects model, a more conventional estimator, which allowed more standardised diagnostic testing. The remainder of this section provides a more detailed discussion of the chosen techniques.

### 6.4.1 Transformation issues and GLM

As noted above the raw scale cost data have the usual rightward skew. Generalised linear models (GLM) estimation was used to address this problem. GLM involves the *a priori* specification of a link function that relates a linear combination of coefficients and independent variables with the dependent variable. A log-link function was used in the present analysis:

$$\text{(6-15)} \quad \log(E(y_i)) = \log(\mu_i) = \eta_i = x_i'\beta$$

and therefore, $E(y_i) = \exp(x_i'\beta)$. The GLM formulation has the *expected* value, $\mu$, of the data transformed according to the link function. The data $Y$ are assumed to be drawn from a particular distribution with an expected value $\mu$ and variance $\text{var}(Y_i)$, conditional on $x$. The coefficients $\beta$ are then chosen to maximise the likelihood, given the fixed $x_i$'s, that $\mu_i$ is the expected value of specified distribution from which $Y_i$ was drawn. Hence the expected value is a systematic (non-stochastic) function of the optimal $\beta$ estimates. Wedderburn suggests the estimation of parameters by maximising *quasi-likelihood* functions because it avoids specific *a priori* parameterisation of the distribution (Wedderburn, 1974). These functions are otherwise known as generalised estimating equations (GEE) and take the form:

$$\text{(6-16)} \quad \sum_{i=1}^{N} \left( \frac{\partial \mu}{\partial \beta} \right)' \text{var}(Y_i)^{-1} (Y_i - \mu_i) = 0$$

This parameter vector $\beta$, the solution to (6-16), is found by iteration using Fisher's scoring method (McCullagh and Nelder, 1989). The advantage of using GEE estimation is that re-transformation of (6-15) to the raw scale gives an expected value of the cost variable as a non-stochastic function of only the independent variables: $E(y_i) = \exp(x_i'\beta)$. Compare this to an equivalent OLS regression of a log-transformed dependent variable: $\log(Y_i) = x_i'\beta + \varepsilon_i$, which on re-transformation is

(6-17) $E(Y_i) = \exp(x_i'\beta)E(\exp(\varepsilon_i))$

Calculating the expected value of the exponentiated error term (which is not equal to zero) is difficult, particularly in the presence of heteroscedasticity[48] (Manning, 1998; Manning and Mullahy, 2001).

#### 6.4.1.1 Choosing distribution functions for cost estimates

The drawback of GEE estimation is the need to specify a distribution for the cost variable. In general the variance can be written as $\text{var}(Y_i) = a(\phi)V(\mu_i)$. The use of extended quasi-likelihood estimation of GLM obviates the need to specify exactly the distribution of $Y$, that is, the value of the dispersion or 'scale' parameter, $\phi$ (indeed, the $a(\phi)$ term drops out of the estimation function). However, an appropriate *variance function*, $V(\mu)$, must be specified. The class of (exponential) distributions from which this choice is made in GLM estimation can be summarised as: (see Forder, 2005)

(6-18) $\text{var}(Y) = a(\phi)V(\mu_i) = \kappa_0\mu_i^\kappa = \kappa_0\left(g^{-1}(\eta_i)\right)^\kappa$

where $\kappa$ is an integer on the support $\kappa \in [0,3]$. Manning (1999) and Blough (1999) suggest that estimates of $\kappa$ can be derived from an OLS estimation of (6-18), when taking the following form, and given a log-link function:

(6-19) $\log\left((Y_i - \hat{\mu}_i)^2\right) = \log(\kappa_0) + \kappa\log(\exp(x_i'\beta)) + \varepsilon_i = \alpha + \kappa\eta_i + \varepsilon_i$

where $\hat{\mu}$ is the predicted expected value. This is a form of Park Test for heteroscedasticity. Since we are interested in $\partial[\log(\text{var}(Y_i))]/\partial\eta_i = \kappa$, re-transformation issues are not relevant in this Park estimation. The variance (6-18) was also estimated

[48] The *smearing* estimator has been suggested as a robust, non-parametric alternative (Duan, 1983). It is based on the estimated residuals $\hat{\varepsilon} = \log(Y_i) - x_i'\hat{\beta}$ and takes the form: $s = N^{-1}\sum_{i=1}^{N}\exp(\hat{\varepsilon}_i)$. The estimated value of the dependent variable then becomes: $E(Y_i) = \exp(x_i'\beta)s$. Again however, this estimator is problematic when the error exhibits non-constant variance, that is, when $\varepsilon = \varepsilon(x)$. (In practice, log transformation may remove non-constant error variance problems that often arise with cross-sectional analysis performed on the raw scale).

by non-linear least squares, although this carries more assumptions than the OLS model. The NLS model produced generally equivalent results although the estimates of $\kappa$ were slightly smaller.

The estimated variance will depend on the predicted value $\hat{\mu}_i$, which in turn depends on the chosen variance function. Nonetheless, whilst the inappropriate choice of variance function will lead to some imprecision of the estimates of $\hat{\mu}_i$, in the main the shape of the $\mathrm{var}(Y)$ will be maintained (for relevant conditions, see Manning and Mullahy, 1999). On the whole the three models produced consistent Park Test values of $\kappa \approx 2$ for each of following estimators: (i) OLS on log dependent variable with assumed homoscedastic log-normal errors, (ii) GLM with Gaussian variance function (iii) GLM with Poisson variance function, (iv) GLM with gamma variance function, and (v) GLM with inverse Gaussian variance function. These tests supported our adopted assumption of quadratic variance, that is, the gamma distribution.

### 6.4.2 Panel data – random effects

In this study two consecutive years of cross-sections of local authorities were used:

$$(6\text{-}20)\quad Y_{it} = \alpha + x_{it}'\beta + \varepsilon_{it},\ i = 1,\ldots,N; t = 1,\ldots,T$$

A panel data set has the advantage of controlling for some omitted variable effects, although clearly with only two years of data this is limited in the current analysis. A two-way model can also tackle some of the problems of omitted time effects that are, in addition, relatively invariant across observational units:

$$(6\text{-}21)\quad Y_{it} = \alpha + x_{it}'\beta + u_t + v_i + \varepsilon_{it}$$

Such omissions can result in heterogeneity bias (Hsiao, 1986).

### 6.4.3 Skewed data and panel effects

In light of the re-transformation problem discussed above the primary estimation method used is GEE. The use of a panel data is addressed using a population-averaged GEE model, which is an extension of (6-16):

(6-22) $\sum_{i=1}^{N}\left(\frac{\partial\mu}{\partial\beta}\right)' V_i^{-1}(\alpha)(Y_i-\mu_i)=0$

where $V_i(\alpha)=A_i^{\frac{1}{2}}R_i(\alpha)A_i^{\frac{1}{2}}$ and $A_i = diag\{V_{i1},...,V_{in_i}\}$ (see Forder, 2005). Here $R$ is the working correlation matrix and assumptions are required concerning its structure (for details see StataCorp, 1999). The primary modelling choice for this analysis was the "unstructured" distribution, although in practice, with a limited number of time periods, the one-way and this two-way model are little different.

To provide some comparison, a more conventional estimator was also used, the (one-way) GLS random-effects model (Greene, 1993):

(6-23) $Y_{it}=\alpha+x_{it}'\beta+u_i+\varepsilon_{it}$

A number of diagnostics are available for this estimation. In random effects estimations the group effects are assumed to be uncorrelated with the other regressors. Should this assumption not hold, the estimator would be inconsistent (i.e. mis-specified due to omitted variables). This assumption is not however needed for fixed effect estimators, and this characteristic is used as a basis for a test of the former (Hausman, 1978). Hausman's test is that for the random effects estimator to be consistent it should not differ systematically from the fixed effects estimator, and this forms the null hypothesis for a chi-squared test. For none of the models described below could this null hypothesis be rejected. To test that the variance of additional error $u_i$ is significantly different from zero, the Breusch and Pagan Lagrangian multiplier test was used. Again this condition was satisfied for all the models reported below.

## 6.5 Results

### 6.5.1 Transaction costs

As noted above two models of transaction costs were estimated: the ratio model (6-12) and the translog model (6-13). GEE was the primary method used to estimate both these models. As noted, GLS with a log-transformed dependent variable was also estimated.

### 6.5.1.1 Transaction cost ratio model

Table 6-11 lists the variables in the ratio model and both the GEE and GLS estimates. The cost functions utilise variables from five categories: output, input costs, population, dependency and process factors – see Table 6-6 for a definition and sections 6.2.1 and 6.2.2. Both models were highly significant overall.

**Table 6-11. Transaction costs models – Linear, panel and Generalised Estimating Equations**

| | | GLS, random effects | | GEE, random effects | |
|---|---|---|---|---|---|
| **Variable** | | **Coefficient** | **Std error** | **Coefficient** | **Std error** |
| Output ratio (BC) | | 1.204E+00 | 3.594E-01* | 1.267E+00 | 3.471E-01* |
| Output ratio (BC) * wage (log) | | -3.252E-01 | 9.321E-02* | -3.420E-01 | 9.002E-02* |
| Total output (log) | | 1.414E-01 | 8.301E-02* | 1.591E-01 | 8.117E-02* |
| Domiciliary output (log) | | 2.617E-01 | 1.264E-01* | 3.032E-01 | 1.216E-01* |
| Domiciliary output | | -1.440E-08 | 1.800E-08 | -1.900E-08 | 1.740E-08 |
| Wage (log) | | 4.357E-01 | 1.445E-01* | 3.973E-01 | 1.392E-01* |
| Pop 65-75 (log) | | 7.648E-01 | 1.095E-01* | 7.724E-01 | 1.061E-01* |
| Pop ratio (log) | | -1.595E-01 | 2.259E-01 | -1.032E-01 | 2.177E-01 |
| SMR | | 6.618E-03 | 3.471E-03* | 6.188E-03 | 3.348E-03* |
| Single room% (squared) | | 2.270E-04 | 1.446E-04 | 1.852E-04 | 1.412E-04 |
| Single room% | | -3.526E-02 | 2.425E-02 | -2.839E-02 | 2.370E-02 |
| Hosp% | | 9.498E+00 | 6.567E+00 | 1.258E+01 | 6.318E+00* |
| Constant | | -3.078E+00 | 1.494E+00* | -3.577E+00 | 1.448E+00* |
| Model | | Linear (GLS), random effects | | GEE population-averaged | |
| Link/dependent variable | | Log (transaction cost) | | log | |
| Family | | | | gamma | |
| Correlation | | | | unstructured | |
| Number of obs | | 230 | | 230 | |
| Number of groups | | 117 | | 117 | |
| Obs per group: | min | 1 | | 1 | |
| | avg | 2.0 | | 2.0 | |
| | max | 2 | | 2 | |
| Wald chi2(9) | | 554.68 | | 610.85 | |
| Prob > chi2 | | <0.0000 | | <0.0000 | |
| Scale parameter | | | | .0717761 | |
| R-sq: | within | 0.0026 | | | |
| | between | 0.8400 | | | |
| | overall | 0.7992 | | | |
| rho | | 0.4819 | | | |
| Hausman specification test | | 9.40 (NS) | | | |
| Breusch and Pagan test | | 17.00 * | | | |

Note * denotes significant at 10 per cent or better

The GEE model was estimated using a gamma error specification; the Park test results (see section 6.4.1.1) are listed in Table 6-12. The table lists the estimate of the specification integer κ for the different assumed error functions. Except for the Gaussian form, the nearest integer is κ = 2, which is the gamma specification. The GLS model satisfied the Hausman and the Breusch and Pagan LM tests to support the choice of random effects. The model also produced a high measure of fit.

**Table 6-12. Park tests – assumed variance functions**

| *Error specifications* | GEE ratio model | | |
|---|---|---|---|
| | **Transaction costs** | **Production costs** | **Total costs** |
| Log OLS - hom, normal | 1.946E+00 | 1.865E+00 | 2.024E+00 |
| Gaussian | 1.426E+00 | 2.006E+00 | 9.967E-01 |
| Poisson | 1.694E+00 | 1.877E+00 | 1.164E+00 |
| Gamma | 1.862E+00 | 1.914E+00 | 2.005E+00 |
| Inverse Gaussian | 1.807E+00 | 2.419E+00 | 2.229E+00 |

A Box-Cox transformation of the ratio $\Omega_i$, the variable *Output ratio (BC)*, was used to address the main hypothesis. Some experimentation suggested that the inclusion of an interaction between the output ratio and the (log of the) input price variable, *wage (log)* produced the best fit. Overall, because the different estimators produced very similar results, further analysis below is of the GEE estimation results.

Table 6-13 summarises in the main results of the estimations. The relevant coefficient is the change in (log) transaction costs associated with a marginal increase in the listed factor. Also, the for GEE model, elasticities are reported. Elasticity in this case is given by

(6-24) $$\varepsilon = \frac{\left(\Delta Y^T / Y^T\right) \times 100}{\left(\Delta q^h / q^h\right) \times 100} = \frac{q^h}{Y^T} \frac{\partial Y^T}{\partial q_h}\bigg|_{q^T}$$

as $\Delta Y^T = \partial Y^T / \partial q^h \big|_{q^T} \Delta q^h$ and where the marginal change in transaction costs for a (relative) increase in hierarchical output (i.e. given total output) is $\partial Y^T / \partial q^h \big|_{q^T} = \beta_R \left(q^h / q^T\right)^{\lambda-1} \left(1/q^T\right) Y^T$, which is equivalent to $\partial Y^T / \partial q^h - \partial Y^T / \partial q^m$.

**Table 6-13. Transaction costs models – Derived marginal effects**

| Variable | GLS, random effects Coefficient | Std error | | GEE, random effects Coefficient | Std error | | Elasticity |
|---|---|---|---|---|---|---|---|
| Output ratio (constant output) | -3.761E-06 | 1.129E-06 | * | -3.962E-06 | 1.081E-06 | * | -0.063 |
| Output (constant output ratio) | 3.144E-06 | 1.846E-06 | * | 3.537E-06 | 1.805E-06 | * | 0.159 |
| Wage | 8.483E-01 | 1.189E-01 | * | 8.313E-01 | 1.145E-01 | * | 0.831 |
| Domiciliary output | 5.405E-05 | 2.310E-05 | * | 6.031E-05 | 2.180E-05 | * | 0.231 |

Note * denotes significant at 10 per cent or better

Holding total output constant, a change in the proportion of hierarchically arranged provision (output ratio) was negatively related to transaction costs. Thus we infer that LAs with a high proportion of hierarchically arranged provision have relatively low transaction costs, and vice versa. The use of an interaction term in the model suggests that this effect is exacerbated in areas of high input costs. In other words, in such areas e.g. London, a small increase in the proportion of hierarchically arranged provision is associated with a greater reduction in transaction costs than for areas of low input costs.

By contrast, holding output ratio constant, an increase in total output is associated with an increase in transaction costs. Although statistically significant the size of the effect is small; greater output means higher transaction costs, but it also suggests that a large part of total transaction costs is a fixed transaction cost element (this inference is explored below). A nursing home output variable was originally specified but was dropped due to a high degree of collinearity with residential care output. The correlation between (total) resident-weeks in residential care and resident-weeks in nursing homes was 0.84 (Spearman's test of independence rejected at $p < 0.001$) (see also below).

Domiciliary care output is also positively related to transaction costs. Since the transaction cost measure includes services for all older people this finding is expected. The estimated elasticity is also comparable with that associated with (total) residential care activity. Other non-residential services were not included due to their high correlation with domiciliary care output; correlation between spending in 1998/9 on domiciliary and other non-residential services was 0.78 (Spearman's test of independence rejected at $p < 0.001$).[49]

[49] Expenditure on other non-residential services could have been included, possibly reducing co-linearity problems, but would have introduced an input cost element as well as activity.

Input prices, with wage as a proxy, are positively related to transaction costs and have an estimated elasticity near to one. This implies that transaction costs are sensitive to input prices as might be expected. If the input price (wage) accounts for much of the average of total transaction costs then its elasticity would be close to one.

A number of control factors were also significant. The size of the population aged 65 to 75 in the local authority had a positive association with transaction cost, with an elasticity of 0.77. This variable might capture some older people client group service output not picked up directly by the residential care and domiciliary care activity measures. It also may be an LA scale factor. Dependency measures, such as the LA's standardised mortality ratio (*SMR*) and the number of hospital admissions of people aged 75 and over with a diagnosis of hypothermia per older people population (*Hosp%*) were also significant. Higher average dependency – reflected by higher values of these two measures – would be expected to increase transaction costs. Highly frail service users would need a more considered and extensive specification of care packages and greater follow-up that would increase costs. This general hypothesis is borne-out by the data: *SMR* and *Hosp%* have elasticities of, respectively of 0.62 and 0.01.

#### 6.5.1.2 Transaction cost translog model

The translog model – as a flexible functional form not limited to a ratio – is a good basis for calculating marginal transaction costs directly (Table 6-14). It is estimated with the same set of population, dependency and process factors, but with a flexible specifications of outputs and input costs that allow second order effects and output interactions with input costs. A gamma error specification was used to be consistent with the ratio model, and the associated Park test value supported this choice. Table 6-15 reports the translog model using GLS with a log dependent variable, which was estimated for comparison.

A marginal change in resident-weeks of hierarchical (LA) output (not holding total output constant) is related to total transaction cost as follows in the translog model:

$$(6\text{-}25)\quad \partial Y^T/\partial q^h = \left(\beta_h + 2\beta_{hh}\tilde{q}_i^h + \beta_{hm}\tilde{q}_i^m + \beta_{hw}\log w_i\right)\left(q^h\right)^{\lambda-1} Y^T = \vartheta^h\left(q^h\right)^{\lambda-1} Y^T$$

Similarly, for market (independent) output, the marginal effect on total transaction cost is:

$$(6\text{-}26)\quad \partial Y^T/\partial q^m = \left(\beta_m + 2\beta_{mm}\tilde{q}_i^m + \beta_{hm}\tilde{q}_i^h + \beta_{mw}\log w_i\right)\left(q^m\right)^{\lambda-1} Y^T$$

**Table 6-14. Transaction costs models – Translog model (Generalised Estimating Equations)**

| Variable | | Linear, random effects Coefficient | Std error |
|---|---|---|---|
| LA output (Box-Cox) | $\beta_h$ | 1.468E-02 | 2.032E-02 |
| Ind output (Box-Cox) | $\beta_m$ | 3.909E-02 | 2.004E-02* |
| Wage (log) | Log $w$ | 2.692E+00 | 4.242E+00 |
| LA output (Box-Cox) squared | $\beta_{hh}$ | -9.680E-06 | 4.180E-05 |
| Ind output (Box-Cox) squared | $\beta_{mm}$ | -6.580E-05 | 6.100E-05 |
| Wage (log) squared | (Log $w$)$^2$ | -1.234E-01 | 5.126E-01 |
| LA output (Box-Cox) * wage (log) | $\beta_{hw}$ | -6.431E-03 | 5.221E-03 |
| Ind output (Box-Cox) * wage (log) | $\beta_{mw}$ | -7.775E-03 | 5.186E-03 |
| LA output (Box-Cox) * Ind output (Box-Cox) | $\beta_{hm}$ | 1.507E-04 | 7.530E-05* |
| Domiciliary output (log) | | 3.182E-01 | 1.528E-01* |
| Domiciliary output | | -2.690E-08 | 1.870E-08 |
| Pop 65-75 (log) | | 6.510E-01 | 9.768E-02* |
| Pop ratio (log) | | -1.628E-01 | 2.245E-01 |
| SMR | | 4.029E-03 | 4.156E-03 |
| Single room% | | 3.216E-03 | 1.751E-03* |
| Hosp% | | 9.951E+00 | 7.066E+00 |
| Constant | | -9.158E+00 | 8.700E+00 |
| Model | | GEE population-averaged | |
| Link/dependent variable | | log | |
| Family | | gamma | |
| Correlation | | unstructured | |
| Number of obs | | 230 | |
| Number of groups | | 117 | |
| Obs per group: min | | 1 | |
| avg | | 2.0 | |
| max | | 2 | |
| Wald chi2(9) | | 794.98 | |
| Prob > chi2 | | <0.0000 | |
| Scale parameter | | .0757969 | |
| Park test parameter | | 1.955901 | |

Note * denotes significant at 10 per cent or better

Using the estimation results these two marginal effects are respectively £6 and £41 based on mean values, giving a (marginal) transaction costs ratio of 0.15 and a difference of £35 (see Table 6-16). The ratio model can also be used to estimate marginal effects. The GEE ratio model generated a ratio of marginal costs of 0.18. The overall results imply that, at the margin an increase in provision secured in the market is associated with higher extra transaction costs than using hierarchical means. For comparison – but noting the re-transformation problem – the marginal effects from the GLS model were very similar: £9 and £37 respectively.

**Table 6-15. Transaction costs models – Translog model (GLS random effects)**

| Variable | | Linear, random effects Coefficient | Std error | |
|---|---|---|---|---|
| LA output (Box-Cox) | $\beta_h$ | 1.108E-02 | 2.493E-02 | |
| Ind output (Box-Cox) | $\beta_m$ | 4.255E-02 | 2.537E-02 | * |
| Wage (log) | Log $w$ | 1.413E+00 | 4.222E+00 | |
| LA output (Box-Cox) squared | $\beta_{hh}$ | 1.810E-06 | 4.910E-05 | |
| Ind output (Box-Cox) squared | $\beta_{mm}$ | -4.280E-05 | 6.380E-05 | |
| Wage (log) squared | (Log $w$)$^2$ | 4.844E-02 | 5.254E-01 | |
| LA output (Box-Cox) * wage (log) | $\beta_{hw}$ | -5.317E-03 | 6.281E-03 | |
| Ind output (Box-Cox) * wage (log) | $\beta_{mw}$ | -9.433E-03 | 6.373E-03 | |
| LA output (Box-Cox) * Ind output (Box-Cox) | $\beta_{hm}$ | 1.273E-04 | 9.000E-05 | * |
| Domiciliary output (log) | | 2.937E-01 | 1.487E-01 | * |
| Domiciliary output | | -2.750E-08 | 2.380E-08 | |
| Pop 65-75 (log) | | 6.661E-01 | 1.124E-01 | * |
| Pop ratio (log) | | -1.624E-01 | 2.485E-01 | |
| SMR | | 4.518E-03 | 3.400E-03 | |
| Single room% | | 3.645E-03 | 2.196E-03 | * |
| Hosp% | | 7.221E+00 | 6.810E+00 | |
| Constant | | -6.819E+00 | 8.698E+00 | |
| Model | | GEE population-averaged | | |
| Dependent variable | | Log | | |
| Number of obs | | 230 | | |
| Number of groups | | 117 | | |
| Obs per group: min | | 1 | | |
| avg | | 2.0 | | |
| max | | 2 | | |
| Wald chi2(9) | | 517.09 | | |
| Prob > chi2 | | <0.0000 | | |
| R-sq: within | | 0.008 | | |
| between | | 0.837 | | |
| overall | | 0.792 | | |
| rho | | 0.505 | | |
| Hausman specification test | | 17.84 | | |
| Breusch and Pagan test | | 19.66 | | |

Standard errors for complicated relationships such as (6-25) and (6-26) can be estimated using a bootstrapping methodology. This technique draws repeated samples from the data. In this case, it runs the model on that drawn sample calculating the coefficients as normal. It also derives the mean values of the $q$ and log$w$ factors for the sample and stores these and the coefficients for each repetition. The marginal effects are derived for each repetition and summarised using central tendency and variance (standard error) statistics for the, in this case, 1000 repetition sampling distribution. Because this sampling procedure is unbiased the derived slopes equal the above marginal values. The estimated standard errors are used to calculate confidence intervals.

**Table 6-16. Marginal transaction costs**

| Output | Unit | Marginal cost |
|---|---|---|
| LA residential | Resident-weeks per year | 6 |
| Independent residential | Resident-weeks per year | 41 |
| Domiciliary | Clients served per year | 303 |

This bootstrapping method was initially tried for the GEE model, but suffered the problem of the likelihood functions failing to converge. Results were obtained, however, for the computationally (much) simpler GLS model. For the hierarchical marginal effect, standard errors of 16.5 were found (compared to a marginal effect of 8.8), indicating that these marginal effects are not significantly different from zero. For the market marginal effect, estimated standard errors were similar in size at 14.5, relative to a marginal effect of 37.4, indicating strong significantly positive marginal transactions cost. These results are consistent with the significance of the governance ratio variable in the ratio models above. They are very likely to be representative of the significance of the GEE model, and in view of this, and the issues around re-transformation, we proceed using the GEE results.

The GEE model was also estimated with the market output variable $q^m$ specified as the sum of independent sector residential and nursing home resident-weeks. The resultant marginal costs differences were very similar.

This model also provides an estimate of the marginal transaction costs of domiciliary care:

$$(6\text{-}27)\quad \partial Y^T / \partial q^d = \left( \frac{\beta_{d1}}{q^d} + \beta_{d1} \right) Y^T$$

which produces a point estimate of marginal costs of £303 per client served.

### 6.5.1.3 Average versus marginal transaction costs

For policy purposes the *average* transaction cost per unit output is a particularly useful measure. Transactions costs are composed of a fixed and variable element. Generally speaking the models above give marginal effects. However, by making some assumptions we can go further and explore the fixed costs as well. To proceed therefore, and given the above results, we make the assumption of *constant marginal costs* (see Annex 6-2 for an analysis of this assumption). Total *variable* cost for the major service groups used by older people – i.e. LA (hierarchical) residential care, independent (market) residential care, nursing home care, domiciliary care and other non-residential – is:

$$(6\text{-}28)\quad V^T = \left(\partial Y^T/\partial q^h\right)q^h + \left(\partial Y^T/\partial q^m\right)q^m + \left(\partial Y^T/\partial q^n\right)q^n + \left(\partial Y^T/\partial q^d\right)q^d + \left(\partial Y^T/\partial q^g\right)q^g$$

i.e. $\dfrac{V^T}{q} = \dfrac{\partial V^T}{\partial q}$ if $\dfrac{\partial^2 V^T}{\partial q^2} = 0$. We do not have a direct estimate of the marginal costs for nursing home output, but we might assume it to have similar marginal transaction costs to market residential care output being provided entirely under market arrangements. Nor do we have an estimate for other non-residential care, $\partial Y^T/\partial q^g$. For day care we assume that the marginal cost per client served is the same as for domiciliary care output: $\partial Y^T/\partial q^{g1} = \partial Y^T/\partial q^d$. For other non-residential services (e.g. meals) we arbitrarily assume marginal costs of half that figure $\partial Y^T/\partial q^{g2} = \frac{1}{2}\partial Y^T/\partial q^d$. In either of these latter cases the size of expenditure is very small relatively speaking, so these assumptions have a limited bearing anyway. Fixed transaction cost for the average local authority is then: $F^T = Y^T - V^T$. To avoid double counting, these fixed costs need to be allocated to service type: residential, nursing homes and non-residential. Without other information this allocation – i.e. $F^{Tk}$ in $F^T = F^{T1} + F^{T2} + \ldots,$ – is assumed to be in the same proportions that each service type contributes to total variable costs, $V^T$, e.g. for residential we have a proportion: $\Psi^r = V^{Tr}/V^T = \left[\left(\partial Y^T/\partial q^h\right)q^h + \left(\partial Y^T/\partial q^m\right)q^m\right]/V^T$. Similarly, we derive $\Psi^n$ and $\Psi^D$ respectively for nursing and non-residential. These estimates provide a breakdown of transaction costs $Y^{Tk} = F^T\Psi^k + V^{Tk}$ for $k = r, n, D$.

Total transaction cost for all services is $Y^T$ = 5363,000 (for the sample average authority – see Table 6-3). Each service category variable costs are calculated from (6-28) using the marginal cost figures in Table 6-16 and their respective outputs. Fixed costs are calculated in total and then allocated according to the $\Psi$s. The results are in Table 6-17.

**Table 6-17. Fixed, variable and total *transaction* costs (£ 000s)**

| Output | Unit | Variable cost | | Fixed cost | | Total cost | |
|---|---|---|---|---|---|---|---|
| Residential | Resident-weeks per year | $(\partial Y^T/\partial q^h)q^h + (\partial Y^T/\partial q^m)q^m$ | 1303 | $F^T\Psi^r$ | 658 | $Y^{Tr}$ | 1961 |
| Nursing | Resident-weeks per year | $(\partial Y^T/\partial q^n)q^n = (\partial Y^T/\partial q^m)q^n$ | 939 | $F^T\Psi^n$ | 474 | $Y^{Tn}$ | 1413 |
| Non-residential | Clients served per year | $(\partial Y^T/\partial q^d)q^d + (\partial Y^T/\partial q^{g1})q^{g1} + (\partial Y^T/\partial q^{g2})q^{g2}$ | 1322 | $F^T\Psi^D$ | 667 | $Y^{TD}$ | 1990 |
| All | | $V^T$ | 3564 | $F^T$ | 1799 | $Y^T$ | 5363 |

The estimates of fixed costs allow us to also calculate total and variable transaction costs for residential care where, first, *all residential care provision is by hierarchical (LA) means* and, second, where *all residential care provision is by market (independent) means*. Fixed transactions costs for residential care remain unchanged (£658K) in these scenarios (Table 6-18).

**Table 6-18. Total and average transaction costs of residential care – all provision by one governance arrangement**

| Output | Total cost (£ 000s) | | Average cost | |
|---|---|---|---|---|
| Hierarchical (LA residential) | $(\partial Y^T/\partial q^h)(q^h + q^m) + F^T\Psi^r$ | 940 | $\dfrac{(\partial Y^T/\partial q^h)(q^h + q^m) + F^T\Psi^r}{q^h + q^m}$ | 21 |
| Market (Independent residential) | $(\partial Y^T/\partial q^m)(q^h + q^m) + F^T\Psi^r$ | 2516 | $\dfrac{(\partial Y^T/\partial q^m)(q^h + q^m) + F^T\Psi^r}{q^h + q^m}$ | 56 |
| Difference | | 1576 | | 35 |

These calculations give some indication of average transaction costs for hierarchical and market means of governance. Average costs in the table are sensitive to estimates of fixed costs. However, the *difference* between hierarchical and market average costs remains: $(\partial Y^Y/\partial q^m)-(\partial Y^Y/\partial q^h)$, which is our estimate at the sample mean. The sensitivity of the results to changes in key factors is demonstrated in Table 6-19.

**Table 6-19. Average transactions costs – sensitivity analysis**

| *Percentage change in listed factor* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Residential care fixed cost proportion, Ψr | 1 | 1 | 1 | 1 | 0.9 | 0.66 | 1.1 | 1.33 |
| LA marginal cost | 0.9 | 0.66 | 1 | 1 | 1 | 1 | 1 | 1 |
| IR marginal cost | 1 | 1 | 0.9 | 0.66 | 1 | 1 | 1 | 1 |
| *Predicted average costs* | | | | | | | | |
| LA | 20 | 19 | 22 | 24 | 19 | 16 | 22 | 26 |
| IR | 56 | 56 | 53 | 45 | 55 | 51 | 57 | 61 |
| Difference | 36 | 37 | 31 | 21 | 35 | 35 | 35 | 35 |

### 6.5.2 Production costs

Table 6-20 reports the production costs ratio model, which assesses the impact of the output ratio variable on total production costs at the local authority level. Again, a GLS and GEE estimator were used, and the models satisfied the relevant diagnostic tests.

Table 6-21 lists some marginal effects associated with LA-level production costs. Output ratio is significant and positive, implying higher production costs under hierarchical arrangements. Marginal production costs can be derived from the results as:

$$\text{(6-29)} \quad \partial Y^C/\partial q^h = \left[\beta_R\left(\frac{q^h}{q^T}\right)^{\lambda-1}\left(\frac{q^m}{q^{T2}}\right)+\frac{\beta_T}{q^T}\right]Y^C$$

and

$$\text{(6-30)} \quad \partial Y^C/\partial q^m = \left[\beta_R\left(\frac{q^h}{q^T}\right)^{\lambda-1}\left(\frac{-q^h}{q^{T2}}\right)+\frac{\beta_T}{q^T}\right]Y^C$$

which work out to be respectively for hierarchy and market: £88.60 and £55.00 (per resident per week).

**Table 6-20. Production costs models – Linear, panel and Generalised Estimating Equations**

| Variable | GLS Random effects model Coefficient | Std error | | GEE model Coefficient | Std error | |
|---|---|---|---|---|---|---|
| Output ratio (BC) | 4.881E-02 | 1.904E-02 | * | 3.852E-02 | 1.775E-02 | * |
| Total output (log) | 2.538E-01 | 6.804E-02 | * | 2.378E-01 | 5.976E-02 | * |
| Wage (log) | -2.342E+00 | 1.241E+00 | * | -2.559E+00 | 1.167E+00 | * |
| Wage | 5.140E-02 | 2.499E-02 | * | 5.497E-02 | 2.351E-02 | * |
| Gender | -2.487E+00 | 6.579E-01 | * | -2.503E+00 | 6.174E-01 | * |
| Pop 65+ | 6.644E-01 | 9.211E-02 | * | 7.063E-01 | 8.347E-02 | * |
| Pop ratio | 7.394E-01 | 2.531E-01 | * | 7.249E-01 | 2.366E-01 | * |
| SMR | 9.101E-03 | 2.790E-03 | * | 8.462E-03 | 2.619E-03 | * |
| Statement (log) | -2.151E-01 | 9.688E-02 | * | -1.902E-01 | 8.431E-02 | * |
| Statement | 5.660E-03 | 2.165E-03 | * | 5.066E-03 | 1.915E-03 | * |
| Single room (%) | -1.875E-04 | 1.788E-03 | | -4.355E-04 | 1.594E-03 | |
| First period | -6.373E-02 | 1.913E-02 | * | -5.749E-02 | 1.614E-02 | * |
| Met | 2.882E-02 | 7.091E-02 | | 1.391E-02 | 6.672E-02 | |
| Shire | 3.963E-02 | 9.869E-02 | | 3.501E-03 | 9.275E-02 | |
| InnLon | 4.531E-02 | 1.969E-01 | | 4.387E-02 | 1.854E-01 | |
| OutLon | -1.284E-01 | 1.225E-01 | | -1.318E-01 | 1.153E-01 | |
| Constant | 7.598E+00 | 3.762E+00 | * | 8.059E+00 | 3.537E+00 | * |
| Model | Linear (GLS), random effects | | | GEE population-averaged | | |
| Link/dependent variable | Log (production cost) | | | log | | |
| Family | | | | gamma | | |
| Correlation | | | | exchangeable | | |
| Number of obs | 230 | | | 230 | | |
| Number of groups | 117 | | | 117 | | |
| Obs per group: min | 1 | | | 1 | | |
| avg | 2.0 | | | 2.0 | | |
| max | 2 | | | 2 | | |
| Wald chi2(9) | 924.76 | | | 1047.14 | | |
| Prob > chi2 | <0.0000 | | | <0.0000 | | |
| Scale parameter | | | | .0496747 | | |
| R-sq: within | 0.1295 | | | | | |
| between | 0.8975 | | | | | |
| overall | 0.8835 | | | | | |
| rho | .67837275 | | | | | |
| Hausman specification test | 7.07 | | | | | |
| Breusch and Pagan test | 39.43 | | | | | |

Note * denotes significant at 10 per cent or better

Assuming constant marginal costs, fixed costs can be calculated as:

$$(6\text{-}31)\quad F^C = Y^C - \left(\partial Y^C / \partial q^h\right) \cdot q^h - \left(\partial Y^C / \partial q^m\right) \cdot q^m$$

With this assumption, we can calculate (a point estimate of) average production costs given, first, that all provision is hierarchical as:

(6-32) ${}^{h}Y^{C} = \left(F^{C}/q^{T}\right) + \left(\partial Y^{C}/\partial q^{h}\right) = 280.5$

and, second, that all provision is under market arrangements:

(6-33) ${}^{m}Y^{C} = \left(F^{C}/q^{T}\right) + \left(\partial Y^{C}/\partial q^{m}\right) = 254.0$

The difference between these point estimates is £26.5.

**Table 6-21. Production costs models – Derived marginal effects**

| | GLS, random effects | | | GEE, random effects | | | |
|---|---|---|---|---|---|---|---|
| Variable | Coefficient | Std error | | Coefficient | Std error | | Elasticity |
| Output ratio (constant output) | 2.775E-06 | 1.082E-06 | * | 2.190E-06 | 1.009E-06 | * | 0.035 |
| Output (constant output ratio) | 5.645E-06 | 1.513E-06 | * | 5.289E-06 | 1.329E-06 | * | 0.184 |
| Wage | 5.980E-03 | 4.015E-03 | | 5.349E-03 | 3.778E-03 | | 0.276 |
| Statement | 2.681E-03 | 1.021E-03 | * | 2.431E-03 | 9.383E-04 | * | 0.175 |

Note * denotes significant at 10 per cent or better

### 6.5.3 Total cost

Table 6-22 lists the variables in the ratio model and both the GEE and GLS estimates. Both models were highly significant overall. The GEE model was estimated using a gamma error specification; the Park test results are listed in Table 6-12. The table lists the estimate of the specification integer κ for the different assumed error functions. Except for the Gaussian and Poisson form, the nearest integer is κ = 2, which is the gamma specification. The GLS model satisfied the Hausman and the Breusch and Pagan LM test to support the choice of random effects. The model also produced a high measure of fit.

In addition to the categories of variables in transaction cost (ratio) model, the total costs model includes LA type dummies. These dummies were found to have some significant effect capturing some LA-level fixed effect, above and beyond the LA level control factors (e.g. population and wage). Nonetheless, as noted the GLS model satisfied the random effects tests. To account for inter-temporal fixed effects, a period dummy variable was also included.

**Table 6-22. Total costs models – Linear, panel and Generalised Estimating Equations**

| Variable | GLS Random effects model Coefficient | Std error | | GEE model Coefficient | Std error | |
|---|---|---|---|---|---|---|
| Output ratio (BC) | 3.683E-01 | 2.064E-01 | * | 3.279E-01 | 1.924E-01 | * |
| Output ratio (BC) * wage (log) | -9.592E-02 | 5.406E-02 | * | -8.541E-02 | 5.032E-02 | * |
| Total output (log) | 9.844E-02 | 3.535E-02 | * | 9.434E-02 | 3.100E-02 | * |
| Domiciliary output (log) | 1.941E-01 | 6.997E-02 | * | 1.963E-01 | 6.796E-02 | * |
| Domiciliary output | -1.480E-08 | 9.510E-09 | | -1.490E-08 | 9.240E-09 | |
| Wage (log) | -1.262E+00 | 7.163E-01 | * | -1.355E+00 | 6.931E-01 | * |
| Wage | 2.871E-02 | 1.432E-02 | * | 3.052E-02 | 1.388E-02 | * |
| Pop 65+ | 9.358E-01 | 6.142E-02 | * | 9.397E-01 | 5.788E-02 | * |
| Pop ratio | 3.076E-01 | 1.465E-01 | * | 3.319E-01 | 1.406E-01 | * |
| Gender | -2.175E+00 | 4.240E-01 | * | -2.237E+00 | 4.079E-01 | * |
| SMR | 7.261E-03 | 1.857E-03 | * | 7.167E-03 | 1.787E-03 | * |
| Statement (log) | -9.218E-02 | 4.961E-02 | * | -8.863E-02 | 4.317E-02 | * |
| Statement | 2.436E-03 | 1.129E-03 | * | 2.383E-03 | 1.001E-03 | * |
| Single room (%) | -1.583E-02 | 2.279E-02 | | -1.488E-02 | 2.095E-02 | |
| First period | -5.936E-02 | 9.344E-03 | * | -5.868E-02 | 7.875E-03 | * |
| Met | 1.251E-01 | 4.096E-02 | * | 1.266E-01 | 3.974E-02 | * |
| Shire | -3.524E-02 | 5.636E-02 | | -3.367E-02 | 5.454E-02 | |
| InnLon | 3.127E-01 | 1.178E-01 | * | 3.141E-01 | 1.141E-01 | * |
| OutLon | 1.815E-02 | 7.116E-02 | | 1.629E-02 | 6.912E-02 | |
| Constant | 3.420E+00 | 2.225E+00 | | 3.728E+00 | 2.157E+00 | * |
| Model | Linear (GLS), random effects | | | GEE population-averaged | | |
| Link/dependent variable | Log (production cost) | | | log | | |
| Family | | | | gamma | | |
| Correlation | | | | exchangeable | | |
| Number of obs | 230 | | | 230 | | |
| Number of groups | 117 | | | 117 | | |
| Obs per group: min | 1 | | | 1 | | |
| avg | 2.0 | | | 2.0 | | |
| max | 2 | | | 2 | | |
| Wald chi2(9) | 2633.99 | | | 2798.08 | | |
| Prob > chi2 | <0.0000 | | | <0.0000 | | |
| Scale parameter | | | | .0163688 | | |
| R-sq: within | 0.3308 | | | | | |
| between | 0.9623 | | | | | |
| overall | 0.958 | | | | | |
| rho | 0.755 | | | | | |
| Hausman specification test | 9.34 | | | | | |
| Breusch and Pagan test | 51.96 | | | | | |

Note * denotes significant at 10 per cent or better

Table 6-23 lists various marginal effects on total costs. Output ratio is negative but not statistically different from zero; in other words, around the sample means, a change in governance arrangement (as conceived as a change in the ratio of provision organised under hierarchical to all provision) has no significant effect on total costs. The model was estimated without area and time dummies, and while the resultant statistical performance was less good, the output ratio variable remained insignificant. In addition, the output ratio variable was also tried without interaction effects (with inputs costs) and was also insignificant in that specification.

