ISSN 2042-2695

CENTRE *for* ECONOMIC PERFORMANCE

CEP Discussion Paper No 1232

July 2013

Regional Growth and Regional Decline

Holger Breinlich, Gianmarco I. P. Ottaviano
and Jonathan R. W. Temple

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract
Since the early 1990s, there has been a renaissance in the study of regional growth, spurred by new models, methods and data. We survey a range of modelling traditions, and some formal approaches to the 'hard problem' of regional economics, namely the joint consideration of agglomeration and growth. We also review empirical methods and findings based on natural experiments, spatial discontinuity designs, and structural models. Throughout, we give considerable attention to regional growth in developing countries. Finally, we highlight the potential importance of processes that are specific to regional decline, and which deserve greater research attention.

Keywords: Regional growth, regional convergence, regional decline
JEL Classifications: O18, R11, R12, R13

This paper was produced as part of the Centre's Globalisation Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements
Prepared for the forthcoming Handbook of Economic Growth, Volume 2, edited by Philippe Aghion and Steven N. Durlauf. We thank Paddy Carter, Klaus Desmet, Adeel Malik, Annalisa Marini, Stephen Redding, Esteban Rossi-Hansberg and Helen Simpson for various helpful contributions. The usual disclaimer applies.

Holger Breinlich is a Research Associate at the Centre for Economic Performance, London School of Economics and Political Science. He is also a Senior Lecturer in Economics at the University of Essex. Gianmarco Ottaviano is an Associate of the Globalisation Programme at the Centre for Economic Performance and a Professor of Economics at the London School of Economics and Political Science. Jonathan Temple is a Professor of Economics at the University of Bristol.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

> Europe, as it has become more integrated, has also become more difficult to write about. — Perry Anderson, *The New Old World*, p. xi

1 Introduction

From 2006 onwards, an exhibition of photographs has toured galleries in Europe and the USA, now titled *The Ruins of Detroit*. The photographs, by Yves Marchand and Romain Meffre, show various scenes from the recent past of America's Motor City: the ruined Spanish-Gothic interior of the United Artists Theater (closed 1984), the abandoned waiting hall of Michigan Central Station (closed 1988), the derelict ballroom of the Lee Plaza Hotel (closed early 1990s), and an abandoned school book depository with its textbooks scattered and covered in debris. The photographs bring home, in a way that statistics do not, what it can mean for a city or region to experience an extended period of decline. In Detroit's case, that decline has been precipitous: from one of America's wealthiest cities, with a city population of around 1.8 million at its peak, to a population that is now around 700,000. It has left the city with falling property values, enough vacant land to accommodate the whole of Paris, and a rate of violent crime among the highest of any American city. Economists sometimes refer to changing patterns of economic activity as 'adjustment', but many of those who have lived through the city's decline will have experienced it chiefly as a tragedy.

One reason Detroit's experience has attracted such attention is that relative decline is rarely so marked, or rapid. The disparities between cities and regions are generally more stable than this, even at times of growth and structural change. But it is also true that disparities can be substantial and persistent, lasting many decades. They can often become an important part of how a country sees itself, and how it evolves over time. As Judt (1996) notes, the divisions and tensions between southern and northern Italy are a theme as old as the Italian state itself. In England, the nineteenth-century writer Elizabeth Gaskell published her novel *North and South* in 1855. More than 150 years later, regional differences in living standards, health outcomes, political beliefs and social norms continue to be summarized as England's 'north-south divide'. Similar phenomena can be seen in the developing world, sometimes on an even larger scale. China's coastal cities are more prosperous than its inland regions. In India, there are substantial disparities across states in terms of literacy rates, life expectancy and living conditions, as well as income (Drèze and Sen 1997). Poverty rates vary widely within Brazil, with low rates in the booming south-east and

much higher rates in the rural north-east (Skoufias and Katayama 2011). The list could easily be multiplied but, as Williamson (1965) remarks, many countries have a tendency to see their own regional imbalances as uniquely pronounced and intractable.

This chapter will describe a range of models and evidence that can be used to understand regional growth and the evolution of spatial disparities over time. This is an unusually complex topic, and one that requires an eclectic approach and general equilibrium reasoning. In a long-run spatial equilibrium, households and firms must not prefer other locations to their current location. As Glaeser and Gottlieb (2009) emphasize, this implies that research on places is vitally different from research on countries, and requires that population, income, and prices are considered simultaneously. Regions are interdependent to an even greater extent than countries, and there is a real sense in which regional growth is a harder topic than national growth.

There is a further departure from the standard competitive paradigm. A long tradition in urban economics and economic geography explains the structure of cities and economic activity in terms of various externalities. These have the potential to generate inefficient and undesirable outcomes. Although some externalities are reasonably well understood, at least in theory, others are not. The example of Detroit shows how one mechanism in regional decline will be changes in crime and social norms, amplifying changes that originated elsewhere. Recent empirical work on local institutional variation within developing countries, such as Dell (2010), suggests remarkably powerful and long-lived effects of this variation. The reasons for this remain unclear, but could include not only the persistence of institutions, but also the intergenerational transmission of social norms and political beliefs. Economists have only just begun to engage with such complex forces.

These points hint that a single 'canonical' model of regional growth is neither likely nor desirable. There are so many interesting research questions that it would be a mistake to seek or impose a single framework. The chapter will discuss how spatial disparities evolve over time; the circumstances in which there is a 'regional problem'; how differences in regional living standards and productivity can arise; the data and methods used to study regional growth; and the forces that drive regional growth and regional decline.

Some of these questions are too intertwined to address sequentially. As in the literature on national economic growth, it can be a mistake to attempt a sharp distinction between 'growth' and 'levels', as if these two phenomena necessarily

require separate models. In practice, it makes little sense to write about regional growth without taking a view on what determines relative levels of income per capita. But we also note that 'regional growth' does not only mean growth in average living standards. In common usage, it often means growth in relative population or total income, as a region outperforms others. One theme of our chapter is that changes in the relative sizes or population densities of regions merit more attention from researchers. For the study of decline, in particular, it is important to analyze depopulation rather than simply relative living standards.

Another complication is even more fundamental. Productivity growth is accompanied by, and to a large extent inseparable from, changing patterns of agglomeration and dispersion. Growth will respond to, and bring, changes in demand patterns, sectoral and occupational structure, skill levels, transport costs and infrastructure, financial development, and even local institutions and political economy. All of these could reconfigure the spatial structure of population and production. Yet modelling growth and agglomeration as outcomes of a joint process is far from straightforward, as Krugman (1995) noted. We call this 'the hard problem' of regional economics, and review some models that seek to address it.

When we turn to the evidence, we depart from existing surveys by considering a wider range of countries. The New Economic Geography literature has tended to focus on Europe, Japan and the USA, but the study of regional prosperity is even more important for contemporary developing countries, as Venables (2005) emphasizes. After all, some Chinese provinces and Indian states exceed many countries in population and land area. These include India's Uttar Pradesh (population around 200 million), and Maharashtra and Bihar (both in excess of 100 million). Guangdong province in China has a population of more than 100 million once migrants are included. The intrinsic importance of this should be clear, and the consideration of developing countries has a further benefit, widening the scope of the available evidence. Recent work has opened up some startling research possibilities, not least through the use of satellite data on light density at night to map activity at the sub-national level.

A final theme will be the formidable identification problems that arise in studying regional data. Some can be seen in narrow terms as spatial dependence; for example, errors in a regression model will often be correlated across regions. But more fundamentally, the requirements of spatial equilibrium will link regional outcomes and characteristics so tightly that it is rarely clear how to recover causal effects. This continues to be a major obstacle to understanding

regional growth or making policy recommendations. We will discuss a range of empirical methods, which recover causal effects with varying degrees of plausibility, and the complementary role of structural models.

With all this in mind, the coverage of the chapter is intentionally broad, not to say sprawling. Over the past two decades, after long years of neglect, economists have developed a rich theoretical and empirical literature on economic geography, chiefly inspired by Krugman (1991). But concurrently, and largely independently, researchers working on growth and development have studied the effects of policy reforms and institutional variation using sub-national data. Influential papers include Banerjee and Iyer (2005), Besley and Burgess (2000), Holmes (1998), Jayaratne and Strahan (1996), and Tabellini (2010), but these are just a few examples from an increasingly extensive literature. An important aim of this chapter is to bring these various strands of research together, and use them to interpret each other.

To keep the scope of the chapter manageable, we also need to set some limits to what we cover. We emphasize work in economics, and especially recent work that takes a general equilibrium approach. This is a significant limitation, because the study of regional outcomes extends well beyond economics, to include research in geography, urban planning, sociology, statistics, and demography. From the mid-1950s onwards, elements of these approaches began to coalesce in the interdisciplinary field of regional science. This field has sometimes drawn on ideas from economics, such as applied general equilibrium modelling, adapted to include a spatial dimension.[1] It is clear, however, that the traditional methods of regional science are rarely well-adapted to the study of regional growth and convergence dynamics. Their roots lie in static models, or in empirical methods that will rarely identify causal effects within a spatial equilibrium. It is also noticeable that, when textbooks on regional economics turn to growth, the approaches presented lack coherence. They range across basic trade-theoretic analyses, through closed-economy one-sector growth models, to an emphasis on the demand-side role of regional exports and trade balances. Each time, it is all too easy to see what is missing: interesting dynamics, an explicit spatial dimension, a central role for supply adjustments and constraints. For all the benefits of an eclectic or interdisciplinary approach, many of the interesting questions demand general equilibrium reasoning, and the task is lost without it.

This should be clear if we consider one of the strongest associations in the data. Gennaioli et al. (2013a) emphasize that regional output per capita is

[1] An overview of quantitative methods in regional science can be found in Isard et al. (1998).

strongly correlated with average human capital. In a regression of output per capita on average years of education and country dummies, using 1500 regions across 105 countries, they find that education explains 38% of the variation in output per capita within countries. This is striking, but nobody would propose that regions within a country are each endowed with a fixed stock of skilled workers. Investigating the association requires models in which regional factor supplies are endogenous to the location decisions of workers and firms; Gennaioli et al. construct one such model, and alternatives will be considered below.

In summary, the study of regional growth often requires structural models that draw heavily on economic ideas; an understanding of the various forms of interdependence between regions; and empirical methods that can overcome the identification problems raised by that interdependence. We take these endeavours as our central focus, rather than the much wider literature in regional science and geography.[2] Nor do we provide a discussion of regional policies: these require an understanding of the mechanisms at work in regional growth and regional decline, but we do not develop the links explicitly. Finally, we note that some of the analytical issues overlap with those of urban economics; for discussion of ideas specific to urban economics and the growth of cities, see the chapter by Duranton and Puga in this volume.

The remainder of the chapter considers different perspectives in turn. Section 2 looks at convergence and polarization, the associated methods, and some stylized facts. Section 3 will discuss the nature and interpretation of spatial disparities, and when they matter. The remaining sections, which are really the heart of the chapter, investigate the drivers of regional growth and decline. Section 4 covers an array of relevant models. Section 5 sets out two classes of models that consider growth and agglomeration jointly. We then review empirical methods (section 6) and some of the main findings (section 7). Section 8 briefly discusses regional decline as a distinct phenomenon, while section 9 concludes.

2 Convergence, divergence, polarization

Do regional economies have a tendency to move closer together, grow in parallel, or move further apart? This question has spurred many empirical studies,

[2]For a broader perspective, see Clark et al. (2000). The relationship between work in economics and geography has been extensively discussed, as in Brakman et al. (2009, chapter 12), Krugman (1995) and Ottaviano and Thisse (2005). Brakman et al. also survey regional growth, as does Harris (2010). For a discussion of the regional policy implications of recent work by economists, see Combes (2011).

but is surprisingly difficult to answer. Part of the problem may lie with the question. We are interested in how a distribution of outcomes evolves over time, but distributions can behave in complex ways, and much is lost by collapsing this behaviour into a crude binary opposition between convergence and divergence (Durlauf et al. 2009). As in the literature on national growth, part of the interest in these questions revolves around more complex possibilities. These include distinct 'convergence clubs', or the emergence of polarization. A related question is that of mobility within the distribution. Since a detailed survey of regional convergence is already available (Magrini 2004) we emphasize the studies that are especially relevant to understanding regional growth, or that have emerged over the last decade of research. The details of some of the more technical methods, using time series concepts, transition matrices, or mixtures of densities, can be found in the appendix to the chapter.