The size of output ratio elasticity is about 8 per cent of the output ratio elasticity with regard to transaction costs (see Table 6-25). This finding is therefore consistent with the above results, that transaction costs are higher and production costs lower for markets compared with hierarchies.

**Table 6-23. Total costs models – Derived marginal effects**

| | GLS, random effects | | | GEE, random effects | | | |
|---|---|---|---|---|---|---|---|
| Variable | Coefficient | Std error | | Coefficient | Std error | | Elasticity |
| Output ratio (constant output) | -3.685E-07 | 6.827E-07 | | -3.340E-07 | 6.167E-07 | | -0.005 |
| Output (constant output ratio) | 2.189E-06 | 7.862E-07 | * | 2.098E-06 | 6.895E-07 | * | 0.094 |
| Wage | 4.226E-03 | 2.348E-03 | * | 4.248E-03 | 2.381E-03 | * | 0.219 |
| Domiciliary output | 1.300E-05 | 3.830E-06 | * | 3.644E-05 | 1.245E-05 | * | 0.139 |
| Statement | 1.159E-03 | 5.758E-04 | * | 1.155E-03 | 4.709E-04 | * | 0.083 |

Note * denotes significant at 10 per cent or better

## 6.6 Conclusion

The last chapter developed theoretical hypothesis concerning, primarily, the relative size of transaction and production costs of market and hierarchical provision. This chapter seeks to determine whether these hypotheses are supported by the data. The empirical analysis began with the definition and calculation of total transaction cost. The available financial data report both total expenditure to the local authority and the amount consumed specifically in provision/production of all services for older people (where 'production costs' include the prices paid to independent sector providers in markets). We take a broad definition of transaction cost such that total costs are a sum of transaction and production costs. Total purchaser transaction costs are then found by subtraction. The aim, however, is to explore the size of transaction costs by governance arrangement. As outlined in chapter 4, an 'ownership' definition of governance arrangements is sufficient for our purposes. In other words, services provided by 'in-house' providers can be regarded as being deployed under hierarchical arrangements.

Those supplied by independent sector providers must be under market arrangements as implied by the conceptual definition of governance in chapter 3. The only case where this relationship might not hold is if authorities adopt full *internal* markets. The commissioning survey explored this issue (see chapter 4) and found that whilst there are some differences as to how in-house provision is organised in the sample, councils with significant in-house provision utilised hierarchical arrangements. Generally, where the authority was a substantial user of market arrangements (i.e. a large proportion of service outsourced), their in-house service showed indications of being more internal-market-like. If anything, this supports the specification of the governance indicator as used.

In practice, local authorities use a mix of in-house and independent sector providers, and therefore a mix of governance arrangements. The local rationale for this mix varies, as does the mix itself – quite substantially across the country – in terms of the numbers and proportion of all supported placements that go to either in-house or independent sector providers. Exploiting this variation, we can model the relationship between supported placement mix and total purchaser transaction costs to determine (marginal) transaction cost per place by type of placement, and therefore, by governance arrangement. According to the theoretical hypotheses, it should follow that where authorities have a high proportion of independent sector providers, their average and marginal transaction costs will be higher than authorities with a low proportion.

The results were highly consistent between the models and they provided support for the theoretical hypothesis. As regards the ratio models, the ratio variable was significant with the expected sign. The translog models produced marginal transaction costs (point) estimates of £6 per place per week for hierarchical placements; estimated confidence intervals do contain zero. For markets, the point estimate was £41 per place for market places and this was significantly different from zero. The results also suggest that fixed transaction costs were sizeable. Further analysis estimated average transaction costs at £21 and £56 per place per week respectively.

For production costs, a significant difference was found in the other direction; marginal costs of £89 for hierarchy and £55 for markets. Overall, the total (production + transaction) costs model could not find a significant difference between market and hierarchy, which is consistent with the findings for transaction and production costs individually.

Analysis elsewhere at home level has found greater differences in production costs between in-house and market providers (e.g. Netten et al., 1998). Nonetheless, the implications of the current analysis are clear. Much of the pro-market rhetoric that underpinned social care policy in the late 1980s and 1990s, as discussed in chapter 4, focused on production unit cost differences between markets and hierarchies. And these comparisons do show the benefits of markets. A more comprehensive comparison should also consider the differences in transaction costs, which the above analysis suggests, work in the other direction.

Since local authorities decide governance arrangements, we have focused on the costs to purchasers. Nonetheless, we can also take a societal perspective, adding provider transaction costs into the equation. According to the theory of the last chapter, we would expect this addition to reinforce the above conclusions. After subtracting planning costs incurred in both market and hierarchies, the theory suggests that only providers in markets will bear transaction costs. Moreover, because at least some of these provider transaction costs will have been passed on to purchasers as costs of production, it means that adjusted market production costs will be even, underlining hypothesis H2. Of course, this re-allocation of costs will not change the total costs (production and transaction) incurred in markets compared to hierarchies.

Indeed, the comparison of total costs suggests that whilst markets may still have an advantage in total cost terms, the gap is far smaller than that implied by looking at production costs alone. Small enough, perhaps, to justify the continued use of hierarchical arrangements to some degree, if other non-cost considerations are also brought to bear.

## Annex 6-1

**Table 6-24. Data Sources**

| | Variable | Source |
|---|---|---|
| Output | • *Local authority supported residents*, 31 March 1998 and 1999<br>• *Independent sector supported residents*, 31 March 1998 and 1999<br>• Total supported residents, 31 March 1998 and 1999 | *Community Care Statistics*, Bulletin 2000/2, Table S3<br>*Community Care Statistics*, Bulletin 1999, Table S3<br>(see http://www.doh.gov.uk/ public/stats3.htm |
| | • *Local authority resident weeks*, 1997/8 and 1998/9<br>• *Independent sector resident weeks*, 1997/8 and 1998/9<br>• Total resident weeks, 1997/8 and 1998/9 | Department of Health, *Social Services Performance*, Performance Assessment Framework/Key Indicator statistics, KS1, 1998/9 (11 May 2000) and 1997/8.<br>(see http://www.doh.gov.uk/ paf/index.htm |
| | Domiciliary output<br>*Numbers of people helped to leave at home* 1998/9 and 1997/8 | PAF key indicators (see above) |
| Expenditure | Local Authority Personal Social Services Gross Expenditure, 1998-99:<br>• Central strategic (SSR) expenditure<br>• Older people/EMI expenditure:<br>➢ Assessment & Commissioning<br>➢ Residential costs – LA homes<br>➢ Residential costs – Indep. homes<br>➢ Non-residential | Social Services Performance in 1998-99, Bulletin 2000/10 (11 May 2000): R03 statistics, Department of Health and DETR.<br>Local authority breakdowns for 1998/9 and 1997/8<br>(see http://www.doh.gov.uk/ public/pss_stat.htm) |
| Input costs | Capital<br>• Property prices, 1998 | Regional Trends 34, 1999, Office of National Statistics (adapted from StatBase, dataset RT34610, source HM Land Registry). |
| | Labour<br>• WTE social services staff by local authority, 1999<br>• WTE management staff | Local authorities staffing statistics, Statistical bulleting, SSDS001.<br>(see http://www.doh.gov.uk/ public/psstaff.htm) |
| Dependency | Population statistics by local authority, 1997<br>• Population 65+,<br>• 65-75<br>• total population<br>• SMR<br>• Population by gender | Regional Trends 34, 1999, Office of National Statistics (adapted from StatBase, dataset RT341401, Source ONS). |
| | Hospital admissions, 1997/8 | PAF key indicators (see above) |
| Process factors | Statement of needs provision, 1997/8 & 1998/99 | PAF key indicators (see above) |
| | Single room | PAF key indicators (see above) |

## Annex 6-2

Differentiating marginal costs gives:

$$\begin{aligned}(6\text{-}34)\qquad \partial^2 Y^T/\partial q^h \partial q^h &= \partial\left[\left(\partial Y^T/\partial \log Y^T\right)\left(\partial \log Y^T/\partial q^h\right)\right]/\partial q^h \\ &= \left(\partial Y^T/\partial \log Y^T\right)\partial^2 \log Y^T/\partial q^h \partial q^h + \left(\partial Y^T/\partial \log Y^T\right)\left(\partial \log Y^T/\partial q^h\right)^2 \\ &= \left[\partial^2 \log Y^T/\partial q^h \partial q^h + \left(\partial \log Y^T/\partial q^h\right)^2\right]Y^T\end{aligned}$$

or

$$(6\text{-}35)\quad \partial^2 Y^T/\partial q^h \partial q^h = \left[2\beta_{hh}\left(\left(q^h\right)^{\lambda-1}\right)^2 + \vartheta^h(\lambda-1)\left(q^h\right)^{\lambda-2} + \left(\left(q^h\right)^{\lambda-1}\vartheta^h\right)^2\right]Y^T$$

Similarly for independent sector marginal costs:

$$(6\text{-}36)\quad \partial^2 Y^T/\partial q^m \partial q^m = \left[2\beta_{mm}\left(\left(q^m\right)^{\lambda-1}\right)^2 + \vartheta^m(\lambda-1)\left(q^m\right)^{\lambda-2} + \left(\left(q^m\right)^{\lambda-1}\vartheta^m\right)^2\right]Y^T$$

Second order interactive effects on marginal costs are:

$$(6\text{-}37)\quad \partial^2 Y^T/\partial q^h \partial z = \left[\left(\partial\vartheta^h/\partial z\right)\left(q^h\right)^{\lambda-1} + \left(q^h\right)^{\lambda-1}\vartheta^h\left(\partial \log\left(Y^T\right)/\partial z\right)\right]Y^T$$

and equivalent for market provision. Using the estimated parameters gives second order cost effects as reported in Table 6-25. Both the own-second order effects were negative, although both are very small in size, largely supporting out assumption of constant marginal transaction costs. Hierarchical and market marginal costs are negatively related to input costs $w$.

**Table 6-25. Changes in marginal transaction costs**

| Second order effect | Value |
|---|---|
| $\partial^2 Y^T/\partial q^h \partial q^h$ | -5.04E-04 |
| $\partial^2 Y^T/\partial q^m \partial q^m$ | -3.05E-03 |
| $\partial^2 Y^T/\partial q^h \partial w$ | -9.46E-01 |
| $\partial^2 Y^T/\partial q^m \partial w$ | -2.89E-01 |
| $\partial^2 Y^T/\partial q^h \partial q^m$ | 4.88E-02 |
| $\partial^2 Y^T/\partial q^m \partial q^h$ | 3.29E-02 |

# Chapter 7. Contracts and profits

## 7.1 Introduction

This chapter considers the effects of reimbursement structure on providers operating under market governance. It considers how the choice of contracts between commissioners and providers operating in market-like governance arrangements can have implications for provider behaviour and so for outcomes.

As described in chapter 3, choices regarding contracting are multi-dimensional, covering incentive 'power', timing, contingency and so forth (see section 3.3.3.1). The empirical relevance of these choices in social care varies however (see chapter 4). Moreover, these choices tend to be correlated with other non-contracting dimensions of governance, especially with ownership. For example, although both high and low powered contracts are used, high-powered incentives are almost exclusively used in markets (i.e. dispersed ownership) and low powered contracts in hierarchies (unified ownership). Contract contingency varies between, but also within governance archetypes. In order to isolate ownership from contract contingency effects we concentrate on this choice within market governance arrangements.

A change in the degree of contingency of the contracting system will impact on provider behaviour as considered here in three ways, via stakeholders' potential exploitation of information advantages, their response to risk and their response to cost shocks. The investigation of information problems relating to contracting choices is addressed in the health economics literature. This work has tended to focus on cream-skimming and 'upcoding' behaviours in hospital markets, especially Medicare reimbursement of hospitals under prospective payments systems, but there are also important examples relating to England (e.g. Propper, 1995). There is evidence of opportunistic use of information (see Culyer and Newhouse, 2000). Analyses in social care – as discussed in chapter 2 – find evidence of similar information opportunism. In particular, Forder (1997) finds evidence of upcoding type behaviour in residential care markets.

In Chapter 5, section 5.8.1.3, the core proposition was elaborated; that given sufficient price contingency on user cost parameters (e.g. β in the model), we would expect providers to exaggerate costs and so push up prices, other things equal. If contracts are not contingent in this way, cream-skimming will occur instead and prices will remain unchanged. Furthermore, if cost exaggeration does happen, utility and so profit under contingent contracts will be higher than under non-contingent contracts.

To sum up, the hypothesis regarding this information effect is that prices and profitability will not be less under contingent contracts compared with non-contingent contracts. The current chapter refines this result, and adds the potential effects of risk and unaccounted-for cost increases (cost shocks). We would expect risk averse providers to require a risk premium to cover the addition risk exposure that occurs under non-contingent contracts (i.e. that do not account for cost related contingencies). This risk effect on price and profitability potentially offsets the information effect. However, below we argue that there is very little evidence of risk premiums being paid in social care. Cost drivers have generally been pushing costs up in residential care in recent years (see chapters 3 and 4), but there has been little account made in terms of compensatory prices rises (Forder and Netten, 2000a). Therefore, cost shocks effects are aligned with the information effects. Overall, in this chapter we develop a main hypothesis that prices and profitability will be higher for residential care providers under contingent contracts compared with non-contingent contracts, other things equal.

This chapter provides an in-depth look at the types of contract used to conduct social care transactions. After this introduction the chapter has five substantive sections. The next section (7.2) develops a formal specification of contracts. It defines the concept of contract contingency. The section also lays out general objective functions for providers that allow for broader provider motivation (see also chapter 3). Section three (7.3) considers how providers can exploit information advantages under different types of contract. Section four (7.4) looks at the risk properties of these contracts. Section five (7.5) analyses how cost shocks affect providers under different contracts. These three effects are key elements in understanding how providers behave in terms of pricing, costs and profitability. Together they are the basis for hypothesis about provider behaviour that is tested in chapter 8. Section six (7.6) specifies hypothesis for empirical testing. Policy implications are considered in chapter 9.

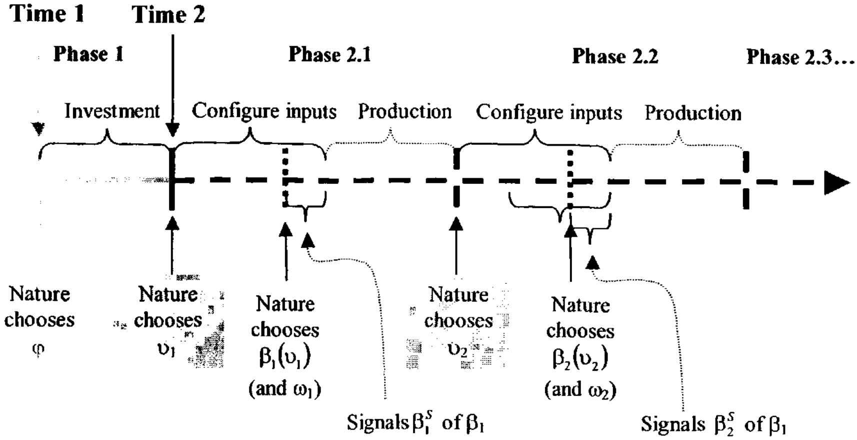
## 7.2 Contract choice

To show the implications of contract contingency choices we use the basic model outlined in chapter 1. Provider utility is given by (1.4). In previous chapters the focus has been on the distinction between the investment stage, period 1, and the effort stage, period 2 (see figure 1.1). Investment decisions are not relevant to information issues discussed in the present chapter and therefore we consider only period 2 activities.

The focus on market governance makes it relevant to explicitly allow more than one provider. This focus also prompts a structure where period 2 has multiple production stages. In particular, in period 2 we assume repeated transactions denoted $t = 1, \ldots, T$, an assumption that allows us to consider a more comprehensive range of information-related behaviour such as reputation effects. The actual number of sub-periods $T$ may be uncertain to stakeholders. With reference to chapter 1, uncertainty about the state-of-the-world $k$ is characterised by a number of parameters (see section 1.4 of chapter 1), however in this chapter the focus is on $\beta$. The parameter $\varphi$ has a primary bearing on investment, which is taken as given in this chapter. Potentially $\upsilon$ is relevant, but because we assume that purchasers always undertake assessments and providers always measure $\beta$, then both parties have the same full information embodied in $\upsilon$. Purchasers as social services departments have a statutory duty in law to undertake user assessments. And because $\beta$ is largely about a user's experience of care, the cost of its measurement for providers is low enough that it is always worth the measurement cost. Finally, the $\omega$ parameter has asymmetric measurement costs and is determined in period 2. However, its misrepresentation in markets is of little consequence because it affects provider costs – which are also unknown to purchasers. In any case purchasers base their decisions on prices not costs and so misreporting of $\omega$ does not change purchaser behaviour – see section 5.3.2.2 of chapter 5 for details.

As a consequence Figure 1.1 can be adapted for our purposes here. Figure 7-1 highlights the events in phase 2, the effort phase, and the relevance of the signal of $\beta$.

**Figure 7-1. Information structure – market governance**




### 7.2.1 Objective functions

Utility is given by (1.4) in chapter 1. For each state *k*, provider *i*'s utility is (suppressing the *i* and *k* subscripts):

$$U = \gamma[v(\pi - y) + \breve{\psi}(\rho)] = \gamma[v(\pi - y) + r\psi(\rho)] \tag{7-1}$$

with, profit:

$$\begin{aligned} \pi &= r(\theta)D(\hat{\beta})(p(\hat{\beta}) - c(\beta)) + \frac{r(\theta)r(\theta_1)}{1+b}D(\hat{\beta}_1)(p(\hat{\beta}_1) - c(\beta_1)) \\ &+ .. + \frac{r(\theta)\prod_{t=1}^{T} r(\theta_t)}{(1+b)^t} D(\hat{\beta}_T)(p(\hat{\beta}_T) - c(\beta_T)) + ... \end{aligned} \tag{7-2}$$

We assume that for the future (time $t > 1$) providers operate with an initial expectation of $\beta$ at its mean level: $\breve{\beta}$; there is extensive uncertainty about referrals and other pertinent factors and providers are boundedly rational (see chapter 2).

Thus (7-2) can be written:

$$\pi = r(\theta)D(\hat{\beta})(p(\hat{\beta}) - c(\beta)) + r(\theta)\sum_{t=1}^{\infty} \frac{\breve{r}^t}{(1+b)^t} \breve{D}(\breve{p} - \breve{c})$$

$$(7\text{-}3) \quad = r(\theta)\left[D(\hat{\beta})(p(\hat{\beta}) - c(\beta)) + \frac{\breve{D}(\breve{p} - \breve{c})}{1 - \breve{r} + b}\right]$$

$$= r(\theta)\left[D(\hat{\beta})(p(\hat{\beta}) - c(\beta)) + \breve{\pi}\right]$$

where the embellishment $\breve{e}$ denotes expected (average) values of the corresponding variable $e$. In order to have the infinite sum converge as indicated we assume that $1 - \breve{r} + b < 1$. Hence the interest rate $b$ must be small. Uncertainty, especially far into the future, and the independence of transactions via new referrals means that providers do not expect current *realised* actions to directly bear on future profits. They may form expectations that the discounted stream of future profits may embody information rents, but that the size of current information rents is not a good indicator; some prior expectation is used instead. Continuity into the future is directly relevant, however, in that the loss of a contract now also implies a loss of future profit streams. This modelling approach therefore mirrors the reputation models outlined in chapter 2.

Furthermore, in (7-1) the present value of service price, $\rho$, is written:

$$(7\text{-}4) \quad \rho = p + \frac{p_1}{1+b} + \frac{p_2}{(1+b)^2} + \ldots = p + \frac{\breve{p}}{b}$$

where $\breve{p}$ is the expected future price. Providers suffer disutility from having actual prices at 'excessive' levels. However, if the contract does not proceed this disutility is generally not incurred – provider reservation utility is zero.[50] Hence we have: $\breve{\psi}(\rho) = r\psi(\rho)$. The utility function for each provider is:

$$(7\text{-}5) \qquad U = \gamma v\left(r(\theta)D(\hat{\beta})\left[p(\hat{\beta}) - c(\beta)\right] + r(\theta)\breve{\pi} - y\right) + \gamma r(\theta)\psi(\rho)$$

[50] Although loss of contract potentially implies a loss of access for users, providers in that situation would expect some alternative arrangement or provider to instead provide the service and hence restore access for users.

### 7.2.2 Contract contingency

We need a robust definition of contingency. The intensity and range of components of a care package – and so the package's cost – is positively related to a person's care needs.[51] The service-cost-characteristic, $\beta$, is a good summary statistic of cost. Let $\beta \in [\bar{\beta}, \underline{\beta}]$ be the actual cost-parameter of clients served by the provider. As outlined in previous chapters, $\hat{\beta} \in [\bar{\hat{\beta}}, \underline{\beta}]$ is the reported cost-parameter, where $\beta \in [\bar{\beta}, \underline{\beta}] \subset [\bar{\hat{\beta}}, \underline{\beta}]$ and $\theta = \hat{\beta} - \beta$. Furthermore, $\beta^{I} \in [\bar{\beta}, \underline{\beta}]$ is the cost-parameter of the client *initially* referred to the provider – which the provider can, with varying ease, reject – and $\tilde{\beta}$ is the purchaser's estimate of the actual cost-parameter $\beta$. Finally, $\breve{\beta}$ is the sample average value of $\beta^{I}$. This notation is summarised in Table 7-1.

**Table 7-1. Cost parameters - notation**

| Term | Defintion |
|---|---|
| $\beta$ | *Actual* cost-parameter |
| $\hat{\beta}$ | *Reported* cost-parameter |
| $\beta^{I}$ | Cost-parameter of *initially referred* client |
| $\breve{\beta}$ | Sample average value of $\beta^{I}$ |
| $\tilde{\beta}$ | Purchaser's estimate of the actual cost-parameter $\beta$ |

#### 7.2.2.1 Contingent contracts

For each provider in each state, demand under a *contingent* contract is $D^{C} = D^{C}(\hat{\beta})$, being twice differentiable with $D^{C}_{\hat{\beta}} \geq 0$ and $D^{C}_{\hat{\beta}\hat{\beta}} = 0$. Also:

*Assumption 7-1.* $D(\underline{\beta}) \geq D_{\hat{\beta}}(\beta)(\bar{\beta} - \underline{\beta})$.

This assumption means that the purchaser's willingness to pay is high even for the least dependent clients relative to changes in demand. We are specifically interested here in

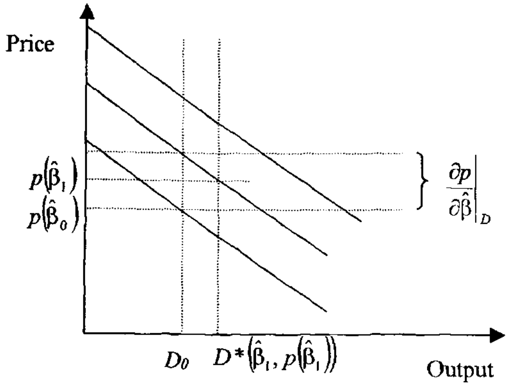
[51] It is worth noting that whilst it is generally true that people with higher levels of dependency will need more support to manage activities of daily living etc. than people with lower dependency, the relationship between cost and dependency in practice whilst valid is less straightforward that we approximate here.

price contingency with respect to $\beta$ and not just with $\upsilon$, although the two parameters are highly correlated. Prices that are contingent on costs related to $\beta$ allow some scope for providers' reports of $\beta$ to differ from its actual value. Contingency strictly limited to $\upsilon$ would still allow some scope, but for our purposes here, $\beta$ is synonymous with cost changes due to case-mix (as opposed to say, the price of labour).

The size of the relationship between demand and reported client characteristics defines contingent contracts. The exact formulation used in practice by the purchaser could vary enormously and is arbitrary. Nonetheless, some purchasers will have prices changing by a larger amount in response to a unit change in the cost-parameter than others. To make this 'degree' of contingency – defined as the size of price changed in response to a change in the reported cost-parameter – more meaningful we can measure it relative to a change in costs. In particular, it is useful to define the degree of contingency, $p_{\hat{\beta}}$, relative to how the same change in the actual cost-parameter would affect *marginal cost* $c_{\beta}$. This corresponds to the underlying aim of using contingent contracts, which is to compensate providers for higher costs associated with higher cost clients. There is no implication that purchasers use a reimbursement formula that is explicit in $c_{\beta}$, even if they could measure marginal costs. This comparison is just to provide a baseline from which to gauge the degree of contingency.

Price change in response to a change in reported dependency results from an increase in demand/willingness and ability to pay by the purchaser at each output. In turn, such a demand change may elicit a change in output by an optimising provider that depends on the slope of the demand function – see Figure 7-2.

**Figure 7-2. Price contingency**




Differentiating the profit function (7-3) with respect to price for any given state $k$, we have the first order condition:

$$(7\text{-}6)\quad \Omega \equiv p - c + \frac{D}{D_p} = 0$$

and so,

$$(7\text{-}7)\quad D_{\hat{\beta}} = -\frac{\Omega_{\hat{\beta}}}{\Omega_D} = -\frac{\left.\frac{\partial p}{\partial \hat{\beta}}\right|_D}{\frac{1}{D_p} + \frac{1}{D_p}} = -\frac{\left.\frac{\partial p}{\partial \hat{\beta}}\right|_D D_p}{2} = -\tfrac{1}{2} p_{\hat{\beta}}\big|_D D_p$$

We assume that the term $D_p$ is not a function of $\hat{\beta}$ in a symmetric equilibrium (and that $D_{pp} = 0$). Suppose that the reimbursement arrangement is such that:

$$(7\text{-}8)\quad p_{\hat{\beta}} = \frac{1}{h} c_{\beta}$$

Commissioners may have pre-determined what price increases/decreases they are prepared to make for users of higher/lower dependency. Thus, if the provider's report of dependency is accepted (given the information that commissioners also have about the service user) then $p_{\hat{\beta}}$ indicates the price premium they are prepared to pay. We can usefully define contingency 'categories' in terms of ranges of these parameters. For example, a *fully contingent* contract is one where $p_{\hat{\beta}} = c_\beta$ or equivalently, $h = 1$ in the relationship (7-8). A *partially contingent* contract is where $p_{\hat{\beta}} < c_\beta$ or equivalently, $h >$ 1. For a non-contingent contract $p_{\hat{\beta}} = 0$, or equivalently, $h = \infty$. For shorthand we will say that fully contingent contracts i.e. $h = 1$ are called C1 contracts. For contracts where $1 < h < \infty$, i.e. partial contingency, these are referred to as C2 contracts.

In practice, under different contract types ($h$), *reported* dependency changes need not be equivalent to *referred* dependency changes. Indeed, $p_{\beta'} = p_{\hat{\beta}} \dfrac{\partial\hat{\beta}}{\partial\beta'} = p_{\hat{\beta}} m(h)$ where $\dfrac{\partial\hat{\beta}}{\partial\beta'} = m(h)$ is a function of contract type $h$. Similarly: $c_{\beta'} = c_\beta \dfrac{\partial\beta}{\partial\beta'} = c_\beta w(h)$. Taken together, the relationship between actual and reported dependency is:

$$\frac{\partial\hat{\beta}}{\partial\beta} = \frac{\partial\hat{\beta}}{\partial\beta'} \Big/ \frac{\partial\beta}{\partial\beta'} = m/w = M .$$

The significance of this result is that if $M \neq 1$ for derivatives from sample average referred dependency then for a contract of contingency $h$ in (7-8), we have $p_{\beta'} \equiv p_{\hat{\beta}} m(h) = \frac{1}{h} c_\beta m(h) \neq \frac{1}{h} c_\beta w(1) \equiv \frac{1}{h} c_{\beta'}$. For example, in this case, for a fully contingent contract where $h = 1$, we could find that $p_{\beta'} \neq c_{\beta'}$, whilst $p_{\hat{\beta}} = c_\beta$.

#### 7.2.2.2 Non-contingent contracts

For *non-contingent* contracts, $D^N = D^N(\breve{\beta})$, where $\breve{\beta}$ is some average value of $\beta$ such that $\breve{\beta} \in (\underline{\beta}, \overline{\beta}) \equiv \{\breve{\beta} \in R^1 : \underline{\beta} < \breve{\beta} < \overline{\beta}\}$. Also $D^N_{\hat{\beta}} = 0$ and $D^N_{\hat{\beta}\hat{\beta}} = 0$. Therefore, for non-contingent contracts, $p_{\hat{\beta}} = 0$, or in general, $1/h = 0$.

## 7.3 Information

The asymmetry of information inherent in social care markets potentially allows providers to push up their price-cost margins. Two types of information problem are relevant as outlined in chapter 2 and chapter 5 (section 5.3.2.1): *cost exaggeration*, a form of moral hazard, and *cream-skimming*, a form of adverse selection (Rasmusen, 1992). The extent of these problems is sensitive to the *ex post* contingency of the contract.

*Cost exaggeration* involves providers having an opportunity to overstate the intensity and range of components of the individual's care package – and so the package's cost – i.e. $\hat{\beta} > \tilde{\beta} = \beta$. Providers engaging in *cream-skimming* selectively choose residents who have relatively low expected service costs (i.e. below the purchaser's estimate) but claim that these clients have a service-cost-characteristic in the region of the purchaser's estimate i.e. $\hat{\beta} = \tilde{\beta} > \beta$. Although the provider's report does not differ from the purchaser's estimate, it is still greater than the *actual* service-cost-characteristic because high service-cost clients are rejected. If purchasers are paying for care on the basis of service-cost-characteristic $\tilde{\beta}$ then cream-skimming will mean that providers are only providing care at level of $\beta$, which is less than $\tilde{\beta}$.

The incidence of these two information problems is highly sensitive to the (financial) incentives embodied in the transaction between purchaser and provider (Forder, 1997a). Reimbursement arrangements that reflect the purchaser's willingness to pay more for the care of more dependent people – i.e. contingent contracts – will accommodate the moral hazard behaviour described, but provide no incentive for cream-skimming. However, if demand is not expressed over individual clients – i.e. non-contingent contracts – then the exaggeration of clients care needs is not going to affect demand and so provider reimbursement. Such arrangements may promote cream-skimming however, because by doing so providers can lower costs relative to a fixed income.

### 7.3.1 Probability of detection

It is the asymmetry of information that lies at the heart of this issue. We assume that the cost to providers of collecting information $\beta$ is zero and that purchasers have a sufficiently large cost as to make full measurement of $\beta$ inefficient. Purchasers do however get a signal of $\beta$ to assess a probability that providers' reports of $\beta$ are inaccurate e.g. from the user assessment, which yields parameter $\upsilon$.

The probability $r$ that the purchaser will be satisfied with the provider's report is defined as $r = r(\hat{\beta} - \tilde{\beta})$. However, we assume that the purchaser's predicted value $\tilde{\beta}$ equals the actual value $\beta$. So we can write $r = r(\hat{\beta} - \beta) = r(\theta)$ where $r : R \to I$ with $I \equiv \{r : 0 \le r \le 1\}$.

*Assumption 7-2.* We assume that $r_{\hat{\beta}} = r_\theta \dfrac{\partial\theta}{\partial\hat{\beta}} = r_\theta \le 0$. Similarly, $r_{\hat{\beta}\hat{\beta}} < 0$, $r_{\hat{\beta}\hat{\beta}\hat{\beta}} = 0$, $r_{\hat{\beta}}(0) = 0$ and $r(0) = 1$ i.e. when $\hat{\beta} = \beta$. Also $r_\beta = r_\theta \dfrac{\partial\theta}{\partial\beta} = -r_\theta \ge 0$ and $r_{\beta\beta} > 0$, $r_{\beta\beta\beta} = 0$, $r_\beta(0) = 0$.

The chance of detecting small deviations of reported from the actual service-cost-characteristic $(1 - r)$ initially increases slowly from zero, but latter more rapidly. An inverse quadratic relationship fulfils the criteria of assumption 7-2.

### 7.3.2 Cost exaggeration

The model can be used first to demonstrate the basic moral hazard problem (of $\hat{\beta} > \tilde{\beta} = \beta$) discussed above.

*Lemma 1. Misrepresentation of the client's service-cost-characteristic. This lemma has two parts.*

*(a) It is optimal for profit maximising providers to misrepresent by overstating the client's service-cost-characteristic i.e.* $\hat{\beta} > \tilde{\beta} = \beta$ *under a contingent contract and where there is no client selection.*

*(b) This incentive does not exist under a non-contingent (fixed price) contract such that* $\hat{\beta} \le \tilde{\beta} = \beta$.

*Proof*

(a) Since we are considering profit maximisers, differentiating the profit function (7-3) gives:

$$(7\text{-}9)\quad \pi_{\hat\beta} = rDp_{\hat\beta} + rD_{\hat\beta}(p-c) + Dr_\theta(p-c) + r_\theta\breve{\pi}$$

From the profit function (7-3), differentiating by the time 1 price, $p$, gives the first order condition: $z \equiv r\left(p - c + \frac{D}{D_p}\right) = p - c + \frac{D}{D_p} = 0$ and so, $D_{\hat\beta} = -\frac{p_{\hat\beta}\big|_D D_p}{2}$ (see (7-7)). We can now write (7-7) as:

$$(7\text{-}10)\quad \pi_{\hat\beta} = D\left[-\frac{rD_{\hat\beta}}{D_p} - \frac{Dr_\theta}{D_p} + rp_{\hat\beta}\right] + r_\theta\breve{\pi}^C$$

At $\hat\beta = \beta$ we can write this differential as:

$$(7\text{-}11)\quad \pi_{\hat\beta} = D\left[-\frac{D_{\hat\beta}}{D_p} + p_{\hat\beta}\right] = D\left(\tfrac{1}{2}p_{\hat\beta}\big|_D + p_{\hat\beta}\right) = D\left(\tfrac{1}{2}p_{\hat\beta}\big|_D + p_{\hat\beta}\big|_D + \frac{D_{\hat\beta}}{D_p}\right) = Dp_{\hat\beta}\big|_D > 0$$

using assumption 7-2. Because output is adjusted following a shift in the demand curve (to maximise profits), the actual change in price, $p_{\hat\beta}$ is less than $p_{\hat\beta}\big|_D$ i.e. $p_{\hat\beta} = p_{\hat\beta}\big|_D + \frac{\partial p}{\partial D}\frac{\partial D}{\partial\hat\beta} = p_{\hat\beta}\big|_D + \frac{D_{\hat\beta}}{D_p}$ (see Figure 7-2). With the assumed linear demand curves, the price adjustment due to a change in optimal output is equal to half of the original demand shift i.e. $\tfrac{1}{2}p_{\hat\beta}\big|_D$.

It follows that at $\hat\beta = \beta$ we do not have a turning point and so $\hat\beta = \beta$ cannot be a solution. Let us write the solution – the optimal reported value of β under contingent contracts – as $\hat\beta^{*C}$. For values $\hat\beta^{*C} > \beta$ we have $r < 1$ and $r_\theta < 0$, which, in view of $D_{\hat\beta} > 0$ is not a contradiction with (7-10).[52]

[52] We have assumed for exposition that $D_{pp} = 0$. However, if we were to allow for $D_{pp} > 0$ then proposition 1 still holds unless price elasticity of demand is very close zero.

(b) Under a non-contingent contract, $D_{\hat{\beta}} = 0$ and so we have:

$$\pi_{\hat{\beta}} = D\left[-\frac{Dr_{\theta}}{D_p}\right] + r_{\theta}\breve{\pi} < 0$$

for all values of $\theta > 0$. There is no incentive to exaggerate the value of the client service-cost-characteristic. □

Under contingent contracts therefore the reported client-costs parameter will be higher than under non-contingent contracts. In the appendix, Lemma A1 checks the conditions under which $\hat{\beta}^{*C}$ is a global maximum. These conditions are consistent with the assumed form of the detection function (assumption 7-2).

Lemma 1 establishes that $r < 1$. Intuitively, misrepresentation is likely to be modest when providers value future business and wish to secure their reputation (Kreps and Wilson, 1982). Providers are not willing to risk much current period exploitation because it endangers potentially sizeable future profits as denoted by $\breve{\pi}$ in (7-9).

What are the implications for the size of misrepresentation when providers are satisficers rather than profit maximisers? The expectation is that satisficing providers will misrepresent to a lesser degree than their profit maximising counterparts because they suffer disutility from excessive pricing. The term $\psi$ in the utility function is negatively related to (the present value of) price. Generally then profit maximising prices will produce some disutility and at the margin we would expect satisficers to reduce $\hat{\beta}$ below profit maximising levels, $\hat{\beta}^{*C}$. However, because they do not suffer disutility of this type if they provide nothing, this problem does offer the possibility of a rather counter-intuitive result. If total disutility of price increases was very high, and reputation effects were minimal then, in theory at least, a provider might wish to *increase* misrepresentation in order to reduce *expected* prices. Nonetheless, in this case, marginal profit would also fall along with disutility, and since the former is assumed to have a greater value, in practice we should not get the counter-intuitive result.

Differentiating the utility function (7-5) for given state $k$ yields:

$$(7\text{-}12)\quad U_{\hat{\beta}} = v'\pi_{\hat{\beta}} + \frac{\partial[r\psi]}{\partial\hat{\beta}} = v'\pi_{\hat{\beta}} + \frac{\partial[rD\Psi]}{\partial\hat{\beta}} = v'\pi_{\hat{\beta}} + rD\Psi'\rho_{\hat{\beta}} + r\Psi D_{\hat{\beta}} + D\Psi r_{\hat{\beta}}$$

where, as in chapter 5, $\psi(\rho_k) = D\Psi(\rho_k)$.[53] At $\hat{\beta}^{*C}$, this function reduces to: $rD\Psi'\rho_{\hat{\beta}} + r\Psi D_{\hat{\beta}} + D\Psi r_{\hat{\beta}}$ where the first and second terms are negative, but the last is positive ($\Psi < 0$, $\rho_{\hat{\beta}} = p_{\hat{\beta}} > 0$, $r_{\hat{\beta}} = r_\theta < 0$). If the objective function was slightly different in that providers felt disutility associated with excessive prices even if they lost the contract[54], (7-12) reduces to: $U_{\hat{\beta}} = v'\pi_{\hat{\beta}} + \frac{\partial[\psi]}{\partial\hat{\beta}} = v'\pi_{\hat{\beta}} + \frac{\partial[D\Psi]}{\partial\hat{\beta}} = v'\pi_{\hat{\beta}} + D\Psi'\rho_{\hat{\beta}} + \Psi D_{\hat{\beta}} < 0$. This result is clear-cut. It means that at $\hat{\beta}^{*C}$, where $\pi_{\hat{\beta}} = 0$, we have $U_{\hat{\beta}} < 0$ and so reductions in reported β will increase utility.

Returning to the case where the differential is as (7-12), then at $\hat{\beta}^{*C}$, although this function suggests the possibility of the sign of the differential to be positive, further investigation undermines this possibility. To show this, assume that marginal production costs are normalised to zero. This allows us to focus on the disutility of excess prices rather then the disutility of having to bear production costs. Moreover, we should note that we have assumed excessive price disutility to be modest in size.

We draw on the differential (7-9), which with zero marginal costs is, $\pi_{\hat{\beta}} = rDp_{\hat{\beta}} + rD_{\hat{\beta}}(p-c) + Dr_\theta(p-c) + r_\theta\breve{\pi} = rDp_{\hat{\beta}} + rD_{\hat{\beta}}p + Dr_\theta p + r_\theta\breve{\pi}$. At $\hat{\beta}^{*C}$, $\pi_{\hat{\beta}} = 0$ and so this function can be re-arranged to be:

$$(7\text{-}13)\quad Dr_\theta = -\frac{Drp_{\hat{\beta}}}{p} - rD_{\hat{\beta}} - \frac{r_\theta\breve{\pi}}{p}$$

---

[53] i.e. that disutility of excessive prices is felt per unit output.

[54] because losing the contract will reduce access for users, at least during transition to a new provider

Substituting into (7-12) gives:

$$(7\text{-}14)\quad U_{\hat{\beta}} = v'\pi_{\hat{\beta}} + rD\Psi'\rho_{\hat{\beta}} + r\Psi D_{\hat{\beta}} + \Psi\left(-\frac{Drp_{\hat{\beta}}}{p} - rD_{\hat{\beta}} - \frac{r_{\theta}\breve{\pi}}{p}\right)$$

or

$$(7\text{-}15)\quad U_{\hat{\beta}} = rD\Psi'\rho_{\hat{\beta}} - \Psi\frac{Drp_{\hat{\beta}}}{p} - \Psi\frac{r_{\theta}\breve{\pi}}{p} = \frac{1}{p}\left[rDp_{\hat{\beta}}(\Psi'p - \Psi) - \Psi r_{\theta}\breve{\pi}\right]$$

Now by assumption there is no disutility when prices are low i.e. $\psi'' = 0$ and $\Psi = 0$ for $p \leq \breve{p}$. Therefore we can suppose, in general, that $\Psi = \Psi'(p - \breve{p}) = \Psi'p - \Psi'\breve{p} < 0$ for $p > \breve{p}$ and so $(\Psi'p - \Psi) = \Psi'\breve{p} < 0$. Hence, at $\hat{\beta}^{*C}$,

$$(7\text{-}16)\quad U_{\hat{\beta}} = \frac{1}{p}\left[rDp_{\hat{\beta}}(\Psi'p - \Psi) - \Psi r_{\theta}\breve{\pi}\right] < 0$$

Moreover, when reputation effects are comparatively strong i.e. when $\breve{\pi}$ is larger (e.g. due to a smaller rate of discount), then $U_{\hat{\beta}}\left(\hat{\beta}^{*C}\right)$ will be more negative.

We can therefore safely assume that $\frac{\partial[r\psi]}{\partial\hat{\beta}} < 0$ and so advance the following straightforward lemma,

*Lemma 2. Other things being equal, when expected disutility of price falls with respect to an increase in* $\hat{\beta}$ *i.e.* $\frac{\partial[r\psi]}{\partial\hat{\beta}} < 0$, *then information rent/profit level differences for profit maximising providers using contingent contracts will be higher than for satisficing providers using contingent contracts:*

$\Lambda^{IC}_{\psi'<0} = \pi\left(\hat{\beta}^{*C}_{\psi'<0}\right) - \pi\left(\beta^{C}_{\psi'<0}\right) < \Lambda^{IC}_{\psi=0} = \pi\left(\hat{\beta}^{*C}_{\psi=0}\right) - \pi\left(\beta^{C}_{\psi=0}\right)$.

*Proof*

Differentiating the utility function gives:

(7-17) $$U_{\hat{\beta}}\left(\hat{\beta}^{*C}\right)=v'\pi_{\hat{\beta}}+\frac{\partial[r\psi]}{\partial\hat{\beta}}<0$$

A solution, $\hat{\beta}^{*}_{\psi'<0}<\hat{\beta}^{*C}$, may exist or a corner solution may exist where $U_{\hat{\beta}}<0$ for $\hat{\beta}=\beta$. In either case however, $\pi_{\hat{\beta}}\left(\hat{\beta}^{*}_{\psi'<0}\right)>0$ and so there is a lessened misuse of information than in the profit maximising case. □

We have considered the optimal reported cost-parameter with respect to the sample mean value of the cost-parameter of the referred client. How then will the optimal reported parameter change for changes in the cost-parameter of the referred client? In other words, what is $\partial\hat{\beta}/\partial\beta'$, which we denote *m*? Intuitively, we would expect this value to be close to 1 because we are only changing the baseline from which cost exaggeration occurs.

*Lemma 3. (a) For fully contingency (C1) contracts, for any* $\psi$, *m = 1. (b) For partially contingent (C2) contracts* $m\in[\underline{m},\overline{m}]$ *where* $\underline{m}<1$ *and* $\overline{m}>1$.

*Proof*

See appendix 7.7.3. □

This lemma shows that for partially contingent contracts the size of *m* depends closely on the effective value of $D_p$ as associated with changes in referred dependency.

### 7.3.3 Cream-skimming

Cream-skimming involves providers selecting clients with low cost characteristics. Assume initially that there is free selection – providers can effectively choose a value for $\beta$. Providers are initially referred clients with characteristic $\beta' \in [\overline{\beta},\underline{\beta}]$. Is there any incentive to re-select clients with a lower value of $\beta \in (\beta',\underline{\beta}]$?

Purchasers are concerned that providers are not vertically selective – providers can select among clients of the same dependency, but to select lower dependency/lower cost clients when the purchaser is under the impression that no selection occurred is opportunistic. Consequently, if providers are vertically selectively they must misrepresent the dependency-cost characteristics of the selected client as being of the original level. Hence, $r = r(\beta' - \tilde{\beta}) = r(\beta' - \beta) = r(\theta)$ because $\tilde{\beta} = \beta$. As in the above case, $\theta$ is the degree of misrepresentation of $\beta$.

In general, differentiating the profit function with respect to $\beta$ gives:

$$(7\text{-}18)\quad \pi_\beta = \left.\frac{\partial\pi(\beta')}{\partial\beta}\right|_{\hat{\beta}} + \frac{\partial\pi}{\partial\hat{\beta}}\frac{\partial\hat{\beta}}{\partial\beta}$$

Using the first order condition, (7-6), of the profit function, demand changes by:

$$(7\text{-}19)\quad D_\beta = -\frac{\Omega_\beta}{\Omega_D} = -\frac{-c_\beta}{\dfrac{1}{D_p} + \dfrac{1}{D_p}} = \tfrac{1}{2} c_\beta D_p$$

This function indicates how optimal demand changes in response to a change in marginal costs. Furthermore,

$$(7\text{-}20)\quad \left.\frac{\partial\pi(\beta')}{\partial\beta}\right|_{\hat{\beta}} = \pi_\beta\big|_{\hat{\beta}} = rD\frac{\partial(p-c)}{\partial\beta} + rD_\beta(p-c) - Dr_\theta(p-c) - r_\theta\breve{\pi}$$

Using the $\beta$ analogy to (7-70) (see page 263) this becomes:

$$(7\text{-}21)\quad \pi_\beta\big|_{\hat{\beta}} = -rD\left(c_\beta - \frac{D_\beta}{D_p}\right) + rD_\beta(p-c) - Dr_\theta(p-c) - r_\theta\breve{\pi}$$

where

$$(7\text{-}22)\quad \frac{\partial\pi}{\partial\hat{\beta}}\frac{\partial\hat{\beta}}{\partial\beta} = \left[rDp_{\hat{\beta}} + rD_{\hat{\beta}}(p-c) + Dr_\theta(p-c) + r_\theta\breve{\pi}\right]\frac{\partial\hat{\beta}}{\partial\beta}$$

using (7-9). Hence:

$$\pi_\beta = -rD\left(c_\beta - \frac{D_\beta}{D_p}\right) + rD_\beta(p-c) - Dr_\theta(p-c) - r_\theta\breve{\pi}$$

(7-23)

$$+\left[rDp_{\hat{\beta}} + rD_{\hat{\beta}}(p-c) + Dr_\theta(p-c) + r_\theta\breve{\pi}\right]\frac{\partial\hat{\beta}}{\partial\beta}$$

*Lemma 4. Under non-contingent contracts, selection occurs such that* $\beta < \beta^I$.

*Proof*

For a non-contingent contract, $p = p(\breve{\beta})$ so $p_{\hat{\beta}} = 0$ and $D_{\hat{\beta}} = 0$. Differentiating the profit function (7-3) with respect to β gives in this case:

(7-24) $\pi_\beta = -rc_\beta D - r_\theta\left[\left(p(\breve{\beta}) - c(\beta)\right)D + \breve{\pi}\right]$

Under non-contingent contracts there is no cost exaggeration and so $\hat{\beta} = \beta^I$. What incentives for cream-skimming exist at $\hat{\beta} = \beta^I = \beta$? The differential (7-24) then reduces to $\pi_\beta(\beta^I) = -Dc_\beta < 0$ which cannot be a turning point, so $\hat{\beta} = \beta^I = \beta$ is not a solution. The optimal value, if it exists, is thus $\beta^{*N} < \hat{\beta} = \beta^I$. □

In the appendix it is shown that $\beta^{*N}$ is a global maximum – given our assumptions about detection (Lemma A2).

*Lemma 5. Under a fully contingent contract, at the optimal cost exaggeration solution,* $\hat{\beta}^*$*, incentives to cream-skim for a profit maximiser are absent.*

*Proof*

At the optimal cost exaggeration solution:

(7-25) $\pi_{\hat{\beta}}(\hat{\beta}^*) = rDp_{\hat{\beta}} + rD_{\hat{\beta}}(p-c) + Dr_\theta(p-c) + r_\theta\breve{\pi} = 0$

Here $\hat{\beta}*$ is the optimal level of exaggeration when referral dependency is $\beta' = \breve{\beta}$. Cream-skimming involves the selection of clients to reduce actual dependency β to a level below $\beta' = \breve{\beta}$. If providers have already exaggerated dependency to suggest to purchasers that referred dependency in this transaction was $\hat{\beta}*$, then any subsequent cream-skimming does not involve a change in reported dependency levels i.e. $\hat{\beta} = \hat{\beta}*$ and so $\partial\hat{\beta}/\partial\beta = 0$. Furthermore, the purchaser expects demand to remain at the level $\hat{\beta} = \hat{\beta}*$ and so $D_\beta = 0$. The marginal benefits of cream-skimming are given by (7-23), which in the case where $\partial\hat{\beta}/\partial\beta = 0$ and $D_\beta = 0$ becomes:

$$(7\text{-}26)\quad \pi_\beta = -rc_\beta D - \left[Dr_\theta(p-c) + r_\theta\bar{\pi}\right]$$

Re-arranging (7-25) such that $rDp_{\hat{\beta}} + rD_{\hat{\beta}}(p-c) = -\left[Dr_\theta(p-c) + r_\theta\bar{\pi}\right]$ and substituting it into (7-26), using (7-6) gives:

$$(7\text{-}27)\quad \pi_\beta = -rc_\beta D + rDp_{\hat{\beta}} - rD_{\hat{\beta}}\frac{D}{D_p}$$

Using (7-71) and (7-7) this becomes:

$$(7\text{-}28)\quad \pi_\beta\left(\hat{\beta}*\right) = -rDc_\beta + rD\tfrac{1}{2}p_{\hat{\beta}}\Big|_D + rD\tfrac{1}{2}p_{\hat{\beta}}\Big|_D = rD\left(p_{\hat{\beta}}\Big|_D - c_\beta\right) = 0 \text{ for } h = 1.$$

Hence profit maximising providers with fully contingent contracts do not gain from cream-skimming at the cost exaggeration optimal.

□

Provider motivations as we have cast them are not an issue with regard cream-skimming because there is no direct price implication; prices do not change if providers cream-skim clients. This conjecture is the basis of lemma 6.

*Lemma 6. Other things being equal information rent for maximsing providers under non-contingent contracts will be equal to that for satisficing providers using non-contingent contracts:* $\Lambda^{IN}_{\psi'<0} = \pi(\hat{\beta}^{N}_{\psi'<0}) - \pi(\beta^{*N}_{\psi'<0}) = \Lambda^{IN}_{\psi=0} = \pi(\hat{\beta}^{N}_{\psi=0}) - \pi(\beta^{*N}_{\psi=0})$ *or* $\Lambda^{IN}_{\psi'<0} = \pi(\breve{\beta}_{\psi'<0}) - \pi(\beta^{N}_{\psi'<0}) = \Lambda^{IN}_{\psi=0} = \pi(\breve{\beta}_{\psi=0}) - \pi(\beta^{N}_{\psi=0})$.

*Proof*

The utility function under non-contingent contracts is:

$$U = v(\pi - y) + \psi(\rho(\breve{\beta})) = v(rD[p(\breve{\beta}) - c(\beta)] + r\bar{\pi} - y) + r\psi(\rho) \tag{7-29}$$

where $\rho(\breve{\beta})$, the (present value) of price, is a constant set by the local authority purchaser. At this value, $\psi = 0$ so (7-29) reduces to: $U = v(rD[p(\breve{\beta}) - c(\beta)] + r\bar{\pi} - y)$ i.e. the same as the maximising case. Effectively providers lose control over pricing and therefore disutility stemming from charging 'excessive' prices is irrelevant. □

This result is slightly removed from the spirit of satisficers as somehow having an altruistic component to their utility functions. What drives the above result is that $\Psi = \Psi'(p - \breve{p})$ and that $\Psi = 0$ for $p \leq \breve{p}$. We have assumed thus far that $\breve{p}$ is fair price for the population average dependency. Individual providers may feel that actually $\breve{p}$ refers to the home average dependency, and therefore when they cream-skim $\breve{p}$ falls. It follows that the non-contingent market price is then above $\breve{p}$ when cream-skimming occurs. In this case we have, with analogy to Lemma 2,

$$U_\beta = v'\pi_\beta + \frac{\partial[r\psi]}{\partial\beta} = v'\pi_\beta + r\frac{\partial\psi}{\partial\breve{p}}\frac{\partial\breve{p}}{\partial\beta} + \psi r_\beta = v'\pi_\beta - r\psi'\frac{\partial\breve{p}}{\partial\beta} + \psi r_\beta \tag{7-30}$$

where $\frac{\partial\breve{p}}{\partial\beta} > 0$ by construction. Again referring to Lemma 2, it follows that for a small reduction in $\psi'$ from $\psi = 0$, $\left.\frac{\partial U_\beta(\beta^{*N})}{\partial\psi'}\right|_{\psi=0} = -r\frac{\partial\breve{p}}{\partial\beta} < 0$ i.e. a lessened cream-skimming would occur in this case.

How does cream-skimming behaviour change with changes in the referred client cost-parameter: $\frac{\partial\beta}{\partial\beta^{I}}$?

*Lemma 7.* $\frac{1}{2} < w = \frac{\partial\beta}{\partial\beta^{I}} \leq 1$.

*Proof*

See appendix 0.

Unlike cost parameter exaggeration under contingent contracts, Lemma 7 holds for both maximising and satisficing providers. This result stems from Lemma 6, which indicates that both types of providers have the same cream-skimming incentives.

Together Lemma 1 and Lemma 4 show that for basic reimbursement systems (of the sort described above), under both types of contract, asymmetric information gives rise to information rents, in the form cream-skimming or cost-misrepresentation. What then are the conditions when contingent contracts will generate greater rents than non-contingent contracts?

Let $\pi^{C}\left(\theta^{C*}_{\psi=0}\right)$ be the (optimal) rent from a pure strategy of cost-exaggeration, and $\pi^{N}\left(\theta^{N*}_{\psi=0}\right)$ be the (optimal) rent from a pure strategy of cream-skimming. Then we have the following proposition.

*Proposition 1. For fully contingent contracts (h = 1), for profit maximisers with initial referrals* $\beta^{I} = \breve{\beta}$, $\pi^{C}\left(\theta^{C*}_{\psi=0}\right) > \pi^{N}\left(\theta^{N*}_{\psi=0}\right)$. *For some h > 1,* $\bar{h}$ *exists such that* $\pi^{C}\left(\theta^{C*}_{\psi=0}\right) \leq \pi^{N}\left(\theta^{N*}_{\psi=0}\right)$.

*Proof*

See appendix 7.7.2.

What essentially drives this result is that although marginal gains per unit of output are the same under the two strategies i.e. $\pi_\theta^N(\theta)\big|_D = \pi_\theta^C(\theta)\big|_D = D\left[rc_\beta - \frac{D}{D_p}r_\theta\right] + r_\theta\bar{\pi}$ , under contingent contracts, demand can change to its optimal level, whilst with non-contingent contracts it cannot. In the latter case the prevailing level of demand will not be the optimising level and consequently profit gain will be lower.

Providers have three available strategies: to exaggerate costs, to cream-skim or to opt for a mix of the first two strategies. Lemma 1 shows that providers with non-contingent contracts have no incentives to exaggerate costs and therefore they will only cream-skim. Proposition 1 and Lemma 5 show that providers with contingent contracts will choose between either of the pure strategies depending on the degree of contract contingency, $h$. In particular, when $h \geq \bar{h}$ the pure cream-skimming strategy will yield the highest rent. In that respect a provider with a contingent contract with a low degree of contingent will act in the same way as a provider facing a non-contingent contract. In fact $h \geq \bar{h}$ ensures that $\pi^C\left(\theta_{\psi=0}^{C*}\right) < \pi^N\left(\theta_{\psi=0}^{N*}\right)$. But since at $h = \bar{h}$ we have $\pi_\theta^C(0) < \pi^N(0)$, we can instead define $1 < h^N < \bar{h}$ as the level of contingency that gives $\pi^C\left(\theta_{\psi=0}^{C*}\right) = \pi^N\left(\theta_{\psi=0}^{N*}\right)$.

The proposition implies that providers either cream-skim or exaggerate cost reports depending on the degree of contingency. And it is possible that cream-skimming can occur under low contingency contracts (i.e. where $h \geq \bar{h}$ ). Intuitively, because the chance of detection is defined on the difference between the reported and actual cost-parameter, $\theta$, the chance being the same for an equal difference resulting from either cream-skimming or cost exaggeration, the dominance of the marginal benefits at $\theta = 0$ of one type of asymmetric information behaviour or the other will apply through the range of values of $\theta$. In other words, either cream-skimming but not cost exaggeration or vice versa is the provider's optimal strategy. And this choice hinges on the degree of contingency, $h$. If the degree of contingency of a contingent contract is high then providers will earn greater information rent by exaggerating cost reports compared to rents accrued from cream-skimming. Below the critical value it would be optimal to cream-skim regardless of contract type.

Proposition 1 refers to profit maximising providers. How does a change in provider motivations affect these relationships? To generalise we can write the difference in information rent as $\delta_{\psi}^{I} = \pi^{C}\left(\theta_{\psi}^{C*}\right) - \pi^{N}\left(\theta_{\psi}^{N*}\right), \forall\psi \leq 0$. The next proposition shows that the difference is reduced when providers suffer some excess price disutility i.e. $\psi < 0$.

*Proposition 2.* $\delta_{\psi=0}^{I} > \delta_{\psi'<0}^{I}$ *for* $\delta_{\psi=0}^{I} > 0$ *and* $\delta_{\psi=0}^{I} = \delta_{\psi'<0}^{I}$ *for* $\delta_{\psi=0}^{I} = 0$

*Proof*

First,

$$
\text{(7-31)} \quad
\begin{aligned}
\delta_{\psi}^{I} &= \pi^{*C}(\breve{\beta}) - \pi^{*N}(\breve{\beta}) \\
&= \pi^{C*}\left(\theta_{\psi}^{C*}\right) - \pi^{N*}\left(\theta_{\psi}^{N*}\right) \\
&= \pi\left(\hat{\beta}_{\psi}^{*C}\right) - \pi\left(\beta_{\psi}^{*N}\right) - \pi\left(\breve{\beta}_{\psi}\right) + \pi\left(\breve{\beta}_{\psi}\right) \\
&= \Lambda_{\psi}^{IC} - \Lambda_{\psi}^{IN}, \qquad\qquad \forall\psi \leq 0
\end{aligned}
$$

Hence,

$$
\text{(7-32)} \quad
\begin{aligned}
\delta_{\psi=0}^{I} - \delta_{\psi'<0}^{I} &= \Lambda_{\psi=0}^{IC} - \Lambda_{\psi=0}^{IN} - \Lambda_{\psi'<0}^{IC} + \Lambda_{\psi'<0}^{IN} \\
&= \Lambda_{\psi=0}^{IC} - \Lambda_{\psi'<0}^{IC}
\end{aligned}
$$

which follows from Lemma 6, i.e. $\Lambda_{\psi'<0}^{IN} = \Lambda_{\psi=0}^{IN}$. Second,

$$
\text{(7-33)} \quad \delta_{\psi=0}^{I} - \delta_{\psi'<0}^{I} = \Lambda_{\psi=0}^{IC} - \Lambda_{\psi'<0}^{IC} > 0
$$

from Lemma 2 when $\delta_{\psi=0}^{I} > 0$ i.e. $h = 1$.

In addition, for $\delta_{\psi=0}^{I} = 0$, this implies providers under all contract types cream-skim, and again from Lemma 6, $\delta_{\psi=0}^{I} - \delta_{\psi'<0}^{I} = -\Lambda_{\psi=0}^{IN} + \Lambda_{\psi'<0}^{IN} = 0$. □

In other words the difference in information rents between profit maximisers and satisficers comes down to the difference in rents from cost exaggeration since there is no difference from cream-skimming. The implication is that when providers are satisficers they opt for the cream-skimming strategy at higher degrees of contract contingency than profit maximisers.

Cream-skimming as modelled above allows providers to freely select possible residents. In effect, providers are able to select residents from a sufficiently large pool of potential users, choosing low dependency, low cost residents (but claiming to provide service to the average mix). However, given the high level of competition and apparent over-supply of residential care, we might expect the referral of potential residents to be much more limited (see also Pauly, 1984). As a result, rejecting a potential resident may leave the home with a vacant place (attracting no revenue, but still incurring some costs) until a suitably low dependency person is referred. Homes may set themselves up to attract low dependency people, but need to do so without this strategy becoming overly apparent to purchasers. In any case they will be subject to competition from other homes trying the same strategy. Whatever the exact mechanism, homes will not have available an inexhaustible supply of low dependency potential residents. Without simultaneous replacement, the costs of cream-skimming behaviour rise significantly (Forder, 1997a).

The relaxation of this free replacement assumption can be modelled as a factor $g$ that reduces expected demand. Thus we expand the profit function (7-3) to be:

$$(7\text{-}34)\quad \pi = [r(\theta) - g(\theta)]D(\hat{\beta})(p(\hat{\beta}) - c(\beta) + \breve{\pi})$$

where $g_\theta(\theta) > 0$, $g(0) = 0$ and so $r(0) - g(0) = 1$. Any increase above zero of $\theta$ due to cream-skimming not only yields a positive chance of punitive action (loss of all demand) but also a reduction in demand. In practice, purchasers may offer some replacement of potential residents so that selection does not imply a one-for-one reduction in output, but we assume the probability of replacement is less than 1. So what are the implications?

*Lemma 8.* $\pi_\beta(\theta,0) < \pi_\beta(\theta,g) < 0$

*Proof*

The counterpart to (7-24), the first order condition for pure cream-skimming is:

$$(7\text{-}35)\quad \pi_{\beta}(g,\theta)=D\left[-(r-g)c_{\beta}-\left(p(\breve{\beta})-c(\beta)+\frac{\breve{\pi}}{D}\right)(r_{\theta}-g_{\theta})\right]$$

Hence, $\pi_{\beta}(\theta,g)>\pi_{\beta}(\theta,0)$ because $g_{\theta}(\theta)>0$ and $g>0$. □

This result implies that cream-skimming is less rewarding and so reinforces Proposition 1.

*Proposition 3. If replacement is costly such that* $g_{\theta}>0$ *then at* $h$ = $h^{N}$ *cost exaggeration yields greater profits than cream-skimming i.e.* $\delta^{I}_{\psi=0}>0$.

*Proof of proposition 3.*

This proposition follows from propositions 1 and 2 and Lemma 8. □

The lack of free replacement means that even with low contingency contracts cost exaggeration can be the optimal choice.

## 7.4 Risk and Uncertainty

Relaxing the assumption of risk neutrality, when providers are risk averse, we would expect a *risk effect* to will impact on profit rates, with the size of which is dependent on the type of contract a provider uses. As we have noted above, provider costs vary according to the characteristics of those placed. Demand is also assumed to vary from state to state and together the cost and demand effects will have implications for profit in any given state. When the provider is uncertain of cost and demand, and so profitability, then the choice of contract has implications for prices, the effect being influenced by the degree of provider risk aversion (Shavell, 1979). To fix this idea we assume that $v''<0$ and $v'''=0$ and also that $v''_{\psi=0}=v''_{\psi>0}$.

Profits are state-contingent and so provider utility is: (the provider identifier $i$ is suppressed for clarity)

$$U = \gamma_1[v(\pi_1 - y) + r_1\psi(\rho_1)] + \gamma_2[v(\pi_2 - y) + r_2\psi(\rho_2)] + \dots + \gamma_s[v(\pi_s - y) + r_s\psi(\rho_s)] \tag{7-36}$$

where $\gamma_1 + \gamma_2 + \dots + \gamma_s = 1$ and $\pi(\breve{\beta}) - y = \sigma$ is the provider's expected (period 2) profit, which is unchanged for any state $k$. We can write expected profit as:

$$\gamma_1\pi_1 + \gamma_2\pi_2 + \dots + \gamma_s\pi_s = \sigma + y \tag{7-37}$$

and expected price:

$$\gamma_1\rho_1 + \gamma_2\rho_2 + \dots + \gamma_s\rho_s = \rho \tag{7-38}$$

Let expected profit in each state vary: $\pi_k - y = \sigma + \phi_k$ which means that $\gamma_1\phi_1 + \gamma_2\phi_2 + \dots + \gamma_s\phi_s = 0$. Also let $\rho_k = \rho + \mu(\phi_k)$, again with $\gamma_1\mu(\phi_1) + \dots + \gamma_s\mu(\phi_s) = 0$.

State 1 – with profits $\pi_1 - y = \sigma + \phi_1$ – is the lowest profit state. State $s$ – with profits $\pi_s - y = \sigma + \phi_s$ – is the highest profit state. Therefore we have: $\phi_1 < \phi_2 < \dots < \phi_{s-1} < \phi_s$.

Expected utility is therefore:

$$U = \gamma_1 v(\sigma + \phi_1) + \gamma_1 r\psi(\rho + \mu(\phi_1)) + \dots + \gamma_s v(\sigma + \phi_s) + \gamma_s r\psi(\rho + \mu(\phi_s)) \tag{7-39}$$

Since $v''' = 0$ and $\psi'' = 0$ utility will be given by the second-order Taylor series expansion:

$$U = \gamma_1[v(\sigma) + v'(\sigma)\phi_1 + \tfrac{1}{2}v''\phi_1^2 + r\psi(\rho) + r\psi'(\rho)\mu(\phi_1)] + \dots + \gamma_s[v(\sigma) + v'(\sigma)\phi_s + \tfrac{1}{2}v''\phi_s^2 + r\psi(\rho) + r\psi'(\rho)\mu(\phi_s)] \tag{7-40}$$

$$U = (v(\sigma)+r\psi(\rho))[\gamma_1+\gamma_2+\ldots+\gamma_s]+v'(\sigma)[\gamma_1\phi_1+\gamma_2\phi_2+\ldots+\gamma_s\phi_s] + r\psi'(\rho)[\gamma_1\mu(\phi_1)+\gamma_2\mu(\phi_2)+\ldots+\gamma_s\mu(\phi_s)]+\tfrac{1}{2}v''[\gamma_1\phi_1^2+\gamma_2\phi_2^2+\ldots+\gamma_s\phi_s^2] \tag{7-41}$$

$$U = v(\sigma)+\tfrac{1}{2}v''[\gamma_1\phi_1^2+\gamma_2\phi_2^2+\ldots+\gamma_s\phi_s^2]+r\psi(\rho) \tag{7-42}$$

Each state-of-the-world $k$ is characterised by the client-cost-parameter of the referred client, that is, $\beta^l$. What are the profit consequences of the deviation of this parameter from the sample average value $\breve{\beta}$? Profit is $\pi = \pi(\hat{\beta}(\beta^l), \beta(\beta^l))$. Expanding this function around $\breve{\beta}$ gives:

$$\phi_k = \left[\frac{\partial\pi}{\partial\hat{\beta}}\frac{\partial\hat{\beta}}{\partial\beta^l}+\frac{\partial\pi}{\partial\beta}\frac{\partial\beta}{\partial\beta^l}\right]d\beta^l \tag{7-43}$$

where $d\beta^l = \beta^l - \breve{\beta}$. The two differentials $\partial\hat{\beta}/\partial\beta^l$ and $\partial\beta/\partial\beta^l$ describe respectively how reported and actual dependency change with the cost parameter of the referred client (see lemmas 3 and 7). These differentials reflect implications for cost exaggeration behaviour and cream-skimming respectively, and are functions of the degree of contract contingency. Because this behaviour occurs after the state $k$ is known, it is completely described by first order conditions.

The size of $\phi_k$ will depend on the degree of contract contingency measured, which affects the relationship between $p_{\hat{\beta}}$ and $c_\beta$:

*Lemma 9.* $\phi_k^{C1} = (\pi_{\hat{\beta}}+\pi_{\beta^l})d\beta^l = 0$

For fully contingent contracts i.e. $h = 1$, the above propositions tell us that providers do not cream-skim and therefore: $\beta \equiv \beta^l$ and hence $\frac{\partial\pi}{\partial\beta}\frac{\partial\beta}{\partial\beta^l} \equiv \pi_{\beta^l}\Big|_{\hat{\beta}}$. Lemma 3 indicates that $m = 1$ and that $\pi_{\hat{\beta}}+\pi_{\beta^l} = 0$. Hence from (7-43) we have Lemma 9. □

For non-contingent contracts, $\hat{\beta} \equiv \beta^{I}$ and so $\frac{\partial \pi}{\partial \hat{\beta}} \frac{\partial \hat{\beta}}{\partial \beta^{I}} \equiv \pi_{\beta^{I}}\big|_{\beta}$. Differentiating the profit function (7-3) with respect to $\beta^{I}$ holding $\beta$ constant is equivalent to the effect of a pure increase in the degree of misrepresentation of $\theta = \hat{\beta} - \beta \equiv \beta^{I} - \beta$ by $\beta^{I}$ i.e. $\frac{\partial \pi}{\partial \hat{\beta}} \frac{\partial \hat{\beta}}{\partial \beta^{I}} \equiv \pi_{\beta^{I}}\big|_{\beta} = r_{\theta}(D(p-c)+\breve{\pi}) < 0$. In this case, with cream-skimming, the size of (7-43) depends on the value of $w$:

$$(7\text{-}44)\quad \phi_k^N = \left[\frac{\partial \pi}{\partial \hat{\beta}} \frac{\partial \hat{\beta}}{\partial \beta^{I}} + \pi_{\beta} w\right] d\beta^{I}$$

$$(7\text{-}45)\quad \phi_k^N = \left[r_{\theta}(D(p-c)+\breve{\pi}) + \pi_{\beta} w\right] d\beta^{I}$$

or using (7-24),

$$(7\text{-}46)\quad \phi_k^N = \left[-D^N c_{\beta} r w + (1-w) r_{\theta}(D(p-c)+\breve{\pi})\right] d\beta^{I}$$

Hence, we have:

*Lemma 10.* $\phi_k^N < 0$ *for* $d\beta^{I} > 0$

All the terms in square brackets in (7-44) are negative (see Lemma 7). □

*Lemma 11. Regarding partially contingent contracts, a positive change in referred dependency, $d\beta^{I} > 0$, will yield a reduction in profits:* $0 > \phi_k^{C2} > \phi_k^N$.

The impact on profits will be smaller than for providers with non-contingent contracts because revenue changes – at least partially – in line with positive cost shocks. So we have: $\frac{\partial \pi^{C2}}{\partial \beta^{I}}(\breve{\beta})\Big|_{\hat{\beta},D} = \frac{\partial \pi^{N}}{\partial \beta^{I}}(\breve{\beta})$, but $\frac{\partial p}{\partial \beta^{I}}(\breve{\beta}) > 0$ and, under C2 contracts, demand is allowed to vary to its profit maximising level. □

Writing $\frac{1}{2}v''[\gamma_1\phi_1^2 + \gamma_2\phi_2^2 + ... + \gamma_s\phi_s^2] = \frac{1}{2}v''\Phi$ we can establish the following proposition:

*Proposition 4.* $\frac{1}{2}v''\Phi^N < \frac{1}{2}v''\Phi^C < 0$ *if* $h = 1$.

*Proof.*
The first part requires that $\Phi^C < \Phi^N$. It is sufficient therefore to show that $\gamma_k(\phi_k^C)^2 < \gamma_k(\phi_k^N)^2, \forall k = 1,..,s$, which can be reduced to $|\phi_k^C| - |\phi_k^N| < 0, \forall k = 1,..,s$. This latter condition is satisfied as a result of Lemma 9 and Lemma 10. □

Utility in each case is (from (7-42)):

$$U^j = v(\pi^{*j}) + \frac{1}{2}v''\Phi^j + r\psi(\rho^j), \quad j = C, N \tag{7-47}$$

The *risk-effect* on utility is the term $\frac{1}{2}v''\Phi^j(\pi_1^j,.., \pi_s^j)$. The difference between contingent and non-contingent risk effects is: $\Delta = \frac{1}{2}v''(\sigma)\Phi^C - \frac{1}{2}v''(\sigma)\Phi^N > 0$. Suppose that this difference is the only difference in the utility of representative providers under different contracts, that is, all providers would have the same expected profit in the absence of risk. Providers need to be compensated for comparatively higher disutility of expected profit fluctuations such that $U^C - U^N = 0$ and this can be achieved by paying a risk premium to add to expected profit. The way in which the compensation factor is awarded to providers is important. We assume that the risk premium takes the form of a higher price on each unit sold, i.e. $\delta^R = D(p^R - \breve{p})$ so that $\frac{\delta^R}{D} + \breve{p} = p^R$ and $\frac{\partial p^R}{\partial \delta^R} = \frac{1}{D}$. Conceivably, the compensation could be awarded as more sales, assuming that each sale accrues a positive profit, thus giving $p_{\delta^R} = 0$. However, this option is unlikely in practice as it presupposes some excess service need exists as well as spare capacity or zero costs in expanding output.

*Proposition 5. For profit maximisers, with C1 and C2 contingent contracts,* $\delta^R < 0$.

*Proof.*

Using (7-47) the expected utility difference is reduced to zero:

$$(7\text{-}48)\quad U^C - U^N = v(p^R) - v(\breve{p}) + r\psi(p^R) - r\psi(\breve{p}) + \Delta = 0$$

or

$$(7\text{-}49)\quad \Delta^R \equiv v(p^R) - v(\breve{p}) + r\psi(p^R) - r\psi(\breve{p}) = -\Delta$$

Differentiating we have: $\frac{\partial \Delta^R}{\partial p^R} = v' + r\psi' > 0$ (by assumption 1-1 and since $r \leq 1$) where the premium $p^R$ does not affect the expected profit variation under each contract type i.e. $\Delta_{p^R} = 0$. Hence, $\frac{\partial \Delta^R}{\partial \delta^R} = \frac{\partial \Delta^R}{\partial p^R}\frac{\partial p^R}{\partial \delta^R} = (v' + r\psi')\frac{1}{D} > 0$, which means that the premium must be negative, $\delta^R < 0$, to satisfy (7-49). □

### 7.4.1 Risk and provider motivation

How will risk effects of different contracts compare as between profit maximisers and satisficers? There are two aspects to this problem, which can be seen with reference to (7-48). First, how does profit change following a cost shock, and therefore what the size of $\Delta$ will be in the function. Second, what the implications are for how the premium is paid to providers with least protecting contracts. We can compare full contingent (C1) contracts with non-contingent (NC) contracts. As to the first aspect, under C1 contracts all providers are fully protected and Lemma 9 indicates that cost shocks do not change profits. Hence, we have: $\Delta = -\frac{1}{2}v''(\sigma)\Phi^N > 0$. Lemma 6 indicates that under contingent contracts there are no behavioural differences between provider types, and therefore $\Delta_{\psi'} = 0$. As to the second aspect, it is clear from (7-49) that $\frac{\partial^2 \Delta^R}{\partial p^R \partial \psi'} > 0$. However, this result is predicated on the way the risk premium is paid. Satisficing providers are somewhat resistant to price increases (above $\breve{p}$) and so would be less happy with a unit price based premium than their profit maximising counterparts, unless they considered the risk premium to be part of the fair price under the non-contingent contract i.e.

$\frac{\partial \Delta^R}{\partial p^R} = v' + r\psi' - r\psi' \frac{\partial \breve{p}}{\partial p^R}$. In particular, for $\frac{\partial \breve{p}}{\partial p^R} = 1$, $\frac{\partial \Delta^R}{\partial p^R} = v' > 0$. In this case, there is no difference in $\delta^R$ between provider types.

Having considered the case where $\breve{p}$ is provider specific, it seems also reasonable to assume that $\breve{p}$ also changes on a per provider basis with regard to cream-skimming – see Lemma 6 and (7-30). With this assumption, $\Delta_{\psi'} \neq 0$. As in proposition 4 the risk implications can be ascertained from $|\phi^C| - |\phi^N|$, the difference in the absolute size of the profit change following a cost shock. Since, moreover, $\phi^C = 0$ for all states under C1 contracts, our attention can be focused on $|\phi^N|$. Using (7-24) in (7-30) gives the optimal condition:

$$(7\text{-}50)\quad U_\beta = v'\pi_\beta + \frac{\partial[r\psi]}{\partial\beta} = -rc_\beta D - r_\theta\left[\left(p(\breve{\beta}) - c(\beta)\right)D + \breve{\pi}\right] + \frac{\breve{\psi}_\beta}{v'} = 0$$

This function can be substituted into (7-46) to give an approximation for expected change in profit following a cost shock $d\beta^I$:

$$(7\text{-}51)\quad \phi_k^N = \left[-D^N c_\beta r w + (1-w)\left(-Dc_\beta r + \frac{\breve{\psi}_\beta}{v'}\right)\right] d\beta^I$$

or

$$(7\text{-}52)\quad \phi_k^N = \left[-Dc_\beta r + \frac{\breve{\psi}_\beta}{v'}(1-w)\right] d\beta^I$$

At the optimal for satisficers, cream-skimming is at a lower level in this alternative case and therefore, the term $-Dc_\beta r$ will be lower for satisficers than maximisers. However, the second term in (7-52) is higher for satisficers than maximisers. There is a balancing effect when compared to the situation with profit maximisers. Satisficers operate at higher output with a better chance of having to actually proceed with the contract after time 2 and therefore incur costs. But the disutility of cream-skimming in this case reduces the amount of profit made and therefore how much it will change following a change in referral dependency. When $w$ tends to 1, however, the former effect is greater and satisificers would find themselves experiencing greater profit variability.

The above results are collected in the following proposition.