2.1 Beta-convergence

Do regions converge to the same level of income per capita? Our starting point is Figure 1, which shows annualized growth against initial income per capita for 47 contiguous US states, using data for 1880 and 1990.[3] This is similar to figure 11.2 presented in Barro and Sala-i-Martin (2004) and uses their data. At first glance, the strength of the correlation between growth and initial income per capita is decisive evidence for the absolute convergence of regions; they draw attention to the high R^2, 0.92, of a regression close to that shown. At first glance, this suggests that regional disparities are transitory phenomena.

The figure is a little deceptive, however. To see this, consider what would happen if two regions with the same initial income differed in their growth rates, so that one region was slightly above the regression line and one slightly below. Over the 110 years between 1880 and 1990, this small difference in growth rates would compound to imply a large difference in relative levels. Hence, even the strong negative correlation in figure 1 does not imply that spatial disparities will be eliminated. To show this, figure 2 presents the same data in a different way, showing income per capita relative to the median region in 1990 against that in 1880, together with a 45-degree line. The shallow slope of the (dashed) regression line is consistent with mean reversion, but using the vertical axis, it is also clear that significant differences in living standards across US states remain. The richest state has more than twice as much income per capita as the poorest.

[3]The state missing from the 48 contiguous states is Oklahoma, which lacks data for 1880 given its late statehood.

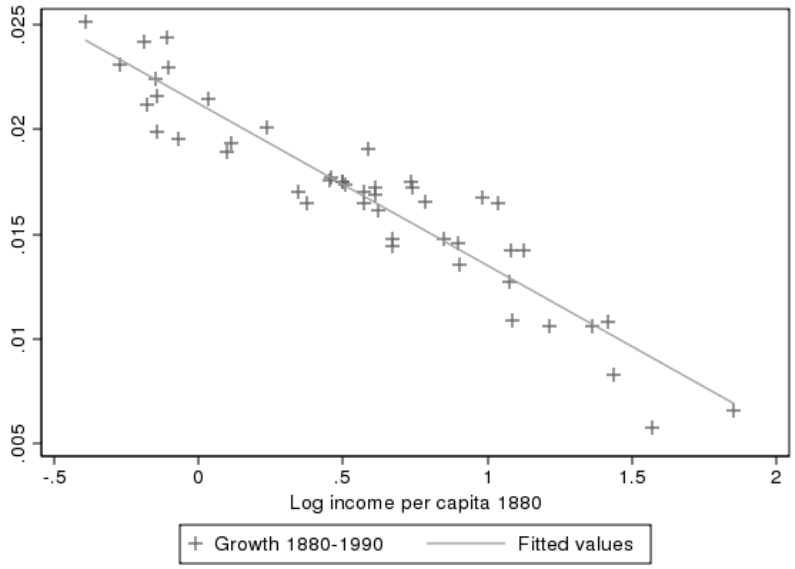

Figure 1: Absolute convergence?

This is contrary to what might have been expected from figure 1, but in line with the common sense view that regional disparities persist over time.[4]

Worldwide, regional disparities are pervasive and substantial, especially in poorer countries. Using data on 1537 regions across 107 countries, for the year 2005, Gennaioli et al. (2013a) report that the average ratio of income per capita in the richest region to the poorest region is 4.41. The ratio is 3.77 for Africa, 5.63 for Asia, 3.74 for Europe, 4.60 for North America, and 5.61 for South America. The ratios are substantially higher in some cases, including Indonesia, Mexico and Russia. Most of these figures do not correct for price levels, which are often higher in richer regions. Nevertheless, it seems unlikely that price differentials could account for the majority of the variation in nominal incomes.

The early literature on beta-convergence was highly successful in drawing attention to the rich interest of regional growth, and spurred a major research effort in the area. But the specific approach has also drawn criticism, partly on econometric grounds that we discuss in the appendix, and partly because the results can be hard to interpret. This is because they rely on viewing regional data through the prism of the neoclassical growth model. Indeed, it is often asserted that regional data provide an ideal testing ground for that model. These claims are misplaced, because the neoclassical model typically rules out cross-

[4]There are some changes in rankings — the Spearman rank correlation is 0.47 — and beta-convergence is sometimes argued to be informative about mobility; but it is not straightforward to map an estimated convergence rate onto a readily-interpreted scale for a mobility index.

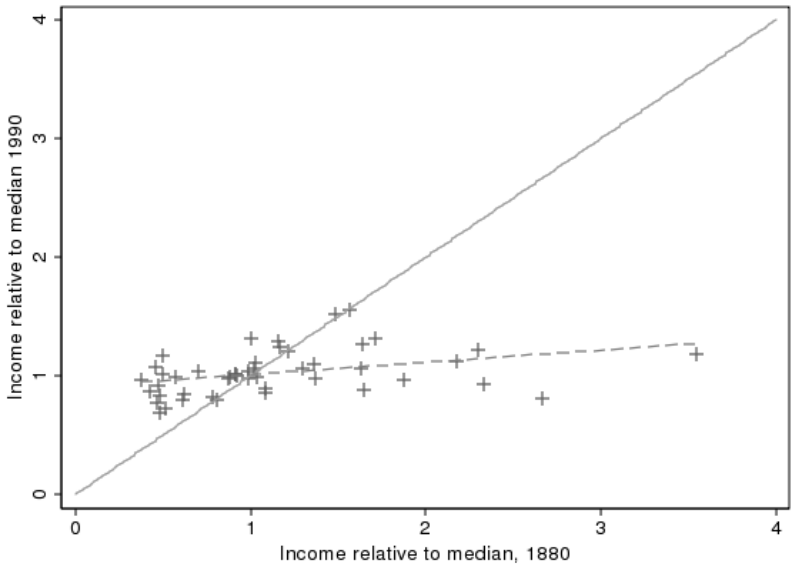

Figure 2: Long-run disparities

border flows of goods, services, capital and labour. The assumption that regions are closed to such flows is hardly attractive. But allowing for these flows is not straightforward, and general equilibrium models rarely lead to simple regression specifications. These perspectives suggest that beta-convergence studies will miss a great deal, and other methods are needed.

2.2 Inequality and polarization

When considering regional convergence, a useful starting point is to ask whether the cross-section variability of income per capita is increasing, stable, or falling over time. The most prominent version of this is sigma-convergence, which considers the evolution of the standard deviation (σ) of the logarithm of income per capita; regions are said to be converging if the standard deviation is falling over time (Barro and Sala-i-Martin 1991). This measure of regional inequality is not Lorenz-consistent, but could be replaced with the Gini coefficient, the coefficient of variation (as in Williamson 1965), or the Theil measures. The Theil measures, and other members of the generalized entropy class, have the significant advantage for regional analysis that they are decomposable; see Cowell (2011) for a textbook treatment.

A further question is whether or not to weight regions by their populations; Milanovic (2005b) provides a discussion of this. If the aim is to capture the spatial inequality perceived by a randomly-drawn individual, then weighting by

population is natural, as in Williamson (1965) and many subsequent papers. But for the analysis of regional growth, a researcher might be interested in the effects of physical geography, institutions and policies. In that case, it might be sensible to give regions equal weight even when they vary in size, rather than allow the results to be dominated by the characteristics of the largest regions.

As a first step in a descriptive exercise, the study of inequality measures is often valuable. But a given time path for regional inequality could correspond to a variety of underlying processes, with different long-run implications. A small group of regions may diverge from a larger group; as a result, measured inequality could increase even while a large number of regions grow in parallel. A related possibility is that the distribution becomes polarized. This term is often used rather loosely, to indicate some degree of high or rising inequality across regions. The view that deregulated market economies give rise to excessive polarization, in various senses, is especially common on the political left (for example, Dorling 2011); others use the term to indicate a 'disappearing middle' or 'clustering around extremes'.

On a more formal definition, polarization can be seen as concerned with multiple modes, the distance between these modes, and the distribution of probability mass around them. Drawing on Duclos et al. (2004), imagine that there is reduced variation in living standards at two different ranges of the regional income distribution. This is likely to reduce inequality, but polarization increases, since the contrast between the two groups is made sharper and more visible. More generally, polarization - in this technical sense - can increase even as the cross-section dispersion in living standards falls. These arguments suggest the need to look beyond inequality, using methods reviewed in the appendix to this chapter.

2.3 Findings

It seems inevitable that regional disparities will sometimes be compounded by growth and agglomeration. Seen against the long span of human history, current disparities may be a comparatively recent phenomenon; Bairoch (1993) argued that there was considerable uniformity in development levels in the early modern period (say, 1500-1800). One empirical approach relates regional inequality to the national level of development, as in the classic paper by Williamson (1965). He hypothesized an inverse-U relationship, with regional inequality rising and then falling as development proceeded. More recently, Barrios and Strobl (2009) examine the relationship using data for 12 European countries over 1975-2000.

The data plotted in their figure 2 suggest that regional inequality is increasing at lower levels of development, before either levelling off or reducing somewhat, but rarely returning to its initial level. For a much larger set of countries, Lessmann (2011) finds some evidence for the inverse-U relationship over 1980-2009, with regional inequality peaking at a development level close to that of, say, Mexico or the Czech Republic. He also finds some evidence that regional inequality increases at very high levels of GDP per capita (roughly, Canada's level).

A more common approach in the literature, following Barro and Sala-i-Martin (1991), has been to consider the evolution of regional inequality over time. Sometimes data on regional output are used, and sometimes data on income. For a few countries, including China and Indonesia, region-specific price deflators are available. For developing countries in particular, the treatment of natural resource revenues can be important to the results. Differences across regions in age structure, employment rates and part-time work are another complicating factor. We defer a more thorough discussion of regional data until section 6 below.

A mixture density approach (see the appendix) has been applied to European regions by Pittau (2005) and Pittau and Zelli (2006). Their work suggests a multimodal structure for the 1970s and early 1980s, and represents the distribution as a mixture of two well-separated normal densities. These two clusters later converge. They also find that, from the mid-1990s, a small group of very rich regions (Brussels, Hamburg, Île de France and Luxembourg) moves further ahead, a result also highlighted by Enflo (2010). This suggests recent polarization, and also indicates that using a benchmark region or weighted average to assess convergence is risky. Evidence for polarization also emerges from other studies, including Canova (2004).

An alternative approach is based on transition matrices or stochastic kernel densities. In applications of these methods, the stationary distribution for the US states appears to be unimodal (Johnson 2000) while the stationary distribution for European regions is more likely to appear bimodal. But one of the most sophisticated time-series studies of the US, that by Carvalho and Harvey (2005), finds that the two richest 'macro-regions', New England and the Mid-East, have pulled away from others over time. This indicates some degree of polarization may be emerging for the US as well as Europe.

For Japanese prefectures, Sala-i-Martin (1996) found a sharp decline in the dispersion of average personal income between 1940 and 1955, and a smaller

decline in the 1970s. At a more disaggregated level, Seya et al. (2012) found a decline in the log variance across Japanese municipalities over the 1990s, and a slight increase in the 2000s. For Russia, the usual finding is that regional disparities increased sharply in the first years of the post-Soviet era (for example, Fedorov 2002) and were high in comparison to many other countries; more recently, over the course of the 2000s, they seem to have fallen (Guriev and Vakulenko 2012).

The experiences of many developing countries are at least as interesting. The literature on China, in particular, is extensive. It typically finds divergence for the 1970s, followed by a period of convergence in the wake of agricultural reforms, and then further divergence during the rapid industrialization of the 1990s (for example, Weeks and Yao 2003). For the 1990s onwards, the fast growth of the coastal provinces is often emphasized, consistent with a story in which market access has promoted industrial development. Démurger et al. (2002) note the importance of three exceptionally rich provinces, Beijing, Tianjin, and Shanghai, in raising the overall degree of regional inequality. But even excluding these provinces, regional inequality rose over the 1990s. It is substantially higher than India's, a finding that is conventionally explained in terms of barriers to mobility within China (for example, Gajwani et al. 2006).