*Proposition 6. (a) If satisficing providers suffer disutility when prices exceed the population mean, then comparing C1 contracts with NC contracts, the risk premium is the same for both satisficers and maximisers. (b) If satisficing providers suffer disutility when prices exceed the mean price given actual dependency, then when* w $\rightarrow$ *1,* $\delta^{R;C1}_{\psi'<0}(\breve{\beta}) < \delta^{R;C1}_{\psi=0}(\breve{\beta}) < 0$. *Overall, given a mix of C1 and C2 contracts, we can say that:* $\delta^{R}_{\psi'<0}(\breve{\beta}) \leq \delta^{R}_{\psi=0}(\breve{\beta}) \leq 0$.

## 7.5 Cost shocks

The above propositions regarding risk effects refer to *expected* mark-up rates evaluated before production (made at time 2), modified by a pre-determined risk premium. Thus, absent other factors, the term $\delta^R_\psi$ is the expected or average difference between the profit rates of providers operating with contingent rather than non-contingent contracts. But when we make an observation of the actual state of nature $k$ after time 2, it is probably different from that average state. Actual profits will vary depending on the contingency of the contract.

Actual net mark-up rates are state-of-the-world dependent: $\pi^i_k - y = \sigma^i + \phi^i_k$, $i = C$, $N$. We know that $\sigma^C - \sigma^N = \delta^R_\psi$ so:

$$(7\text{-}53)\quad \pi^C_k(\hat{\beta} = \beta) - \pi^N_k(\hat{\beta} = \beta) = \phi^C_k - \phi^N_k + \delta^R_\psi(\breve{\beta})$$

In the absence of cost shocks then $\pi^C_k = \pi^N_k + \delta^R_\psi$. However, suppose that a large upwards cost shock occur in state $k$ such that $d\beta^I = \beta^I - \breve{\beta} > 0$. Then we have:

*Lemma 12. For contingent contracts,* $\phi^{C1}_k - \phi^N_k > 0$.

*Proof.*

C1 contracts fully compensate a change in costs and so $\phi_K^{C1} = 0$. NC contracts do not. Therefore $\phi_K^N < 0$. C2 contracts partially compensate so will be have an intermediate effect between NC and C1 contracts. □

We can write the upwards cost shock effect as $\delta^S(\tilde{\beta}) = \phi_k^C - \phi_k^N$. For any contingent contract, $\delta^S(\tilde{\beta}) > 0$ by Lemma 12. Therefore, if positive costs shocks are sufficiently large then the actual profits of contingent contract holders could be, *ceteris paribus*, higher than non-contingent contract profits, despite the latter receiving a risk premium. In any case actual increases in payment following an upward cost shock would offset the *ex ante* risk premium.

The effects for satisficers depends on our treatment of access pricing. Satisficers do not like to increase prices above a baseline and so limit access. If the baseline is sample average dependency at the beginning of the year then following an average increase during the year, even at $\theta = 0$, price will be above access price. However, if the baseline is the average throughout the year then access price will rise proportionately and costs shock effects will be little different as between maximisers and satisficers.

## 7.6 Conclusions

### 7.6.1 Net effects

The set of theory above can be interpreted to give us a number of empirical propositions. Our main aim is to assess empirically the impact of contract choices, defined as the degree of contingency, on provider price and price-cost margins. In the theory we distinguished between *contingent* and *non-contingent* contracts. Choices have profit implications in respect to the expected size of: information rents, risk premiums and additional profits due to cost shocks. How then do these individual effects of contract choice combine to give an overall expected profit implication for contract choice? Table 7-2 summarises the individual effects, distinguishing degree of contingency ($h$). It shows the effects on profit of using, respectively, a C1 or C2 contract *compared to using a non-contingent contract.* Thus for example, regarding information rents, profit maximisers with C1 contracts are predicted to have larger rents that profit maximisers with non-contingent contracts, *ceteris paribus*. The latter do accrue information rents but – due to being unable to select optimal output – cannot accrue as much at the margin for a given degree of misrepresentation of β.

**Table 7-2. Individual effects**

| | *C1* **contracts:** $h = 1$ | | *C2* **contracts:** $h > \bar{h}$ | |
|---|---|---|---|---|
| | **Max** | **Satis** | **Max** | **Satis** |
| Information | $\delta^{I}_{\psi=0}(\breve{\beta}) > 0$ Proposition 1 | $\delta^{I}_{\psi'<0}(\breve{\beta}) < \delta^{I}_{\psi=0}(\breve{\beta})$ Proposition 2 | $\delta^{I}_{\psi=0}(\breve{\beta}) = 0$ Proposition 1 | $\delta^{I}_{\psi'<0}(\breve{\beta}) = \delta^{I}_{\psi=0}(\breve{\beta}) = 0$ Proposition 2 |
| Information - no free replacement | $\delta^{I}_{\psi=0}(\breve{\beta}) > 0$ Proposition 4 | $\delta^{I}_{\psi'<0}(\breve{\beta}) < \delta^{I}_{\psi=0}(\breve{\beta})$ Proposition 2 | $\delta^{I}_{\psi=0}(\breve{\beta}) > 0$ Proposition 4 | $\delta^{I}_{\psi'<0}(\breve{\beta}) < \delta^{I}_{\psi=0}(\breve{\beta})$ Proposition 2 |
| Risk premium | $\delta^{R}_{\psi=0}(\breve{\beta}) < 0$ Proposition 5 | $\delta^{R}_{\psi'<0}(\breve{\beta}) < \delta^{R}_{\psi=0}(\breve{\beta}) < 0$ Proposition 6 | $\delta^{R}_{\psi=0}(\breve{\beta}) < 0$ Proposition 5 | $\delta^{R}_{\psi'<0}(\breve{\beta}) \leq \delta^{R}_{\psi=0}(\breve{\beta}) \leq 0$ Proposition 6 |
| Cost shock | $\delta^{S}_{\psi}(\breve{\beta}) > 0$ Lemma 12 | | $\delta^{S}_{\psi}(\breve{\beta}) > 0$ Lemma 12 | |

The aim is develop some hypotheses as to the relationships between profit rate and contract choice overall and how these relationships differ in various circumstances. Because the individual effects change in different circumstances we can use the theory to predict which circumstances are likely to have aligned individual effects, and which will generate contradictions.

The above table does not demonstrate any completely aligned effects. However, the respective propositions do provide some, albeit limited, indication of strength of effect. Table 7.3 maps this additional information about individual effects. Moreover, the table crudely combines this information about individual effects to propose a net effect. Strictly speaking there is only a limited basis for making these comparisons since they rest on – in some cases – independent parameters e.g. the degree of risk aversion. This is essentially an empirical question. What is clear is that the availability of financial compensation for risk requires action on the part of commissioners – including recognition that this is an issue. The evidence on commissioning from the commissioning survey reported in chapter 4, suggests in practice this action is limited. recognition of risk issues. We conclude that only modest risk premiums are paid, if at all, implying a limited risk related profit difference between those with contingent and non-contingent contracts (but as a result a shortfall in utility as between providers with these different types of contract). There is more anecdotal support for the selective reporting of information, however. For example, a significant minority of people do find their condition improves, but this is generally not reported to purchasers.

### 7.6.2 Contracts with lower values of $h$ ($\vec{h} < 1$)

C1 and C2 contracts are provided for illustration. With limited information about dependency and about provider marginal cost-dependency relationships, commissioners could easily set up contingent contracts with $h$ values that are less than $h = 1$ i.e. where $p_{\hat{\beta}} > c_{\beta}$. Generally speaking the smaller the value of $h$, the higher are information rents.

Moreover, if $h < 1$ then contracts are overcompensating providers for changes in β so much so that profits do fluctuate with the state of the world, and compared to a $h = 1$ contract, risk averse providers would need a risk premium. For contract with $h$ values that are modestly below 1, this risk premium is still likely to be a lot smaller than for non-contingent contracts. But clearly as $h$ tends to zero, so information rents grow and risk premium effects would also go positive. For some value of $h$, $\vec{h} < 0$ as $h \rightarrow 0$ so $\delta^{I}_{\psi=0}(\tilde{\beta}) > 0$, $\delta^{R}_{\psi=0}(\tilde{\beta}) \geq 0$ and $\delta^{S}_{\psi}(\tilde{\beta}) > 0$. In this case providers with contracts of contingency $\vec{h} < 0$ are predicted to unambiguously have higher profits that providers with non-contingent contracts.

**Table 7-3. *Ex post* contingent contracts compared with non-contingent contracts – effects on mark-up rate differences**

| | *C1* contracts: $h = 1$ | | *C2* contracts: $h > \vec{h}$ | | *Contracts* $h < \vec{h}$ |
|---|---|---|---|---|---|
| | Max | Satis | Max | Satis | |
| Information | ++ | + | + | 0/+ | ++ |
| Risk premium | – – | – – | – | –/– – | 0 |
| Cost shock | ++ | ++ | + | + | ++ |
| **Net effect?** | ++ | + | + | 0 | +++ |

Overall, therefore the 'net effect' in the table summarises our hypotheses – used as the basis for our empirical analysis – of the association between observed profit or mark-up levels and use of contingent compared with non-contingent contracts. In particular, profit maximisers with C1 contracts are predicted to have the greatest comparative mark-up, and satisficers with C2 contracts the lowest.

## 7.7 Appendix

### 7.7.1 Appendix lemmas

*Lemma A1. A sufficient condition for* $\hat{\beta}^*$ *to be a global maximum is for* $r_{\theta\theta} \leq -1$.

*Proof.*

A sufficient condition is that $\pi_{\hat{\beta}\hat{\beta}} < 0$.

(7-54) $\pi_{\hat{\beta}\hat{\beta}} = \left(2 r_\theta D_{\hat{\beta}} + D r_{\theta\theta}\right)\left(p - c\right) + 2 p_{\hat{\beta}}\left(r D_{\hat{\beta}} + D r_\theta\right) + r_{\theta\theta}\breve{\pi}$

(using assumption 7-2)

(7-55) $\pi_{\hat{\beta}\hat{\beta}} = D r_{\theta\theta}\left(p - c\right) + 2 p_{\hat{\beta}} r D_{\hat{\beta}} + \left[2 r_\theta D_{\hat{\beta}}\left(p - c\right) + 2 p_{\hat{\beta}} D r_\theta + r_{\theta\theta}\breve{\pi}\right]$

The term in square brackets in (7-55) is negative, so our condition $\pi_{\hat{\beta}\hat{\beta}} < 0$ reduces to showing:

(7-56) $D r_{\theta\theta}\left(p - c\right) + 2 p_{\hat{\beta}} r D_{\hat{\beta}} < 0$

or

(7-57) $\frac{-r_{\theta\theta}}{r} D^2 > D_{\hat{\beta}}^2$

If $r_{\theta\theta} \leq -1$ and $r \leq 1$ then (7-57) reduces to $D > D_{\hat{\beta}}$ which holds for the class of demand functions under consideration. As $\theta$ increases from 0 so the right-hand-side of (7-57) increases, whilst the left-hand-side stay constant – thus higher values of $r_{\theta\theta}$ will be sufficient to ensure that $\pi_{\hat{\beta}\hat{\beta}} < 0$.

A quadratic of the form: $r = 1 - \theta^2$ would fit the conditions of assumption 7-2 and produce a second order derivative of –2. □

*Lemma A2.* $\beta^{*N}$ *is a global maximum.*

*Proof.*

A sufficient condition is that $\pi_{\beta\beta} < 0$. Since as $r_{\theta\theta} = -r_{\theta\beta} < 0$, we have:

$$(7\text{-}58)\quad \pi_{\beta\beta} = r_{\theta\theta}\breve{\pi} + 2r_{\theta}c_{\beta}D + (p - c)Dr_{\theta\theta} < 0$$

### 7.7.2 Proof of proposition 1

Consider the first order effect with respect to θ under each contract. For pure cream-skimming change in rent (N) from (7-24) is:

$$(7\text{-}59)\quad \pi_{\theta}^{N}(\theta) = -\pi_{\beta} = D^{N}\left[r^{N}c_{\beta} - \frac{D^{N}}{D_{p}}r_{\theta}^{N}\right] + r_{\theta}^{N}\breve{\pi}^{N}$$

From (7-10) for cost exaggeration (C) we have:

$$(7\text{-}60)\quad \pi_{\theta}^{C}(\theta) = \pi_{\hat{\beta}} = D^{C}\left[r^{C}p_{\hat{\beta}}\Big|_{D} - \frac{D^{C}}{D_{p}}r_{\theta}^{C}\right] + r_{\theta}^{C}\breve{\pi}^{C}$$

We proceed in three parts.

(a) Consider first the marginal profit at $\theta = 0$. We have from Lemma 4 $\pi_{\theta}^{N}(0) = D^{N}c_{\beta} = D(\breve{\beta})c_{\beta}$ and, from Lemma 1, $\pi_{0}^{C}(0) = D^{C}p_{\hat{\beta}}\Big|_{D} = D(\breve{\beta})p_{\hat{\beta}}\Big|_{D}$. Hence with a fully contingent contract, $\pi_{\theta}^{C}(0) = \pi_{\theta}^{N}(0)$. (Note, that using a second order expansion for $\theta = 0$, $d\pi^{C} > d\pi^{N}$ for $d\theta > 0$ as $\pi_{\theta\theta}^{C}(0) = 2D_{\hat{\beta}}p_{\hat{\beta}} + r_{\theta\theta}[D(p-c) + \breve{\pi}] = 2D_{\hat{\beta}}p_{\hat{\beta}} + \pi_{00}^{N}(0) > \pi_{\theta\theta}^{N}(0)$ bearing in mind that $D_{\beta} =$ 0).

(b) Now consider marginal profit at $\theta^{*N}$ (which is the same for maximisers and satisficers – see Lemma 6). Clearly, $\pi_{\theta}^{N}(\theta^{*N}) = 0$. Suppose that $\theta^{C} = \theta^{*N}$ and also that $D^{C} = D^{N} = D$. Then it is clear from (7-59) and (7-60) that $\pi_{\theta}^{N}(\theta^{*N}) = \pi_{\theta}^{C}(\theta^{*N})$ and that $p^{C} - c = p^{N} - c$, so that $\pi^{N} = \pi^{C}$.

However, at $\theta^{*N}$ with contingent contracts $D^C$ is not limited to a value $D^N$. Hence $\theta^{*N}$ need not be a turning point for contingent contracts. The first order condition in its full form, i.e.(7-9), at $\theta^{*N}$ and $D^N$ is:

$$\pi_\theta^C(D^N) = r\left[D^N p_{\hat\beta} + D_{\hat\beta}(p^N - c)\right] + D^N r_\theta (p^N - c) + r_\theta \breve{\pi} = r\Pi_\theta + D^N r_\theta (p^N - c) + r_\theta \breve{\pi}$$

where $\Pi = D(p-c)$ is actual profit (rather than expected profit). Note also that we have assumed that in the future the contracting regime may change, so that the size of $\breve{\pi}$ is independent of contract type. For non-contingent contracts, $\Pi_\theta = -c_\theta D^N = D^N p_{\hat\beta}\big|_D h$, in analogy with contingent contracts. However, $D^N \neq \arg\max(\Pi)$ since with $D$ free, from (7-7) the optimal demand response for a change in $\theta$ is $D_{\hat\beta} = -\frac{1}{2} p_{\hat\beta} D_p > 0$. It follows that $\Pi^N(\theta^N + \Delta\theta) - \Pi^N(\theta^N) < \Pi^C(\theta^N + \Delta\theta) - \Pi(\theta^N)$, i.e. that $\Pi_\theta^C(\theta^N) > \Pi_\theta^N(\theta^N)$, or that $\frac{\partial \Pi_\theta^C}{\partial D}(D^N(\theta^N)) > p_{\hat\beta}\big|_D$.

Now at $\theta^{*N}$,

$$\text{(7-61)} \quad \frac{\partial \pi_\theta^C(D^N(\theta^N))}{\partial D} > r\frac{\partial \Pi_\theta^C}{\partial D}(D^N(\theta^N)) + r_\theta(p^N - c) + r_\theta \frac{\breve{\pi}}{D^N}$$

This inequality arises because the right hand side does not include any price reduction due to an increase in $D$. Now at the turning point for the non-contingent contract problem, $\pi_\theta^N(D^N) = r\Pi_\theta^N + D^N r_\theta (p^N - c) + r_\theta \breve{\pi} = 0$, so,

$$r_\theta(p^N - c) = -r\frac{\Pi_\theta^N}{D^N} - r_\theta \frac{\breve{\pi}}{D^N} = -r p_{\hat\beta}\big|_D h - r_\theta \frac{\breve{\pi}}{D^N}$$

Hence for fully contingent contracts ($h$ = 1), (7-61),

$$\text{(7-62)} \quad \frac{\partial \pi_\theta^C(D^N(\theta^N))}{\partial D} > r\left(\frac{\partial \Pi_\theta^C}{\partial D}(D^N(\theta^N)) - p_{\hat\beta}\big|_D\right) - r_\theta \frac{\breve{\pi}}{D^N} + r_\theta \frac{\breve{\pi}}{D^N} > 0$$

since $\frac{\partial \Pi_{\theta}}{\partial D}\left(D^{N}\left(\theta^{N}\right)\right) > p_{\hat{\beta}}\Big|_{D}$. It follows that at $\theta^{*N}$, $\pi_{\theta}^{C}\left(\theta^{*N}\right) > 0 = \pi_{\theta}^{N}\left(\theta^{*N}\right)$.

(c) When $h = 1$, from parts (a) and (b) $\pi_{\theta}^{C}(0) > \pi_{\theta}^{N}(0)$ and $\pi_{\theta}^{C}\left(\theta^{*N}\right) > \pi_{\theta}^{N}\left(\theta^{*N}\right)$. It is then sufficient to posit that $\pi_{\theta}^{C}\left(\theta^{N}\right) > \pi_{\theta}^{N}\left(\theta^{N}\right)$ for all $\theta^{N} \in \left[0, \theta^{N*}\right]$ given our assumed continuous, differentiable functions.

For some $h > 1$, called $\bar{h}$ there will be equality between the marginal profit rates: $\pi_{\theta}^{C}\left(\theta^{*N}\right) = \pi_{\theta}^{N}\left(\theta^{*N}\right)$. This can be seen with respect to (7-61) where $\frac{\partial \Pi_{\theta}^{C}}{\partial D}\left(D^{N}\left(\theta^{N}\right)\right) - \bar{h}p_{\hat{\beta}}\Big|_{D} = 0$. At $\bar{h}$ we have $\pi_{\theta}^{C}(0) \leq \pi_{\theta}^{N}(0)$.

(d) Integrating the first order condition on the support $\theta^{N} \in \left[0, \theta^{N*}\right]$ gives:

(7-63) $\pi^{N}\left(\theta^{N*}\right) = \int_{0}^{\theta^{*N}} \pi_{\theta}^{N}(\theta)d\theta + \pi(\breve{\beta})$

and

(7-64) $\pi^{C}\left(\theta^{N*}\right) = \int_{0}^{\theta^{*N}} \pi_{\theta}^{C}(\theta)d\theta + \int_{\theta^{*N}}^{\theta^{*C}} \pi_{\theta}^{C}(\theta)d\theta + \pi(\breve{\beta})$

and noting that $\pi^{C}(0) = \pi^{N}(0) = \pi(\breve{\beta}) \equiv \pi(\beta')$. When $h = 1$ the second term in (7-64) is positive in view of $\pi_{\theta}^{C}\left(\theta^{*N}\right) > 0$ from part (c). Also, on the support $\theta^{N} \in \left[0, \theta^{N*}\right]$ we have that $\pi_{\theta}^{C}\left(\theta^{N}\right) > \pi_{\theta}^{N}\left(\theta^{N}\right)$ and so $\int_{0}^{\theta^{*N}} \pi_{\theta}^{C}(\theta)d\theta > \int_{0}^{\theta^{*N}} \pi_{\theta}^{N}(\theta)d\theta$. It then follows that $\pi^{C}\left(\theta_{\psi=0}^{C*}\right) > \pi^{N}\left(\theta_{\psi=0}^{N*}\right) = \pi^{N}\left(\theta_{\psi>0}^{N*}\right)$ as required.

To use analogous argument, when $h \geq \bar{h} > 1$, $\pi^{C}\left(\theta_{\psi=0}^{C*}\right) < \pi^{N}\left(\theta_{\psi=0}^{N*}\right) = \pi^{N}\left(\theta_{\psi>0}^{N*}\right)$ as required. □

### 7.7.3 Proof of lemma 3

Using the envelope theorem we can write: $m(h) = \partial\hat{\beta}/\partial\beta' = -U_{\hat{\beta}\beta'}/U_{\hat{\beta}\hat{\beta}}$. Since $\rho_{\hat{\beta}\hat{\beta}} = 0$, $\psi'' = 0$, the denominator

(7-65) $U_{\hat{\beta}\hat{\beta}} = v'(\pi)\pi_{\hat{\beta}\hat{\beta}} + \pi_{\hat{\beta}}v''(\pi)\pi_{\hat{\beta}} + \psi'(\rho)\rho_{\hat{\beta}\hat{\beta}} + \rho_{\hat{\beta}}\psi''(\rho)\rho_{\hat{\beta}}$

reduces to

(7-66) $U_{\hat{\beta}\hat{\beta}} = v'(\pi)\pi_{\hat{\beta}\hat{\beta}} + \pi_{\hat{\beta}}v''(\pi)\pi_{\hat{\beta}}$.

Similarly the numerator is:

(7-67) $U_{\hat{\beta}\beta'} = v'(\pi)\pi_{\hat{\beta}\beta'} + \pi_{\hat{\beta}}v''(\pi)\pi_{\beta'}$.

Using the first order condition, (7-6), on the profit function,

(7-68) $D_{\beta'} = -\dfrac{\Omega_{\beta'}}{\Omega_D} = -\dfrac{-c_{\beta'}}{\dfrac{1}{D_p}+\dfrac{1}{D_p}} = \tfrac{1}{2}c_\beta D_p$

In addition, we have from differentiating (7-9):

(7-69)

$$-\pi_{\hat{\beta}\beta'} = Dp_{\hat{\beta}}r_\theta + (p-c)D_{\hat{\beta}}r_\theta + D(p-c)r_{\theta\theta} + \chi_{\beta'}rD_{\hat{\beta}} + D\chi_{\beta'}r_\theta - D_{\beta'}rp_{\hat{\beta}} - D_{\beta'}r_\theta(p-c) + r_{\theta\theta}\bar{\pi}$$

where

(7-70) $\chi_{\beta'} = \dfrac{\partial(p-c)}{\partial\beta'} = -c_\beta + p_{\beta'} = -c_\beta + \dfrac{1}{D_p}D_{\beta'} = -\tfrac{1}{2}c_\beta$

Put another way, for profit maximisers, $p-c = -\dfrac{D}{D_p}$ and $\dfrac{\partial[-D/D_p]}{\partial\beta'} = -\tfrac{1}{2}c_\beta$. Also, note that

(7-71) $\dfrac{\partial(p-c)}{\partial\hat{\beta}} = p_{\hat{\beta}} = p_{\hat{\beta}}\big|_D + \dfrac{1}{D_p}D_{\hat{\beta}} = \tfrac{1}{2}p_{\hat{\beta}}\big|_D$.

(7-72) $\pi_{\hat{\beta}\hat{\beta}} = Dp_{\hat{\beta}}r_\theta + (p-c)D_{\hat{\beta}}r_\theta + D(p-c)r_{\theta\theta} + r_{\theta\theta}\bar{\pi} + 2rD_{\hat{\beta}}p_{\hat{\beta}} + D_{\hat{\beta}}(p-c)r_\theta + Dp_{\hat{\beta}}r_\theta$

(7-73) $\pi_{\hat{\beta}\hat{\beta}} = -\pi_{\hat{\beta}\beta'} - D\chi_{\beta'}r_\theta - \chi_{\beta'}rD_{\hat{\beta}} + 2rD_{\hat{\beta}}p_{\hat{\beta}} + D_{\hat{\beta}}(p-c)r_\theta + Dp_{\hat{\beta}}r_\theta + D_{\beta'}rp_{\hat{\beta}} + D_{\beta'}r_\theta(p-c)$

(7-74) $\pi_{\hat{\beta}\hat{\beta}} = -\pi_{\hat{\beta}\beta'} + Dr_\theta(p_{\hat{\beta}} - \chi_{\beta'}) + rp_{\hat{\beta}}(D_{\hat{\beta}} + D_{\beta'}) + rD_{\hat{\beta}}(p_{\hat{\beta}} - \chi_{\beta'}) + (p-c)r_\theta(D_{\hat{\beta}} + D_{\beta'})$

(7-75) $\pi_{\hat{\beta}\hat{\beta}} = -\pi_{\hat{\beta}\beta'} + Dr_\theta \tfrac{1}{2}\left(p_{\hat{\beta}}\Big|_D - c_\beta\right) + rp_{\hat{\beta}}(D_{\hat{\beta}} + D_{\beta'}) + rD_{\hat{\beta}}\tfrac{1}{2}\left(p_{\hat{\beta}}\Big|_D - c_\beta\right) + (p-c)r_\theta(D_{\hat{\beta}} + D_{\beta'})$

(7-76) $\pi_{\hat{\beta}\hat{\beta}} = -\pi_{\hat{\beta}\beta'} - \varphi$

For fully contingent contracts where $h = 1$, then $p_{\hat{\beta}}\Big|_D - c_\beta = 0$, and so $\pi_{\hat{\beta}\hat{\beta}} = -\pi_{\hat{\beta}\beta'}$. However, for partially contingent contracts when $h > 1$, $p_{\hat{\beta}}\Big|_D - c_\beta < 0$, the value of $\varphi$ is unclear. At $r = 1$, $\varphi > 0$ and at $r = 0$, $\varphi < 0$. When $h < 1$, $p_{\hat{\beta}}\Big|_D - c_\beta > 0$, so that at $r = 1$, $\varphi < 0$ and at $r = 0$, $\varphi > 0$.

Now,

$$(7\text{-}77)\quad \begin{aligned} U_{\hat{\beta}\hat{\beta}} &= -v'(\pi)\pi_{\hat{\beta}\beta'} - v'(\pi)\varphi + \pi_{\hat{\beta}}v''(\pi)\pi_{\hat{\beta}} \\ &= -U_{\hat{\beta}\beta'} - v'(\pi)\varphi + \pi_{\hat{\beta}}v''(\pi)(\pi_{\hat{\beta}} + \pi_{\beta'}) \end{aligned}$$

which means:

$$(7\text{-}78)\quad m(h) = \frac{-U_{\hat{\beta}\beta'}}{U_{\hat{\beta}\hat{\beta}}} = 1 + \frac{v'(\pi)\varphi(h) - \pi_{\hat{\beta}}v''(\pi)(\pi_{\hat{\beta}} + \pi_{\beta'})}{U_{\hat{\beta}\hat{\beta}}}$$

Furthermore,

$$(7\text{-}79)\quad \pi_{\beta'} = rD\frac{\partial(p-c)}{\partial\beta'} + rD_{\beta'}(p-c) - Dr_\theta(p-c) - r_\theta\bar{\pi}$$

$$(7\text{-}80)\quad \pi_{\beta'} = -rD\left(c_{\beta'} - \frac{\beta'}{D_p}\right) + rD_{\beta'}(p-c) - Dr_\theta(p-c) - r_\theta\bar{\pi}$$

Hence using (7-68),

$$(7\text{-}81)\quad \pi_{\beta'} = -\tfrac{1}{2}rDc_\beta + rD_{\beta'}(p-c) - Dr_\theta(p-c) - r_\theta\bar{\pi}$$

and so for fully contingent contracts, for any $\psi$, $\pi_{\hat{\beta}} + \pi_{\beta'} = 0$ as $p_{\hat{\beta}} = \tfrac{1}{2}p_{\hat{\beta}}\big|_D > 0$ .

Hence, for fully contingent contracts for any $\psi$, $m = 1$.

For partially contingent contracts in the profit maximising case $m = 1 + \dfrac{v'(\pi)\varphi}{U_{\hat{\beta}\hat{\beta}}}$,

because $\pi_{\hat{\beta}} = 0$. Since $U_{\hat{\beta}\hat{\beta}} < 0$ the usual second order condition of a first order maximum, we have $m = 1 + \dfrac{v'(\pi)\varphi}{U_{\hat{\beta}\hat{\beta}}} < 1$ at $r = 1$ and $m > 1$ at $r = 0$. At the optimal $r^*$, the value of $m$ is ambiguous, although its possible values do include $m = 1$. □

### 7.7.4 Proof of lemma 7

Noting that we can write $\theta = \beta' - \beta \equiv \hat{\beta} - \beta$, differentiating (7-24) gives:

$$(7\text{-}82)\quad \pi_{\beta\beta} = Dr_\theta c_\beta + D(p-c)r_{\theta\theta} + r_{\theta\theta}\bar{\pi} + Dr_\theta c_\beta < 0$$

and

$$(7\text{-}83)\quad \pi_{\beta\beta'} = -Dr_\theta c_\beta - D(p-c)r_{\theta\theta} - r_{\theta\theta}\bar{\pi} - Dr_\theta p_{\beta'} - D_{\beta'}rc_\beta - D_{\beta'}(p-c)r_\theta$$

However, since prices are fixed at the sample average dependency then $p_{\beta'} = 0$. Moreover, optimal demand is a function of price and actual marginal cost i.e. $D^* = D(p(\hat{\beta}), c(\beta))$, and the former is fixed. Actual marginal costs are a function of the actual level, not the referred level, of dependency (where these might differ). Therefore, $D_{\beta'} = 0$ and (7-83) then becomes:

(7-84) $\pi_{\beta\beta'} = -Dr_\theta c_\beta - D(p-c)r_{0\theta} - r_{\theta\theta}\breve{\pi}$

We then have:

(7-85) $\pi_{\beta\beta'} = -\pi_{\beta\beta} + Dr_\theta c_\beta$

Twice differentiating utility gives:

(7-86) $U_{\beta\beta'} = \pi_\beta v''(\pi)\pi_{\beta'} + v'(\pi)\pi_{\beta\beta'} = v'(\pi)\pi_{\beta\beta'}$

as at the optimal $\pi_\beta = 0$ for any $\psi$ by Lemma 6. Similarly,

(7-87) $U_{\beta\beta} = v'(\pi)\pi_{\beta\beta}$

Therefore, with analogy to the proof for lemma 3, part 1 of this proof is:

(7-88) $w = -\dfrac{U_{\beta\beta'}}{U_{\beta\beta}} = 1 - \dfrac{v'(\pi)r_0 c_\beta D}{v'(\pi)\pi_{\beta\beta}} = 1 - \dfrac{r_\theta c_\beta D}{\pi_{\beta\beta}} \le 1$

Furthermore, $\dfrac{r_\theta c_\beta D}{\pi_{\beta\beta}} < 1$ or $r_0 c_\beta D - \pi_{\beta\beta} > 0$ because $r_\theta c_\beta D - \pi_{\beta\beta} = \pi_{\beta\beta'} > 0$.

Part 2 is as follows. To begin with write: $\dfrac{\pi_{\beta\beta}}{r_\theta c_\beta D} = 2 + \dfrac{r_{00}D(p-c) + r_{00}\breve{\pi}}{r_\theta c_\beta D}$ (where $D = D^N(\breve{\beta})$). For $\beta' = \breve{\beta}$, $\breve{\pi} = \dfrac{D(p-c)}{1-r+b}$. Thus $\dfrac{\pi_{\beta\beta}}{r_0 c_\beta D} = 2 + \dfrac{r_{00}(p-c)\left(\dfrac{2-r+b}{1-r+b}\right)}{r_0 c_\beta}$. Now at the optimal value of $\beta$, $\pi_\beta = 0$, implying $c_\beta = -\dfrac{r_\theta}{r}(p-c)\left(\dfrac{2-r+b}{1-r+b}\right)$ from (7-24). Therefore, $\dfrac{\pi_{\beta\beta}}{r_0 c_\beta D} = 2 - \dfrac{r_{\theta\theta}r}{r_\theta^2}$ or $w = \dfrac{r_0^2 - r_{\theta\theta}r}{2r_\theta^2 - r_{\theta\theta}r}$. When $r = 0$, then $w = ½$. When $r = 1$, then $r_0 = 0$ and so $w = 1$. Hence $½ < w \le 1$ as required. □

# Chapter 8. Empirical analysis of contract effects

## 8.1 Introduction

This chapter describes the empirical investigation of the main propositions of the previous chapter i.e. the proposition summarised in table 7.3. The aim is to determine the relationship between contract types used by providers – that is the degree of contract contingency – and both prices and price-cost margins (profit rates), controlling for other relevant factors. The literature describes the options and challenges of empirical investigations of profitability (a particularly useful overview is Bresnahan, 1988). This chapter describes how the chosen approach – a residual demand elasticities method – is applied. It draws on Forder (2000). The chapter then discusses the derivation of the various profit 'mark-up' indicators. The results follow, and then conclusions are presented about how well the data square with the theory.

## 8.2 Empirical Foundations

A general imperfect competition model with product differentiation (Bresnahan, 1988; Baker and Bresnahan, 1985) of the following type is used. Consider a market with $n$ heterogeneous providers of residential care. To be clear about market power at the provider level we start with the case where each provider supplies one service. Later we develop the model to allow providers to sell more than one product.

Demand at period 2 for service 1 is given by (1.1), which in inverse form is:

(8-1) $$p_1 = h_1(x_1, x_2, ..., x_n, y_1, y; \sigma_1)$$

where, $p_1$ is the service 1 demand price, $x_1$ to $x_n$ are the (expected) demands faced by the $n$ service/providers in the market, $y_i$ is a vector of exogenous variables that shift demand – and would include types of contract $Y$ used – and $y$ are non-firm specific (e.g. buyer) shift factors. Finally, $\sigma_1$ is the parameter vector. The usual assumption that $\frac{\partial x_1}{\partial p_1} < 0$ and

$\frac{\partial x_1}{\partial p_j} > 0$ for $j = 2 \ldots n$ is made. Buyers are distributed throughout the market according to their tastes for each care product.

The partial derivative, $\partial h_1/\partial x_1$, indicates how provider 1's demand changes when the prices of all other providers are kept constant – the Nash-in-quantity case or Cournot case (see Martin, 2002). However, to allow a more general specification of provider conjectures we can be explicit about the dependence of the other providers' demands on provider 1's demand:

$$(8\text{-}2) \quad p_1 = h_1(x_1, x_2(x_1), \ldots, x_n(x_1), y_1, y; \sigma_1)$$

Allowing $x_2, \ldots, x_n$ to vary, the *total* derivative of provider 1's price with regard its demand (excluding changes in the exogenous variables) is:

$$(8\text{-}3) \quad \frac{dp_1}{dx_1} = \frac{\partial p_1}{\partial x_1} + \sum_{j=2}^{n} \frac{\partial p_1}{\partial x_j}\frac{\partial x_j}{\partial x_1} = \frac{\partial p_1}{\partial x_1} + \sum_{j=2}^{n} \frac{\partial p_1}{\partial x_j}\lambda_{j1}$$

where $\lambda_{j1}$ is the conjectural variation. In the Bertrand (price competition) case each provider expects its rival's price to remain constant for changes in its own price. Then, with $p_j(x_1) = const$, we have: $\lambda_{j1} = -\frac{\partial p_j/\partial x_1}{\partial p_j/\partial x_j}$. Provider 1's *residual* price differential – the relationship between its price and demand net of responses of other providers – is dependent on that provider's conjectures about the other providers. This conjecture is therefore part of the provider's optimal price function. Indeed, it is the residual price differential that we wish to determine because such an estimate would encompass the effects of competitive interaction as it prevails in the sample market. It tells us about provider market power, which we turn to next.

To proceed in the derivation of the provider's optimal price function we need to be explicit about provider objectives and cost functions. We allow for marginal costs to change, albeit by relatively small amounts because both the capital (mainly property) and labour inputs are quite divisible for the industry. We write the marginal cost function as:

$$(8\text{-}4) \quad c_i = c(x_i, z_i, z; \theta_i)$$

or more specifically

$$(8\text{-}5) \quad c_i = \bar{c}(z_i, z; \theta_i) + c_x x_i$$

where $z$ is a vector of cost shift elements including factor prices such as the price of labour and the price of capital. It also includes the choice of contract, which can affect the cost function if it induces cream-skimming and so changes the balance of client types served by the provider. Also, $z_i$ are provider specific factors such as home characteristics. The elements of the vector $\theta_i$ are the unknown parameters of the cost function. Total costs are given by the integral of (8-5):

$$(8\text{-}6) \quad C_i = \int_0^{x} c_i dx_i = \int_0^{x} \left[\bar{c}(z_i, z; \theta_i) + c_x x_i\right] dx_i = \left[\bar{c} x_i + \tfrac{1}{2} c_x x_i^2\right]_0^{x} = \bar{c} x_i + \tfrac{1}{2} c_x x_i^2 + F_i$$

Average costs are thus:

$$(8\text{-}7) \quad \frac{C_i}{x_i} = \bar{c} + \tfrac{1}{2} c_x x_i + \frac{F_i}{x_i} = c - \tfrac{1}{2} c_x x_i + \frac{F_i}{x_i}$$

and average variable costs are:

$$(8\text{-}8) \quad \frac{C_i^V}{x_i} = \bar{c} + \tfrac{1}{2} c_x x_i = c - \tfrac{1}{2} c_x x_i$$

We do not have information on the specific incentives/motivations of individual organisations, that is, whether particular providers are profit maximisers or satisficers, and if the latter, what value of $\psi'$ applies. We will assume that provider types are defined by two limit cases. The first and most straightforward is the profit maximiser, i.e. $\psi' = 0$.

Profit maximisers have the usual period 2 profit function (at time $t$ with time subscripts suppressed) (see (1.5)):

$$(8\text{-}9) \quad \pi_{it} = p_i(x_i)x_i - C_i(x_i)$$

The first order condition is:

$$(8\text{-}10) \quad \frac{\partial \pi_i}{\partial x_i} = p_i - c_i + \left( \frac{\partial p_i}{\partial x_i} + \sum_{j \neq i}^{n-1} \frac{\partial p_i}{\partial x_j} \lambda_{ji} \right) x_i(p, x_{-i}, y_i, y; \sigma_i) = 0$$

or with more manageable notation,

$$(8\text{-}11) \quad p_i = c_i - p_x x_i$$

where $p$ is the vector of all providers prices. The other limit case is the satisficer with the constraint that providers break-even, that is prices at least equal average variable cost. We can assume that provider types are uniformly distributed between these limit cases, such that on average price setting is a weighted sum of these extremes. With a weighting factor $\alpha(\psi')$, the average price - cost relation is:

$$(8\text{-}12) \quad p_i = \alpha(c_i - p_x x_i) + (1-\alpha)\frac{C_i^V}{x_i}$$

$$(8\text{-}13) \quad \begin{aligned} p_i &= \alpha(\bar{c} + c_x x_i - p_x x_i) + (1-\alpha)(\bar{c} + \tfrac{1}{2} c_x x_i) \\ &= (\tfrac{1}{2}\alpha c_x + \tfrac{1}{2} c_x - \alpha p_x)x_i + \bar{c} \end{aligned}$$

The weighting or *motivations* parameter, $0 \leq \alpha \leq 1$, measures the average propensity of providers to use market power to make profit. It is assumed to be exogenously determined, and not directly observable.

Overall then (8-13) indicates that a provider's (expected) price setting will depend on the extent of product differentiation and competition as they act on provider own price elasticity, on providers' conjectures regarding price responses of competitors (i.e. the conjectural variations parameter, $\lambda_{ji}$), and also on the value of motivations parameter ($\alpha$). The latter implies that this measure is of *actual* rather than *potential* price-cost margins (see Forder, 2000).

### 8.2.1 Estimating price-cost margins

Three estimates are required to calculate price-cost margins as specified by (8-13): the 'perceived' price differential $p_x \equiv dp_i / dx_i$ ; the change in marginal costs; and the motivations parameter. Undertaking a direct estimation of $dp_i/dx_i$ i.e. (8-3) presents considerable practical difficulties because the number of own- and cross-elasticities is very large. Baker and Bresnahan (1985) offer a more manageable approach that involves removing the dependence of the other $n-1$ provider's prices on the demand function of a representative provider.