Milanovic (2005b) studies regional inequality in the five federations of Brazil, China, India, Indonesia and the US, over the 1980s and 1990s. For Brazil, he finds no clear trend; Azzoni (2001) studies Brazil for a longer period (1939-1995) and finds an overall decline, although one interrupted by a sharp increase in the 1970s. For India and Indonesia, Milanovic finds regional inequality to have increased. Hill et al. (2008) also study Indonesia, but find that the coefficient of variation of non-mining output per capita was broadly stable over 1975-2004, despite fast growth.[5] The reason for the inconsistent findings is not clear, and more generally, Milanovic (2005b) notes that the field lacks a consistent terminology and approach.[6]

Convergence methods have sometimes been applied to variables beyond income or output. Evans and Karras (1996) found evidence for rapid conditional convergence of TFP for the contiguous US states, and similarly rapid mean reversion for the returns to capital, computed using data on capital-output ratios and factor shares. Overall, their results point to inter-region capital mobility, but

[5] If mining is included, the coefficient of variation declined sharply in the 1970s and 1980s.

[6] The picture may be worse than he suggests: some papers on convergence make basic errors in the use of inequality measures, such as calculating the standard deviation of income rather than log income, or calculating the coefficient of variation using log income.

also that US states are converging to region-specific steady-states determined by long-run differences in productivity.

The convergence behaviour of regional house prices has also been studied. This is interesting in its own right, but also because housing costs will be determined jointly with incomes, and hence informative about the mechanisms generating spatial disparities. Using a model-based approach to US data from 1975, Van Nieuwerburgh and Weill (2010) emphasize that inequality in (quality-adjusted) house prices across US regions has risen substantially. By calibrating a model with a spatial equilibrium, they argue that increased dispersion of regional productivity is needed to explain the US data. Their analysis demonstrates the benefits of studying movements in cross-section dispersion using a structural model, rather than treating the study of convergence as solely an econometric problem; we return to this point later.

3 Do regional disparities matter?

At first glance, differences in average living standards across regions contribute to overall inequality, and involve some degree of unfairness or injustice, with social and political consequences. An individual born in a depressed region may have fewer opportunities and poorer life chances than an individual born in a more prosperous region of the same country. A widely-held view is that uneven regional development can undermine social cohesion and generate political tensions. This seems especially likely in those developing countries where regional disparities coincide with the spatial distribution of ethnic groups or natural resources. But the issue is not confined to developing countries, as witnessed by regional tensions within Belgium and Italy, among other countries. Judt (1996) prophesized that the project of European integration would eventually be undermined by polarization between dominant 'super-regions' — such as Baden-Württemberg, Rhône-Alpes, Lombardy and Catalonia — and an economically depressed periphery.[7] The European Union has made strengthening 'cohesion' across the regions of its member states a major priority, enshrined by treaty, and around a third of the central EU budget is directed at policies to reduce spatial disparities.

[7] In his view, the richer regions would be likely to identify themselves ever more closely with Europe, while the less successful periphery would see increasing scepticism and resentment about the European project, and perhaps a resurgence of nationalism.

3.1 Composition effects and welfare

Much of this is familiar, but care is needed over the meaning of 'living standards', and their comparison across space. In simple models, if workers can move freely between regions, then returns to worker characteristics are equalized across space, at least in the long run.[8] Spatial differences in average income per capita do not map straightforwardly into spatial inequality in life chances, welfare differences, or life satisfaction. In a spatial equilibrium, regional disparities will often reflect composition effects. An agricultural region may have relatively low income per capita not because it is 'depressed', inefficient, or its workers underpaid, but because the relatively unskilled account for a high share of its employment, or because its activities are relatively labour-intensive.

These points apply more broadly, and their importance is supported empirically. Acemoglu and Dell (2010) show that approximately half of the within-country, between-region inequality in labour income in the Americas (Canada, U.S. and Latin America) can be accounted for by differences in workers' education and experience. Another branch of the literature uses household-level data to account for regional differences in household composition, and to estimate how the returns to characteristics vary across locations; relevant studies include Nguyen et al. (2007) for Vietnam and Skoufias and Katayama (2011) for Brazil. Nguyen et al. find that, in their data for 1993, the urban-rural gap in household consumption per capita is primarily due to differences in covariates such as education, ethnicity, and age, and this is true throughout the distribution. In survey data for 1998, however, there is more evidence for differences in returns to covariates between the urban and rural sectors. In the case of Brazil, Skoufias and Katayama find that differences in household composition account for most of the inter-regional differences in consumption, but find some evidence for differences in returns across metropolitan areas, and between metropolitan and non-metropolitan urban areas.[9]

The importance of composition effects for income comparisons is clear. Less widely appreciated, a similar argument applies to spatial comparisons of happiness or well-being. The literature often assumes that a spatial equilibrium requires utility to be equalized across locations, but this will only happen if in-

[8] For a model that does give rise to spatial variation in skill premia, see Davis and Dingel (2012).

[9] They attribute these return differences to agglomeration economies, but this argument is not complete, because labour mobility should equalize returns across locations even in the presence of such economies. As in the case of Vietnam, alternative explanations could include a short-run disequilibrium, or unmeasured differences in characteristics (such as those in ability or the quality of schooling); Young (2013) emphasizes this latter possibility.

dividuals are homogeneous. When individuals differ, average well-being is also likely to differ in equilibrium (Deaton and Dupriez 2011). To give a concrete example, if retirees are happier than the middle-aged, and especially likely to move to Florida, then average self-reported happiness will tend to be relatively high in Florida. Again, compositional effects give rise to differences in average outcomes.

Although these effects make interpretations uncertain, Pittau et al. (2010) document some interesting differences in self-reported life satisfaction across European regions, with especially wide variation in Belgium, Germany, Italy, Portugal and Spain. For reasons that are unclear, residents of capital cities are especially likely to report low levels of life satisfaction. Knight and Gunatilaka (2011) summarize their work on happiness in China, which finds that (perhaps unexpectedly) mean urban happiness is slightly below mean rural happiness, while households of rural migrants living in the cities report lower average happiness than other urban households. But the compositions of these populations may differ in terms of characteristics that are hard to observe.

Oswald and Wu (2010, 2011) study differences in self-reported well-being across US states, and emphasize that these differences are not modest. Their data allow them to control for a variety of personal characteristics, and therefore address composition effects. The overall argument in Oswald and Wu (2010) is that well-being, conditional on individual characteristics including income category, is correlated with measures of local amenities (measuring non-income aspects of quality-of-life) extracted separately from a compensating-differentials approach, in earlier work by Gabriel et al. (2003). This correlation is seen as independent validation of the well-being measures. For the present chapter, it is more noteworthy that the correlation is consistent with a spatial equilibrium under labour mobility, in which the income received by a given type of worker will be lower in those states with good amenities.

Composition effects have a stark consequence. If heterogeneous workers are allowed to sort across locations within a market economy, then regional disparities in average incomes and even average life satisfaction are inevitable, and a condition of efficiency. This line of argument seems to conflict with common sense: surely equality across regions is obviously desirable? What the common-sense view misses is the need to follow Sen (1980), and ask 'equality of what?'. One possible criterion is to compare the utilities of a given type of individual across space. But, for the reasons just explained, a long-run equilibrium which equalizes these utilities will rarely equalize *average* outcomes.

3.2 When is there a regional problem?

Although the logic of a spatial equilibrium is powerful, the perception remains that uneven regional development is problematic. The literature often suggests that a spatial equilibrium will be inefficient, because externalities play a central role. Outside economics, observers often suggest that regional disparities are a form of social injustice. At a minimum, economists risk underestimating the burdens of adjustment; the experience of Detroit is a salutary reminder of a harsher reality.

One concern is that growth and agglomeration in a core area could make those living in the periphery worse off, even in absolute terms. This will be a particular concern when there are barriers or frictions that restrict the mobility of individuals or firms, and labour mobility may be especially difficult for the poor.[10] But there are ways in which injustice could arise even when mobility is unrestricted. It may be that human capital is relatively costly to acquire in poorer regions; since children cannot choose where to locate, regional disparities would contribute to differences in life chances and inequality. Olivetti and Paserman (2013) argue that, given the tendency for children to remain in the same region as their parents, regional disparities help to explain the decline in social mobility seen in the US in the first part of the twentieth century. Durlauf (2012) discusses the possible relevance of poverty trap ideas at the regional level.

Another argument is that spatial equilibrium takes time to achieve. There could be lengthy periods for which utility differences persist. Those who leave declining regions are likely to experience significant disruption in their lives, relative to the residents of more prosperous regions. Moreover, life chances may be influenced, in ways that economists have rarely analyzed. This perspective could imply less emphasis on comparisons of averages, and more on regional differences in labour markets, as in the work of Overman and Puga (2002) on the spatial polarization of European unemployment rates. But even this argument is not straightforward; in models based on matching frictions, if workers can move freely between locations, the asset value of unemployment will be the same across locations.[11]

[10] Guriev and Vakulenko (2012) find evidence for poverty-related immobility in Russia in the 1990s, and Phan and Coxhead (2010) for Vietnam.

[11] See Satchi and Temple (2009) and Kline and Moretti (2013) for related analyses. In a dual economy model with matching frictions, the 'urban' region could mean one city, or urban areas at multiple locations, each with the same labour market tightness, but of indeterminate relative size.

3.3 The origins of regional disparities

What are the underlying causes of regional differences in prosperity? Later sections of the chapter will consider this question in detail. But a useful first step is to set history to one side, and investigate the relative contributions of proximate influences, such as physical or human capital. For the purpose of an accounting exercise, we can adopt the simplifying device of a regional production function:

$$p_i Y_i = p_i A_i K_i^{\alpha} \left(h_i L_i\right)^{1-\alpha} \tag{1}$$

where Y_i denotes the aggregate level of output of region i, p_i the price of output net of trade costs incurred by local firms, A_i is total factor productivity, K_i is the region's capital stock, and $h_i L_i$ is the region's supply of effective units of labour, where h_i is the average human capital level. Our primary interest in this section is to see how much of the variation in $p_i Y_i$ can be explained by variation in the factors on the right-hand side of (1).

This exercise is a regional counterpart to the cross-country literature on development accounting reviewed by Caselli (2005) and Hsieh and Klenow (2010). The main finding of that literature is that international output differences are only partially explained by differences in physical capital and educational attainment, with the majority accounted for by total factor productivity, measured as a residual term. Hsieh and Klenow (2010) suggest that TFP accounts for 50-70% of cross-country output differences, with human and physical capital only accounting for 10-30% and 20% respectively. Should we expect similar findings at the regional level?

Aiello and Scoppa (2000) and Scoppa (2007) investigate this for Italy, and Tamura (2012) for the US. For some countries, regional data are more detailed than cross-country data; human and physical capital may be available over longer periods of time, and at lower levels of aggregation. A disadvantage is that regional price levels (p_i) are often unobserved, although Tamura (2012) uses (limited) information on regional price dispersion in the US.[12] Broadly speaking, these studies tend to confirm the cross-country finding that TFP is more important than human and physical capital in explaining output differences. However, this conclusion is sensitive to the way in which human capital is measured. For example, Scoppa (2007) finds that using quality-adjusted education levels can raise the contribution of human capital to over 50%.

[12] We discuss regional price deflators in the appendix to this chapter. Another issue for some studies is that the assumption of Cobb-Douglas technologies is restrictive; Bernard et al. (2013) find evidence against this assumption for the US.

Other papers also find that regional prosperity and average human capital are strongly associated. This includes work on spatial sorting, drawing on the urban economics literature. Combes, Duranton and Gobillon (2008) use a large panel of French employees (close to 20 million observations for the years 1976-1998) to isolate the importance of skill composition in explaining local variation in wages. They find that up to half of the variation in wages across 341 French employment areas can be explained by differences in skills. For 119 areas of Great Britain, Rice, Venables and Patacchini (2006) find a smaller, but still substantial, role for occupational composition.[13]

The evidence points in a consistent direction. In the study of regional prosperity, a key question is why skilled individuals are more likely to choose to locate in some regions than others. To answer this question needs general equilibrium models that can map regional characteristics into endogenous outcomes such as the supply of skills in each region, other factor supplies, and (in some cases) the endogenous determination of total factor productivity. Much of the rest of the chapter will be about this endeavour.

Before we describe the relevant theories, there is another point to note. We have discussed differences in outcomes across locations, but we have not allowed the nature of those locations to play a determining role. At least since Adam Smith's *Wealth of Nations*, it has been acknowledged that regional outcomes are related to their physical geography. It seems equally clear that outcomes are related to the outcomes of neighbours, and more broadly, to economic geography. To illustrate this, we plot regional GDP per capita against distance from Luxembourg, for a large number of European regions; see Figure 3, reproduced from Breinlich (2006). The strong correlation suggests that models of regional prosperity will need to engage with physical and economic geography.

A range of other observations support the idea that regional prosperity is connected with geography. There are clear spatial correlations of activity within most countries. Activity tends to gravitate towards areas with relatively good transport links, and close to large markets. It is unlikely to be an accident that much of China's industrial development has been concentrated on its coast, or that Brazil's interior is poorer than its coastal cities. Even for a well-integrated, developed economy such as the US, much activity is located on the coast, while

[13]These findings are based on a decomposition of earnings differences, but as in cross-country variance decompositions, it is not clear how covariance terms should be treated. Duranton and Monastiriotis (2002) study changes in regional wage inequality in the United Kingdom, between 1982 and 1997, and especially the divergence between London and other regions. This was mainly driven by a stronger rise in education levels in the capital and an increase in relative skill premia.

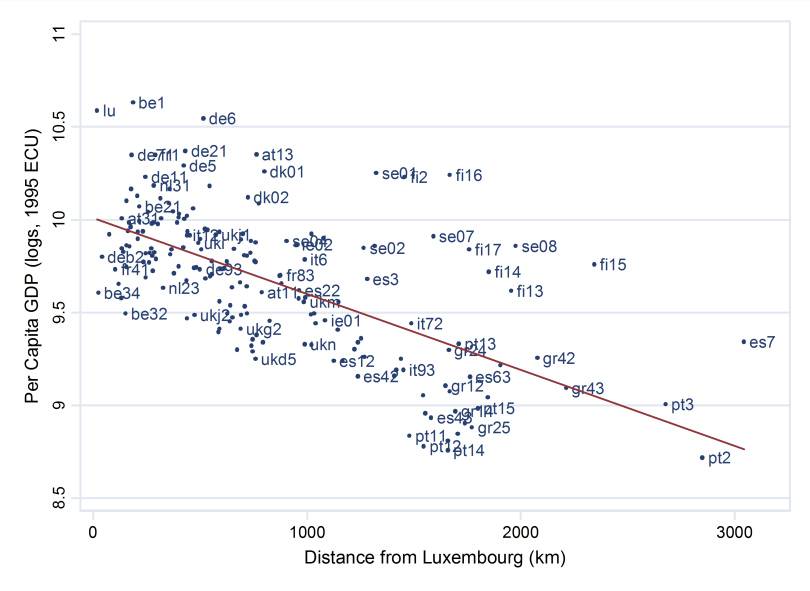

Figure 3: Regional GDP and proximity to Luxembourg

population movements are influenced by aspects of physical geography, such as climate.

4 Models of regional prosperity

Geography clearly matters, but understanding its implications often requires formal models. As in other general equilibrium contexts, such as the study of international trade, informal reasoning can easily go astray. Yet for a long time, the study of regional prosperity was rather overlooked by economists. In the post-war period, it tended to attract those who were sceptical about some of the tools and findings of conventional economic theory. The complexities of the regional growth process were interesting in themselves, but could also be used to illustrate reservations about economic theory (and general equilibrium theory in particular) that applied more widely.

Perhaps the best-known contribution along these lines is that of Myrdal (1957). He argued that market forces would lead to divergence between regions and ongoing disequilibrium, driven by 'cumulative causation', or the tendency for a change in a given direction (such as an increase in one region's productivity) to instigate further changes in the same direction. Kaldor (1970) took up this theme, emphasizing that localized industrial development generates cumulative advantages, due to various forms of increasing returns. But Hirschman (1958) had already countered that such views went too far. Although some degree of

uneven development was inevitable, it would be self-limiting: as regions moved apart, there would be powerful forces working to limit further divergence.

The further development of the literature has revealed some truth in both sets of positions. A contemporary treatment will rarely see an inherent contradiction in the coexistence of feedback effects and the study of an equilibrium; but it is likely to follow Myrdal in stressing the importance of these feedback effects for analysing regional prosperity. Agglomeration may promote further agglomeration, and initiate other changes that are part of a larger, self-sustaining process. Further, Myrdal and Kaldor were interested in the idea that the forces which drive growth and agglomeration are not readily separable, and this recognition continues to pose a major challenge for theorists.

4.1 A theoretical challenge

The formal general equilibrium models of the 1950s and 1960s typically assumed perfect competition, so that firms are price takers in markets for output and factors. But when firm and individual location decisions are introduced in realistic ways, perfect competition can rarely accommodate an interesting equilibrium. Consider what happens when output depends on capital and labour, and these factors can move freely between regions. In that case, workers and firms will all move to whichever region has the highest total factor productivity.

This argument can be expressed in more formal terms. For location decisions to be non-trivial, imagine that individuals and firms must choose an address, and these choices affect their utility and profitability respectively. This assumes some indivisibility in the way they spread their consumption and production activity across space, and also assumes — for space to play a determining role — that there are transaction or transport costs involved when consumption and production are geographically separated. These are realistic assumptions, but they cause the competitive paradigm to break down (Starrett 1978). A sufficient condition for this breakdown is that different locations have the same characteristics. Then, either there is no equilibrium with perfect competition, or all individuals and firms gather in a single location. More formally, if preferences are monotone, space is homogeneous, and transport is costly, there is no competitive equilibrium which involves transportation.

To generate an interesting location problem, localized externalities and indivisibilities at the level of individual firms and workers are necessary. But it is costly transportation that ultimately gives substance to the effects of geography. In the absence of transport costs, space is immaterial even when individuals

have distinct locations. Scotchmer and Thisse (1992) call this the folk theorem of spatial economics. It underpins two laws of economic geography set out by Prager and Thisse (2012): not all activities are available everywhere (the first law) and what happens close to us is more important than what happens far from us (the second law).[14]

These ideas play a fundamental role in the economic analysis of spatial equilibrium. In the literature that followed Krugman (1991), they have been studied by allowing for increasing returns and market size effects. This has been achieved by sacrificing generality, using various simplifying devices. But making even these models dynamic is not straightforward. Theorists face a trade-off, balancing the demands of realistic geographic and spatial considerations against the simplicity needed for a manageable dynamic analysis. Since agglomeration and growth are complex phenomena on their own, often models with proper geography lack interesting dynamics, while models with interesting dynamics lack proper geography. We use this idea to organize our discussion of various formal models. We first consider growth models that are largely without spatial considerations ('growth without geography'); multi-sector models with some limited spatial content or implications ('growth with limited geography'); and spatial models largely without dynamics ('geography with limited growth'). The hardest problem, to model growth and agglomeration jointly, is deferred until section 5; we call this 'geography and growth'.

4.2 Growth without geography

Asked to consider regional prosperity, some economists have settled on a default approach, which is to see whether progress can be made by ignoring space altogether. At its extreme, this approach treats regions as if they are separate countries. Their locations may differ, but space has no determining role. Trade in goods, and the movement of factors between regions, are either frictionless or (more commonly) ruled out altogether. We call this form of approach 'growth without geography'.

In particular, various neoclassical growth models remain an organizing framework for some research. They are used to think about the determinants of productivity levels, and to motivate many of the convergence studies discussed in section 2 above. This might be a useful place to start, but it is no place to end. By construction, the models cannot account for the patterns of spatial dependence

[14] Also note Tobler's first law of geography: everything is related to everything else, but near things are more related than distant things.

seen in the data, or changes in the spatial distribution of economic activity. The neoclassical growth models have dynamics — capital accumulation, productivity growth — but, to borrow the words from a popular song, they don't know much about geography.

The most sophisticated defence of the neoclassical growth model is that given by Barro and Sala-i-Martin (2004). They argue that, provided capital and labour are not perfectly mobile, the main consequence of factor flows between regions is to modify the rate at which regional economies converge to their steady-states. For example, Shioji (2001) develops a growth model with exogenous public capital, and private capital that is mobile between regions but subject to adjustment costs. This leads to a conditional convergence equation where the steady-state depends on the equilibrium return to private capital, and the region's stock of public capital. This is a more detailed treatment than many in the literature, but problems arise if workers and firms make location decisions based on the spatial distribution of activity, or other characteristics of distinct locations. The logic of a spatial equilibrium requires location decisions and growth to be analyzed jointly, and the neoclassical growth model rules this out.

4.3 Growth with limited geography

Recent models of agglomeration and growth often imply a core-periphery structure, sometimes corresponding to a division between rural and urban regions. This connects with an older class of models, the dual economy tradition, where urban manufacturing and services coexist with rural agriculture. These models have long been studied within development economics, and on the borders of growth economics and trade theory. Historically, dual economy models have been studied mainly in terms of comparative statics: for example, the effect of a given productivity change, or factor accumulation, in changing the equilibrium.[15] Recent work has given greater emphasis to structural transformation as an ongoing, dynamic process. These models provide some insights into spatial equilibrium and the relative growth of rural and urban regions; they also provide a laboratory for developing some basic intuitions, before turning to richer models with more geographical content.

Strictly speaking, in the traditional approach to dual economies, the goods of the respective sectors are labelled, but not the locations of consumption and production, and firm location decisions are not modelled.[16] In the newer models

[15] See Temple (2005) for a survey that emphasizes their empirical applications.

[16] In those dual economy models which incorporate migration costs for workers, it would often

from economic geography, the core-periphery structure emerges endogenously. In contrast, the older models can be seen as reduced forms, in which urban locations happen to have substantial advantages.

Typically, the agricultural and urban sectors are each modelled as producing an homogeneous good, under conditions of perfect competition. The relative price of the agricultural good is either determined exogenously (by the world prices facing a small open economy) or determined by utility maximization (in a closed economy). Less often, the agricultural and urban goods are treated as perfect substitutes. The location decisions of workers play a key role, so that one endogenous variable is the allocation of workers across the two sectors, and another is the equilibrium wage differential between the sectors/regions.

In the simple case where wages equal marginal products, and labour mobility equalizes wages, this maximizes aggregate output, in the absence of externalities or distortions. At the same time, the average product of labour will typically differ across sectors; regional differences in average productivity are a condition of efficiency rather than a sign of its absence. In richer dual economy models, however, inefficiency can easily arise, and urban regions may be too small or too large relative to rural regions.[17]

Some of the most interesting extensions to these models start with the urban labour market. In the model of Harris and Todaro (1970), a fixed urban wage leads to urban unemployment, and migration takes place unless expected utilities are equalized across the rural and urban sectors.[18] In many dual economy models with urban unemployment, productivity growth in the urban region will induce a migration response that increases the number of urban unemployed — the 'Todaro paradox'. In contrast, productivity growth in the rural region will increase rural wages and also relieve the pressure on cities, leading to better outcomes in the urban labour market and smaller regional disparities.

An especially rich approach to dual economies has been developed by Lagakos and Waugh (2013). They consider a general equilibrium Roy model, in which heterogeneous workers sort across sectors according to their comparative advantage. One implication of a Roy model is that all but the marginal worker will strictly prefer their current sector to the alternative. A calibrated version of the model can explain a large wage gap between agricultural and non-agricultural

be most natural to interpret agricultural production as taking place at a single point, and urban manufacturing and services production all taking place at another single point.