Consider the pricing decisions of a given provider, say provider 1. The other firms in the market are denoted $j = 2,..,n$ and have demand functions:

$$\text{(8-14)} \quad p_{-i} = h_{-i}(x_1, x_{-i}, y_{-i}, y; \sigma_{-i})$$

where the subscript $-i$ refers to the vector of providers $j$. The associated first order conditions are given by (8-13), or in this case:

$$\text{(8-15)} \quad p_{-i} = c_{-i} - \alpha\mu_{-i}x_{-i}$$

Using the $(n-1)$ inverse demand equations and the $(n-1)$ supply equations in the respective vector relations (8-14) and (8-15) we can solve simultaneously for $(n-1)$ prices and outputs (as functions of market level shift factors $y$ and $z$) to derive:

$$\text{(8-16)} \quad x_{-i} = E_{-i}(x_1, z, y; \alpha, \lambda_{-i}, \sigma_{-i}, \theta_{-i})$$

Following Baker and Bresnahan's (1985) formulation, for each of the $(n-1)$ providers $j$, the differential of $E_j$ with respect to $x_1$ is firm $j$'s reaction function (to provider 1). Equation (8-16) defines a reduced form equation (for $i \neq 1$) written in terms of provider 1's output ($x_1$). These optimal demands can then be substituted into the inverse demand function to give the equation to be estimated:

$$(8\text{-}17)\quad p_1 = h_1\left(x_1, E_{-i}\left(x_1, z, y; \alpha, \lambda_{-i}, \sigma_{-i}, \theta_{-i}\right), y, y_1; \sigma_1\right)$$

or

$$(8\text{-}18)\quad p_1 = h_1^R\left(x_1, z, y, y_1; \alpha, \lambda, \sigma_1, \sigma, \theta\right)$$

This function is the residual demand curve for product 1. The elasticity of $h^R$ with respect to $x_1$ tells us about the market power provider 1 has over its price, taking into account the adjustment of all other providers' prices and quantities (Bresnahan, 1988, p1049). Specifically,

$$(8\text{-}19)\quad \frac{dp_i}{dx_1} \equiv \frac{\partial h_1^R}{\partial x_1}$$

which can be substituted directly into the optimal price function (8-13) and is a key component in determining price-cost margins. It remains to show how $\partial h_1^R / \partial x_1$ can be estimated. The stochastic form of (8-18) for provider $i$ is (chosen to balance flexibility with parsimony):

$$(8\text{-}20)\quad p_i = b_0 + b_{11}\log x_i + b_{12}x_i^2 + b_2 y + b_3 y_i + b_4 z + u_i$$

Natural logarithms are used to allow for a more general functional form (see chapter 6 regarding the implications). To allow for the impact of (expected) demand on price-cost margins (i.e. $\partial M_i / \partial x_i$), a more general specification is:

$$(8\text{-}21)\quad p_i = b_0 + b_{11}\log x_i + b_{12}x_i^2 + b_{13}Yx_i^2 + b_2 y + b_3 g_i + b_4 z + b_5 Y_i + u_i$$

which uses an interaction term with coefficient $b_{12}$, and extracts contract type $Y_i$ from vector $y_i$ to leave $g_i$. Therefore:

$$(8\text{-}22)\quad \partial p_i^D / \partial x_i \equiv h_{x_i}^R = \frac{b_{11}}{x_i} + 2b_{12}x_i + 2b_{12}\bar{Y}x_i$$

Equilibrium price for provider $i$ is given by the interaction of (residual) demand (8-21) and the supply relation (8-13):

$$(8\text{-}23)\quad p_i = \bar{c}(z_i, z; \theta_i) + \left(\tfrac{1}{2}\alpha c_x + \tfrac{1}{2}c_x - \alpha h_{x_i}^R(x_i, y, z; \sigma_i, \sigma, \alpha, \lambda, \theta)\right)x_i(p_i, x_{-i}, y, y_i; \sigma_i)$$

The stochastic counterpart is,

$$(8\text{-}24)\quad p_i = a_0 + a_1x_i + a_2z_i + a_3z + e_i = a_0 + a_{10}x_i + a_{11}x_iY_i + a_2z_i + a_3z + e_i$$

and so the differential of the supply price is:

$$(8\text{-}25)\quad \partial p_i^S / \partial x_i \equiv a_1 = \tfrac{1}{2}\alpha c_x + \tfrac{1}{2}c_x - \alpha h_{x_i}^R$$

Estimating (8-21) presents a simultaneity problem because it is simultaneously determined by (8-13). Nonetheless, those factors in $z_i$ that are *cost-only* can be used to instrument $x_i$ in (8-21). Since conceivably some variables such as those describing a home's characteristics could enter both $y_i$ and $z_i$, those variables would be unsuitable as instruments. Therefore only the subset of factors $\breve{z}_i \subset z_i : \breve{z}_i \not\subset y_i$ are used to instrument demand. Similarly, when estimating (8-24) provider specific demand factors $\breve{y}_i \subset y_i : \breve{y}_i \not\subset z_i$ can be used to instrument $x_i$. The interaction is included to allow for different values of $h_x(Y)$ for each group $Y$. To account for possible shift effects supply (8-24) includes contract type in the vector $z$, i.e. $Y_i \in z$. Note also that since the vector $z_i$ contains dependency variables $\beta$ we would not expect much of an effect of $Y_i$ alone because, as noted in the previous chapter, cream-skimming works through the reduction of $\beta$, i.e. $c_i = c(x_i, z_i^{-\beta}, \beta(Y), z; \theta_i)$.

To be explicit about the instrumentation in the model suppose that factors common to $y_i$ and $z_i$ form the vector $H_i \subset y_i : y_i \cap z_i$. Then we can write:

$$(8\text{-}26) \quad p_i = \bar{c}(\breve{z}_i, H_i, z) + \left(\tfrac{1}{2}\alpha c_x + \tfrac{1}{2}c_x - \alpha h^R_{x_i}\right)x_i(\breve{y}_i, H_i, y)$$

The above models provide estimates with which to calculate both potential and actual mark-up rates and Lerner indices.

#### 8.2.1.1 Marginal cost mark-up

Re-arranging (8-11) gives the potential marginal cost mark-up rate – that is, the rate that maximises profits.

$$(8\text{-}27) \quad p_i^{\alpha=1} - c_i = M_i = -h^R_{x_i} x_i$$

Actual mark-up rates are provided by estimation of the supply relation and are:

$$(8\text{-}28) \quad p_i^{\alpha} - c_i = m_i = \left(\tfrac{1}{2}\alpha c_x + \tfrac{1}{2}c_x - \alpha h^R_{x_i}\right)x_i - c_x x_i = a_1 x_i - c_x x_i$$

#### 8.2.1.2 Average cost mark-up

Since providers operate with economies of scale mark-up based on average variable cost is a better guide to profitability. Using (8-8), potential average cost mark-up is:

$$(8\text{-}29) \quad p_i^{\alpha=1} - \frac{C_i^V}{x_i} = \Phi_i = p_i^{\alpha=1} - c + \tfrac{1}{2}c_x x_i = M_i + \tfrac{1}{2}c_x x_i$$

and an actual average cost mark-up of:

$$(8\text{-}30) \quad p_i^{\alpha} - \frac{C_i^V}{x_i} = \phi_i = p_i^{\alpha} - c + \tfrac{1}{2}c_x x_i = m_i + \tfrac{1}{2}c_x x_i = a_1 x_i - \tfrac{1}{2}c_x x_i$$

### 8.2.2 Actual mark-up calculation

The calculation of actual mark-up – either (8-28) or (8-30) – requires us to have estimates of $a_1$ and $c_x$. The former is directly estimated (8-24) but the latter is not and need to be inferred from using the supply price condition $\frac{1}{2}\alpha c_x + \frac{1}{2}c_x - \alpha h_{x_i}^R = a_1$. The estimation of the demand equation directly provides a value for $h_{x_i}^R$. However, this still leaves the unknown value of $\alpha$ to be estimated.

The predictions from the demand and supply equations regarding mark-up rates under the two contract types are made given equal values of he other variables, $z$ and $y$. Therefore, costs $\bar{c}(z)$ and the derivative $c_x(z)$, which are functions of $z$ will be equal for both groups. Note in this comparison we are assuming that both groups of providers treat the same types of clients. Cream-skimming effects on the price charged by the non-contingent contracts group will work through demand, in that purchasers believe they are getting a service for higher dependency people than is actually the case. Suppose that for both groups mean dependency of referred clients was the same. Where cream-skimming was occurring, the actual average dependency in the non-contingent contract groups would be lower (hence costs) although demand prices would still be at the sample average level. By comparing supply prices where the dependency of residents is the same between the two groups implies the same costs but predicted prices for the non-contingent contract group that are above the sample average dependency level to reflect the extra profit associated with cream-skimming. The relevant point here is that when the two groups treat the same types of people their costs must be the same; cream-skimming works not by lowering costs given dependency but by lowering dependency – and so costs – given price.

We can estimate the supply price differential of (8-24) for each sub-population by contract type – i.e. being explicit about the dependence on $Y$ – as: $a_1^Y = \frac{1}{2}\alpha c_x + \frac{1}{2}c_x - \alpha h_{x_i}^Y$. It follows from the above (i.e. $c_x$ and $\alpha$ are not functions of $Y$) that $a_1^{Y=1} + \alpha h_{x_i}^{Y=1} = \frac{1}{2}\alpha c_x + \frac{1}{2}c_x = a_1^{Y=0} + \alpha h_{x_i}^{Y=0}$ (also noting that $c_{xx} = 0$ by assumption) and therefore, that:

$$(8\text{-}31)\quad \alpha = \frac{a_1^{Y=1} - a_1^{Y=0}}{h_{x_i}^{Y=0} - h_{x_i}^{Y=1}}$$

and

$$(8\text{-}32)\quad c_x = \frac{a_1^Y - \alpha h_{x_i}^Y}{\frac{1}{2}\alpha + \frac{1}{2}}$$

In addition to these estimates, to calculate mark-up rates by contract type we need also to determine output level and prices for the two groups. We can solve for equilibrium (equal) $p^Y$ in the non-stochastic demand (8-21) and supply (8-24):

$$(8\text{-}33)\quad b_0 + b_{11}\log x_i + b_{12}x_i^2 + b_{13}Yx_i^2 + b_2 y + b_3 g_i + b_4 z + b_5 Y_i = a_0 + a_{10}x_i + a_{11}x_i Y_i + a_2 z_i + a_3 z$$

for each $Y$. Solving the above equation for output and substituting back into the demand curve gives a reduced-form equation for each provider.

$$(8\text{-}34)\quad p_i = h_i^F\left(z, z_i, y, y_i; \alpha, \lambda, \sigma_i, \sigma, \theta_i, \theta\right)$$

In the reduced-form model we have:

$$(8\text{-}35)\quad p_i^{Y=1} - p_i^{Y=0} = \frac{\partial h_i^F}{\partial Y} = s$$

However, the difference in mark-up rates is:

$$(8\text{-}36)\quad \left(p_i^{Y=1} - c_i^{Y=1}\right) - \left(p_i^{Y=0} - c_i^{Y-0}\right) = s + c_i\left(x_i^{Y=0}, z\right) - c_i\left(x_i^{Y-1}, z\right) \approx s + c_x\left(x_i^{Y=0} - x_i^{Y=1}\right)$$

Only therefore if providers are operating with constant returns to scale (i.e $c_x = 0$) will $s$ be equal to the change in mark-up. Whilst the reduced-form is very useful in providing a way of estimating the motivation paramater with the marginal cost differential, the foregoing analysis does serve to highlight the limitations of reduced-form analysis for gauging the impact of contract choice on profits. The likely error is compounded with the size of economies or diseconomies of scale.

The advantage of a reduced-form model is that it avoids the need for identification of the endogenous variables in the structural equations. This requirement can be problematic in that the estimation can be quite sensitive to the specification of the instruments used. Moreover, instrument specification tests can support quite a wide range of specifications. As a result, some experimentation in required.

### 8.2.3 The impact of contract choice

The impact of contract choices on actual mark-up can now be calculated. At the margin, profit difference is

$$(8\text{-}37)\ \Delta m_i = \left(\tfrac{1}{2}\alpha c_x - \tfrac{1}{2}c_x - \alpha h_{x_i}^{Y=1}\right)x_i^{Y=1} - \left(\tfrac{1}{2}\alpha c_x - \tfrac{1}{2}c_x - \alpha h_{x_i}^{Y=0}\right)x_i^{Y=0}$$

According to the theory, we would expect this difference to be positive; providers with contingent contracts derive more information rent from cost exaggerations than those with non-contingent contracts from cream-skimming. The theory indicates that with replacement costs cream-skimming (CS) is unlikely to be large-scale.

The question of the size of CS rents is an empirical one. With cream-skimming actual dependency is less than average dependency of all referred users: $z^0 < \breve{z}$ where the superscript refers to contract contingency type. Apropos (8-26) the corresponding supply relationship will be:

$$(8\text{-}38)\quad p^0(z^0) = \bar{c}(\breve{z}) + a_1^0(x^0)x^0(z^0)$$

or

$$(8\text{-}39)\quad p^0(z^0) = \bar{c}(z^0) + \varepsilon + a_1^0(x^0)x^0(z^0)$$

with, by definition of CS, $\varepsilon < 0$. The problem is that we observe actual dependency $z^0$ not the 'pre-selection' or referred dependency $\breve{z}$. For the contingent contracts group, were cost exaggeration (CE) is observed, actual dependency (but not reported dependency) is equal to referred dependency levels on average (proposition 1 of chapter 7 rules out simultaneous CE and CS). In practice, there are empirical issues in using the contingent contracts group as a benchmark for determining $\breve{z}$ since that group may have other exogenous characteristics, $y$ and $z$, that affect, by chance, the mean referral dependency of users in that group. Below this problem is addressed by estimating the relationship between dependency and characteristics, $y$ and $z$, and then cross-predicting dependency characteristics of contingent contract providers, given those characteristics. Then a like-for-like comparison can be made – see section 8.7.4.

## 8.3 Estimation

### 8.3.1 Specification

The PSSRU cross-sectional survey of residential care – as described in chapter 4 – is the main data source for price-cost margins estimations. The two structural equations to be estimated are (8-21) and (8-24) respectively. The residual demand function is:

$$(8\text{-}40)\quad \begin{aligned} p_i &= b_0 + b_{11}\log x_i + b_{12}x_i^2 + b_{13}Yx_i^2 + b_2 y + b_3 g_i + b_4 z + b_5 Y_i + u_i \\ &\equiv b_0 + b_{11}\log x_i + b_{12}x_i^2 + b_{13}Yx_i^2 + b_2 y + b_{31}\breve{y}_i + b_{32}H_i + b_4 z + b_5 Y_i + u_i \end{aligned}$$

and the supply function is:

$$(8\text{-}41)\quad \begin{aligned} p_i &= a_0 + a_{10}x_i + a_{11}x_i Y_i + a_2 z_i + a_3 z + e_i \\ &\equiv a_0 + a_{10}x_i + a_{11}x_i Y_i + a_{21}\breve{z}_i + a_{22}H_i + a_3 z + a_4 Y_i + e_i \end{aligned}$$

Finally, the demand function on the sample including public sector providers is:

$$(8\text{-}42)\quad p_i = \hat{b}_0 + \hat{b}_{11}\log x_i + \hat{b}_{12}Yx_i^2 + \hat{b}_2 y + \hat{b}_{31}\breve{y}_i + \hat{b}_{32}H_i + \hat{b}_4 z + \hat{b}_5 Y_i + \hat{b}_6 L_i + u_i$$

This function was used primarily to gauge the existence of cream-skimming.

Empirical proxy variables for the demand and supply functions are given in Table 8-1 and descriptive statistics are given in Table 8-2. Output, $x_i$, is specified as the number of places sold in the home (number of places filled). Output information is supplied for the survey date by the home manager. On average 30.6 places were purchased with a range from 3 to 159. Capacity – i.e. the number of available places – averaged 34.4 for an occupancy of 88 per cent.

**Table 8-1. Empirical specification**

| Variable definition | | Demand | Supply | Theory var. |
|---|---|---|---|---|
| *Output* | | | | |
| Output (number of filled places) | Lin | | Endog | $x_i$ |
| | Log | Endog | | $\ln x_i$ |
| | Sqrd | Endog | | $x_i^2$ |
| Output × contingent contract | Lin | | Endog | $Y_i x_i$ |
| | Sqrd | Endog | | $Y_i x_i^2$ |
| Output, lagged | Lin | Instr | | $x_i^{t-1}$ |
| | Sqrd | | Instr | $\left(x_i^{t-1}\right)^2$ |
| | Log | | Instr | $\ln\left(x_i^{t-1}\right)$ |
| *Demand and supply shift factors* | | | | |
| Barthel score (inverse) | Sqrd | Struc | | $H$ |
| | Cbd | | Struc | $H$ |
| Staff with nursing qualification rate | Lin | | Struc | $H$ |
| | Log | Struc | | $H$ |
| Home has no en suite toilets | Dummy | Struc | Struc | $H$ |
| Home size/capacity | Lin | Struc | | $H$ |
| Home has modular layout | Lin | Struc | | $H$ |
| Home has modular layout × size/capacity | Lin | Struc | | $H$ |
| Home purpose built | Dummy | | Struc | $H$ |
| Business started from scratch w/ purpose built home | Dummy | | Struc | $H$ |
| Nursing home | Dummy | Struc | Struc | $H$ |
| Single home organisation | Dummy | | Struc | $H$ |
| Home makes specific EMI provision | Dummy | | Struc | $H$ |
| *Demand only factors* | | | | |
| Resident Funding (% residents privately funded) | Lin | | Instr | $\breve{y}_i$ |
| | Sqrd | Struc | Instr | $\breve{y}_i$ |
| Resident Funding (% residents LA funded) | Lin | | Instr | $\breve{y}_i$ |
| Home flexible regime score | Lin | Struc | Instr | $\breve{y}_i$ |
| Home self expression score | Lin | Struc | Instr | $\breve{y}_i$ |
| Home has no contracts with other LAs | Dummy | | Instr | $\breve{y}_i$ |
| Home has no contracts with other LAs × LA list only | Dummy | Struc | Instr | $\breve{y}_i$ |
| Home is planning to expand | Dummy | | Instr | $\breve{y}_i$ |
| Volunteers aid in organised activity | Dummy | | Instr | $\breve{y}_i$ |
| *Supply only factors* | | | | |
| Provides meals on wheels | Dummy | Instr | | $\breve{z}_i$ |
| Provides meals on wheels × Provides home care | Dummy | Instr | | $\breve{z}_i$ |
| Provides home care | Dummy | Instr | | $\breve{z}_i$ |
| Price fails to cover cost | Dummy | Instr | | $\breve{z}_i$ |
| Price fails to cover cost × Private sector home | Dummy | Instr | | $\breve{z}_i$ |
| Home wage rate (basic), care staff | Sqrd | Instr | | $\breve{z}_i$ |
| *Local demand, supply chars (not home specific)* | | | | |
| Wages: female, manual gross wage | Lin | Struc | | $z$ |
| | Log | | Struc | $z$ |
| Property prices | Lin | | Struc | $z,y$ |
| Area cost adjustment | Lin | Struc | | $z,y$ |
| *Contracts* | | | | |
| Contingent price contract | Dummy | Struc | Struc | $Y_i$ |
| Contingent price contract × qualified staff ratio | Dummy | Struc | | $Y_i d$ |

**Table 8-2. Descriptive statistics**

| Variable definition | | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Price | Linear | 276.07 | 69.01 | 139 | 705.26 |
| Output: number of filled places | Linear | 30.56 | 19.74 | 3 | 159 |
| | Log | 3.25 | 0.58 | 1.10 | 5.07 |
| *Resident characteristics* | | | | | |
| Cognition: | % moderate impairment | 44.66 | 20.27 | 0 | 100 |
| Barthel score home average, inverse | | 8.86 | 3.92 | 0.45 | 18.20 |
| Resident Funding | % residents privately funded | 30.31 | 23.37 | 0 | 100 |
| | % residents LA funded | 35.96 | 27.82 | 0 | 100 |
| *Home characteristics* | | | | | |
| Nursing home | | 0.37 | 0.48 | 0 | 1 |
| Staff with nursing qualification rate | Linear | 0.13 | 0.14 | 0 | 0.63 |
| | Log | 0.11 | 0.12 | 0 | 0.49 |
| Purpose built home | | 0.34 | 0.47 | 0 | 1 |
| Business started from scratch | | 0.48 | 0.50 | 0 | 1 |
| % of single rooms to total beds | Linear | 0.72 | 0.24 | 0 | 1.06 |
| Single home organisation | | 0.51 | 0.50 | 0 | 1 |
| Home has no en suite toilets | | 0.38 | 0.49 | 0 | 1 |
| No. care staff | Dummy: over 40 | 0.15 | 0.36 | 0 | 1 |
| | Dummy: less 15 | 0.20 | 0.40 | 0 | 1 |
| | Number | 27.19 | 19.15 | 2 | 170 |
| Home has own transport | | 0.19 | 0.39 | 0 | 1 |
| Flexible home regime | | 28.33 | 14.33 | 0 | 68.25 |
| Volunteers aid in organised activity | | 0.13 | 0.34 | 0 | 1 |
| Home is only on the approved list of its LA | | 0.32 | 0.47 | 0 | 1 |
| Home has no contracts with other LAs | | 0.19 | 0.39 | 0 | 1 |
| Home is planning to expand | | 0.11 | 0.31 | 0 | 1 |
| Provides meals on wheels | | 0.11 | 0.32 | 0 | 1 |
| Provides home care to non-residents | | 0.09 | 0.29 | 0 | 1 |
| Price fails to cover cost | | 0.56 | 0.50 | 0 | 1 |
| *LA characteristics* | | | | | |
| Wages: female, manual gross wage, LA | linear | 386.80 | 35.45 | 339.60 | 512.80 |
| | squared | 150865.80 | 28222.49 | 115328.20 | 262963.80 |
| *Property prices* | | 89468.99 | 26322.83 | 52613 | 169766 |
| *Regional characteristics* | | | | | |
| Prosperity index (ACA), not London | | 0.90 | 0.32 | 0 | 1.10 |
| London *LA dummy* | | 0.62 | 0.49 | 0 | 1 |
| *Contracts* | | | | | |
| Contingent price contract | Any home | 0.38 | 0.49 | 0 | 1 |

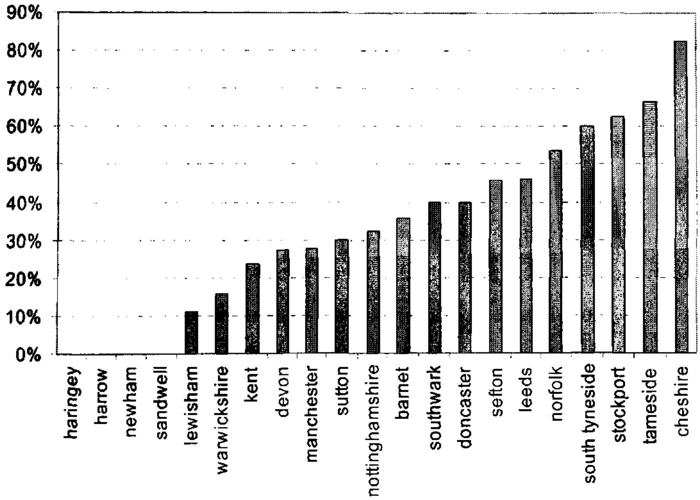
Managers were also asked about the types of contracts they had with purchasers from a number of options that covered both contingency with respect to quantity (i.e. block contracts) and also contingency relative to resident (needs-related) characteristics. Of the former, a very small proportion of providers indicated having quantity contingent contracts (over 95% of providers in the sample had only case-by-case – spot – contracts), an insufficient number for the statistical analysis. As regards the latter the indicator was a response to the question of whether they operated with a price that was "pre-set by local authorities dependent on the type of resident (e.g. level of dependency)". Some 38 per cent of providers indicated that they did have contracts that made allowance for the service needs in this way. It is important to underline that this variable is based on provider *perceptions* of the form of the contracts they had with councils. For example, in some cases the intake range of resident dependency of a home may be sufficiently small that even if the local authority operated with a limited tariff of prices related to dependency, that the intake range only falls into one category. Indeed, as noted in chapter 4, a legacy of the social security voucher system was that many authorities were operating only two prices for older people – a standard and an enhanced rate – but that the enhanced rate only applied to people with very significant co-morbidities e.g. not only physically frail, but also with substantial cognitive impairment. Furthermore, councils may not have a uniform policy for contract types so that contract contingency can vary by provider intra-LA. Figure 8-1 shows that in most local authorities a mix of contract contingency was reported by providers. Only for Haringey, Harrow, Newham and Sandwell did no provider report a contingent contract.[55]

Ideally, we would like to use an indicator of contingency that clearly distinguished between contingency with respect to dependency-related cost i.e. parameter $\beta$ and contingency with respect to only the initial assessment. This is because we wish to identify the degree to which contingency rests on the private information of providers. In fact, because needs profiles change after initial assessment and follow-up is limited, even the later case still relies to some extent on provider-sourced information. In any case, this data does not exist and so in testing for information effects, the contract contingency variable defined above is noisy, embodying information that is partly common to both purchaser and provider. Whilst, this does make finding information effects more difficult, we should also note that for capturing risk and cost shock effects,

[55] However, sample sizes in these authorities were small (less than five in each) and so we cannot be certain that all providers in these four authorities have zero contingency contracts.

we would not want to make the above distinction. In these cases any contingency with respect to costs – stemming from private or common information – is relevant.

**Figure 8-1. Percentage of providers reporting contingent contracts**




A number of providers reported experiencing multiple contract types and in a small number of cases (16) providers simultaneously had contracts for residents that both did and did not make such an allowance. To avoid the problem of contaminating the contract effect, these cases were dropped. As a result 36 per cent of providers had contingent contracts. A dummy variable, $Y_i$, was generated to reflect this use of contingent contracts. This dummy variable was also used interactively with output as specified above.

A range of resident level, home level and market level factors shift demand and supply and need to be accounted for in the price analysis. A number of these factors enter both the demand and supply estimation and form the vector $H_i$. Characteristics such as the size and format of the home, whether it has en-suite facilities, its staffing levels and qualification rates, whether it is organisationally part of a chain, its registration status (whether it is a nursing home) and whether it caters specifically for people with mental health problems are factors influencing the perceived quality/value of the home. These factors clearly also bear directly on costs. Resident characteristics, in particular average levels of dependency, are also taken to influence demand in that they imply certain intensities and types of in-home services and thus differentiate homes. Home size as measured by the number of available beds, i.e. capacity, is closely related to the number of places filled. Consequently, this variable was removed from the vector $z_i$ and so out of the supply function to ensure that output variables do not appear in $\bar{c}_i$ of supply relation (8-23).

Factors that concern only demand for specific providers constitute the vector $\breve{y}_i$. As well as (a subset) entering the demand function, these variables are natural choices as instruments for the supply relation estimation. Ideally from an econometric perspective instruments ought to be highly correlated with demand but not supply and in practice with cross-sectional analyses these conditions are rather difficult to meet. Nonetheless, a number of proxies can be found that, at best, have only indirect, tenuous and ambiguous relationships with costs. First, resident funding source, here specified as the proportions of residents on the home that are privately funded (28%), and those that are local authority funded (38%) (as opposed to funding by social security, health authorities or other means). Second, two quality of life indicators – whether the home has adopted a flexible regime with respect to the organisation of the residents' day and whether residents have scope for self expression. These variables are part of the *sheltered care environment* scale (Moos and Lemke, 1992). It might be argued that a highly regimented lifestyle could reduce care costs, but this effect would be expected to be very small and far less than the impact on demand. Third is whether volunteers help in the organisation of resident activities. Again, volunteers may substitute for paid staff but this would normally just be a bonus, improving quality with little or no additional cost. The fourth and fifth variables are dummies reflecting whether, respectively, providers are on the approved list of the local authority and if they have contracts with local authorities other than their own. A sixth instrument concerns responses from home managers about their plans or otherwise to expand the home. Expansion plans here might indicate excess demand. To this list we also add *lagged* output as an instrument.

The counterparts to the above are variables that shift only supply/costs and not demand for places, which form the set $\bar{z}_i$. The first group of such variables relate to economies of scope and are whether the home also provides a day care service, a meals-on-wheels service and/or a home care service. These variables are anticipated as having minimal influence on the demand for residential care. Providers were also asked if their costs exceeded or not the price of care, which should be directly correlated with actual costs. Actual payment rates for minimum qualified staff for the home were also included. Lagged output was also used as an instrument for the endogenous demand price.

The demand function (8-40) is a residual or partial reduced-form demand function, that is, the influence of competitor homes is netted out using their reduced-form price functions. This process introduces (non-home specific) input cost and demand factors, i.e. $z$ and $y$ (as counterparts to $z_i$ and $y_i$) into home $i$ demand. To capture these terms empirically we use variables measured at the local authority level (rather than at the home level). Using regional variables means avoiding an arbitrary aggregation of other providers' cost factors, an aggregation that would otherwise need to reflect cross-demand elasticities. The relevant variables in the estimation are: local wage rates (female, manual gross weekly wage rates – New Earnings Survey), the Area Cost Adjustment (ACA), and local property prices (HM Land registry), which potentially enter the demand function as well as the cost (supply) function.

### 8.3.2 Estimation methods

*Sample size and missing values*

The total sample of homes was 673 cases, which is disaggregated by home type in Table 8-3. Since the focus of the current analysis is on independent sector providers, LA homes were (initially) excluded. The potential sample is therefore 506 cases. Information on price charged was available for 396 cases (78 per cent) and data on output was available for 505 cases (and for all of the 396 cases with price data). Of the 396 cases, only 366 had data on contract type. As noted above, a number of providers had multiple contracts with mixed incentives and these were dropped to reduce the sample to 350. A further 4 cases were lost due to missing information on home characteristics (including whether the business was started from scratch and home size), which reduces the sample to 346. Missing data on client dependency (i.e. Barthel scores) reduced the working sample to 343 cases (68 per cent of the potential sample). Two providers had 10 or more short-term residents. *A priori*, it is difficult to decide whether these short-term placements constitute (long-term) demand. Yet they do take places which could be filled by long-term placements. In view of these difficulties, and the very small number of such cases, these two providers were dropped from the sample. The final working dataset therefore contained 341 observations (67 per cent). Local authority areas defined market level groups of which there were 21, given an average of 16.2 providers per group. The minimum was 2 providers and the maximum was 55 per group.

**Table 8-3. Home types**

| Home type | Number | Per cent |
|---|---|---|
| LA homes | 167 | 24.81 |
| Nursing home | 159 | 23.63 |
| Private dual registered | 62 | 9.21 |
| Voluntary dual registered | 15 | 2.23 |
| Private residential | 148 | 21.99 |
| Voluntary residential | 122 | 18.13 |

*Estimation procedures*

The data offer two main challenges for estimation. First, to accommodate the strong theoretical case for the endogeneity of price and output. The second to allow local authority level effects to work though to home level behaviour. The model specification includes a number of fixed effects variables at the local authority level (i.e. the vector $y$ and $z$), but given that the sample was drawn from a subset of all local authorities in England, an allowance for random effects was made.

An error components two stage least squares random effects model was used to simultaneously address these concerns. The implementation due to Baltagi was used on the basis of its good performance with unbalanced panel data sets (Baltagi, 1995; Baltagi and Chang, 2000). As indicated above there is an overwhelming theoretical case that price and output are endogenous in the demand and supply equations (8-42) and (8-43). Tests support this hypothesis. Performing a Hausman specification test (see (Pindyck and Rubinfeld, 1981)) on the demand function (8-42) yielded test F statistic of 6.74 ($p = 0.0002$). On the supply function, the test statistic was 2.20 ($p = 0.11$). Both the supply and demand models also demonstrated random effects (respectively the error component of the variance *rho* values 0.28 and 0.097, which were found even with the presence of strong fixed effects terms).

Three other estimation issues arise: the possible skewed and leptokurtic nature of the dependent variable, common in analysis of this type; the censoring of short-term demand at the home's current capacity; and, the possible endogeneity of contract type. As to the former, the dependent variable, price, showed a very slight rightward skew and was modestly leptokurtic (Skewness of 0.84 and Kurtosis of 3.93). Demand theory suggests that price demand relationships are not linear, but this can be, and is, addressed using power transformations of the output variables on the right-hand-side. Hence, un-transformed price was used as the dependent variable for the demand and supply models.

Regarding the issue of possible censoring, the cross-sectional nature of the data means that coefficient estimates relate to the long-term and hence capacity constraints should be less of problem. Some 16% of the sample were operating at capacity and had waiting lists. To cast light on the significance of censoring the demand model was estimated using a tobit, censored-normal estimator, with predicted values of the endogenous variables. The results of this procedure were broadly similar to the non-censored model.[56] Hence, a non-censored model was employed, although home capacity was used as an independent variable in the estimation (partly to also reflect home size and format – see above).

Given home and client characteristics, contract choice is hypothesised to not be a function of cost and price. All contracts are prospectively determined. Arguably, providers might attempt to negotiate for contingent contracts if they expect an intake of high cost residents. Purchasers, however, have a more balanced set of incentives; whilst they wish to keep expenditure down, they also have a concern for provider stability. Moreover, political and historical factors will also be important in contract choice. And it is market dominant local authorities that can impose contract choices on the market (Forder, 2001; Forder and Netten, 2000b). *A priori*, there is only a weak argument for the endogeneity of contract choice. This hypothesis was tested by instrumenting the contract contingency dummy. The Hausman test statistic was 0.12 ($p > 0.5$) rejecting endogeneity.

---

[56] Tobit models are problematic in assuming arbitrary distributions (regarding 'censored' observations), particularly with endogenous variables.

The error distributions give a good indication of the validity of inferences from the two models. The residuals of both models largely conformed to the normal distribution. The demand model error had a skew of 0.21 and kurtosis of 4.55. The supply model error had a very slight rightward skew of 0.42 but was less leptokurtic with a kurtosis value of 3.76. Given the normal distribution of the dependent variable these findings indicate good specification.

Functional form choices for the output variable on the right-hand-side of the equation are not sensitive to these problems. Some experimentation was conducted to determine best fit functional forms. The final demand specification (8-40) produces an elasticity of:

$$(8\text{-}43)\quad \varepsilon^D = \frac{\partial x_i}{\partial p_i}\frac{p_i}{x_i} = \frac{p_i}{b_{11} + 2(b_{12} + b_{13}Y)x_i^2}$$

which we cannot sign *a priori* (below we characterise this function using parameter estimates).

In the supply model a linear specification of output as used. This specification implies a price elasticity of supply of :

$$(8\text{-}44)\quad \varepsilon^S = \frac{\partial x_i}{\partial p_i}\frac{p_i}{x_i} = \frac{p_i}{(a_{10} + a_{11}Y)x_i} = \frac{a_0 + a_{21}\breve{z}_i + a_{22}H_i + a_3 z + a_4 Y_i}{(a_{10} + a_{11}Y)x_i}$$

which, *ceteris paribus* is decreasing with output. In other words, small output providers are more sensitive to price changes, a result with some intuitive appeal.

## 8.4 Results and derived parameters

Table 8-4 reports the main demand and supply structural models, respectively (8-40) and (8-41). Relevant model diagnostics and specification tests – as discussed above – are reported at the foot of each table. A number of interaction effects and non-linear terms were used in the models. The interaction effects allow for factors such as contract type to shift not only the model's intercept but also the slope of relationships between price and other variables.

**Table 8-4. Results – demand and supply models**

| Variable definition | | Demand Co-eff | Demand t-stat | Supply Co-eff | Supply t-stat |
|---|---|---|---|---|---|
| *Output* | | | | | |
| Output (number of filled places) | Lin | | | 0.25 | 1.69 |
| | Log | -54.42 | -1.82 | | |
| | Sqrd | 1.04E-02 | 1.30 | | |
| Output × contingent contract | Lin × Lin | | | 1.01 | 1.65 |
| | Sqrd × Lin | -0.03 | -1.68 | | |
| *Demand and supply shift factors* | | | | | |
| Barthel score (inverse) | Sqrd | 4.90 | 4.95 | | |
| | Cbd | | | 9.81E-03 | 5.36 |
| Staff with nursing qualification rate | Lin | 98.79 | 3.45 | | |
| | Log | | | 158.34 | 8.22 |
| Home has no en suite toilets | Dummy | -24.28 | -4.02 | -12.47 | -3.38 |
| Home size/capacity | Lin | 1.76 | 2.97 | | |
| Home has modular layout | Lin | 54.02 | 1.93 | | |
| Home has modular layout × size/capacity | Lin × Lin | -1.30 | -1.80 | | |
| Home purpose built | Dummy | | | -13.18 | -2.10 |
| Business started from scratch w/ purpose built home | Dummy | | | 14.69 | 2.28 |
| Nursing home | Dummy | 33.25 | 4.31 | 33.83 | 6.43 |
| Single home organisation | Dummy | | | -4.72 | -1.30 |
| Home makes specific EMI provision | Dummy | | | 4.05 | 1.17 |
| *Demand only factors* | | | | | |
| Resident Funding (% residents privately funded) | Sqrd | 2.34E-03 | 1.76 | | |
| Home flexible regime score | Lin | 0.34 | 1.40 | | |
| Home has no contracts w/ other LAs × LA list only | Dummy | -8.28 | -1.12 | | |
| *Local demand, supply characteristics (not home specific)* | | | | | |
| Wages: female, manual gross wage | Lin | 0.27 | 1.77 | | |
| | Log | | | 65.26 | 0.95 |
| Property prices | Lin | | | 7.49E-04 | 3.76 |
| Area cost adjustment | Lin | 298.66 | 3.11 | | |
| *Contracts* | | | | | |
| Contingent price contract | Dummy | 35.34 | 1.95 | -23.64 | -1.26 |
| Contingent price contract × qualified staff ratio | Dummy | 45.97 | 1.13 | | |
| *Constant* | | | | | |
| Constant | | -99.83 | -0.89 | -217.69 | -0.55 |
| Model | | EC 2SLS | | EC 2SLS | |
| Dependent variable form | | Linear | | Linear | |
| R-sq: | within | 0.50 | | 0.68 | |
| | between | 0.72 | | 0.83 | |
| | overall | 0.56 | | 0.72 | |
| Number of obs. (groups) | | 341 (21) | | 341 (21) | |
| | | ***Stat*** | ***DF*** | ***Stat*** | ***DF*** |
| Overall fit | Wald | 443.72 | 17 | 745.87 | 13 |
| Normality | Bowman-Shenton | 36.97 | 2 | 18.00 | 2 |
| Specification | Basmann-Sargan | 4.80 | 6 | 16.28 | 11 |

Table 8-5 lists marginal effects where the point of departure is the sample mean value, unless otherwise noted. Clearly it is important to bear in mind that (a) estimated marginal effects as calculated from other points could be very different, and (b) strictly estimation results apply to small changes around the sample average so that extrapolations significantly removed from the mean should be treated with caution.