[17] Different versions of this can be seen in Graham and Temple (2006) and Satchi and Temple (2009).

[18] Approaches with endogenous wages often have similar implications; see Bencivenga and Smith (1997), Moene (1988) and Satchi and Temple (2009), among others.

workers, without having to appeal to barriers to labour mobility, the traditional approach in dual economy models. It can also explain why international variation in agricultural productivity is usually found to be much larger than international variation in non-agricultural productivity (Caselli 2005).

Even simple two-sector models demonstrate the importance of general equilibrium reasoning, while allowing for multiple sources of growth. In most of the models, the respective paths of rural and urban regions depend partly on rates of technical progress, and partly on capital accumulation. The accumulation of capital often leads to relative expansion of the sector/region which uses it more intensively (usually, but not always, the urban region). For the basic 2×2 model of trade theory, with two goods and two factors, where the factors are both mobile between sectors, this result is the standard Rybczynski effect. A version of that effect reappears in models with alternative labour market assumptions, such as the open economy version of the Harris-Todaro model studied by Corden and Findlay (1975).

In recent years, attention has shifted to dynamic versions of these small-scale general equilibrium models. These are to study structural transformation, and primarily the shift out of agriculture, as part of a transition towards a balanced growth path. Kongsamut et al. (2001) showed that this required strong assumptions that are unlikely to hold in practice. The approach of Ngai and Pissarides (2007) shows how to combine differential productivity growth rates with a balanced growth path, at the expense of restrictive assumptions on production technologies. Their (closed economy) model can explain ongoing declines in the relative price of the manufacturing good and the employment share of that sector.

Other recent work has started to combine dual economy ideas with models from urban and regional economics. Murata (2008) introduces a new mechanism for structural transformation, which draws on the New Economic Geography literature. In his model, a fall in transport costs increases the size of the market for non-agricultural goods, and also - by lowering prices and raising real incomes - prompts a demand shift towards non-agricultural goods. Henderson and Wang (2005) construct a model of the rural-urban transformation which draws on dual economy ideas, but extended to consider the endogenous evolution of distinct cities, and allowing the formation of new cities. Michaels et al. (2012) study the US evolution of populations in rural and urban areas from 1880 to 2000, explaining the observed patterns partly in terms of structural transformation.

Rural-urban income differences can make a substantial contribution to over-

all inequality.[19] Dual economy models can be used to study this, and changes in the relative productivity of different sectors as development proceeds. There is long-standing evidence that structural change is associated with increases in the relative labour productivity of agriculture; see, for example, Temple and Woessmann (2006). Not all dual economy models readily generate this pattern, which makes it a useful test. Gollin et al. (2004) argue that introducing home production leads to a better explanation of the data. More broadly, one weakness of dual economy models is that not much attention has been paid to the modernization of agriculture; Yang and Zhu (2013) is a recent exception.

Although dual economy models have spatial implications, the locations are only differentiated by rural or urban activity, which limits their usefulness for understanding regional prosperity. The model of Gennaioli et al. (2013a) moves further in the required direction: workers and firms have distinct addresses, and decide where to locate. There are two possible types of region, productive and unproductive; at each location, there is a fixed supply of land and housing, and hence some part of the population remains in the less productive regions. The regions all produce the same good, which is freely traded internally. A key margin in the model is that especially able workers will self-select into entrepreneurship, and more able entrepreneurs run larger and more productive firms. Relative to most dual economy models, this gives greater importance to the stock of human capital, an idea that Gennaioli et al. investigate empirically. But the tractability of the model inevitably comes at a price. Although locations exist as discrete points with fixed stocks of land and housing, there is no role for transport costs, and hence the model cannot explain the spatial correlation of activity that is so apparent in the data.

Most dual economy models contrast agriculture with non-agriculture, but the divergent paths of manufacturing and services are increasingly important, for developing countries as well as developed. Desmet and Rossi-Hansberg (2009) argue that the age of a sector matters for the dynamics of agglomeration, defining a sector's age as the time that has elapsed since the last major innovation, such as electrification (for manufacturing) or IT (for services). In its early stages, a major innovation will spur geographic concentration, because knowledge spillovers are important as the technology is refined. But as further development of the technology slows down, concentration gives way to dispersion. Desmet et al. (2012) use this framework to analyze the evolution of employment density across districts of India, where growth has been associated

[19] See Milanovic (2005a, 2005b) and Young (2013), among many others.

with the rapid expansion of services in particular.

Much of the work discussed thus far originates in development economics. A parallel literature describes the interactions between growth and the size distribution of cities, drawing on work in urban economics. In Eaton and Eckstein (1997) and Black and Henderson (1999), localized externalities sustain the emergence of cities and generate increasing returns at the aggregate level, so that agglomeration triggers growth. Gabaix (1999) and Eeckhout (2004) show how models featuring exogenous localized growth and localized externalities can generate the stable distribution of city sizes observed in the data. In particular, they seek to explain why the upper tail of the distribution is approximately Pareto - so Zipf's law holds - although both very small and very large cities are systematically under-represented.

The 'random growth' approach, revived in the recent literature by Gabaix (1999), starts with an initial arbitrary distribution of city sizes and lets each city grow at an arbitrary mean rate, around which cities are hit by period-to-period shocks. It then allows the cities to evolve freely and studies the conditions under which their limit size distribution mimics the observed one. Eeckhout (2004) assumes that total factor productivity in a city is determined by a positive localized externality that increases with city size and an exogenous process of localized technological change. In particular, letting $A_{i,t}$ be the productivity parameter reflecting the technological advancement of city i at time t, Eeckhout assumes that the law of motion of $A_{i,t}$ is given by $A_{i,t} = A_{i,t-1}(1 + \sigma_{i,t})$ with each city experiencing an exogenous technology shock $\sigma_{i,t}$. City-specific shocks are symmetric as well as identically and independently distributed with mean zero and $1 + \sigma_{i,t} > 0$. This law of motion implies that $\log(A_{i,t})$ follows a unit root process. There is clearly no growth in productivity in aggregate but, under appropriate functional forms, the model converges to a long-run distribution of city sizes whose upper tail is Pareto. City growth is proportionate, as also observed in reality.

Here, growth determines agglomeration, but the growth process itself is treated as exogenous. To fill this gap, Duranton (2007), Rossi-Hansberg and Wright (2007) and Córdoba (2008) propose models that generate growth processes consistent with specific features of the observed invariant distributions of city sizes. In all three contributions, growth leads to agglomeration. Duranton notes that it may be easier to match the city size distribution than it first appears, suggesting more attention is needed to the empirical relevance of the different possible mechanisms. Rossi-Hansberg and Wright address a particular

conundrum for spatial theories of growth: the inherent tension between local increasing returns, implied by the existence of cities, and aggregate constant returns, implied by balanced growth. They show that variation in the urban structure through the growth, birth, and death of cities can be seen as the margin that eliminates local increasing returns, to yield constant returns to scale in the aggregate. Their model produces a distribution of city sizes that is consistent with the real one, and whose dispersion is also consistent with the dispersion of productivity shocks found in the data.

The close connection between these models and specific features of the data is attractive. But their usefulness for studying regional growth is ultimately constrained, because the models do not investigate how cities will be distributed across space. As a result, key features of the observed geographical distribution of economic activities, and their evolution through time, are absent.

4.4 Geography with limited growth

We now turn to more complex models, which draw heavily on ideas from international economics. The models we review are predominantly static. Nevertheless, it is often argued that static models can be used to understand the steady-state implications of dynamic processes, otherwise too complex to analyze. In the wake of a major change, such as a fall in transport costs or an improvement in total factor productivity, the outcome of regional adjustment processes can be understood in terms of the changing steady-state of a static model. We call this approach 'geography with limited growth'.

Traditionally, this approach has been the backbone of the economics of agglomeration (Fujita and Thisse, 2002). These models address a fundamental question: can economic interactions generate spatial patterns of activity that are not determined solely by differences in exogenous fundamentals? Will asymmetric patterns of activity emerge even when locations are symmetric? As discussed by Ottaviano and Thisse (2004), the fact that economic activities are unevenly distributed in space is hardly surprising, given that locations differ in their climates, degrees of accessibility, and endowments of productive resources. All these features can be classified under the common label of *first nature*. These features have undoubtedly played an important role in explaining economic history, not least in the early stages of economic development. Exogenous spatial heterogeneity is the cornerstone of neoclassical models of international trade, and land use models in the tradition of von Thünen.

But another driving force of economic history has been the ongoing search

for safe and cheap ways to move materials and products from one location to another. One consequence is that the spatial distribution of economic activity will not map directly against the spatial distribution of natural advantages. When workers maximize utility and firms chase profits, this will generate endogenous patterns of economic activity across space. This idea is captured by the concept of *second nature*, the forms of economic geography that emerge as the outcome of human actions. Modern theories of agglomeration study the relevant forces, unveiling how spatial patterns of activity depend partly on exogenous spatial heterogeneity, and partly on a range of other variables, not least transport costs.

Second nature geography is the outcome of an inherently dynamic process but, as already noted, it can be understood partly by means of tractable static models. At least since Marshall (1920), various second-nature forces have been studied by economists, geographers and regional scientists, stemming from different types of localized technological and pecuniary externalities. For instance, technological externalities associated with production are stressed by modern urban economics, while pecuniary externalities associated with imperfect competition are stressed by spatial competition theory and work in economic geography (see Rosenthal and Strange, 2004).

The literature on these questions is vast, and a thorough assessment is beyond the scope of this chapter. Extended discussions can be found in Fujita and Thisse (2002), Combes, Mayer and Thisse (2008), Neary (2001), Prager and Thisse (2012), and various volumes of the *Handbook of Regional and Urban Economics*. Here we want to highlight the findings most relevant to understanding regional prosperity. These can be summarized in terms of the so-called 'spatial question' in economic theory (Ottaviano and Thisse, 2001). This question has two sides, one positive and one normative. On the positive side, the question at stake is whether the agglomeration of economic activities can be explained in terms of an explicitly-defined market mechanism. On the normative side, if observed patterns of economic activity can be seen in terms of market outcomes, the question at stake is whether such outcomes are likely to be efficient. The answers are 'yes' on the positive side and (usually) 'no' on the normative one. A range of models link agglomeration and market forces, but typically these models are built on externalities and distortions that lead to some degree of inefficiency (for example, Ottaviano and Thisse 2005).

As discussed by Ottaviano and Thisse (2005), the relative importance of technological and pecuniary externalities depends on the spatial scale of the analysis. According to Anas, Arnott and Small (1998), cities are replete with technological

externalities. The same holds in local production systems (Pyke, Becattini and Sengenberger, 1990). Besides local public goods, 'communication externalities' are of particular interest. These could be critical in services such as management, administration, research, and finance. Knowledge, ideas and, above all, tacit information, can be considered as impure public goods that generate spillover effects from one firm or organization to another. If economic agents possess different pieces of information, pooling them through informal communication channels can benefit many, hence the importance of proximity (Feldman, 1994). Thus, to explain geographical clusters of somewhat limited spatial dimension, such as cities and industrial districts, it seems reasonable to appeal to technological externalities. In modelling terms, these can often be accommodated in the competitive paradigm. Future work is likely to draw on the economics of networks, to consider the dissemination of information within and across regions.

But when one turns to a larger geographical scale, alternative mechanisms come into play, and ones that are not easily accommodated in conventional general equilibrium models. Direct physical contact seems unlikely to explain major agglomerations such as the US manufacturing belt, or Western Europe's concentration of economic activity along the 'Hot Banana' urban corridor, stretching from northern England to northern Italy. Instead, economists have sought to explain large-scale agglomeration in terms of pecuniary externalities. These arise from imperfect competition, in the presence of market-mediated linkages between firms and consumers/workers. The relevant models are often grouped under the banner of the New Economic Geography (NEG), which emerged in the early 1990s. This approach draws heavily on analytical tools and ideas from the theory of international trade (see, in particular, Helpman and Krugman, 1985). These tools are used to study the movements of goods, services and factors within countries, and to explain agglomeration as the outcome of endogenous processes in which cumulative causation often plays a role.