**Table 8-5. Demand and supply - net marginal effects**

| Variable | Change | Demand | Supply |
|---|---|---|---|
| *Output* | | | |
| Output | | -1.77 | 0.61 |
| Demand and supply shift factors | | | |
| Barthel score (inverse) | +1 inverse score | 4.90 | 2.53 |
| Staff with nursing qualification | +1 staff with nurs. Qual. | 3.30 | 27.85 |
| Nursing home | | 33.25 | 33.83 |
| Home has no en suite toilets | | -24.28 | -12.47 |
| Home size/capacity | +1 beds | 1.41 | |
| Home has modular layout | | 8.61 | |
| Home purpose built | | | -6.03 |
| Home purpose built & started from scratch | | | 1.51 |
| Business started from scratch | | | 4.91 |
| Single home organisation | | | -4.72 |
| Makes specific provision for EMI | | | 4.05 |
| *Demand only factors* | | | |
| Resident Funding | + 10% private payers | 1.32 | |
| Flexible home regime | + 10% score | 3.39 | |
| *Local demand, supply characteristics* | | | |
| Wages: female, manual gross wage | + 10% average gross wages | 10.25 | 6.53 |
| Property prices | + 10% average property prices | | 6.69 |
| Area cost adjustment | +10% ACA | 30.68 | |
| *Contracts* | | | |
| Contingent price contract | | 14.87 | 7.85 |

### 8.4.1 Demand and supply prices and elasticity

The demand function produced the expected downwards-sloping demand curve. The elasticity of price with respect to changes in output was -0.23 for the sample average, -0.20 for nursing homes and -0.25 for residential care homes. Conversely, the home-level supply curve was upward sloping with a price elasticity with respect to supply of 0.0028 for the sample average.

The demand curve slope $h_x^R$, from which these elasticities are calculated, is –1.77. This figure can be used in (8-27) to calculate a potential mark-up, *M*, or 'market power' for the average provider of £55. This result corresponds to a Lerner index of 19.5%. The demand curve was very slightly non-linear, although according with the usual shape, characterised by positive second order price effects on demand; at the mean $x_{pp}$ is 0.01.

### 8.4.2 Shift effects

The various demand and supply shift factors produced the expected signs and magnitude of effect. In particular, the results suggest that purchasers are willing to pay more for homes catering for more dependent people, requiring a more intensive service, as measured by the Barthel physical dependency, the ratio of nursing qualified staff, and the nursing home dummy variable. These services are also more costly and therefore have a higher supply price. Homes without en suite toilets were less attractive to buyers, but cost less and so have lower supply prices *ceteris paribus*.

Homes with modular architecture, that is where homes are divided into a number of functionally separate living units, are also more appealing to buyers. A number of potential demand and supply factors were found only to demonstrate a supply side effect. Purpose built homes have lower costs, but those starting the business from scratch had higher prices, probably reflecting the lingering effect of a lack of initial expertise. Single home organisation were not less appealing to purchasers, but were associated with higher costs. Homes making specific provision for older people with mental health problems were similarly more costly, but with demand not significantly affected.

Some factors only shifted demand. For example, homes with a greater proportion of more lucrative private payers had higher prices, as did homes that offered a more flexible living regime, a quality indicator. Supply (cost) only factors include, for example, home specific labour costs. However, also included in the model were regional (non-home-specific) input cost factors: the price of labour and property in the local authority area. A high degree of multicollinearity resulted in using these sets of variables together. The regional level variables were found to offer a far better fit and so supply only factors were dropped from the model. The impact of non-home-specific input costs is exemplified by the calculation that a 10% increase in average wages (extrapolating from the margin) would push supply prices up by £6.53. For a 10% change in property prices, supply price on average increased £6.69. Regional input prices also appear in the partial reduced-form (or residual) demand function (to account for the influence of competitor prices). In the demand model an average wage increase corresponded with a demand increase of £10.25. Also included was a local authority cost index – the Department of Health area cost adjustment (ACA) – which is derived from labour and capital prices. Ten per cent increases in the ACA were associated with prices some £30.68 higher at the margin. These findings accord with expectations. From (8-16), $\partial x_{-i}/\partial z = \partial E_{-i}/\partial z < 0$: higher input prices mean higher supply prices and so lower demand $x_{-i}$. Also, from (8-17), since $\partial p_1/\partial x_{-i} = \partial h_1/\partial E_{-i} < 0$ (service 1 prices are inversely related to the output of substitutes as an increase in substitutes reduces the demand for product 1), it follows that $\partial p_1/\partial z = \partial h_1/\partial E_{-i} \cdot \partial E_{-i}/\partial z > 0$.

### 8.4.3 Demand prices and outputs by contract type

The marginal effect of a change in contract type on price in the demand function is £14.87 (see Table 8-5). This figure is derived holding constant other factors in the demand function, including *actual* client dependency (as measured by the Barthel score) and output. It is statistically significant with a derived t-stat of 2.44.

### 8.4.4 Supply price-dependency relationship

What are the effects of dependency on costs? The supply equation can be used to ascertain the impact of dependency on marginal costs. With reference to (8-26) for given $x$ the marginal effect of a change in dependency on the supply price estimation is equivalent to $\bar{c}_\beta$. We use the estimation without home type indicators (including staff qualification ratios) since we are concerned to estimate the total underlying cost response to changes in client dependency. Table 8-6 reports the results of this estimation. Adapting (8-24), we estimate

$$(8\text{-}45)\quad p_i = a_0 + a_{10}x_i + a_{11}x_iY_i + a_{21}\tau_i + a_{221}\beta_i + a_{222}\ln\beta_i + a_3 z + e_i$$

where $\tau_i$ is the sub-vector of $z_i$ with home-type and Barthel score removed. This alternative specification of the supply function accounts for the high collinearity of the nursing qualified staff ratio and (inverse) client dependency score (correlation $p < 0.01$) by dropping the former. The net effect at the mean of changes in (inverse) Barthel on supply price, $a_{22} = a_{221} + a_{222}/\bar{\beta}$, is £8.54 per point.

**Table 8-6. Results – Supply models, no home-type indicators**

| Variable definition | | Supply Co-eff | t-stat |
|---|---|---|---|
| *Output* | | | |
| Output (number of filled places) | Lin | 0.34 | 2.00 |
| Output × contingent contract | Lin × Lin | 0.90 | 1.32 |
| Demand and supply shift factors | | | |
| Barthel score (inverse) | Lin | 11.18 | 7.00 |
| | Log | -24.48 | -2.13 |
| Confusion: % with moderate and severe confusion | Lin | 0.20 | 1.73 |
| Home has no en suite toilets | Dummy | -12.35 | -2.89 |
| Home has modular layout | Lin | -0.20 | -0.04 |
| Home purpose built | Dummy | -27.22 | -3.77 |
| Business started from scratch w/ purpose built home | Dummy | 27.25 | 3.63 |
| Nursing home | Dummy | | |
| Single home organisation | Dummy | -2.76 | -0.65 |
| Home makes specific EMI provision | Dummy | 7.41 | 1.83 |
| *Local demand, supply characteristics (not home specific)* | | | |
| Property prices | Lin | 9.22E-04 | 5.70 |
| *Contracts* | | | |
| Contingent price contract | Dummy | -24.80 | -1.19 |
| *Constant* | | | |
| | | 131.53 | 6.28 |
| Model | | EC 2SLS | |
| Dependent variable form | | Linear | |
| R-sq: | within | 0.57 | |
| | between | 0.79 | |
| | overall | 0.60 | |
| Number of obs | | 341 | |
| Number of groups | | 21 | |
| | | ***Stat*** | ***DF*** |
| Overall fit | Wald | 458.22 | 13 |
| Normality | Bowman-Shenton | 24.48 | 2 |
| Specification | Basmann-Sargan | 12.50 | 9 |

### 8.4.5 Demand and supply slope by contract type

The slope of the demand curve is given by the differentiation of (8-40) i.e. $\partial p_i/\partial x_i \equiv h_{x_i}^R = b_{11}/x_i + 2b_{12}x_i + 2b_{12}\bar{Y}x_i$. Differentiating (8-41) gives the slope of the supply curve $\partial p_i/\partial x_i = a_{10} + a_{11}Y_i$. Using the estimates reported in Table 8-4, these slopes are calculated for the whole sample and for the two contract groups, and are reported in Table 8-7.

**Table 8-7. Slopes of demand and supply**

| Slope | Whole sample | Contingent contracts | Non-contingent contract |
|---|---|---|---|
| Demand | -1.77 | -2.94 | -1.11 |
| Supply | 0.61 | 1.26 | 0.25 |

Bootstrapping was used to calculate the standard errors of these derived estimates, with 1000 repetitions. This technique draws repeated samples from the data. In this case, it runs the model on that drawn sample calculating the *b* coefficients as normal. It also derives the mean values of the *x* and *Y* factors for the sample and stores these and the coefficients for each repetition. The slopes are derived for each repetition and summarised using central tendency and variance (standard error) statistics for the 1000 repetition sampling distribution. Because this sampling is unbiased the derived slopes tend to the values as given in Table 8-7. The estimated standard errors are used to calculate 90% confidence intervals (normal values are used). These results are given in Table 8-8. Both the demand and supply slopes are significantly different from zero at the 90% (and also the 95%) level. The supply slope is especially precisely estimated.

**Table 8-8. Slopes of demand and supply -- confidence intervals**

| Slope | Coefficients | Standard error | 't-stat' | Confidence intervals | |
|---|---|---|---|---|---|
| | | | | Positive | Negative |
| Demand | -1.77 | 0.62 | -2.86 | -2.78 | -0.75 |
| Supply | 0.61 | 0.02 | 30.32 | 0.58 | 0.64 |

### 8.4.6 Equilibrium prices and outputs

Demand and supply prices can be calculated for a range of possible output levels using the relevant regression model results from Table 8-4 for the two cases where the contract type variable *Y* is equal to zero and where it is equal to one, with all other exogenous variables at their mean values. Equilibrium output for the two groups, $x^{*Y}$ is where the predicted demand and supply prices are equal, i.e. drawing on (8-33),

$$(8\text{-}46)\quad b_0 + b_{11}\log x^{*1} + (b_{12} + b_{13})(x^{*1})^2 + b_2\bar{y} + b_3\bar{g} + b_4\bar{z} + b_5$$
$$= a_0 + (a_{10} + a_{11})x^{*1} + a_2\bar{z} + a_3\bar{z}_h$$

for contingent contracts and

$$(8\text{-}47)\quad b_0 + b_{11}\log x^{*0} + b_{12}(x^{*0})^2 + b_2\bar{y} + b_3\bar{g} + b_4\bar{z} + b_5 = a_0 + a_{10}x^{*0} + a_2\bar{z} + a_3\bar{z}_h$$

for non-contingent contracts. In these expressions the bar embellishment indicates the sample mean value of the relevant variable. Note also that $\bar{z}_h$ is sample average across the local authority level variables. The resultant equilibrium values (using point estimates), *all other things being equal,* are given in Table 8-9. These values are calculated by solving for $x^{*Y}$ in the above two functions with all the shift variables at their sample mean and using the corresponding estimated parameter. Demand for contingent contract providers is more price inelastic than for non-contingent contract providers meaning that they optimise at relatively lower outputs and higher prices, all other things being equal.

**Table 8-9. Equilibrium prices and outputs – by contract group**

| | All contract types | Contingent contract | Non-contingent contract |
|---|---|---|---|
| Equilibrium price (£s) | 281 | 297 | 272 |
| Equilibrium output | 31 | 27 | 39 |

The values of $x^{*Y}$ are dependent on the slopes and shapes of the demand and supply curves, as well as many shift variables. To explore the sensitivity of these estimates, solution values were re-calculated for the case where the slopes of the demand and supply curves respectively take their plus and minus confidence interval values – see Table 8-8. For the contingent contracts case, the resultant output range was: 27.0 to 30.3

places per week. For the non-contingent contracts the output range was: 30.7 to 36.5 places per week.[57]

### 8.4.7 Motivation

The actual price mark-up over intercept costs as given by (8-13) i.e. $a_1$ (see (8-24)) can be compared with potential mark-up rates $-h_x^R$ to give an indication of provider motivation. As noted above, the complicating factor is the possibility of economies of scale, that is, $c_x < 0$. To overcome this problem we compare the potential and actual mark-up differences for the two contract-type groups. Since they both operate by definition with the same value of $c_x$, this value drops out and we are left with (8-31). This function is then used to estimate the motivation parameter $\alpha$. Our estimates generate a point value of the motivation profit weight $\alpha$ of 0.55, which would imply that on average providers only take 55% of the potential maximum amount of profit that the current market would sustain.

It is worth noting that we cannot say unequivocally that providers would make the current potential level of profit if all decided to maximise profits. The prevailing market conditions might then lead to a different set of maximising prices. It is clear however, that providers are not taking as much profit as they could given their current circumstances, and without hyper-rational expectations about the response of all other providers if the individual provider decided to attach more weight to profit making.

The nature of the underlying model and the available data mean that we are unable to test whether providers with different contracts operate with different profit weights. However, given that there are not substantial differences between the groups in terms of home and business characteristics (see Table 8-10 for mean comparison t-tests), we assume that motivation does not differ by contract type. Moreover, the comparison of mark-up by contract type is made assuming all other exogenous factors are equal. Since it is not an onerous assumption to make that the underlying motivation function does

---

[57] There is no particular reason why these values should correspond very closely with those in the above tables since we are looking simultaneously at the upper and lower confidence intervals of both demand *and* supply and because these functions are non-linear.

not vary with contract type, then for purposes of this comparison both contract-type groups of providers have this value for profit weight.

These results provide some insight to the sensitivity of the motivation parameter to the estimated coefficients. We can also explore more directly the degree of sensitivity. First, we used combinations of the demand and supply slope plus and minus 90% confidence interval values. Second, proportional sensitivity was conducted using +/- 20% of the relevant point estimate. The former produced a range of values from 0.41 to 0.93, the latter from 0.37 to 0.70 – see Annex 8-1.

**Table 8-10. 'Motivation relevant' home characteristics**

| | **Difference in means (p-value)** |
|---|---|
| Single home organisation | 0.68 |
| Home size/capacity | 0.18 |
| Proportion LA funded | 0.72 |

### 8.4.8 Marginal costs and scale economies

Scale effects on costs are measured by the size of the derivatives of marginal cost with respect to output i.e. $\frac{\partial^2 C_i}{\partial x_i^2} = \frac{\partial c_i}{\partial x_i} = c_x$. They, in other words, indicate downwards or upwards sloping marginal cost functions. The specification of the supply relation (8-24) is such that scale effects on costs, $c_x$, only appear in the marginal effect of output changes on supply price as indicated by i.e. $\frac{1}{2}\alpha c_x + \frac{1}{2}c_x - \alpha h_{x_i}^R = a_1$ (see (8-13)). This specification precludes the need to find proxies for scale (which could appear in the vector $\bar{c}_i$) that are not correlated with output. We should caveat however, that this specification does assume that scale effects are captured adequately by output variations rather than, say, capacity variations (although of course these two variables are highly correlated (Pearson correlation of 0.90).

Equation (8-32) can be used to derive $c_x$ (with these caveats in mind). The estimation results produced a point value of $c_x$ of –0.47. In other words, marginal costs reduce by 47p for every extra place sold (at sample average output). This is a modest value given that output averages around 30 places and price nearly £300 per week. Using the respective confidence intervals of the determining parameters, $c_x$ ranged from 0.30 to –1.24, which is still modest in the context of average prices.

The impact on average *variable* cost is smaller as it is spread over the range of places already sold, being $\frac{\partial(C_i^V/x_i)}{\partial x_i} = \frac{1}{2}c_x$. However, the impact on average total costs is $\frac{\partial(C_i/x_i)}{\partial x_i} = \frac{1}{2}c_x - \frac{F_i}{x_i^2}$, which is greater in absolute terms (more negative) than $\frac{\partial(C_i^V/x_i)}{\partial x_i}$.

We have assumed that the underlying cost structure of the home does not vary by contract type. Therefore, our *ceteris paribus* comparison of mark-up by contract type utilises the same estimate of $c_x$ for both groups of providers

## 8.5 Results: mark-up rates and costs

### 8.5.1 Price-cost margins

Section 8.2.1 provides a number of different specifications of mark-up. Table 8-11 gives the estimated *actual* mark-up rates per week by cost type. The table gives the equation from which the respective mark-up estimate is derived and lists mark-up in pounds per week and as a percentage of the equilibrium price (the latter as given in Table 8-9). Table 8-19 in Annex 8-2 provides details about how the mark-up rates in Table 8-11 (and Table 8-12) are calculated. Cost type refers to whether mark-up is taken over marginal or average variable cost. The table also reports derived 'confidence intervals', which are calculated by using all determining parameters respectively at their upper and lower confidence interval value.

**Table 8-11. Actual mark-up rates - £s per week and % of average weekly price**

| Mark-up (£s per week) over... | | Deriving equation | All contract types | | Contingent contract | | Non-contingent contract | |
|---|---|---|---|---|---|---|---|---|
| Actual, average variable cost | Point | (8-30) | 26.29 | 9.4% | 40.94 | 13.8% | 18.98 | 7.0% |
| | + CI | | 15.43 | | 34.81 | | 4.11 | |
| | – CI | | 37.15 | | 49.96 | | 30.44 | |
| Actual, marginal cost | Point | (8-28) | 33.59 | 12.0% | 47.37 | 16.0% | 28.21 | 10.4% |
| | + CI | | 10.84 | | 30.33 | | -0.42 | |
| | – CI | | 56.35 | | 66.68 | | 53.03 | |

Marginal cost mark-up is the cost per week of serving one more user (assuming that there is spare capacity) and as such does not depend on fixed costs (which are already sunk). However, since in many cases marginal costs are falling with output – homes have economies of scale as suggested by the data – mark-up over marginal cost does not indicate actual profit, only profitability. Mark-up relative to average costs is a better indicator. We specifically use average *variable* costs (8-8) (which avoids a need for an estimate of fixed costs). Average variable costs fall at half the rate of marginal costs (with constant scale economies, $c_{xx} = 0$), and therefore at the margin, mark-up is lower than in the marginal cost case. This indictor of mark-up is a good indicator of market stability since it is these costs that must be covered for providers to stay in business, at least in the short term. As is clear from the results reported in the table, actual mark-up over average cost is relatively modest; non-contingent contract providers, for example, can expect only 7% mark-ups. Moreover, if there are any true fixed costs that are not amortised into variable costs, then subtracting their net present value away from average variable costs mark-up will give an even lower result.

The derived confidence intervals are only a guide, but do give some idea of the possible range the estimates cover. The differences between upper and lower values regarding average cost mark-up are not particularly big; moreover, the ranges for contingent and non-contingent contracts do not overlap.

The low level of average variable cost mark-up partly reflects providers' apparent choice not to take more profit from the business. Suppose that the observed pattern of demand and supply in the data was actually produced by transactions between purchasers and *profit maximising* providers. In other words, observed motivation is $\alpha = 1$. Then as given by (8-27) and (8-29) mark-up over marginal and average costs, respectively, would be at levels given in Table 8-12, levels that are considerably higher than rates based on $\alpha = 0.55$, as estimated. These rates are not, however, those we would find if providers had $\alpha = 0.55$ motivation but decided to increase prices to the level of profit maximisers. It is difficult to know whether in times of increased pressure on profits, providers would try to take more profit or simply quit the market, if with current motivation, they were unable to cover costs. Nonetheless, it is important to be clear that if sample prices were based on $\alpha = 0.55$, then in order to increase profits, providers would have to raise prices beyond these sample values, which would mean reductions in demand relative to sample values. The maximum amount of profit that could be made would therefore be lower than the rates given in Table 8-12.

**Table 8-12. Observed profit maximising mark-up rates**

| Mark-up (£s per week) over... | Deriving equation | All contract types | | Contingent contract | | Non-contingent contract | |
|---|---|---|---|---|---|---|---|
| Observed α = 1, marginal cost | (8-27) | 55 | 19.5% | 91 | 30.8% | 34 | 12.7% |
| Observed α = 1, average var cost | (8-29) | 48 | 16.9% | 85 | 28.6% | 25 | 9.3% |

### 8.5.2 Costs

Marginal and average variable costs are inferred directly from the estimation results. The above mark-up estimates can be subtracted from the equilibrium prices for each group (Table 8-9) to give the relevant cost. In particular, using (8-26) we have:

$$\text{(8-48)}\quad \bar{c} = p_i^{*Y} - \left(\tfrac{1}{2}\alpha c_x + \tfrac{1}{2}c_x - \alpha h_x^R\right)x_i^{*Y} = p_i^{*Y} - a_1 x_i^{*Y}, \quad \forall Y$$

where * indicates equilibrium prices. By construction $\bar{c}$ is not a function of price or output, and therefore when we infer it from equilibrium price and outputs of each contract type group, other things (the exogenous variables) being equal, we will find the same value for each group. This result follows from our derivation of equilibrium prices and outputs for each contract group, which as is clear from (8-46) and (8-47) share the same supply curve whose origin is $\bar{c}$ when others things are equal. The estimated value of $\bar{c}$ is £262 per week.

**Table 8-13. Costs per week**

| Cost (£s per week) | | Deriving equation | All contract types | Contingent contract | Non-contingent contract |
|---|---|---|---|---|---|
| Marginal | Point | (8-5) | 247.39 | 249.14 | 243.54 |
| | + CI | | 270.15 | 269.93 | 270.04 |
| | - CI | | 224.64 | 229.60 | 217.87 |
| Average variable | Point | (8-8) | 254.70 | 255.57 | 252.77 |
| | + CI | | 265.56 | 265.45 | 265.50 |
| | - CI | | 243.84 | 246.32 | 240.45 |

Marginal costs are a function of output (8-5) and so do differ by contract group to the extent that equilibrium output differs by contract group. Similarly, average variable costs (8-8) are a function of output. Table 8-13 lists the respective point estimate values. There will be some variation of these values within the 90% confidence intervals as shown in the table.

Average costs (8-7) are dependent on the size of fixed costs, which could cover capital costs (mortgage), tax, maintenance, etc. The first is the predominate factor. Whilst we do not have any specific information on capital it is nonetheless worth speculating a little as to impact of fixed costs on to average costs and so overall profitability. In particular, we use Laing and Buisson figures of approximately £20000 per place (1996 prices) for capital costs. At this rate the sample average 34.9 bed home would have a capital cost of £697,537. Amortising this figure over 25 years at 7% produces a fixed cost per week of $F$ = £1138 or per week per head, $F/x$ = £37. Assuming the same capacity to output ratios, $F/x$ = £37 regardless of contract group. As $C_i/x_i = C_i^V/x_i + F_i/x_i$, mark-up over average cost will be lower than it is over average variable cost. Note that since we assume that satisficers break-even with respect to average variable costs, then the motivation parameter $\alpha$ and the change in marginal cost, $c_x$, remain unchanged. In this case, average provider mark-up over full average cost would be negative at £-10.38.

Suppose however that satisficers sought to break-even over full average costs instead of average variable costs. Then $\alpha$ and $c_x$ are functions of $F_i$. In particular, in this case the data would give values of $\alpha$ of 0.43, and $c_x$ of $-1.15$ (which are lower than the above estimates). The corresponding mark-up over full average cost for the sample average motivation of 0.43 (i.e. comparable with the above) would be £0.14 at the mean. We cannot test the exact motivation of satisficers, but it is clear that mean mark-up is close to zero or even negative. The difference between mark-up rates by contract type does not change very much (by about £4) in this case.

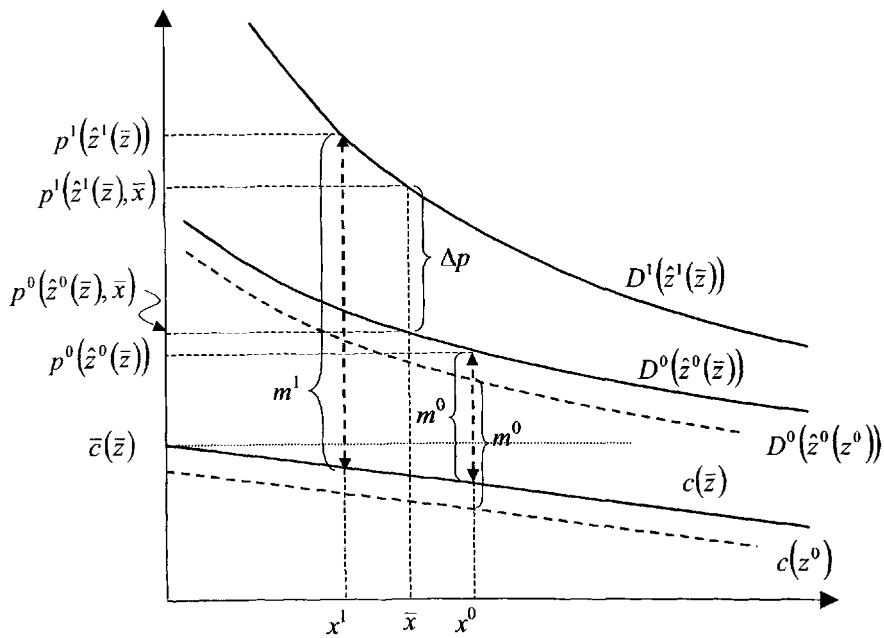
## 8.6 The implications of contract choice

What are the implications for contract choice? The above sets of estimates underpin the main empirical indicators with which to assess our theoretical hypotheses. First, we look at the impact on prices, and then on mark-up rates of moving between contingent and non-contingent contracts, other things equal.

#### 8.6.1.1 Effects on prices

The marginal effect of contract choice on price (of £14.87, significant at the 5% level, see section 8.4.3) indicates that purchasers are willing to pay more for the same care package purchased under a contingent contract. Hence, although the (average) price for the non-contingent contract group may embody an element of cream-skimming, the effect of cost exaggeration in the contingent contracts group is greater. We can use Figure 8-2 to explain this result.

**Figure 8-2. Interpreting contracting results: demand and supply effects**



The figure depicts (stylised) demand curves for the two contract groups. It also shows the corresponding supply relationship i.e. marginal cost. Since this function is calculated on the same home and user cost characteristics, $z = \bar{z}$, and contract choice does not affect costs given these characteristics, there is no cost difference between the contract groups. The marginal impact of contract choice on demand prices (i.e. £14.87) is the difference $\Delta p$ in the figure, that is, the difference between prices: $p^1\left(\hat{z}^1(\bar{z}), \bar{x}\right)$, where the 1 superscript denotes contingent contracts and $p^0\left(\hat{z}^0(\bar{z}), \bar{x}\right)$, where the 0 superscript refers to non-contingent contracts.

These are prices calculated at sample average values of $z$ and $x$, and need not correspond to actual prices for the two contract type groups. Since contingent contract providers do not select users with lower cost characteristics, user cost characteristics in this group should average the same as mean referral level i.e. $\beta^1 = \beta^I$ (where $\beta$ are the user cost characteristics elements of $z$). Non-contingent contract providers, if they cream-skim will have users with lower cost characteristics i.e. $\beta^I > \bar{\beta} > \beta^0$. Absent cream-skimming and with all other relevant characteristics at their means, the non-contingent contract group price would be: $p^0\left(\hat{z}^0(\bar{\beta})\right)$. In practice, the equilibrium value of $\beta$ of non-contingent (NC) contract providers will be below the sample average value of $\beta$ (because they cream-skim) i.e. $\bar{\beta} > \beta^0$ with a price is $p^0\left(\hat{z}^0(\beta^0)\right)$, which is less than $p^0\left(\hat{z}^0(\bar{\beta})\right)$. The NC provider equilibrium price is simply the NC contract price set by the purchaser and therefore must be at this level i.e. $p^0\left(\hat{z}^0(\beta^0)\right) \equiv p^0\left(z^I\right)$. Hence the difference between the notional price and the actual price: $p^0\left(\hat{z}^0(\bar{\beta})\right) - p^0\left(z^I\right) > 0$ is an indication of cream-skimming rents. With reference to the figure, the demand and cost functions for NC providers at the sample average values of $z$ are shifted up from their actual values at $z^0$ (the dashed demand and cost functions).

Given the possibility that non-contingent contract providers are cream-skimming however, the difference $\Delta p$ is not the total size of the CE effect. For contingent contract holders: $p^1\left(\hat{z}^1(\bar{\beta})\right) < p^1\left(\hat{z}^1(\beta^1)\right) \equiv p^1\left(\hat{z}^1(\beta^I)\right)$. Hence, other things equal, $p^1\left(\beta^I, \bar{x}\right) - p^0\left(\beta^0, \bar{x}\right) > p^1\left(\bar{\beta}, \bar{x}\right) - p^0\left(\bar{\beta}, \bar{x}\right) = \Delta p$. Moreover, as the two groups have differently sloped demand curves their equilibrium output levels (where supply price equals demand price) will differ (from each other and the sample mean output level). The differences in prices at these two equilibrium outputs is not $\Delta p$ as it clear from Figure 8-2. Moreover, scale effects on costs will introduce a further adjustment. Price difference between providers with different contracts is not sufficient. However, mark-up rates as given by $m^0$ and $m^1$ are the same regardless of the actual position of the demand and cost functions. So the difference in (price-cost) difference is more informative.

#### 8.6.1.2 Effects on mark-up

In section 8.5.1 mark-up rates over various cost measures were given for each contract type. Here we are interested in the *difference* in mark-up between contract types. Table 8-14 lists the differences as derived from Table 8-11. The actual mark-up differences are most relevant, the change in motivation mark-up differences are mainly for information. The choice of contracts therefore has ramifications for mark-up of around 7% of average weekly price. In terms of the above figure, the actual marginal cost difference is equal to $m^1 - m^0 = \Delta m$. Change in mark-up accounts for both differences in demand elasticity and scale effects on costs.

**Table 8-14. Mark-up rates, difference by contract type**

| Mark-up type difference (£s per week) | difference (£s per week) | |
|---|---|---|
| Actual, marginal cost | 19.16 | 6.82% |
| Actual, average variable cost | 21.96 | 7.82% |

## 8.7 Mark-up: information, risk and shocks

Differences in mark-up by contract type can, according to the theory of the last chapter, be due to differential information effects, risk premiums or within-financial-period cost shocks. Added to this list is the alternative hypothesis that higher mark-up with contingent contracts is due simply to higher payment rates.

### 8.7.1 Higher payment rates

The most plausible payment arrangement that corresponds to this hypothesis is where, in addition to the standard rate paid for low dependency users under contingent and non-contingent contracts alike, contracts pay a premium rate for high dependency users. Homes with a relatively greater proportion of high dependency users will be paid more on average. Furthermore, for this arrangement to replicate our results, additional payment must more than cover the extra costs of serving higher dependency users. Here we look directly at mark-up rates for different levels of dependency by contract type. The demand function (8-40) was re-estimated with an additional dependency interaction on the output (squared) by contract type variable:

$$(8\text{-}49)\quad p_i = b_0 + b_{11}\log x_i + b_{12}x_i^2 + b_{131}\beta_i^H Yx_i^2 + b_{132}Yx_i^2 + b_2 y + b_{31}\breve{y}_i + b_{32}H_i + b_4 z + b_5 Y_i + u_i$$

where $\beta_i^H$ is the proportion of users that are in the top 20% of dependency scale within the home. Under the higher payment hypothesis the difference in mark-up between contingent and non-contingent contracts should be higher for home with many high dependency users than a home with mostly low dependency users.

In terms of (8-49) to support this hypothesis we should see $b_{131}$ to be significantly negative: the greater the proportion of high dependency users within contingent contract homes, the greater the inverse demand elasticity, and so mark-up rate. Table 8-15 provides the results of this estimation, showing that $b_{131}$ is not significant.

This result is perhaps not surprising; for higher payment to be consistent with the data, we would need to find contingent contracts that have to reward providers of high dependency clients at a rate proportionately higher than changes in marginal costs for high dependency users. Or, we would need to find that contingent contracts only partially compensate for high dependency related costs but also that payment is higher under contingent contracts even for low dependency users. This latter is inconsistent with our definition of the higher payment hypothesis.

**Table 8-15. Demand price – mark-up, by dependency**

| Variable definition | | Demand, full sample Co-eff | Co-eff |
|---|---|---|---|
| *Output* | | | |
| Output (number of filled places) | Log | -50.26 | -1.90 |
| | Sqrd | 8.80E-03 | 1.21 |
| Output × contingent contract | | -2.36E-02 | -0.87 |
| Output × contingent contract × high dep % | Sq× Ln × Ln | -4.15E-04 | -0.62 |
| *Demand and supply shift factors* | | | |
| Home has no en suite toilets | | -25.70 | -4.16 |
| Nursing home | | 41.45 | 5.00 |
| Staff with nursing qualification rate | | 85.58 | 2.06 |
| Staff with nursing qualification rate × contingent contract | Ln × Ln | 81.34 | 1.29 |
| High dependency % | | 0.36 | 1.25 |
| Inverse Barthel score | | 20.95 | 2.49 |
| Home size/capacity | | 1.93 | 3.26 |
| Home has modular layout × capacity | Ln × Ln | -1.28 | -1.93 |
| Home has modular layout | | 51.94 | 2.03 |
| *Demand only factors* | | | |
| Flexible home regime | | 0.24 | 0.91 |
| Home has no contracts with other LAs × LA list only | Ln × Ln | -6.92 | -0.90 |
| Resident Funding (% residents privately funded) | | 0.13 | 1.13 |
| *Local demand, supply characteristics* | | | |
| Wages: female, manual gross wage | | 0.27 | 2.38 |
| Area cost adjustment | | 318.04 | 4.18 |
| *Contracts* | | | |
| Contingent price contract | | 35.85 | 1.84 |
| *Constant* | | | |
| Constant | | -144.41 | -1.57 |
| Model | | EC 2SLS | |
| Dependent variable form | | Linear | |
| R-sq: | within | 0.47 | |
| | between | 0.73 | |
| | overall | 0.53 | |
| Number of obs | | 341 | |
| Number of groups | | 21 | |
| | | ***Stat*** | ***DF*** |
| Overall fit | Wald | 431.45 | 19 |
| Normality | Bowman-Shenton | 54.16 | 2 |
| Specification | Basmann-Sargan | 3.01 | 6 |

### 8.7.2 Risk premiums

According to the theory, contingent contracts have more attractive risk properties than non-contingent contracts – propositions 5 and 7 in chapter 7. Given our results therefore, if risk premia are available they are more than offset by the other effects on mark-up rates.

### 8.7.3 Cost shocks

Cost shocks could account for some of the total difference in mark-up. In particular, where client dependency has risen over the contract period ahead of expectations then costs will have also risen. For contingent contracts the rise in costs would be (partially) compensated for by an increase in prices on average across the contract period, and so the relative reduction in profits would be smaller than for the non-contingent contracts group. On average, profits would be higher with contingent contracts in this case, a result consistent with our findings. Would cost shocks explain all this difference? Suppose that at the beginning of the contract period average expected mark-up is no different, other things equal, between the contract groups. Forder and Netten (2000a) show that in the 10-year period before 1996 dependency levels for the residential and nursing care market for older people did indeed increase. However, the total change in that period was 2.37 (inverse) Barthel points.

To exemplify the extreme case suppose that contingent contract providers are fully compensated. Suppose also that demand levels remained constant i.e. that whole population dependency rises. Then the impact of dependency increases on non-contingent contract provider costs is $c_\beta = \bar{c}_\beta + c_x \frac{dx}{d\beta} = \bar{c}_\beta \equiv a_{22} = a_{221} + a_{222}/\bar{\beta}$ in (8-45), that is £8.54 per point. If non-contingent contract prices were revised on average only every 3 years (which is probably again an overestimate) then for comparison with the contingent contract case, we might apply 30% of the 10-year dependency change i.e. 0.71 Barthel points to give Δc of £6.07. Non-contingent contract mark-up should then be different by £6.07 compared to the fully-compensating contingent contracts. Even assuming circumstances that maximise the effects of cost shocks, they are not sufficiently large to account for the mark-up differences we estimate.

### 8.7.4 Cream-skimming

The existence of cream-skimming cannot be inferred from our results on profit differences by contract type. Instead, cream-skimming of low dependency clients (rather than specifically, low cost clients) is directly investigated. We can assume that the dependency characteristics of *referred* clients (referrals made by purchasers) are a function of the characteristics of the home, including its price:

$$(8\text{-}50)\quad \beta_i^{lY} = \beta_i^{lY}\left(p_i^Y, g_i^Y, z_i^Y\right)$$

where $\beta_i^{lY}$ is the initial referral dependency $\beta^l$ for home $i$ with contract $Y$. Also, $g_i$ are home demand characteristics and $z_i$ are home cost characteristics. We also substitute in the reduced-form function for price to give:

$$(8\text{-}51)\quad \beta_i^{lY} = \beta_i^{lY}\left(g_i^Y, y^Y, z_i^Y, z^Y\right)$$

We do not have a direct observation of referral dependency, $\beta^l$. We lean on the theory – proposition 1 of chapter 7 – and our finding of cost exaggeration to assume that *contingent contract* providers do not cream-skim. Therefore, the actual dependency of clients of contingent contract providers is equal to the referred dependency. For this group and this group alone, we can estimate (8-50), i.e,

$$(8\text{-}52)\quad \beta_i^{l1} = \beta_i^{l1}\left(g_i^1, y^1, z_i^1, z^1\right)$$

The relationships between the variables of this function and dependency should themselves be unaffected by the type of contract under which the home operates. As such this function can be used to cross-predict the referral dependency levels of homes with non-contingent contracts, i.e.,

$$(8\text{-}53)\quad \beta_i^{l0} = \beta_i^{l1}\left(g_i^0, y^0, z_i^0, z^0\right) + e$$

with a mean error $e$ of zero. In this context we can advance the following hypothesis: homes with non-contingent contracts are cream-skimming if $\beta_i^{l0} > \beta_i^0$, that is, if predicted referral dependency is greater than actual dependency. The null hypothesis is, specifically, that the sample mean difference is equal to zero: $\frac{1}{N^{Y=0}} \sum_{i\in\{Y=0\}} \left(\beta_i^{l0} - \beta_i^0\right) = 0$.

**Table 8-16. Dependency (Barthel score) estimation**

| Variable definition | | Barthel Score Co-eff | t-stat |
|---|---|---|---|
| *Demand and supply shift factors* | | | |
| Home has no en suite toilets | | 0.13 | 0.22 |
| Nursing home | | 5.91 | 2.71 |
| Home has no contracts with other LAs × LA local list | | 6.18 | 2.42 |
| Home is only on the approved list of its LA | | 0.67 | 0.85 |
| Home has no contracts with other LAs | | -6.46 | -2.75 |
| Home size/capacity | | 0.06 | 0.68 |
| Home size/capacity | sqrd | 5.78E-04 | 0.50 |
| Home size/capacity × nursing home | | -0.06 | -1.11 |
| Home purpose built & business started from scratch | | 0.20 | 0.16 |
| Home purpose built | | -0.92 | -0.88 |
| Single home organisation | | -0.18 | -0.30 |
| Business started from scratch | | -0.03 | -0.05 |
| *Demand only factors* | | | |
| Resident Funding (% residents privately funded) | | -1.12E-02 | -0.62 |
| Resident Funding (% residents LA funded) | | 3.40E-02 | 2.98 |
| Volunteers aid in organised activity | | 0.22 | 0.27 |
| *Supply only factors* | | | |
| Provides home care | | 1.11 | 0.91 |
| Provides meals on wheels | | -1.38 | -1.37 |
| Provides meals on wheels | | -1.85 | -0.90 |
| Home wage rate (basic), care staff | | 3.16E-02 | 1.14 |
| Home wage rate (basic), care staff | sqrd | -3.89E-05 | -1.28 |
| *Local demand, supply characteristics (not home specific)* | | | |
| Property prices | sqrd | 1.16E-10 | 1.46 |
| Wages: female, manual gross wage | sqrd | -1.55E-05 | -1.05 |
| *Constant* | | | |
| Constant | | 0.23 | 0.04 |
| Model | | Pooled OLS | |
| Dependent variable form | | Linear | |
| R-sq: | | 0.50 | |
| | adjusted | 0.38 | |
| Number of obs | | 122 | |
| | | ***Stat*** | ***DF*** |
| Overall fit | F | 4.44 | 22, 99 |
| Normality | Bowman-Shenton chi$^2$ | 6.27 | 2 |
| Specification | Ramsey's RESET F test | 2.03 | 3, 96 |
| Heteroskedasticity | Cook-Weisberg chi$^2$ | 0.01 | 1 |

Table 8-16 report the results of the estimation of function (8-52) i.e. a reduced-form estimation of client dependency using home characteristics. Cross-prediction for the non-contingent contracts sub-sample produced a mean predicted (inverse) Barthel score, $\beta_i^{10}$, of 9.80. For comparison, the sample average value for that population, i.e. $\beta_i^0$, is 9.40. A paired t-test was performed and rejected the null hypothesis ($p = 0.087$). A paired test was used since we are comparing the actual dependency of each home in the non-contingent contracts group with the value predicted on the basis of its characteristics had it the referral pattern of the contingent contracts group. Looking instead at differences of median values, median actual dependency was 8.95 compared with median cross-predicted dependency of 9.83.

This evidence of (modest) cream-skimming can be expressed in terms of the impact on marginal cost. We have an estimate of the marginal effect of a change in dependency on marginal cost, holding output constant: $\left.\frac{\partial c}{\partial \beta}\right|_x = \frac{\partial \bar{c}}{\partial \beta}$. Since we are only considering cream-skimming with regard to our dependency score, inverse Barthel, we use the estimation without confounding variables, that is (8-45). The marginal effect is therefore £8.54 per point. Applying this estimate to the difference between actual and referred dependency, we have estimates of the effects of cream-skimming on mark-up (prices being unaffected by cream-skimming) of £3.44 for the mean difference and £7.52 for the median difference.

Overall, we do find statistically significant cream-skimming behaviour by non-contingent contract providers. However, the point estimates suggest that cream-skimming is modest in size (only 1.2% and 2.7% respectively of average weekly price). It implies there are barriers that limit free-replacement of users – proposition 3 of chapter 7. Any estimate of cost exaggeration from comparative mark-up rates ought to be increased by this amount of cream-skimming because contingent contract providers could always cream-skim and secure this amount if they so wished.

### 8.7.5 Cost exaggeration

The theory developed in the last chapter suggested that for relatively low contingency contracts the benefits to the provider of cream-skimming might outweigh the benefits of cost exaggeration. This proposition hinges on the actual degree of contingency and also the barriers to selection/re-placement of users involved in cream-skimming. Where the latter are high, cost exaggeration will be comparatively attractive for providers even with minimal contract contingency.

Since the other explanations are unlikely to account for the mark-up premium observed for contingent contracts, we therefore have some support for the cost exaggeration hypothesis. Even assuming away the cost shocks and risk premium effects, the difference in mark-up rates is not an estimate of the total size of cost exaggeration since cream-skimming is possible for providers with non-contingent contracts, despite the existence of barriers to re-selection, unless the later are very high. If cream-skimming effects amounted to around 2% of average weekly price, then adding this amount on to the size of cost exaggeration effects would put the latter at getting on for 10%. Moreover, this figure could be further increased if contingent contract providers were discounting prices in return to the favourable risk properties of their contracts.

## 8.8 Transaction costs

Above we have considered the behavioural consequences of the use of different contract contingencies and so have a view of the relative transaction benefits of these contract choices – see section 3.4 of chapter 3. But for a rounded view we also need to consider the transaction costs consequences. In other words, are contingent contracts more expensive to operate than non-contingent contracts?

Data from the commissioning survey – see chapters 1 and 4 – was used to test this proposition (in broad terms). The commissioning survey indicated whether pricing was typically, first, contingent, and second whether it was contingent on client dependency for each LA in the sample for the purchase of care home services from the independent sector (see section 4.5.2 of chapter 4). Expenditure data for each LA gives aggregate transaction costs for older peoples services (see chapter 6). The approach adopted here is to determine whether any association exists between typical contract contingency and transaction costs. To this end a model was estimated at the local authority level with total transaction costs for older people's service as the dependent variable. The dependent variable was somewhat skewed and leptokurtic as is usual for cost data, but not to a limiting degree, so that simple OLS regression was sufficient. A dummy variable for each LA was created taking a one value if the LA did typically relate prices to client dependency in a systematic way. In addition a second dummy was used that took a one value if the LA used a pricing arrangement such that *a price that can vary by (publicly-funded) client/resident (and so can vary by provider)* i.e. option 3 of the survey question about pricing behaviour – see figure 4-5 of chapter 4 for descriptive information. These data were only available for the year after the expenditure data and so we assume that local authorities did not make wholesale changes to their commissioning arrangements in that period.

A number of contract factors were used including the scale of an authority's service activity and prevailing inputs costs (wage rates). The regression model demonstrated reasonable fit and satisfied the usual diagnostics. The results are shown in Table 8-17. The *dependency contingent contract* dummy was significant and positive and so supports our hypothesis. Local authorities that typically use contingent contracts of this type appear to have greater transaction costs other things being equal. Indeed, it accords closely with intuition that a contract with a single flat rate – generally set on an historical basis – is less costly to operate.

**Table 8-17. OLS regression – Older people care transaction costs (total)**

| | coeff | t-stat |
|---|---|---|
| Dependency contingent contract | 384.78 | 1.96 |
| Contingent pricing | 142.63 | 0.59 |
| Supported indie residential homes | 1.07 | 2.69 |
| Supported LA homes | 2.26 | 2.97 |
| Supported indie nursing homes | 2.23 | 3.48 |
| Home help hours | 908.27 | 1.99 |
| Wage rates (log) | 3,706.22 | 4.39 |
| Year | 221.94 | 1.18 |
| Constant | -22,318.60 | -4.45 |
| Observations | 179 | |
| R-sqd | 0.57 | |
| Het. Test | 57.10 | |
| Specification: Ramsey Reset | 0.83 | |
| Normality: Bowman-Shenton | 18.12 | |

## 8.9 Conclusion

The empirical investigation of contract arrangements as used for the purchase of residential care for older people largely supports the theoretical hypothesis of the chapter 7. Contingent contract providers have higher mark-up rates than non-contingent providers, and whilst there is a range of possible explanations for such a result it seems clear that information rents from cost exaggeration play a part.

The derivation of mark-up rates requires a number of parameters and these are estimated from the empirical models. The derivations use point estimations of these parameters, and although sensitivity analysis is used, there remains a degree of uncertainty about the derived estimates. Nonetheless, the focus here is not prediction but rather hypothesis testing. And although the estimates fall into confidence intervals, these confidence intervals do not encompass the null hypothesis position (usually zero) – in other words, all the relevant estimates are statistically significant.

Overall then the data support the main theoretical hypothesis: that providers with contingent contracts secure greater mark-up than those with non-contingent contracts, other things equal; and that satisficing providers do operate in the market, and in doing so secure lower mark-ups than profit maximising providers. Are there any immediate implications? Contingent contract providers appear to charge the purchaser more than non-contingent contracts, and they secure greater profit mark-ups. In part, this entails a redistribution of rent from purchaser to provider, which from a societal perspective is far less onerous than the deadweight losses that constitute the remaining part of the difference in mark-up we see. Moreover, there may be unmeasured benefits from using contingent contracts, especially for example in regard to the long-run stability of the market. Since prices are heavily regulated by the purchaser (at least through its monopsonist buying position) prices are likely to respond slowly to failing supply. At a time of significant income squeeze on providers, the modest flexibility offered by contingent contracts could be the difference between market contraction and wholesale fallout.

Furthermore, providers with contingent contracts have higher dependency users than non-contingent contract providers (see above). As discussed in section 3.5.2.1, and supported by the results of this chapter, the serving of higher dependency clients is a valued outcome for purchasers. Contingent contract providers will also have higher utility, possibly impacting positively on quality of care. Contingent contracts therefore have lower cost to outcome efficiency (i.e. higher mark-ups), but also outcomes of higher value to purchasers. This latter benefit also has to be set against the higher transaction costs contingent contract incur.

A number of other conclusions can be drawn. First, providers do appear to exploit to some extent the complex interplay of information that characterises social care transactions (by exaggerating costs and skimming clients). Second, that cost exaggeration dominates rents from cream-skimming, where the latter, although appearing to happen, does so on a very small scale. Third, that the potentially beneficial risk properties of contingent contracts do not show up as price discounts, or at least are not as large as information rents. This finding does not imply that contingent contract providers gain no utility from reduced risk, or indeed that the utility they gain is modest. It only suggests that any risk effect does not translate substantially into compensatory risk premiums for non-contingent providers. Of course, where mark-up rates are very low, the opportunity for such premiums to restore utility equilibrium between providers with different contract types, *ceteris paribus* may be limited. Fourth, that not all providers are motivated solely by profit making. Finally, that marginal cost is falling at sample average outputs.

## Annex 8-1. Sensitivity of the motivations parameter

The following table shows different calculated values of the $\alpha$, for combinations of 10% and 20% deviations from point estimates.

**Table 8-18. Proportional sensitivity analysis of $\alpha$**

| Demand and supply slopes by contract type | | | | Values of $\alpha$ | |
|---|---|---|---|---|---|
| Contingent contract | | Non-contingent contract | | +/- CI | +/-20% |
| Supply | Demand | Supply | Demand | | |
| + | 0 | + | 0 | 0.55 | 0.66 |
| 0 | + | 0 | + | 0.46 | 0.46 |
| - | 0 | - | 0 | 0.55 | 0.44 |
| 0 | - | 0 | - | 0.69 | 0.69 |
| + | - | + | - | 0.69 | 0.83 |
| - | + | - | + | 0.46 | 0.37 |
| - | - | + | + | 0.93 | 0.70 |
| + | + | - | - | 0.41 | 0.50 |

# Annex 8-2. Calculating mark-up

**Table 8-19. Derivation of mark-up rates**

| Estimate | Function | Ref | Contract type | | | Diff |
|---|---|---|---|---|---|---|
| | | | | Cont | Non-cont | |
| | | | $Y$ | $Y=1$ | $Y=0$ | |
| Demand price slope | $h_x^{R^Y}$ | (8-22) | -1.77 | -2.94 | -1.11 | |
| Supply price slope | $a_1^Y$ | (8-25) | 0.61 | 1.26 | 0.25 | |
| Equilibrium price | $p^{*Y}$ | (8-46) for $Y=$ 1 & (8-47) for $Y=$ 0 | 280.99 | 296.51 | 271.75 | 24.76 |
| Equilibrium output | $x^{*Y}$ | | 31.03 | 27.31 | 39.21 | -11.89 |
| Motivation | $\alpha = \dfrac{a_1^{Y=1} - a_1^{Y=0}}{h_{x_i}^{R^{Y=0}} - h_{x_i}^{R^{Y=1}}}$ | (8-31) | 0.55 | 0.55 | 0.55 | 0 |
| Marginal cost slope | $c_x = \dfrac{a_1^{\bar{Y}} - \alpha h_{x_i}^{R^{\bar{Y}}}}{\frac{1}{2}\alpha + \frac{1}{2}}$ | (8-32) | -0.47 | -0.47 | -0.47 | 0 |
| Intercept marginal cost | $\bar{c} = p^{*Y} - a_1^{\bar{Y}} x^{*Y}$ | (8-13) | 262.00 | 262.00 | 262.00 | 0 |
| Marginal cost | $c^Y = \bar{c} + c_x x^{*Y}$ | (8-5) | 247.39 | 249.14 | 243.54 | 5.60 |
| Average variable cost | $\left(\dfrac{C^V}{x}\right)^Y = \bar{c} + \frac{1}{2} c_x x^{*Y} = c^Y - \frac{1}{2} c_x x^{*Y}$ | (8-8) | 254.70 | 255.57 | 252.77 | 2.80 |
| Potential p - c | $p^{(\alpha=1)Y} - c^Y = M^Y = -h_{x_i}^{R^Y} x^{*Y}$ | (8-27) | 54.82 | 91.37 | 34.45 | 56.92 |
| Actual p-c | $p^{\alpha^Y} - c^Y = m^Y = a_1^Y x^{*Y} - c_x x^{*Y}$ | (8-28) | 33.59 | 47.37 | 28.21 | 19.16 |
| Pot p - C/x | $p^{(\alpha=1)Y} - \left(\dfrac{C^V}{x}\right)^Y = \Phi^Y = M^Y + \frac{1}{2} c_x x^{*Y}$ | (8-29) | 47.51 | 84.94 | 25.22 | 59.72 |
| Actual p - C/x | $p^{\alpha^Y} - \left(\dfrac{C^V}{x}\right)^Y = \phi^Y = m^Y + \frac{1}{2} c_x x^{*Y}$ | (8-30) | 26.29 | 40.94 | 18.98 | 21.96 |

# Chapter 9. Conclusion

## 9.1 Context and subject

This thesis is concerned with questions of (economic) organisation, particularly those at a strategic level. It has attempted to model and analyse the broad organisational options that apply in the social care system in England. These are options about the processes and activities that facilitate the provision of services like social care. The way in which these enabling processes are structured and configured impacts significantly on how well services can be levered to address people's care needs.

There are broad organisational questions, a key one being the relative merits of adopting market-like arrangements in social care as opposed to using bureaucratic or hierarchical arrangements (or indeed hybrids, like 'networks'). Some might argue that network arrangements are sufficiently distinct as to warrant a third category in this comparison. Section 3.3.3.2 in chapter 3 considers this argument but rejects it here for two main reasons. First, all forms of organisation are embedded in social environments and are lubricated by trust – networks are not unique (Granovetter, 1985). Second, the treatment of markets and hierarchies is slightly different from the treatment in the networks literature (e.g. Rhodes, 1997). The market governance archetype is centrally constructed around the idea of *bilateral* (rather than hierarchical) relationships and reduces the distinctiveness of networks. There are also more specific questions that concern the comparative benefits of different contracting arrangements in markets. These questions are highly relevant to the policy debate, but have not been comprehensively addressed, particularly as regards the social care system in England.

This work adopts a *transaction cost* perspective. The main research hypotheses are as follows. First, that the transaction costs generated by (public sector) hierarchies in social care are *lower* than those generated in (quasi-) markets. Second, that production costs in hierarchies are *greater* than in markets. Third, that the choice between contract type (contingency) under market governance has a significant impact on provider behaviour. Specifically, contingent contracts are associated with higher prices and mark-up rates (and potentially lower transaction benefits) and also greater transaction costs. A corollary to the first and second hypotheses is that the *difference* in total public expenditure on residential care for older people i.e. production *and* transaction costs between hierarchies and markets, will be less than the difference in production costs/expenditures alone. This corollary is especially relevant because much of the pro-market rhetoric has been justified on the basis of the difference in the latter i.e. in (unit) production costs alone between markets and hierarchies.

## 9.2 The motivation

The motivation and hoped-for contribution of this research thesis in attending to these questions is two-fold. The first is to address perceived limitations in the current body of organisational economics theory as it applies to questions of comparative public sector organisational choice. The second, and greater priority, is to inform the empirical and policy debate on social care reform (this work is *applied* in its nature – the primarily point of the first contribution is to underpin the development of the second).

### 9.2.1 Theoretical questions

As described in chapter 2, economics of organisation theory was used to develop the model. This theory as currently developed in the literature provides a sound foundation, but it does have some limitations here. First and foremost, whilst this body of theory is inherently *comparative* with regard to governance choices, and where the comparison is between *feasible* alternatives, the theoretical treatment of how hierarchies work, and in particular how *public sector hierarchies* work is not well developed. The literature concentrates on private hierarchies in the form of capitalist firms (operating in markets).

Second is the issue of the treatment of stakeholder motivation and the influence of social context on transactions, as outlined in section 2.4.2 of chapter 2. In social care the relevance of non-profit motivations is high (even for private providers profit seeking appears to be only part of the story – see Le Grand, 1997; Kendall, 2001; Forder, 2000). There are a number of specific limitations to address. In particular, how to specify individuals' objective functions, and especially how to move away from assuming profit maximisation. There is also the treatment of individual's expectations about the behaviour of others, and especially how much individuals are willing to *trust* each other.

Finally, incomplete contract theory leans heavily on the concept of bounded rationality, but is somewhat selective in applying it. Essentially, bounded rationality rules out some behaviour e.g. the forming of complete contracts, but is absent in respect of other behaviour.

### 9.2.2 Policy questions

The policy reforms that have shaped the welfare state in England – including the social care system – have concerned, in significant part, decisions about system organisation. Many commentators see reform of the welfare state as resulting from an eclipsing of the social democratic tradition by neo-liberalism (Le Grand, 2003). The characteristic difference between these doctrines is not so much about the goals of the system[58], but rather the means employed to achieve them. Neo-liberalism advocates choice and competition amongst providers – government steering but not rowing – rather than municipal bureaucratic means (see chapter 4). Moreover, neo-liberalism is not just about a more diverse supply side, but instead an active push for the textbook model of competitive markets and entrepreneurial behaviour wherever possible (Forder, 2002).

And as outlined in chapter 4, social care is no exception, and indeed has gone further than most areas of the English welfare state in embracing markets. New Public Management (NPM) – the applied face of neo-liberalism – was in the ascendancy in social care from the 1980s. As fuelled by the availability of income support to fund independent sector care home placements, the use of markets was firmly established by the early 1990s. From then on the importance of *quasi*-markets has increased. In the care homes sector there has been a steady decline since 1993 of the proportion of placements made to in-house providers under public hierarchical means.

Furthermore, where markets have been adopted, at least in social care, they have tended to be modelling on the textbook market. There is, for example, still a significant reliance on short-term, fixed price spot contracts. Serious questions can be raised about whether markets are not only appropriate in general, but also more specifically, whether *textbook* markets are the best model.

A change in the means of securing services is consistent with a change in *governance structure* as outlined in chapter 3 (see Table 3-1). The NPM approach involves a change

[58] Although neo-liberalism places more emphasis on individual freedom than social justice.

in ownership and control over provision from government to private auspices. Choice and diversity is accommodated by devolution of responsibilities away from the centre. The approach emphasises formal contract use, and high-powered incentives. The literature review of chapter 2, however, indicates a paucity of *comparative* economic evaluation of organisational changes of these kinds in social care.

## 9.3 Methods

The methodology employed here has strong analogies with that used in the mainstream economics literature (see Friedman, 1953). In brief, a model (of social care governance choices) was developed. The modelling produced a number of testable hypotheses. Empirical counterparts of the theoretical hypotheses were then developed. Finally, regression analyses were used to estimate the empirical specifications and test the relevant hypotheses.

There are a number of salient methodological points. To begin with, the broad methodological approach used here includes having strong theoretical foundations, both to develop precise theoretical hypotheses and also to support a well-specified empirical analysis. This approach helps to avoid circularity or tautology i.e. letting function lead form, or in other words to allow specifications to anticipate and accommodate the expected empirical findings.

A bespoke theoretical model was developed for this purpose, building closely on the contract theory literature as reviewed in chapter 2. The model focuses on transactions between a public sector, publicly-funding purchaser and a provider, assuming that whilst each strive to make optimal decisions, they are heavily constrained by the transaction costs of measurement and contracting/bargaining. In addition, providers can have non-profit motivations to balance their profit-seeking goals.

The model is multi-period with uncertainty, allowing for reputation effects to modify behaviour, particularly in the presence of asymmetric information. One of the main features of the model (see chapter 5 in particular) is the treatment of production 'effort' in both quasi-markets and public sector hierarchies. In the former, the product contract has a key place in inducing 'directed' effort from providers. In hierarchies, effort is forthcoming to the extent that employees are willing to follow the instructions of managers. But the model explicitly recognises that managers may knowingly allow some 'shirking' on effort (in addition to 'hidden' shirking) in certain circumstances. In order to produce unambiguous hypotheses concerning transaction costs in quasi-markets and public sector hierarchies respectively (reasonable) assumptions were required about the shape the transactions cost function.

The theory underpins a systematic empirical analysis of care home services – at local authority and care home level – for older people. A broad definition of transaction cost was deemed appropriate. Total expenditure/cost, data available for older people, is the sum of transaction and production costs. Since, the latter are also available, transaction costs are then found by subtraction (implying that all other non-production expenditure is transaction cost). Transaction costs, as used here, are those that apply to the local authority. For hierarchies, this definition means that all transaction costs are included. For markets, providers may also incur (unmeasured) transaction costs that would be added to (local authority) purchaser transaction costs to give a total. Nonetheless, it follows that if purchaser transaction costs in markets exceeds hierarchy transaction costs, then the same must apply for total transaction costs (see section 5.7 of chapter 5 and section 6.2.2 of chapter 6). In addition, although this residual definition of transaction costs may include some expenditure that stretches a stricter definition of transaction cost, this analysis is comparative, and requires only a consistency if definition as it applies to both market and hierarchy arrangements.

The main empirical task in this study is the attribution of total transaction costs to governance alternatives. The empirical indicator for 'governance archetype' in the local authority level analysis – as described in chapters 5 and 6 – was based on ownership distribution. A placement was made under hierarchical arrangements if ownership of the means of purchasing and provision were *integrated* (i.e. both public sector), and market if ownership was dispersed. Although this focus is only on one element of organisational structure (see table 3-1 in chapter 3), synthetic regression using data on control and incentives, as well as ownership, at local authority level – but with a smaller sample size – supported the use of ownership distribution as a summary statistic. In practice, local authorities use a mix of governance arrangements. Exploiting this variation, the relationship between supported placement mix by governance type and total transaction costs was modelled (see chapter 6).

As regards contract contingency (chapters 7 and 8) the data were more straightforward – homes were asked with which contract types they typically operated, and a dummy variable was created on that basis. The contingent contract analysis explored how transaction and production costs differ by contract type. In addition to a comparison of prices – as an indicator of production cost – between different contract types, the analysis also compared mark-up rates. If the underlying efficient relationship of cost to outcome is a technological relationship that is the same for all homes, then mark-up describes the cost of securing this relationship i.e. is a better indicator of transaction benefits.

The attribution of transaction costs to governance structures at the local authority level may incur the problem of 'ecological fallacy'. Methodologically, the regression analyses and also diagnostics, especially the specification tests, minimise this risk by allowing multiple simultaneous explanatory factors to be included in the estimation. Conceptually, moreover, decisions about governance archetypes – market or hierarchy – *are local authority-level strategic choices* and apply equally to all transactions with the authority.

The theoretical model provides a general foundation on which to base the empirical specification of the regression model. As regards the transaction cost analysis two specific functional forms were estimated. The first was a flexible functional form, specifically a generalised translog cost function. The second specified transaction costs as a function of the ratio of places by governance archetype.

The home level analysis of contract contingency estimated a demand and supply system with contract type able to both shift demand and affect the slope of the demand curve. Various regression techniques were used to interrogate the data. They were chosen to address both the characteristics and usual nuances of the data e.g. skewed cost data and panel data, and also, *a priori* considerations such as endogenous variables e.g. in the demand and supply analysis.

The applied econometrics literature has developed considerably in recent years. Whilst, most have been incorporated in this analysis – e.g. generalised linear models with panel data – one interesting development not included is the spatial econometric approach. This approach explicitly allows for some un-specified correlation of behaviour to exist between neighbouring areas e.g. neighbouring authorities tending to adopt similar governance arrangements. Further analysis may be of merit along these lines.

Generally, support or rejection of theoretical hypotheses could be reduced to the checking of the statistical significance and sign of the relevant estimated parameter. In other places statistically significant differences in parameter estimates were important. In these cases, in particular, sensitivity analysis was carried out. There were no cases where the findings of the sensitivity analysis materially changed our understanding of the relationships in question. Furthermore, we need to recognise that the estimates, although derived whilst accounting for many relevant control factors, can still mask some heterogeneity – hence the actual market and hierarchy costs will differ between authorities (Stevens and Normand, 2004). So we need to be cautious in making too wide a generalisation. Overall, regression analysis is at least part 'art' as well as science.

The general positivist methodology used in economics and largely mirrored in the present study has been criticised. Whilst this is not the place for a methodological debate, it is worth noting the criticisms that have been made in the context of the institutional economics literature. In particular, Hodgson (1988) argues that empirical analysis and theoretical hypothesis generation cannot be fully divorced. This can be interpreted as saying that empirical testing is not atheoretical – that we look for what we expect to find. Hodgson does not have a compelling alternative manifesto in methodological terms, but does argue that we need to synthesise according to the weight of evidence. Little comparable work to the current study has been undertaken – as outlined in chapter 2. In Hodgson's terms, this work can therefore be seen as a contribution to the weight of evidence.

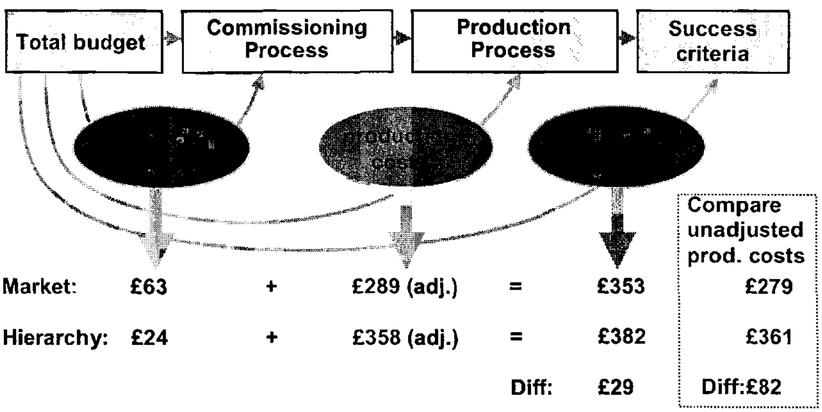
## 9.4 Local authority level markets and hierarchies analysis

### 9.4.1 Results

The estimation results were highly consistent between the different empirical models and they all supported the theoretical hypotheses. For both transaction and production cost estimations the outputs ratio indicator of governance archetype was significant with the expected sign. The translog models suggest that marginal transaction costs for hierarchical placements are low: the point estimate was £6 per place per week with estimated confidence intervals including zero. For markets, the point estimate was £41 per place for market places and this was significantly different from zero. The results suggest that fixed transaction costs are sizeable. Further analysis found that average transaction costs were £21 and £56 per place per week respectively. For production costs, a significant difference was found in the other direction: £89 for hierarchy and £55 per place per week for markets at the margin. Overall, the total (production + transaction) costs model could not find a significant difference between market and hierarchy, which is consistent with the findings for the individual costs.

The results are given in Figure 9-1 with prices inflated for the 2002/3 financial year. Using point estimates from the empirical analysis, average transaction costs per placement per week in markets were £39 per place per week higher than in hierarchies. Production cost estimates were derived from home level regression analysis of the PSSRU residential care survey (see Netten et al., 1998). These analyses controlled for differences in home input costs, user characteristics and outcome indicators.

**Figure 9-1. Production and transaction costs – residential care for older people (2002-3 prices)**




The costs reported in the figure are the predicted costs for market and hierarchical providers holding constant all other variables in the regression; the resulting difference of £69 in the figure is due only to provider (i.e. governance) type. Adding these values together gives total costs, with a difference of £29 higher in hierarchies than markets.

The same residential care survey provided unadjusted 'production costs' with a difference of £82. Much of the *marketisation* (NPM) literature has focussed on this latter difference as a justification of the comparative efficiency of market-based provision.

Overall, the results support the view that after accounting for the full range of costs, the efficiency shortfall associated with hierarchical organisational arrangements may be far less than often imagined. Indeed, the findings of chapter 6 indicate that total costs may not be significantly different.

### 9.4.2 Policy implications

There are direct policy implications from this analysis concerning organisational arrangements in publicly funded care home services in England. There are also implications looking at the wider markets-versus-hierarchies debate for public services.

#### 9.4.2.1 Implications for care homes sector

The immediate implication of these results is that policy makers should take a pause in continuing to shift the organisation of social care to market auspices. Central to the community care reforms of the early 1990s, as outlined Chapter 4, was the explicit separation of purchaser and provider and use of market arrangements (see section 4.2.2). Figure 4-2 shows the steady decline in the number of local authority places and so decline of the use of hierarchical arrangements – to less than 10% of the total by 2003. The analysis above suggests that the total cost savings from doing so, have been modest at best.

The analysis suggests that local authorities need to downwardly revise any cost saving they expect from using markets. At the very least, councils should drop any pre-conceptions that markets are *automatically* more efficient, as implied by much of the New Right/NPM and, indeed, current Labour rhetoric. The analysis highlights the importance of a comprehensive account of total (gross) cost/efficiency differences.

Comparative *total* cost has certainly not been the only criteria that bears on the decision to out-source provision to the market (which is why we cannot read the observed marketisation trend as implying that markets are more efficient). Aside from the many political and other factors, we need to also be clear that the total cost of publicly supported placements is not equal to the net cost that local authorities have to meet from their own budgets (because there are service co-payments – see section 4.4.2.2). Moreover, because there have been differences between the charging and income rules for people in local authority care homes compared to private care homes, the net impact on local authority budgets of market or hierarchy choices has differed from the gross cost impact.

The now abolished (in 2003) residential allowance is an important example. Other things equal, this social security benefit subsidised the local authority for placements in the independent sector care homes. Public sector rules on capital expenditure also affect the markets and hierarchies cost comparison. Institutional public sector accounting rules can in some cases artificially increase the costs of investment in hierarchies as discussed in section 5.5.2 (although there are methods to address this issues – see below). In any case, central policy makers need to look carefully at any practice which subsidies the *net* total cost of market-based (or conversely hierarchical) placements.

The analysis suggests that any efficiency advantage of markets will be highest when markets are competitive, i.e. where the purchaser has a good choice of potential providers. Purchasers do, nonetheless, need to be careful about over-zealous use of monopsony market power. Moreover, there is evidence (in healthcare) of an inverse relationship between quality and competitiveness (Propper, Burgess, and Green, 2004).

Excessive downward pressure on prices from significant monopsony market power could reduce the overall sustainability of market supply. The analysis of chapter 8 (also see below for a summary) suggested that profitability has been very low in the care homes market, and particularly low in the period to which the data here apply (Laing, 2004). The contention has been that market prices were artificially low, and indeed the prices that local authorities have paid in the market has increased (in real terms) in the recent period (Laing, 2004, chapter 1). For example, between 2000/1 to 2003/4 residential care prices (*unadjusted* production costs in terms of the analysis above) increased by over 10% in real terms[59]. These figures are unadjusted for client dependency mix, but the implication is that if market prices at the time of the analysis were lower than their long-run stable values – and that hierarchical unit costs were not similarly depressed – then actual total cost differences may be even less than those reported in Figure 9-1.

The production costs used in the comparisons above are adjusted for case mix, local input prices and some process aspects of service quality at the individual level. But this is someway short of a full adjustment for different outcomes – of individual users and at system level. There is little systematic comparative evidence of user outcomes between care home sectors. But anecdotally, users favour in-house/hierarchical care service providers. A large scale user satisfaction survey of over 18,000 *home care* users reported service quality that was significantly greater for in-house providers ($p < 0.001$) (Netten et al., 2004). If these results were carried over to the care home sector, it would add further weight to the normative efficiency case for hierarchies. It should be noted that any comparative evaluative implications only strictly apply for small changes in the mix of market and hierarchical provision. For example, even if hierarchies did have a total cost advantage after accounting for the above factors, this certainly need not suggest a wholesale return to in-house provision. At the very least we would expect some fringe competition effects to work across sectors, which would have only minimal effect if the market sector were small.

Aside from the question of the appropriate balance between markets and hierarchies, the analysis also suggests a number of ways to improve performance under either governance arrangement. Responsiveness, productivity and perhaps, innovation – or together in the terms of the analysis above, effort shirking – are weaknesses of hierarchies, especially public hierarchies, but can also affect quasi-markets.

Some part of this problem is on the purchaser side and holds for both provider hierarchy and quasi-market (indeed it is a problem with hierarchical purchasing, not with the supply side arrangements). In social care, the local authority is the purchaser and so

[59] Personal Social Services expenditure and unit costs: England: 2000-2001 and Personal Social Services expenditure and unit costs: England: 2003-2004, Department of Health. Comparison of unit cost indicator 2.6, table 6.

needs to anticipate the range of preferences of users. If this is not done well, then poor outcomes are likely, either in markets via a poor product contract specification, or in hierarchies via a poor effort instruction set.

The other part of this problem is on the supply-side. In markets, providers have the incentive to lead innovation and create new markets, pushing these innovations up to the purchaser. In hierarchies, providers have the incentive to do the opposite. And hierarchical managers will be concerned not only with purchasing but also with the functioning of their provider units.

Strengthening the responsiveness of the purchasing function around user needs is a main objective of the system (Department of Health, 2005, 1998), and it is a particular problem where provision is also hierarchically arranged. Since purchasing is all about ascertaining needs and securing services on behalf of users, this in turn requires the organisation to have a user focused *corporate culture*. With regard to public services, instilling this culture can be by influence, consensus management, a reliance on 'champions' such as the National Directors, and most directly, better leadership (Plsek and Wilson, 2001). Forcing structural change can also help instil this culture. In this case, structural change is about making a stronger purchaser-provider distinction and empowering the purchaser function. In hierarchies, it can be achieved by putting purchasing functions (e.g. care management) in hierarchical superiority to providing functions. Ultimately, a purchaser provider distinction can be forced by separating ownership, out-sourcing provision, or in other words, implementing a (quasi) market.

Whilst distinctive and empowered purchasing is important, there remains the question of how to motivate purchasers. In relation to public services, this is primarily a task for Government. The analysis above has compared social care hierarchies and quasi-markets and so concerns the (principal-agent) relationship between purchaser and provider. Motivating purchasers is about the (principal-agent) relationship between government and purchaser, which is in practice hierarchically configured (whilst it might be possible to out-source social care purchasing to a market, this would require primary legislation). The nature of this latter relationship need not differ as between (supply-side) hierarchies and quasi-markets and to date has not, with purchasing residing with the social services department (or PCT).[60] It is therefore outside the scope of this work – although measures to improve hierarchical functioning between purchaser and provider, which we turn to below, also largely apply to the government purchaser relationship. In particular, performance management via the use of *targets* and *selective incentives* applies in this relationship (see below).

In the conventional model of hierarchies, purchaser managers direct the production activities of providers. These are often undertaken in an adaptive and iterative way until the desired result is achieved. Moreover the provider employee accepts these activity instructions (to a point) and carries them out, ultimately motivated by the sanction of dismissal. In markets, a product specification and being residual claimant (high-powered incentives) motivates provider effort.

Improvement in productivity (i.e. effort) in hierarchies is possible from the introduction of the high-powered incentives. The internal market model involves the *widespread* introduction of these incentives in hierarchies (this model falls short of a full market in that providers – being public sector organisations – are not absolute residual claimants). But a more *selective* introduction could improve effort/productivity without requiring a negotiated and verifiable product contract, hence limiting the extra transaction costs. A straightforward example is to link employees' salaries to target achievements, such as process targets – e.g. from key performance indicators, including user satisfaction surveys results, reductions in delayed discharge, reduced user waits for assessment or care packages and so forth (Commission for Social Care Inspection, 2004) – or even changes in user outcome if a robust measure can be found (Netten et al., 2002).

The performance assessment framework (PAF) for councils – that culminates in the *star ratings* system – is an example relevant to the government (DH and CSCI) and purchaser (council) relationship. In practice, the financial incentives for councils in this regard are limited, and this system tends to be more of a minimum performance

[60] But this is changing now with the growth of user commissioning via direct payments or personalised budgets. In this case, users take on some of the purchasing function and because they are private individuals, they can only have a bilateral i.e. market relationship with formal providers, not a hierarchical bureaucratic relationship.

assurance mechanism with zero-rated councils subject to direct intervention by external management teams.

'Targets' in hierarchies have analogy with the product contract in markets in that they are explicit and (somewhat) verifiable, and relate to the output of the organisation. To a certain extent, transactions costs can be kept down because the purchaser manager unilaterally decides these targets and their achievement. However, the greater the proportion of provider employee's incomes that is contingent on performance, the more they are likely to (a) bargain and negotiate with the purchaser manager over targets and (b) challenge, and demand verification of target achievement. In other words, the widespread use of performance incentives will incur market equivalent transaction costs.

The selective use of incentives to augment a basic (low-powered i.e. salary) income could lead to net benefits, but there are further caveats. A main one is that selective targets might lead to partial and misdirected effort (in the same way as incomplete product contract specifications – see section 5.5.3 in chapter 5) (Propper and Wilson, 2003).

The concept of devolved management structures resonates with the above. Rather than having the top of organisation provide a stream of effort instructions to lower levels, the organisation could adopt a more decentralised structure. Each division is tied back to the top using arrangements that are more like product contracts than effort instructions, but without the high-powered (residual claimant) incentives. To keep transaction costs down, these strategic agreements need not be fully specified. To keep production directed, residual claimant incentives to cut costs are softened. The benefits are more efficient information-instruction flows. In social care terms, this can mean either functional or geographic decentralisation of the social services department.

Care home providers in all sectors are subject to inspection by government regulators (see chapter 4). There are 38 National Minimum Standards (NMS) that apply – with remedial action such as loss of registration status or legal action taken in the case of (gross) failure to achieve the standards. These regulations add to, and overlap with the direction of provider activity that comes from either purchaser instructions in hierarchies or product contracts in markets, and clearly add transaction costs.[61] In the past, in-house providers were exempt from such inspections, and that is indeed more consistent with the implications of the analysis above (i.e. hierarchical providers have few incentives to cut quality unless higher quality requires higher productivity). The planned move to 'lighter touch' regulation would address this issue since providers meeting standards (i.e. more in-house providers?) would be subject to fewer inspections.

Other elements of market governance can be incorporated into the hierarchical model, with potential net benefits. An example is public and private partnership (PPP) schemes in relation to investment. In terms of the model above, they involve investment being outsourced to the market, but production undertaken by hierarchical means, i.e. publicly operated. There is a significant literature on the relative merits of these arrangements, but this is out of the scope of the current work (for example Dawson, 2001, Propper and Wilson, 2001). Nonetheless, an immediate observation is that PPP and PFI schemes do involve very significant transaction costs that would only be justified if public investment opportunities were highly constrained (see section 5.5.2).

Turning to *market arrangements*, a major source of transaction costs is the product contract. The analysis of chapter 5 suggests that poorly specified product contracts in relation to complex services can lead to misdirected effort. Nonetheless, in frequently repeated, long-term contractual relationships, it is reasonable to suppose that purchasers and providers will have become well acquainted with what constitutes directed and misdirected effort. Detailed product contract negotiation and specification could then be scaled back with purchasers more trusting of providers to produce directed efforts. Providers could benefit if some of the savings on transaction costs were available as income. These so-called *relational* contracts (Dore, 1983; Sako, 1992) can be effective and low cost (see chapter 5, especially section 5.3.1.2 and chapter 2). They rely on on-going trust in relationships and the alignment of motivations between purchaser and provider that comes from each sharing the same social networks.

Attempts to move toward a higher trust, lower cost relationship have been attempted (as noted in chapter 4). For example, the Government's 2001 *Building Capacity and*

---

[61] Which at the time was included in local authority transaction costs as used in the analysis above.

*Partnership in Care* (BCPC) charter, which called for working that "promotes mutual trust" (see page 89 above). The aim was to form an 'agreement' between statutory and independent sector. BCPC provided a checklist of appropriate behaviours and therefore offered guidance to the sector. However, it did not come with specific powers or incentives to prompt councils to adopt a more relational stance with independent sector providers. Nonetheless, the Department's Change Agent Team (CAT) has worked on the ground with councils to promote this style of working, and has developed a toolkit, *Commissioning and the Independent Sector*, which checks compliance with BCPC (Change Agent Team, 2004).

In recent years, relationships in the sector have improved, but can hardly be described as relational – see for example the demands of the English Community Care Association (of independent providers). There has been a decade of very tight financial settlements for care homes (see chapter 8 and below) and provider turnover is significant, limiting long-term relationships (Netten, Darton, and Williams, 2003). Without good levels of trust, the use of relational contracts would result in adverse outcomes.

Markets also potentially suffer 'hold-up' problems as described in chapter 5, section 5.3.1.1. The most obvious example is under-investment in specialised physical assets – e.g. purpose build nursing homes in low-income areas that have limited private pay demand. Another related example is staff training, which benefits the user (and so purchaser), but where the provider expects a chance to lose the newly trained staff to another provider. In both cases, cost-sharing (CS) contracts can offer benefits – see section 5.2.1.1. They could involve local authority purchasers sharing or matching a part of the investment cost. Currently there are few examples in the field of these kinds of arrangements. There would be further transaction costs in agreeing a matching contract on investment, training etc., but these could be relatively modest. At the very least, recognition of this problem would lead to improvements in performance.