This literature was founded by Krugman (1991), who develops a model in which agglomeration arises through the mobility of labour. This mobility endogenously generates variations in market size that promote further agglomeration since, in the presence of transport costs, firms want to locate near large markets. Spatial agglomeration can also rise through input-output linkages, in which the location choices of firms influence the size of the market for other firms and/or input costs (Venables 1996). Some ideas have also been borrowed from urban economics: congestion and rising land rents can be introduced to offset the intrinsic advantages of particular regions, as in Helpman (1998) and Gennaioli et

al. (2013a), among others.

Endogenous agglomeration arises because mobile factors like to cluster, and this can polarize the regional landscape between an active 'core', and a 'periphery' in which immobile factors face lower real remuneration. The emergence of a core-periphery structure typically depends on the level of trade frictions. In the absence of congestion in the use of land or other non-tradable and non-replicable resources, low trade frictions foster agglomeration, as immobile demand in the periphery can be serviced from the core (Krugman, 1991; Krugman and Venables, 1995). When congestion matters, the opposite is true: if trade frictions are low, the high local cost of non-tradables pushes mobile factors away from the core (Helpman, 1998). In the general case, agglomeration is more likely to emerge for trade frictions that are neither too low nor too high (Puga, 1999; Ottaviano, Tabuchi and Thisse, 2002).[20]

Early NEG models were mainly aimed at explaining the 'residual variation' of economic activities across locations based on second nature forces, classified as promoting either agglomeration or dispersion. Later models have increasingly brought first nature into the picture. To see what might be learnt about regional growth from this form of approach, we provide a quick sketch of the framework used in Rice and Venables (2003).[21] Assume two types of workers, skilled and unskilled, each with Cobb-Douglas utility functions based on four goods — housing, an international tradable good, a good that can be traded domestically but not internationally (e.g., certain financial services) and a good that cannot be traded domestically (e.g., restaurant meals or haircuts). The goods other than housing are each produced using Cobb-Douglas technologies. The international tradable, and the non-tradable, are treated as homogeneous, produced under constant returns to scale and perfect competition. The nationally-traded good is produced under monopolistic competition as in Dixit and Stiglitz (1977). Finally, assume that workers can freely move between cities. Other things equal, they will migrate to cities with some intrinsic advantage, such as better amenities, until the advantage is offset by higher commuting costs and higher land rents. All rents are distributed to workers as a lump sum, in proportion to wages.

Under these assumptions, the prices of all goods are the same in all locations, but the skill mix of the labour force in each city is indeterminate. Two cities with

[20] See Fujita, Krugman and Venables (1999), Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud (2003), Ottaviano and Thisse (2004), and Combes, Mayer and Thisse (2008) for detailed accounts of NEG models.

[21] See also Overman, Rice and Venables (2010), who develop a diagrammatic approach to economic linkages across space.

different relative endowments of skilled labour will produce different quantities of the two traded goods, but have the same factor prices — a version of the factor prize equalization theorem of trade theory, applied within a country. Cities may then differ in terms of not only skill mix, but also in size and GDP per employee; and many different configurations of these outcomes are possible. But in the model, the free movement of labour implies that the utility of a given type of worker must be the same in all locations. And since wages are equal across locations, housing costs plus commuting costs must also be equal across cities. In this case, cities that are relatively skill abundant and high income must also be low density, so that the relatively high housing demand of skilled workers is offset by lower commuting costs.

At first glance, there is no 'regional problem', because the mobility of labour ensures that the utilities of a given type of worker are equalized across locations. But as new workers enter the labour force, the attainment of equilibrium relies on migration: each new generation has to relocate to restore the balance between the production structure of individual cities and their endowments of skilled relative to unskilled labour. In a richer model, reallocations could involve significant costs.

This simple setup has some counterfactual predictions: for example, the model predicts a negative correlation between GDP per employee and density, but the correlation in the data is often thought to be positive. A richer model gives one city/region an intrinsic advantage that will make it larger in equilibrium, and introduces transport costs for the nationally-traded good, which is assumed to be skill-intensive. The larger market of the dominant city makes it a profitable location for the nationally-traded good, bidding up wages (and GDP per employee) in the dominant city. Given transport costs, the price of the nationally-traded good is lower in the dominant city, and workers are attracted to the city by higher wages and lower prices, until these advantages are offset by higher housing costs. Again, utilities are equalized in equilibrium, but changes in underlying parameters, such as transport costs or the intrinsic advantages of the dominant city, will generate population and asset price movements.

An alternative modification assumes that one city has a productivity advantage in the production of the (skill-intensive) international tradable; no transport costs; and commuting costs that are equalized across cities. Now consider an increase in the traded-sector productivity advantage of a dominant city. This will raise wages in the dominant city, crowding out the nationally-traded sector, and attracting workers until the high wages are fully offset by a higher price of

the non-traded good and higher housing costs. This latter version of the model generates positive correlations between density, GDP per employee, wages, average skills, price levels and housing costs, which may often be the empirically relevant case. But there is no distinctively spatial pattern to the process of agglomeration.

For now it is interesting to consider what has been learnt from the sketch above. There may be disparities in skill endowments and GDP per employee across regions, but regions that appear advantaged may also have higher housing costs and higher prices for goods that are not traded across regions. Utilities are equalized in equilibrium by assumption. But spatial disparities continue to have relevance for policy-makers, not least if there are costs of adjustment. Changes in parameters — such as the productivity advantage of one city — might induce a lengthy transition process that has relatively modest ultimate benefits. In a numerical example in Rice and Venables (2003), a relatively modest change in traded-sector productivity can generate large population movements. This process of transition and adjustment may be associated with equilibrium utility levels that are only modestly higher than before. It is noteworthy that, in a quantitative exercised based on US data, Desmet and Rossi-Hansberg (2013a) find a similar result: eliminating differences in productivity or amenities across US cities would lead to major population movements, but modest welfare gains. For China, the welfare gains are estimated to be larger by an order of magnitude.

To date, the literature on geography and trade is long on models that study the mechanics of agglomeration in abstract landscapes, but remains short on the development of realistic quantitative versions of those models. This is largely because of their complexity, which often restricts the analysis to a small number of symmetric regions. Recent studies have started to fill this gap. In so doing, they have borrowed from the new literature on international trade in which the cross-country productivity distribution is endogenous; the literature reveals new sources of gains from trade under perfect competition (Eaton and Kortum, 2002) and imperfect competition (Bernard, Eaton, Jensen and Kortum, 2003; Melitz, 2003; Melitz and Ottaviano, 2008).[22]

A recent example of this work is Donaldson (2010), who develops a Ricardian trade model to study the effects of the Indian railway network, introduced by the British when India was under colonial rule. The model draws on Eaton and Kortum (2002) and features many regions, many commodities, and costly trade.

[22]For a discussion of these new sources of gains from trade see Arkolakis, Costinot and Rodríguez-Clare (2012) and Melitz and Redding (2013).

Regions are assumed to have different productivity levels across commodities, generating opportunities to exploit comparative advantage through trade. When two regions become linked by a railway, their bilateral trade cost falls and this allows for specialization according to comparative advantage. The empirical implementation of the model allows Donaldson (2010) to quantify the extent to which the railway network improved India's trading environment, in terms of lower trade costs, smaller inter-regional price gaps, and larger trade flows between regions and internationally. Further, he also investigates how much of the estimated reduced-form welfare gains plausibly arise from newly exploited gains from trade. In particular, he finds that those gains account for virtually all of the observed reduced-form impact of railways on real income estimated from the data.

More generally, Redding (2012) also develops a tractable model of regional economic geography based on Eaton and Kortum (2002), suitable for quantitative investigations. He studies the general equilibrium of an economy with an arbitrary number of regions connected by an arbitrary pattern of geographical trade costs, modelled as iceberg transport costs that may differ between all bilateral pairs of regions. Labour is assumed to be mobile across regions. The productivity level of each region is drawn from a Fréchet distribution. Regions with higher productivity pay higher wages, which attracts population until the higher wages are offset by higher living costs. Regions with good market access (low transport costs) will have low prices for traded goods, and again this is offset by population movements that drive up housing costs. Hence, in equilibrium, welfare is equalized across locations, but regions that are productive or well situated have higher nominal wages, larger populations and higher housing costs. The model is especially well-suited for studying the effects on regional growth of an economy-wide trade liberalization, or more generally, a fall in external trade costs. A liberalization of trade will lead to an endogenous internal reallocation of population, implying a combination of regional growth and regional decline.

An alternative vein of research gives more attention to the relationship between first nature geography and the decisions of mobile, and heterogeneous, people and firms under imperfect competition. Melitz (2003) and Melitz and Ottaviano (2008) provide a basis for this class of models. A key distinction is whether the heterogeneous characteristics of agents are assumed to be revealed to them before, or after, their location decisions. Sorting models study how heterogeneous agents, aware of their characteristics *ex ante*, will sort themselves into locations of varying sizes (Nocke, 2006; Baldwin and Okubo, 2006; Davis and Dingel,

2012; Okubo, Picard and Thisse, 2010; Picard and Okubo, 2012).[23] In contrast, selection models study what happens when heterogeneity materializes *ex post*, after agents have already committed to their locations: they can then self-select across whatever economic activities are available in those locations.

Behrens and Robert-Nicoud (2012) present a selection model where *ex ante* identical individuals decide whether to move from a common rural hinterland to cities. Their heterogeneity is revealed after this decision has been made, and the decision itself is assumed to be irreversible, which rules out sorting. They show that larger market size increases productivity partly through a finer division of labour driven by pecuniary externalities (richer availability of intermediates) and partly through a selection process. Meanwhile, higher productivity increases market size by providing incentives for rural-urban migration. Behrens, Duranton and Robert-Nicoud (2010) analyze both sorting and selection in a model where agglomeration is driven by technological externalities. They distinguish between *ex ante* heterogenity ('talent'), known to agents before they decide where to locate, and *ex post* heterogeneity ('luck'), revealed to agents after their location decisions have been made. Agents choose locations based on their talent, while luck influences subsequent occupational choices. More talented agents stand a better chance of finding more productive occupations in larger locations; this complementarity between talent and market size leads to the sorting of more talented agents into larger markets. Then, more demanding selection in more talented locations implies that average productivity is higher in these locations. Higher productivity, in turn, complements the agglomeration benefits of larger locations, and so markets with greater concentrations of talent are larger in equilibrium. Markups are constant, as in Melitz (2003). This implies that, conditional on sorting and agglomeration, selection becomes independent of market size.

Similarly to Behrens and Robert-Nicoud (2012), Ottaviano (2012) dispenses with *ex ante* heterogeneity (and first nature asymmetries) in order to investigate how firm heterogeneity influences the aggregate balance between agglomeration and dispersion forces, in the presence of pecuniary externalities. This is a selection model based on Melitz and Ottaviano (2008). A further departure from the analysis of Behrens and Robert-Nicoud (2012) is that the model allows location decisions to be reversible, and whether regions are characterized as 'urban' or 'rural' is determined endogenously by those decisions.[24] The emer-

[23] While other papers focus on firm heterogeneity on the supply side, in terms of productivity, the distinctive feature of Picard and Okubo (2012) is their study of heterogeneity on the demand side, in terms of tastes.

[24] Behrens, Mion, Murata and Südekum (2011) take a similar approach in their study of spatial

gence of agglomeration is driven by pecuniary rather than technological externalities. Markups are determined endogenously with larger market size leading to lower markups. This implies that, differently from Behrens, Duranton and Robert-Nicoud (2010) but as in Behrens and Robert-Nicoud (2012), selection is still more demanding in larger markets, even after conditioning out sorting and agglomeration.