#### 9.4.2.2 General implications

The more general implication of the results is to question whether there is some inherent inverse relationship between the size of transaction and production costs. For example, do measures to promote competition necessarily increase contracting and monitoring costs? Can providers be made more responsive without increasing the power of the financial incentives they face? The empirical analysis is not designed to address these trade-off questions, although it does support the theory that can be used more directly to cast light on these issues. Up to a point, some trade-off must occur, although the model does suggest that the trade-off will be less pronounced in a high trust environment rather than a low trust one. Trust or reputation effects appear to work well when stable; but when trust breaks down the adverse consequences – in the absence of any safeguards – can be severe.

### 9.4.3 Theoretical and methodological implications

As regards the wider body of theory, this thesis has focused on the working of *public* hierarchies. Significant scope remains for further theoretical development however. Indeed this focus is both a strength and a weakness of this work. In terms of the theory, a number of specific issues have been explored. First, a non-profit utility term was used in the objective function. This changed predicted behaviour in the model, although quite often the effect was just an amelioration of the strength of profit seeking behaviour. As a basis for empirical work, where this profit seeking intensity is an empirical question, the value added is not absolutely clear-cut. Second, although there is a large literature on trust, relatively little interfaces with the comparative governance literature. David Kreps' work on reputation is an important exception, and these ideas have been used extensively in this work e.g. in the way that (care) workers trust their managers sufficiently so as to cede significant control to managers. Furthermore, reputation, being a dynamic concept, allows for people's expectations to change, which is also an empirically appealing feature. Third, the treatment here is to assume that people are rational but constrained by the transaction costs of undertaking relevant activities. Hence, people do not write complete contracts because the transaction costs are prohibitively high. It is in this sense that people are boundedly rational.

The empirical strategy has been to directly measure behaviour and infer outcomes (and so efficiency) on that basis. Ideally, empirical work should compare net outcome for stakeholders (weighted appropriately), but this is difficult for usual reasons (e.g. measuring outcomes) (Knapp, 1984). A focus on behaviour rather than outcomes is arguably the next best approach.

## 9.5 Contracts

### 9.5.1 Results

Differences in price and mark-up by contract contingency type can be due to differential information effects, risk premiums, and within-financial-period cost shocks, according to the theory of chapter 7. The information effects include exaggeration by providers of cost-relevant factors – e.g. the costs of services for users with particular needs – in order to secure higher payment, and also cream-skimming. The theory indicates that these factors produce differential effects on behaviour in markets, including the level of prices, outputs, and mark-up rates.

The homes in the sample had a mix of contract types: just under two-fifths reported operating with contingent contracts. There was also a mix within authorities; typically around a third of sample providers within each authority reported contingent contracts.

The main hypothesis – that cream-skimming and risk effects will be limited, and that cost-exaggeration and upward cost pressures effects for contingent contracts would dominate – was supported by the empirical analysis. Prices were statistically significantly higher for contingent contracts compared with non-contingent contracts, by over 5% of the sample average price, all other things equal. This price difference is calculated when holding constant differences in underlying costs drivers, some of which are under the control of providers via selection behaviour. For providers that cream-skim, these cost factors will be lower than for other providers. By controlling for i.e. equalising cost drivers, we are generating predicted prices for this group than are higher than the observed prices. If we did not control in this way, price differences would be greater, but would over-estimate the information effect (because cream-skimming affects costs, and this is not observed unless prices are adjusted for cost differences).

The above price difference is calculated holding output the same between contract groups. However, because a different contract type means different provider demand functions, equilibrium outputs should also change, bringing, in turn an output-related price adjustment (along the demand curve). Accounting for this output adjustment increases the difference in price (at the margin) due to contract contingency to nearly

9%. Output changes also affect marginal costs (because the marginal cost function is downwards sloping). This effect can be estimated by comparing estimated price-cost margins. When this output-related adjustment to costs occurs, the adjusted price difference becomes just under 7% relative to marginal costs or 8% relative to average variable costs (see Table 8-14 in chapter 8).

Overall, whatever particular comparison we make, total 'production cost' to the purchaser (i.e. the price it pays) is higher under contingent contracts. 'Production cost' as defined in this thesis is the price paid (under different governance arrangements) to secure identical services, net of transaction costs i.e. it is concerned with the comparative efficiency of production. In practice, however, not all the characteristics of a service are observable, including, in particular, some client characteristics. Providers with non-contingent contracts would be catering for lower dependency users on average, something that implies lower value for purchasers (but is not reflected in production cost/price differences). Contingent contracts providers have higher production costs than non-contingent contracts to provide an identical service, but also support people with comparatively higher needs, and in a more appropriate fashion. In other words, contingent contracts are comparatively less *cost*-effective but provide a more highly valued or a more *effective* service.

There is also a question about provider sustainability under non-contingent contracts. Although information rents may be limited, providers with contingent contracts should be (a) compensated for facing more risk and (b) be at a disadvantage in a climate of service cost increases (due, for example, to lower need users making greater use of home care, a policy actively promoted by Government). Where these compensations are not forthcoming – for various reasons – providers may be absorbing short run net income shortfalls, a position not sustainable in the long run.

The analysis showed that mark-up rates in the industry were very low at the time of the survey (see also Laing, 2004). The average for the sample was a mark-up over average variable cost of £26 per place (within a confidence interval of £15 to £37) or just over 9% of average revenue. At the margin, mark-up was slightly higher (12%) because marginal costs were falling (see Table 8-11 in chapter 8). Part of the reason for this low rate is that, on average, providers were not making as much profit as demand conditions would allow. Point estimates suggested that only 55% of maximum mark-up was taken. This finding is important because it means that providers are either not solely motivated by profit, or are not hyper-rational (profit) optimisers, or indeed, both. Further pressure on prices may force providers to become more profit orientated in order to survive.

The mark-up analysis has focused on marginal and average variable costs because these are most telling with regard to (short-run) market profitability. Providers also have fixed costs, most predominantly in the form of property capital costs. Whilst, only a portion of these capital costs is a sunk cost (i.e. the property can be sold), providers still need to service this debt to remain in business. When these costs are figured in, mark-up rates fall to zero or even less.

With either contract type, purchasers could increase prices, and have done so since the survey (see above). This action will improve sustainability under non-contingent contracts as underlying costs increase. But it will not lessen the risk exposure of providers with non-contingent contracts. And, although the distribution of user dependency/need within homes may be narrowing (as lower level needs cases are more likely to be supported in their own home), variability remains significant.

### 9.5.2 Policy implications

The analysis suggests that in practice cream-skimming is relatively modest. In this case, the cost-effectiveness shortfall of contingent contracts is the most important comparative disadvantage. Moreover, although risk is affecting providers, it does not appear to be attracting an overwhelming degree of compensation. In addition, as expected, transactions costs were found to be lower under non-contingent contracts.

Taking these together, this analysis suggests that non-contingent contracts – having higher net transaction benefits than contingent contracts – are the preferred choice. This implication is, however, conditional on either limited cream-skimming and/or on purchasers ascribing only a modest increase in the value of services for high needs people rather than low need people. The latter, is unlikely given the way that *Fair Access to Care* eligibility criteria work in social care on the principle of greater attention to those in greatest need (see chapter 4). Furthermore, any policy conclusion about the continued use of low- or non-contingent contracts should only apply within client group – it does not apply for any placement regardless of registration status of the care home. A non-contingent contract might be use for potential placements to a nursing home, but not the same contract (price) for placement to a care home without nursing.

The main concern with non-contingent contracts (within client group) is provider risk. A small interview survey of care homes in the late 1990s found that just under half of providers reported 'excess risk' (Forder et al., 2000). And yet price premiums are not apparent in non-contingent contracts. Some form of extreme 'stop-loss' provision could be introduced alongside non-contingent arrangements for the majority of service users. This would pass some of the cost of very high need cases back to the purchaser. It would need to interface with NHS continuing care arrangements as people with very extensive service needs may already be eligible for NHS care, which pays a (much) higher rate than local authority social services.

Overall, contracting is a core part of the commissioning activities of local authorities. There has been some practical policy development in this regard. The Department of Health Change Agent Team and its Learning and Improvement Network (LIN) on commissioning has made an important contribution (Change Agent Team, 2004, chapter 3). The Audit Commission has also produced some relevant analysis and toolkits (e.g. Audit Commission, 1997a, 1997b). But there is scope for Government to go further in supporting commissioning and contracting performance in local authorities

At the time of the analysis, mark-up rates in the care homes market were low by almost all standards. Market demand has been falling as alternatives to care homes become more readily available (see chapter 4) and with a contraction of a competitive market we would expect to see low mark-up rates and home closures (Netten, Darton, and Williams, 2003). However, because the prices local authorities will pay are now rising again (see above), there is some indication that prices at the time were inefficiently low. In any case, commissioners will need to become much more adept at understanding demand and supply dynamics in their local markets if they are to avoid 'stop-go' like cycles of market capacity. Information on prices, analysis of closures and a mapping of local input prices are important components in this task. A part of this task is also to appreciate that not all providers by any means are purely motivated by profit. Having this characteristic underlined will help purchasers in developing their relationships with providers.

### 9.5.3 Theoretical and methodological implications

The main methodological contributions of the contracts analysis are two fold. The first is the use of residual demand elasticities to calculate mark-up rates by contract type. This approach avoids the difficult task of estimating marginal costs from accounting cost data. It instead infers mark-ups from actual provider behaviour. The second contribution is the attempt to incorporate non-profit motivation into the empirical analysis, and indeed to estimate the average propensity of providers to seek maximum profits. This extends the approach of Forder (2000), and suggests that providers are not always maximising profits. It was not possible, however, to differentiate this motivation propensity between providers with different contracts.

## 9.6 Final comments

This work has been concerned to promote a research agenda in long-term care that has three key components. First, to focus on the organisational aspects of the delivery of services to people with care needs (rather than just looking at different care production technologies). In doing so, to undertake a comparative analysis of *feasible* alternative organisational arrangements, not just comparisons with textbook ideal-types. Second, for the analysis to have strong theoretical foundations and therefore to provide a systematic basis on which to address the data. Third, to undertake a rigorous, quantitative empirical analysis that is relevant to, and able to directly inform policy-making.

**Word count (excluding only references): 98692**

## References


6, P, and J Forder. 1996. Can Campaigning be Evaluated? *Nonprofit and Voluntary Sector Quarterly* 25 (2):225-247.

Alchian, A, and H Demsetz. 1972. Production, Information Costs and Economic Organisation. 5:777-795.

Arrow, K, and F Hahn. 1971. *General Competitive Analysis*. Edinburgh: Oliver and Boyd.

Atkinson, A. B., and J. E. Stiglitz. 1980. *Lectures on Public Economics*. Maidenhead: McGraw-Hill.

Audit Commission. 1997a. *The coming of age: improving care services for older people*. London: Audit Commission.

Audit Commission. 1997b. *Take Your Choice*. London: Audit Commission.

Baker, G. P, and T. N Hubbard. 2001. Empirical Strategies in Contract Economics: Information and the Boundary of the Firm. *American Economic Review* 91 (2):189-194.

Baker, G. P, and T. N Hubbard. 2003. Make Versus Buy in Trucking: Asset Ownership, Job Design, and Information. *American Economic Review* 93 (3):551-572.

Baker, J, and T Bresnahan. 1985. The Gains from Merger or Collusion in Product-Differentiated Industries. *Journal of Industrial Economics* 33 (4):427-444.

Baltagi, B. 1995. *Econometric analysis of panel data*. New York: John Wiley and Sons.

Baltagi, B, and Y-J Chang. 2000. Simultaneous equation with incomplete panels. *Econometric Theory* 16:269-279.

Baron, D. 1989. Design of Regulatory Mechanisms and Institutions. In *Handbook of Industrial Organisation*, edited by R. Schmalensee and Willig. Amsterdam: North-Holland.

Baron, D, and R Myerson. 1982. Regulating a Monopolist with Unknown Costs. 50:911-30.

Barr, N. 1993. *The economics of the welfare state*. London: Weidenfeld and Nicholson.

Bartlett, W. 1991. Quasi-Markets and Contracts: A Market and Hierarchies Perspective on NHS Reform. *Public money and management* 11 (3):53-61.

Bator, F. 1957. The simple analytics of welfare maximisation. *American economic review* 47:22-59.

Bator, F. 1958. The anatomy of market failure. *Quarterly journal of economics* 72:351-379.

Baumol, W.J., J.C. Panzar, and R.D. Willig. 1982. *Contestable markets and the theory of industrial structure*. New York: Harcourt Brace Jovanovich.

Besley, T, and M Ghatak. 2001. Government versus private ownership of public goods. *Quarterly Journal of Economics* 116 (4):1343–72.

Blair, T. 1998. The third way: New politics for the new century. Fabian Pamphlet 588, The Fabian Society, London.

Blaug, M. 1986. *Economic history and the history of economics*. Hemel Hempstead: Harvester Wheatsheaf.

Blough, D, C Madden, and M Hornbrook. 1999. Modelling risk using generalised linear models. *Journal of Health Economics* 18:153-171.

Bresnahan, T. 1988. Empirical Studies of Industries With Market Power. In *Handbook of Industrial Economics*, edited by R. Schmalansee and R. Willig. Amsterdam: North Holland.

Buchanan, J.M. 1986. Rights, Efficiency, and Exchange: The Irrelevance of Transactions Cost. In *Liberty, Market, and State*. Brighton: Wheatsheaf.

Caves, D, L Christensen, and M Tretheway. 1980. Flexible cost functions for multiproduct firms. *Review of Economics and Statistics* 62 (3):477-481.

Challis, D. 1992. Community Care of Elderly People: Bringing Together Scarcity and Choice, Needs and Costs. *Financial Accountability and Management* 8 (2):77-96.

Change Agent Team. 2004. Changing times: Improving services for older people. Report on the work of the health and social care change agent team 2003/04. London: Department of Health.

Coase, R. 1937. The Nature of the Firm. *Economica* 4:386-405.

Coase, R. 1960. The problem of social cost. *Journal of law and economics* 3:1-44.

Coleman, J. 1994. A Rational Choice Perspective on Economic Sociology. In *The Handbook of Economic Sociology*, edited by N. J. Smeltzer and R. Swedberg. Princeton, N.J.: Princeton University Press.

Coleman, J. S. 1990a. *Foundations of social theory*. Cambridge, MA: Harvard University Press.

Coleman, J. S. 1990b. Interview with James Coleman. In *Economics and Sociology*, edited by R. Swedberg. New Jersey: Princeton University Press.

Comas-Herrera, A, and R Wittenberg. 2003. Joseph Rowntree Foundation.

Commission for Social Care Inspection. 2004. Social services performance assessment framework indicators 2003-2004. London: CSCI and National Statistics.

Culyer, A, and J Newhouse, eds. 2000. *Handbook of Health Economics*. Amsterdam: Elsevier.

Cyert, R, and J March. 1963. *A Behavioural Theory of the Firm*. Englewood Cliffs New Jersey: Prentice-Hall.

Darton, R, A Netten, and J Forder. 2003. The cost implications of the changing population and characteristics of care homes. *International Journal Of Geriatric Psychiatry* 18 (3):236-243.

Davies, B. 1992. *Care management, efficiency and equity: The international experience*. Canterbury: PSSRU.

Davis, L, and D North. 1971. *Institutional change and American economic growth*. Cambridge: Cambridge University Press.

Dawson, D. 2001. The Private Finance Initiative: a public finance illusion? *Health Economics* 10 (6):479-486.

Department of Health. 1990. *Community care in the next decade and beyond: Policy guidance*. London: HMSO.

Department of Health. 1991. *Purchase of service: Practical guidance*. London: HMSO.

Department of Health. 1998. Modernising Social Services: Promoting independence, Improving protection, Raising standards. London: Department of Health.

Department of Health. 2002. Fair Access to Care.

Department of Health. 2005. Independence, Well-being and Choice: Our vision for the future of social care for adults in England, Cm 6499. London.

Dixit, A 1980. The Role of Investment in Entry Deterrence. 90:95-106.

Donaldson, C. 1990. Willingness to pay for publicly-provided goods, a possible measure of benefit? *Journal of Health Economics* 9:103-118.

Donaldson, C, and K Gerard. 1989. Countering Moral Hazard in Public and Private Health Care Systems: A Review of Recent Evidence. *Journal of Social Policy* 18 (2):235-251.

Doney, P. M., J. P. Cannon, and M. R. Mullen. 1998. Understanding the Influence of National Culture on the Development of Trust. In *Special Topic Forum on Trust in and Between Organizations*, edited by S. B. Sitkin, D. M. Rousseau, R. S. Burt and C. Camerer: Academy of Management.

Dore, R. 1983. Goodwill and the Spirit of Market Capitalism. *British Journal of Sociology* 34 (3):459-82.

Dow, G. 1987. The function of authority in transaction cost economics. *Journal of economic behaviour and organisation* 8:13-38.

Dranove, D, and M Satterthwaite. 2000. The industrial organisation of health care markets. In *The Handbook of Health Economics*, edited by A. Culyer and J. Newhouse. Amsterdam: Elsevier.

Duan, N. 1983. Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association* 78:605-610.

Evers, A. 1993. The welfare mix approach. Understanding the pluralism of welfare systems. In *Balancing pluralism New Welfare Mixes in Care for the Elderly*, edited by A. Evers and I. Svetlik. Aldershot: Avebury.

Ferguson, B, and J Keen. 1996. Transaction costs, externalities and information technology in health care. *Health economics* 5:25-36.

Forder, J. 1997a. Contracts and Purchaser-Provider Relationships in Community Care. *Journal of health economics* 16:517-542.

Forder, J. 2000. Mental health: market power and governance. *Journal of Health Economics* 19 (6):877-905.

Forder, J. 2001. Prices, contracts and residential care. Discussion Paper, PSSRU, LSE Health and Social Care, London School of Economics.

Forder, J. 2002. Regulating entrepreneurial behaviour in social care. In *Regulating Entrepreneurial Behaviour in European Health Care Systems*, edited by R. Saltman, R. Busse and E. Mossialos. Buckingham: Open University Press.

Forder, J. 2005. A note on the derivation of Generalised Linear Models. PSSRU discussion paper, London School of Economics.

Forder, J, B Hardy, J Kendall, and M Knapp. 2000. Prices, contracts and competition. PSSRU discussion paper 1580, Personal Social Services Research Unit, London School of Economics.

Forder, J, M Knapp, B Hardy, J Kendall, T Matosevic, and P Ware. 2004. Prices, contracts and motivations: institutional arrangements in domiciliary care. *Policy and Politics* 32 (2):207-222.

Forder, J, and A Netten. 2000a. The effects of population dependency changes on costs, prices and total expenditure. PSSRU discussion paper 1581, Personal Social Services Research Unit, London School of Economics and University of Kent.

Forder, J., M. Knapp, and G. Wistow. 1996. Competition in the Mixed Economy of Care. *Journal of Social Policy* 25 (2):201-221.

Forder, J., and A. Netten. 2000b. The Price of Placements in Residential and Nursing Home Care. *Health economics* 9 (7):643-657.

Forder, J.E. 1997b. Contracts and Purchaser-Provider Relationships in Community Care. *Journal of health economics* 16:517-542.

Frank, R G, and M Gaynor. 1991. Incentives, Optimality and Publicly Provided Goods: The Case of Mental Health Services. WP 3700, National Bureau of Economic Research.

Frank, R, and T McGuire. 2000. Economics and mental health. In *Handbook of Health Economics*, edited by A. Culyer and J. Newhouse. Amsterdam: Elsevier Science.

Frey, B. 1997a. A constitution for knaves crowds out civic virtues. *The Economic Journal* 107:1043-1053.

Frey, B. 1997b. *Not just for the money: An economic theory of personal motivation*. Cheltenham: Edward Elgar Publishing.

Friedman, M. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.

Fudenberg, D, B Holmstrom, and P Milgrom. 1990. Short-term contracts and long-term agency relationships. *Journal of economic theory* 51:1-31.

Fukuyama, F. 1995. *Trust*. London: Penguin.

Fulop, N, G Protopsaltis, A King, P Allen, A Hutchings, and C Normand. 2005. Changing organisations: a study of the context and processes of mergers of health care providers in England. *Social Science and Medicine* 60 (1):119-130.

Fundenberg, D, and J Tirole. 1992. *Game Theory*. Cambridge, Mass.: MIT Press.

Giddens, A. 1990. *The Consequence of Modernity*. Stanford, California: Stanford University Press.

Giddens, A. 1998. *The third way: The renewal of social democracy*. Cambridge: Polity Press.

Glennerster, H, and M Matsaganis. 1993. The UK Health Reforms: The Fundholding Experiment. *Health Policy* 23:179-191.

Granovetter, M. 1985. Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology* 91 (3):481-510.

Granovetter, M. 1995. Coase revisited: business groups in the modern economy. *Industrial and corporate change* 4 (1):93-129.

Greene, W. 1993. *Econometric analysis*. New York: Prentice-Hall.

Gronbjerg, K. 1990. Managing nonprofit funding relations: case studies of six human service organisations. Paper read at Independent Sector Spring Forum, at Washington DC.

Grossman, S, and O Hart. 1983. An Analysis of the Principal-Agent Problem. *Econometrica* 51 (1):7-45.

Grossman, S, and O Hart. 1986. The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy* 94 (4):691-719.

Grout, P. 2003. Public and Private Sector Discount Rates in Public-Private Partnerships. *Economic Journal* 113 (486):C62-C68.

Hamilton, G, and R Feenstra. 1995. Varieties of hierarchies and markets: an introduction. *Industrial and corporate change* 4 (1):51-91.

Hannan, M, and J Freeman. 1984. Structural Inertia and Organizational Change. *American Sociological Review* 49:194-264.

Hart, O. 1995. *Firms, Contracts and Financial Structure, Clarendon Lectures in Economics*. Oxford: Oxford University Press.

Hart, O. 2003. Incomplete Contracts and Public Ownership: Remarks, and an Application to Public-Private Partnerships. *Economic Journal* 113 (486):C69-C76.

Hart, O. . 1985. Monopolistic competition in the spirit of Chamberlin: A general model. *Review of Economic Studies* 52:529-46.

Hausman, J. 1978. Specification tests in econometrics. *Econometrica* 46:1251-1271.

Hey, J, and P Lambert. 1987. *Surveys in the Economics of Uncertainty*. Oxford: Basil Blackwell.

Hirst, M. 2001. Trends in informal care in Great Britain during the 1990s. *Health and Social Care in the Community* 9 (6):348 - 357.

Hodgson, G. 1988. *Economics and Institutions: A Manifesto for a Modern Institutional Economics*. Cambridge: Polity Press.

Hoyes, L, and R Means. 1993. Quasi-markets and the reform of community care. In *Quasi-market and social policy*, edited by J. Le Grand and W. Bartlett. Basingstoke: Macmillan.

Hsiao, C. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.

Jensen, M, and W Meckling. 1976. Theory of the firm: managerial behaviour, agency costs and ownership structure. *Journal of financial economics* 3:305-360.

Joskow, P. 1987. Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets. *American Economic Review* 77 (1):168-185.

Joskow, P. 1988. Asset specificity and the structure of vertical relationships: empirical evidence. *Journal of Law, Economics and Organization* 4:95-117.

Kagel, J, and A Roth. 1994. *The handbook of experimental economics*. Princeton, New Jersey: Princeton University Press.

Kendall, J. 2001. Of Knights, Knaves And Merchants: The Case Of Residential Care For Older People In England In The Late 1990s. *Social policy and administration* 35 (4):360-375.

Kendall, J. 2003. *The Voluntary Sector: Comparative Perspectives in the UK*: Routledge.

Kendall, J, and M Knapp. 1996. *The voluntary sector in the UK*. Manchester: Manchester University Press.

Kendall, J, T Matosevic, J Forder, M Knapp, B Hardy, and P Ware. 2003. The motivation of domiciliary care providers in England: New concepts, new findings. *Journal of Social Policy* 32 (4):489-512.

Kennen, J, and R Wilson. 1993. Bargaining with private information. *Journal of economic literature* 31:45-104.

Kitson, M, J Michie, and M Sheehan Quinn. 2001. Markets, competition, cooperation and innovation. ESRC Centre for Business Research, University of Cambridge Working Paper No. 212, University of Cambridge.

Klein, B, and K B Leffler. 1981. The Role of Market Forces in Assuring Contractual Performance. *Journal of Political Economy* 89 (4):615-42.

Knapp, M 1984. *The economics of social care*. London: Macmillan.

Knapp, M, and J Kendall. 2001. Voluntary Sector Providers of Care for Older People in Comparative Perspective. In *Voluntary Organisations and Social Policy in Britain*, edited by M. Harris and C. Rochester: Macmillan.

Knapp, M, G Wistow, J Forder, and B Hardy. 1994. Markets for Social Care: Opportunities, Barriers and Implications. In *Quasi-markets in the welfare state*, edited by W. Bartlett, C. Propper, D. Wilson and J. Le Grand. Bristol: SAUS Publications.

Kreps, D. 1990a. Corporate Culture and Economic Theory. In *Perspectives on Positive Political Economy*, edited by J. Alt and K. Shepsle. Cambrdige: Cambridge University Press.

Kreps, D. 1990b. *A Course in Microeconomic theory*. London: Harvester Wheatsheaf.

Kreps, D, and R Wilson. 1982. Reputation and Imperfect Information. *Journal of Economic Theory* 27:253-279.

Kreps, D. 1990c. *Game Theory and Economic Modelling*. Oxford: Clarendon Press.

Kreps, D. 1996. Markets and Hierarchies and (Mathematical) Economic Theory. *Industrial and Corporate Change* 5 (2):561-595.

Laffont, J.-J., and J. Tirole. 1993. *A Theory of Incentives in Procurement and Regulation* Cambridge: MIT Press.

Laing and Buisson. 2003. Care of elderly people: market survey 2003. London: Laing and Buisson Publications Ltd.

Laing, W. 1993. *Financing long term care: the crucial debate*. London: Age Concern England.

Laing, W. 2004. *Calculating a fair price for care: A toolkit for residential and nursing care costs (second edition)*. Bristol: Policy Press.

Le Grand, J. 1991. Quasi-markets and Social Policy. *The Economic Journal* 101 (9):1256-67.

Le Grand, J. 1992. Quasi-Markets and Community Care. Paper read at Rowntree Symposium on The Exchange of Experience Across Welfare Sectors, at University of Birmingham.

Le Grand, J. 1997. Knights, Knaves or Pawns? Human Behaviour and Social Policy. *Journal of social policy* 26 (2):149-70.

Le Grand, J. 2003. *Motivation, Agency and Public Policy: Of Knights and Knaves, Pawns and Queens*. Oxford: Oxford University Press.

Le Grand, J., and W. Bartlett. 1993. *Quasi-markets and Social Policy*. Basingstoke: Macmillan.

Levaggi, R. 1996. NHS contracts: An agency approach. *Health economics* 5:341-352.

Lewis, J, and H Glennerster. 1996. *Implementing the new community care*. Buckingham: Open University Press.

Lewis, J., P. Bernstock, and V. Bovell. 1995. The Community Care Changes: Unresolved Tensions in Policy and Issues in Implementation. *Journal of Social Policy* 24 (1):73-94.

Lipsey, R, and K Lancaster. 1956. The General Theory of Second Best. *Review of Economic Studies* 24:11-32.

Lunt, N, R Manion, and P Smith. 1996. Economic discourses and the market: the case of community care. *Public administration* 74:396-391.

Lyons, B. 1996. Empirical relevance of efficient contract theory: inter-firm contracts. *Oxford review of economic policy* 12 (4):27-52.

Lyons, B, and J Mehta. 1997. Contracts, opportunism and trust: self-interest and social orientation. *Cambridge journal of economics* 21:239-257.

Lyons, B, and Y Varoufakis. 1989. Game theory, Oligopoly and Bargaining. In *Current Issues in Microeconomics*, edited by J. D. Hey. London: Macmillan.

Ma, C-T A. 1994. Health care payment systems: cost and quality incentives. *Journal of economics and management strategy* 3 (1):93-112.

MacNeil, I. 1985. Relational contracts: what we do and do not know. *Wisconsin Law Review*:483-525.

Manning, W. 1998. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics* 17:283-295.

Manning, W.G., and J. Mullahy. 1999. Estimating log models: to transform or not to transform?, University of Chicago.

Manning, W.G., and J. Mullahy. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* 20 (4):461-494.

Martin, S. 2002. *Advanced Industrial Economics*. Oxford: Blackwell.

Matosevic, T, M Knapp, J Kendall, J Forder, P Ware, and B Hardy. 2001. Domiciliary Care Providers in the Independent sector. Personal Social Services Research Unit (PSSRU) Monograph, London School of Economics.

McCullagh, P, and J Nelder. 1989. *Generalized linear models*. Boca Raton: Chapman & Hall/CRC Press.

McGuinness, T. 1992. Markets and Managerial Hierarchies. In *The Economics of the Firm*, edited by R. Clarke and T. McGuinness. Oxford: Blackwell.

McNulty, T, and E Ferlie. 2002. *Re-engineering health care: The complexities of organisational transformation*. Oxford: Oxford University Press.

Means, R, H Morbey, and R Smith. 2002. *From Community Care to Market Care: The Development of Welfare Services for Older People*. Bristol: Policy Press.

MEOC Team. 2001. MEOC Commissioning survey 2001. London: PSSRU, London School of Economics.

Milgrom, P, and J Roberts. 1982. Predation, Reputation, and Entry Deterrence. *Journal of Economic Theory* 27:280-312.

Milgrom, P, and J Roberts. 1992. *Economics, Organization and Management*. New Jersey: Prentice-Hall.

Milgrom, P., and J. Roberts. 1990. Bargaining Costs, Influence Costs, and the Organization of Economic Activity. In *Perspectives on Positive Political Economy*, edited by J. Alt and K. Shepsle. Cambridge: Cambridge University Press.

Miller, G. 1992. *Managerial Dilemmas: the Political Economy of Hierarchy*. Cambridge: Cambridge University Press.

Mishan, E. J. 1976. *Cost-benefit analysis*: Praeger.

Moos, R, and S Lemke. 1992. Sheltered care environment scale manual. Palo Alto, California: Department of Veterans Affairs.

Musgrave, A. 1981. "Unreal Assumptions" in Economic Theory: The F-Twist Untwisted. *Kyklos* 34:377-87.

Nash, J. 1953. Two-Person Cooperative Games. *Econometrica* 21:128-140.

Nelson, R, and S Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Belknap/Harvard University Press.

Netten, A, A Bebbington, R Darton, J Forder, and K Miles. 1998. 1996 survey of care homes for elderly people (final report). Discussion paper 1423/2, Personal Social Services Research Unit.

Netten, A, R Darton, and J Williams. 2003. Nursing home closures: effects on capacity and reasons for closure. *Age and Ageing* 32 (3):332-337.

Netten, A, J Francis, K Jones, and A Bebbington. 2004. Performance and quality: user experiences of home care services. PSSRU Discussion Paper 2104/3, PSSRU, University of Kent.

Netten, A, M Ryan, P Smith, D Skatun, A Healey, M Knapp, and T Wykes. 2002. The development of a measure of social care outcome for older people. Discussion Paper 1690/2, University of Kent.

North, D. 1990. *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.

Nyman, J.A. 1994. The effects of market concentration and excess demand on the price of nursing home care. *Journal of industrial economics* 42 (2):193-204.

Nyman, J.A. . 1985. Prospective and cost-plus Medicaid reimbursement, excess Medicaid demand and the quality of nursing home care. *Journal of health economics* 4:237-59.

Nyman, J.A. . 1989. The private demand for nursing home care. *Journal of health economics* 8 (2):209-231.

Osborne, D. E., and T. Gaebler. 1992. *Reinventing Government.* New York: Addison-Wesley.

Osborne, S. 1997. Managing the coordination of social services in the mixed economy of welfare: competition, cooperation or common cause? *British Journal of Management* 8:317-328.

Pauly, M. 1984. Is cream-skimming a problem for the competitive medical market? *Journal of health economics* 3:87-95.

Perrow, C. 1990. Economic theories of organisation. In *Structure of capital: the social organisation of the economy*, edited by S. Zukin and P. DiMaggio. Cambridge: Cambridge University Press.

Pickard, L., R Wittenberg, A Comas-Herrera, B Davies, and R Darton. 2000. Relying on informal care in the new century? Informal care for elderly people in England to 2031. *Ageing and Society* 20 (6):745-772.

Pindyck, R.S, and D.L. Rubinfeld. 1981. *Econometric Models and Economic Forecasts.* 2 ed: McGraw-Hill Company.

Plsek, P, and T Wilson. 2001. Complexity science: Complexity, leadership and management in healthcare organisations. *British Medical Journal* 323:746-749.

Powell, W. 1990. Neither market nor hierarchy: Network forms of organization. In *Research in Organizational Behavior, Vol. 12*, edited by B. Staw and L. Cummings. Greenwich, CN: JAI Press.

Propper, C. 1995. Agency and incentives in the NHS internal market. *Social Science and Medicine* 40 (12):1683-1690.

Propper, C. 1996. Market structure and prices: The response of hospitals in the UK National Health Service to competition. *Journal of Public Economics* 61:307-335.

Propper, C, and D Wilson. 2003. The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy* 19 (2):250-267.

Propper, C, S Burgess, and K Green. 2004. Does competition between hospitals improve the quality of care? *Journal Of Public Economics* 88 (7-8):1247-1272.

Propper, C, and D Wilson. 2003. The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy* 19 (2):250-267.

Putterman, L, ed. 1986. *The economics nature of the firm: a reader.* Cambridge: Cambridge University Press.

Rasmusen, E. 1992. *Games and information: An introduction to game theory.* 2nd ed. Oxford: Blackwell.

Rhodes, R A W. 1997. *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability*. Buckingham: Open University Press.

Rhodes, R.A.W. 1995. The New Governance: Governing Without Government. Paper read at The State of Britain Seminars.

Ridley, N. 1988. *The local right: enabling not providing.*

Roth, A, and F Schoumaker. 1983. Expectations and reputation in bargaining: an experimental study. *American economic review* 73:362-72.

Rubinstein, A. 1998. *Modelling bounded rationality*. Cambridge, Massachusetts: The MIT Press.

Sako, M. 1992. *Prices, Quality and Trust: Inter-firm Relations in Britain and Japan*. Cambridge: Cambridge University Press.

Saltman, R, R Busse, and E Mossialos, eds. 2002. *Regulating entrepreneurial behaviour in social care*. Buckingham: Open University Press.

Saltman, R, J Figueras, and C Sakellarides. 1998. *Critical challenges for health care reform in Europe*. Buckingham: Open University Press.

Savas, S, I Sheiman, E Trgakes, and H Maarse. 1998. Contracting models and provider competition. In *Critical challenges for health care reform in Europe*, edited by R. Saltman, J. Figueras and C. Sakellarides. Buckingham: Open University Press.

Sawyer, M. 1993. The nature and role of the market. In *Transaction costs, markets and hierarchies*, edited by C. Pitelis. Oxford: Blackwell.

Schleifer, A. 1985. A Theory of Yardstick Competition. 16:319-27.

Selten, R. 1975. Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games. *International Journal of Game Theory* 4:25-55.

Selten, R. 1978. The Chain Store Paradox. *Theory and Decision* 9:127-59.

Shavell, S. 1979. Risk-sharing and incentives in the principal and agent relationship. *Bell journal of economics* 10:55-73.

Shelanski, HA, and PG Klein. 1995. Empirical research in transaction cost economics: a review and assessment. *Journal of law, economics, & organization* 11 (2):335-361.

Simon, C, and L Blume. 1994. *Mathematics for economists*. New York: W. W. Norton & Company, Inc.

Simon, H. 1951. A formal model of the employment relationship. *Econometrica* 19:293-305.

Simon, H. 1955. A behavioural model of rational choice. *Quarterly journal of economics* 69:99-118.

Smelser, N, and R Swedberg, eds. 1994. *The handbook of economic sociology*. Princeton, N.J: Princeton University Press.

Smith, S R, and M Lipsky. 1993. *Nonprofits for hire: The welfare state in the age of contracting*. London: Harvard University Press.

Social Services Inspectorate. 2004. SSI performance assessment 2004, operating policies, London.

Spulber, D. 1989. *Regulation and markets*. Cambridge, Mass.: The MIT Press.

StataCorp. 1999. *Stata reference manual: release 6*. College Station, TX: Stata Corporation.

Stevens, W, and C Normand. 2004. Optimisation versus certainty: understanding the issue of heterogeneity in economic evaluation. *Social Science and Medicine* 58 (2):315-320.

Stigler, G. 1971. The Theory of Economic Regulation. 2:3-21.

Swedberg, R. 1994. Markets as social structures. In *The handbook of economic sociology*, edited by N. Smelser and R. Swedberg. Princeton, N.J: Princeton University Press.

Taylor, M, and P Hoggett. 1994a. Quasi-markets and the transformation of the independent sector. In *Quasi-markets in the welfare state*, edited by W. Bartlett, C. Propper, D. Wilson and J. Le Grand. Bristol: SAUS Publications.

Taylor, M, and P Hoggett. 1994b. Trusting in Networks? The Third Sector and Welfare Change. In *Delievering Welfare*, edited by P. 6 and I. Vidal. Barcelona: CIES.

Taylor-Gooby, P. 1997. Markets and motives: trust and egoism in welfare markets, University of Kent.

Timmins, N. 2001. *The Five Giants: A Biography of the Welfare State*. New York: HarperCollins.

Tirole, J. . 1988 *The theory of industrial structure*. Cambridge: MIT Press.

Titmuss, R. 1970. *The gift relationship: from human blood to social policy*. London: Allen & Unwin.

Varian, H. 1978. *Microeconomic analysis*. New York: W. W. Norton and Company.

Vickers, J. 1996. Market power and inefficiency: a contracts perspective. *Oxford review of economic policy* 12 (4):11-26.

Vickers, J., and G. Yarrow. 1988. *Privatisation: An Economic Analysis*. Cambridge: MIT Press.

Wedderburn, R. 1974. Quasi-likelihood functions, generalised linear models and the Gauss-Newton method. *Biometrika* 61:439-447.

Weisbrod, B. 1975. Towards a Theory of the Voluntary Nonprofit Sector in a Three Sector Economy. In *Altruism, Morality and Economic Theory*, edited by E. S. Phelps. New York: Russell Sage.

Weisbrod, B. 1988. *The Nonprofit Economy*. Cambridge, Mass.: Harvard University Press.

Williamson, O. 1975. *Markets and Hierarchies*. New York: Free Press.

Williamson, O. 1979. Transaction cost economics: the governance of contractual relations. *Journal of law and economics* XXII (2):233-61.

Williamson, O. 1985a. *The economic institutions of capitalism*. New York: Free Press.

Williamson, O. 1986. *Economic Organisation: Firms, Markets and Policy Control*. Brighton: Wheatsheaf Books.

Williamson, O, and S Winter, eds. 1993. *The Nature of the Firm: Origins, Evolution, and Development*. New York: Oxford University Press.

Williamson, O. 1994. Transaction Cost Economics and Organization Theory. In *The handbook of economic sociology*, edited by N. Smeltzer and R. Swedberg. Princeton, N.J.: Princeton University Press.

Williamson, O.E. . 1985b. *The economic institutions of capitalism*. New York: Free Press.

Wistow, G, M Knapp, B Hardy, and C Allen. 1994. *Social Care in a Mixed Economy*. Buckingham: Open University Press.

Wistow, G, M Knapp, B Hardy, J Forder, J Kendall, and R Manning. 1996. *Social Care Markets: Progress and Prospects*. Buckingham: Open University Press.

Wittich, G, and G Roth, eds. 1978. *Economy and society: an outline of interpretive sociology (Max Weber)*. Berkeley: University of California Press.