Combes et al. (2012) also extend the model of Melitz and Ottaviano (2008) to allow for agglomeration economies driven by technological externalities. They estimate the relative importance of selection and agglomeration in determining the spatial distribution of firm productivity levels. Following Melitz and Ottaviano (2008), they rule out labour mobility across locations, although extensions to the basic framework can be made.[25] To distinguish between agglomeration and selection effects, they nest a generalized version of Melitz and Ottaviano (2008) and a model of agglomeration in the spirit of Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002). In larger (more dense) locations, the firm productivity distribution is left-truncated due to more demanding selection, but also right-shifted and dilated due to agglomeration. Combes et al. (2012) show how to estimate these effects by studying how the quantiles of the log productivity distribution in a large city will be related to the quantiles of the log productivity distribution in a small city. They estimate the relationship from data on French employment areas, and find no difference in the left-truncation of the log productivity distribution between dense and less dense areas. This suggests that the firm selection mechanism cannot explain spatial productivity differences across these areas. As they acknowledge, this result might not generalize to countries that are less well integrated than France, or where firms charge prices that differ across locations. Even for the French case, it does not rule out selection effects altogether, since their intensity could be the same across locations.

frictions, allowing for the joint determination of location sizes, productivity levels, markups, wages, consumption diversity, and the number and size distribution of firms.

[25] In a separate online appendix (http://diegopuga.org/papers/selectagg_webapp.pdf), they show how their model can be extended to include worker mobility, consumption amenities, and urban crowding costs, without affecting the key equilibrium equations on which their empirical analysis is based. They restrict their attention to a situation in which there exists a unique stable spatial equilibrium with (asymmetric) dispersion. In contrast to Ottaviano (2012), whether heterogeneity fosters agglomeration or dispersion is beyond the scope of their paper.

5 Geography and growth

We now consider growth models that make space for space: dynamic models of the growth process in which space plays a determining role. We have labelled this the 'hard problem' of regional economics. Desmet and Rossi-Hansberg (2012a, p. 2-3) provide a clear statement of the problem:

> Incorporating a continuum of locations into a dynamic framework is a challenging task for two reasons: it increases the dimensionality of the problem by requiring agents to understand the distribution of economic activity over time and over space, and clearing goods and factor markets is complex because prices depend on trade and mobility patterns. These two difficulties typically make spatial dynamic models intractable, both analytically and numerically.

One reason the problem becomes intractable is that, if we think of a dynamic model as one with forward-looking investment decisions, then agents must anticipate the solutions for future prices, and hence the equilibrium patterns of trade and mobility, at all future dates. As Desmet and Rossi-Hansberg note, the only way forward is to simplify the problem. Our review distinguishes between two families of models, dynamic NEG models and dynamic 'sequential market clearing' (SMC) models. These vary in whether locations are ordered in space as in the real world. Though the task of combining agglomeration forces with interesting long-run dynamics and growth paths is far from complete, these two families of models represent the current frontier of theoretical research on regional growth. With this in mind, we set the ideas out in detail.

5.1 Non-ordered space

We first consider non-ordered space. An influential literature in trade theory has developed dynamic models with two or more countries. The relevant contributions include Grossman and Helpman (1991), Young (1991), Ventura (1997), Eaton and Kortum (1999), and Cuñat and Maffezzoli (2007). In these models, either autarky is compared to free trade or, when trade costs are introduced, countries are not ordered in space. From the viewpoint of spatial economics, the most attractive members of this family are the dynamic versions of NEG models. These typically feature a small number of locations (in most cases only two) that exchange goods and ideas in the presence of frictions. In the wake of Krugman (1991), localized pecuniary externalities drive the agglomeration of

production. Endogenous growth is introduced by adding innovation in product variety with technological externalities, as in Grossman and Helpman (1991). The localized nature of these externalities, due to frictions in the exchange of ideas between regions, drive the agglomeration of innovation and can lead to cumulative causation in the location of production and innovation.

Baldwin and Martin (2004) survey several different specifications of these models and tease out their main insights. Cumulative causation implies the joint agglomeration of innovation and production. Aggregate growth is then driven by factor accumulation in a small subset of regions. This leads to 'growth poles' and 'growth sinks'. However, due to the localized nature of the technological externalities in innovation, the endogenous emergence of regional disparities is accompanied by faster aggregate growth and higher welfare in all regions.

Minerva and Ottaviano (2009) present a simple unifying model with two regions that encompasses a variety of insights from this line of research in a parsimonious way. It highlights the implications of geography for the dynamic process of regional growth. In this model, the geographical element arises partly due to costs of trading goods across regions ('transport costs') and partly from barriers to exchanging ideas ('communication costs'). The model illustrates how agglomeration and growth can reinforce each other, giving rise to the cumulative causation that Myrdal envisaged.

An important limitation should be acknowledged at the outset, which is that the model rules out labour mobility across regions. Relaxing this assumption is not straightforward, as we discuss later. It has been relaxed in an alternative class of models, based on sequential market clearing. These models, reviewed later, can accommodate labour mobility, congestion in land use and a large number of regions, without sacrificing analytical tractability. At the same time, the account of growth in such models tends to be more stylized than the one we describe here.

Following Minerva and Ottaviano (2009), let us assume that there are two regions, North and South. To abstract from first nature, the exogenous attributes of the two regions are the same. First, they are populated by an identical number Q of geographically immobile workers. As each worker supplies one unit of labour inelastically, Q is also the regional endowment of labour. Second, regions are endowed with an identical initial stock of knowledge capital K_0. Through time, profit-seeking R&D laboratories create additional knowledge capital that is freely mobile between regions. In so doing, they finance their investments through bonds, with riskless return $r(t)$ at time t, sold to workers in a perfect

inter-regional capital market. Henceforth, in the presentation of the model we will focus on North. Analogous expressions will apply to South.

Transport costs and localized spillovers play a key role in the analysis. Workers consume two goods, a homogeneous 'traditional' good Y and a horizontally differentiated 'modern' good D, with preferences given by the following utility function:

$$U = \int_{t=0}^{\infty} \log \left[D(t)^{\alpha} Y(t)^{1-\alpha} \right] e^{-\rho t} \mathrm{d}t \tag{2}$$

In (2) $D(t)$ represents the CES consumption basket of the different varieties of good D:

$$D(t) = \left[\int_{i=0}^{N(t)} D_i(t)^{1-1/\sigma} \mathrm{d}i \right]^{1/(1-1/\sigma)}, \quad \sigma > 1 \tag{3}$$

where $D_i(t)$ is the consumption of variety i and $N(t)$ the total number of varieties in the economy.

Given a unit elasticity of intertemporal substitution, intertemporal utility maximization determines the evolution of expenditures according to the Euler equation:

$$\frac{\dot{E}(t)}{E(t)} = r(t) - \rho \tag{4}$$

where $E(t)$ is individual expenditure. The Cobb-Douglas instantaneous utility function is then maximized when the shares α and $1 - \alpha$ of individual expenditures $E(t)$ are allocated to the consumption of the modern and traditional goods respectively. In turn, the fraction $\alpha E(t)$ is distributed across the varieties of the modern good depending on their relative prices. This gives individual demand:

$$D_i(t) = \frac{p_i(t)^{-\sigma}}{P(t)^{1-\sigma}} \alpha E(t) \tag{5}$$

In (5) $P(t)$ represents the exact price index associated with the CES consumption basket (3):

$$P(t) = \left[\int_{i=0}^{N(t)} p_i(t)^{1-\sigma} \mathrm{d}i \right]^{1/(1-\sigma)} \tag{6}$$

so that σ measures both the own- and the cross-price elasticities of demand.

The production of the traditional good is characterized by perfect competition and constant returns to scale with labour as its only input. An appropriate choice of units means that the unit labour requirement can be set to 1. This implies that the profit-maximizing price of Y equals the wage. The traditional good is assumed to be freely traded between and within regions. Hence, both its price

and the wage are equalized across regions. Selecting good Y as the numéraire pins down the common wage to 1.

The production of the modern varieties is characterized by monopolistic competition and increasing returns to scale. These arise from the presence of a fixed cost incurred in terms of one unit of knowledge capital per variety. Variable costs are incurred, instead, in terms of β units of labour per unit of output. Due to the fixed capital requirement, at any instant t the total number of varieties available in the economy is determined by the aggregate knowledge capital stock $K^w(t)$. In equilibrium there is a one-to-one relation between firms and varieties, and so the total number of firms $N(t)$ is equal to $K^w(t)$. In turn, due to the free mobility of knowledge capital, the entry decisions of firms will determine where varieties are actually produced, and we use $n(t)$ to denote the number of Northern firms and varieties. Entry is free, and at any given instant there are many potential entrants. These need knowledge capital to start producing. In the presence of a capital supply that is fixed at any given instant, competitive bidding by entrants transfers all operating profits to capital owners.

Geography is introduced in the product market, by assuming that trade flows of differentiated varieties face iceberg transport costs, within and between regions. The size of the internal transport costs differs across regions: in North and South, firms have to ship $\tau_N > 1$ and $\tau_S > 1$ units respectively, in order to deliver one unit to their domestic customers. As for inter-regional trade, the delivery of one unit requires the shipment of $\tau_R > 1$ units, regardless of the direction of trade. Within-region shipments are less costly than inter-regional ones, and this cost advantage is more pronounced for North. Hence, we have $\tau_N < \tau_S < \tau_R$. This ranking of the transport cost parameters identifies North as the developed 'core' and South as the developing 'periphery'.

All firms in both markets face the same constant elasticity of demand σ and the same marginal production cost β. Hence, profit maximization leads to the same producer price ('mill price') for all firms as a constant markup over marginal cost $p = \sigma\beta/(\sigma - 1)$. The corresponding consumer prices ('delivered prices') simply reflect differential transport costs: $p_N = p\ \tau_N$, $p_S = p\ \tau_S$, $p_R = p\ \tau_R$. With these prices, operating profits are $\pi(t) = \beta x(t)/(\sigma - 1)$. Here $x(t)$ denotes firm output inclusive of the quantity lost in transit, and the price index (6) can be rewritten as $P(t) = pN(t)^{\frac{1}{1-\sigma}} \left[\delta_N \gamma(t) + \delta_R (1 - \gamma(t))\right]^{\frac{1}{1-\sigma}}$, where $\gamma(t) = n(t)/N(t)$ is the share of firms located in North and $N(t) = K^w(t)$ is the total number of firms as well as the total stock of knowledge capital. The parameters $\delta_N \equiv (\tau_N)^{1-\sigma}$, $\delta_S \equiv (\tau_S)^{1-\sigma}$ and $\delta_R \equiv (\tau_R)^{1-\sigma}$ measure the efficiency

of internal and external transportation with $0 < \delta_R < \delta_S < \delta_N < 1$.

The national capital stock $K^w(t)$ is accumulated through profit-seeking R&D by perfectly-competitive laboratories facing constant returns to scale. Knowledge spillovers are assumed to increase the productivity of researchers as knowledge accumulates, and this sustains growth in the long run. Geography is introduced in the knowledge capital market through a specification of the R&D technology that encompasses both localized knowledge spillovers (Martin and Ottaviano, 1999) and intermediate business services (Martin and Ottaviano, 2001) as captured by the following constant-returns-to-scale production function:

$$\dot{K}(t) = A(t) \left[\frac{D(t)}{\varepsilon}\right]^{\varepsilon} \left[\frac{Q_I(t)}{1-\varepsilon}\right]^{1-\varepsilon} \tag{7}$$

where $\dot{K}(t) \equiv dK(t)/dt$ is the flow of knowledge created at time t, $Q_I(t)$ is labour employed in R&D, $D(t)$ is the basket of business services, and $\varepsilon \in (0,1)$ is the share of business services in R&D. Note that the basket of business services is assumed to be the same as the consumption basket, for analytical convenience. In (7) $A(t)$ refers to the North's total factor productivity in R&D and is assumed to be an increasing function of the total stock of knowledge $K^w(t)$ as embodied in the operations of modern producers. Specifically, the region-specific level of productivity in R&D is given by $A(t) = A\ K^w(t)^{\mu}\left[\omega_N \gamma(t) + \omega_R(1-\gamma(t))\right]^{\mu}$, where A is a positive constant. Here $\mu \in (0,1)$ measures the intensity of the knowledge spillovers, whose geographical diffusion is hampered by frictional communication costs. Their spatial decay is regulated by the ω's. It is assumed to be steeper between regions than within them, and steeper in South than in North, reflecting their different development stages. Hence $0 < \omega_R < \omega_S < \omega_N < 1$. The larger ω, the lower the corresponding communication costs.

Both transport and communication costs create an incentive for innovation to cluster where production also disproportionately happens. To see this, we can use profit maximizing prices and the equilibrium wage to compute the marginal cost associated with (7) as:

$$F(t) = \frac{P(t)^{\varepsilon} w^{1-\varepsilon}}{A(t)} = \frac{\eta}{N(t)\left[\omega_N \gamma(t) + \omega_R(1-\gamma(t))\right]^{1-\frac{\varepsilon}{\sigma-1}} \left[\delta_N \gamma(t) + \delta_R(1-\gamma(t))\right]^{\frac{\varepsilon}{\sigma-1}}} \tag{8}$$

where $\eta = p^{\varepsilon}/A$ is a positive constant, and we have imposed the constraint $\mu + \varepsilon/(\sigma-1) = 1$ so that in the long run the economy follows a balanced growth path. This constraint preserves the incentive to invest in R&D in the long run, as

the marginal cost of innovation decreases over time at the same rate as its benefit measured by the value of a firm.

Inspecting (8) reveals that, given the rankings of ω's and δ's, the marginal cost of innovation is lower in North provided it hosts a larger number of firms. As we will see, this is indeed the case in equilibrium, as lower internal transport costs increase the size of the local market. Hence, due to perfect competition among laboratories, in equilibrium they will all be located in North. Even though long-run growth is entirely driven by Northern innovators, they are still financed in the inter-regional capital market by both Northern and Southern workers. This implies that, in equilibrium, the value $v(t)$ of a unit of knowledge capital has to satisfy an arbitrage condition. It requires the bond yield $r(t)$ to be equal to the percentage return on investment in knowledge capital, consisting of the percentage capital gain $\dot{v}(t)/v(t)$ and the percentage dividend $\pi(t)/v(t)$:

$$r(t) = \frac{\dot{v}(t)}{v(t)} + \frac{\pi(t)}{v(t)} \tag{9}$$

where $v(t) = F(t)$ as, due to perfect competition in R&D, profit-maximizing laboratories price knowledge capital at marginal cost.

Finally, the model is closed by imposing that in equilibrium product and labour markets clear. Consider the product market first. Substituting the profit maximizing prices into demands (5) allows us to state the market-clearing conditions for Northern and Southern firms as:

$$\begin{aligned} x(t) &= \frac{p^{-\sigma}\delta_N}{P(t)^{1-\sigma}}\left[\alpha E(t)Q + \varepsilon F(t)\dot{N}(t)\right] + \frac{p^{-\sigma}\delta_R}{P^*(t)^{1-\sigma}}\alpha E^*(t)Q \\ x^*(t) &= \frac{p^{-\sigma}\delta_S}{P^*(t)^{1-\sigma}}\alpha E^*(t)Q + \frac{p^{-\sigma}\delta_R}{P(t)^{1-\sigma}}\left[\alpha E(t)Q + \varepsilon F(t)\dot{N}(t)\right] \end{aligned} \tag{10}$$

where an asterisk flags Southern variables. Only Northern demand is augmented by intermediate expenditures $\varepsilon F(t)\dot{N}(t)$ as R&D is active only in North. Turning to the labour market, this clears when the total endowment of labour $2Q$ is fully employed in innovation $Q_I(t) = (1-\varepsilon)F(t)\dot{N}(t)$, in modern production $Q_D(t) = [(\sigma-1)/\sigma]\left[2\alpha E(t)Q + \varepsilon F(t)\dot{N}(t)\right]$, and in traditional production $Q_Y(t) = 2(1-\alpha)E(t)Q$:

$$2Q = \frac{\sigma-\varepsilon}{\sigma}F(t)\dot{N}(t) + 2\frac{\sigma-\alpha}{\sigma}E(t)Q \tag{11}$$

We now study agglomeration and growth. The market clearing conditions for products and labour can be used to highlight how growth affects location

and, vice versa, location affects growth. We focus on a balanced growth path with constant expenditures and a constant growth rate of knowledge capital $g = \dot{K^w}(t)/K^w(t) = \dot{N}(t)/N(t)$. Constant expenditures imply $\dot{E} = 0$ so that, given (4), we have $r = \rho$. Further, FN and γ are constant, and hence the evolution of the value of knowledge capital is determined by the growth rate of knowledge capital through the implied change in the marginal cost of R&D, $\dot{v}/v = \dot{F}/F = -g$. In other words, the marginal benefit of innovation (v) and its marginal cost (F) both fall at the same constant rate.

The arbitrage condition (9) implies that, in equilibrium, all firms achieve the same level of profits and hence the same scale of output wherever they are. Then we can use (10) to determine this common output scale as

$$x = \left[(\sigma - 1)/\beta\sigma\right]\left[(2\alpha EQ + \varepsilon FNg)/N\right] \tag{12}$$

This can be used to rewrite (9) as a function of E, g and FN. The resulting expression can be solved together with labour market clearing (11) to show that, in equilibrium, expenditure equals permanent income:

$$2EQ = 2Q + \rho FN \tag{13}$$

and the growth rate satisfies:

$$g = \frac{\alpha}{\sigma - \varepsilon}\frac{2Q}{FN} - \rho\frac{\sigma - \alpha}{\sigma - \varepsilon} \tag{14}$$

Substituting (8) into (14) shows that *location affects growth* through the marginal cost of innovation FN net of the spillover from accumulated knowledge capital:

$$g = \frac{\alpha}{\sigma - \varepsilon}\frac{2Q}{\eta}\left[\omega_N\gamma + \omega_R(1-\gamma)\right]^{1-\frac{\varepsilon}{\sigma-1}}\left[\delta_N\gamma + \delta_R(1-\gamma)\right]^{\frac{\varepsilon}{\sigma-1}} - \rho\frac{\sigma - \alpha}{\sigma - \varepsilon} \tag{15}$$

In particular, more agglomeration in North makes innovation less costly and hence leads to faster growth.

The joint solution of the product market clearing conditions (10) determines not only the firms' common output scale, but also the share of Northern firms, as:

$$\gamma = \frac{1}{2} + \frac{1}{2}\frac{\delta_R\left(\delta_N - \delta_S\right)}{\left(\delta_N - \delta_R\right)\left(\delta_S - \delta_R\right)} + \frac{\delta_N\delta_S - \delta_R^2}{\left(\delta_N - \delta_R\right)\left(\delta_S - \delta_R\right)}\left(\theta - \frac{1}{2}\right) \tag{16}$$

In (16) $\theta = \left(\alpha EQ + \varepsilon FNg\right)/\left(2\alpha EQ + \varepsilon FNg\right)$ is the Northern share of expendit-

ures in the modern sector, after taking into account that $E = E^*$ since regions share the same initial endowments. It depends on the endogenous variables E, FN and g. However, using (14) it can be expressed as a function of g only, thus allowing us to rewrite (16) as:

$$\gamma = \frac{1}{2} + \frac{1}{2}\frac{\delta_R\left(\delta_N - \delta_S\right)}{\left(\delta_N - \delta_R\right)\left(\delta_S - \delta_R\right)} + \frac{1}{2}\frac{\delta_N\delta_S - \delta_R^2}{\left(\delta_N - \delta_R\right)\left(\delta_S - \delta_R\right)}\frac{\varepsilon}{\sigma}\frac{g}{g+\rho} \tag{17}$$

This shows that *growth affects location* through its influence on the Northern share of expenditures. In particular, faster growth increases the Northern expenditure share as innovation takes place only in North, which leads more firms to locate there.

Expressions (15) and (17) highlight a crucial result: agglomeration (larger γ) and growth (larger g) are jointly determined. Although the two do not interact dynamically, this can still be seen as a form of cumulative causation: forces which promote growth indirectly promote agglomeration, and vice versa. The outcome is a trade-off for policy-makers, between promoting growth and reducing regional disparities. Further insights into the role of geography are readily gained by focusing on two extreme cases that arise when the cost of innovation is determined by communication costs only ($\varepsilon = 0$) or by transport costs only ($\varepsilon = \sigma - 1$) as in Martin and Ottaviano (1999) and Martin and Ottaviano (2001) respectively. If $\varepsilon = 0$, lower communication costs within North foster growth but have no impact on agglomeration. The same applies to lower inter-regional communication costs. In contrast, lower communication costs in South have no impact as long as no innovation takes place there. Moreover, changes in transport costs affect location, but have no impact on growth. If $\varepsilon = \sigma - 1$, reductions in inter-regional and intra-North transport costs promote agglomeration in North as well as growth; reductions in intra-South transport costs promote relocation from North to South, but also hamper growth.

As we noted previously, this analysis has ruled out labour mobility, which is hard to accommodate in multi-region endogenous growth models. In principle, mobility could be introduced as in Fujita and Thisse (2003) but, absent congestion in land use, this would simply lead to the clustering of all factors in a single region. In principle, allowing for congestion in land use could avoid this outcome, but leads to a model that is analytically intractable. Studies that allow for labour mobility in a multi-region endogenous growth model, under perfect foresight, include Walz (1996) and Baldwin and Forslid (2000). As discussed by Fujita and Thisse (2002), the assumption of costless migration in Walz

(1996) leads to bang-bang behaviour that does not accord with reality. Migration is gradual in Baldwin and Forslid (2003), at the expense of analytical complexity. For reasons of tractability, Fujita and Thisse (2003) focus on a steady-state equilibrium in which the spatial distribution of skilled workers is time-invariant. Although they provide a stability analysis, the details of the transition process are not studied.

5.2 Ordered space

Dynamic NEG models enhance our understanding of the common forces underlying growth and agglomeration. As argued by Desmet and Rossi-Hansberg (2010), however, their focus on a small number of locations misses the richness of the observed geography of economic activities, and limits their empirical applications. Generalizing them to more than a few regions introduces problems of analytical tractability, especially when one allows for frictions in the mobility of capital (Baldwin, Martin and Ottaviano, 2001) or labour (Fujita and Thisse, 2003). Some progress could still be made through numerical methods, as shown by Fujita, Krugman and Venables (1999) for static models in a continuous space, but work in this vein remains limited.

A small number of papers study a fully dynamic setup with a continuum of locations: these include Brito (2004), Brock and Xepapadeas (2008, 2009) and Boucekkine, Camacho and Zou (2009). They typically focus on the allocation problem of a social planner but, absent more structure, it is hard to extract general insights. The main problem is that, in order to make decisions, forward-looking agents need to understand the whole distribution of economic activities over space and time implied by each feasible action.

Desmet and Rossi-Hansberg (2010, 2012a) advance an alternative approach, initially proposed in Rossi-Hansberg (2005), that is analytically tractable when space is continuous and one-dimensional. To reduce the complexity of the problem, they model a situation in which agents do not have to consider the future allocation paths, because these paths are beyond their control and do not affect their returns from current decisions. Hence, though forward-looking, agents solve static problems. This is achieved by imposing enough structure on the diffusion of technology or on the mobility of agents and the way property rights over land are allocated among them. This approach generates a dynamic process in which locations continuously change in occupational structure and employment density, but the aggregate economy converges to a balanced growth path.

In a simplified version of their model, Desmet and Rossi-Hansberg (2010)

study an economy in which all markets are perfectly competitive; locations accumulate technology by investing in innovation in one homogeneous-good industry and by receiving spillovers from other locations; factor mobility is frictionless; and trade is the result of agents holding a diversified portfolio of land across locations. Land is given by the unit interval $[0,1]$, time is discrete and total population is $\overline{L}$. The one-dimensional space $[0,1]$ is divided in connected intervals ('counties'), each administered by a local government.

Consumers-workers in location l solve the utility maximization problem:

$$\max_{\{c(l,t)\}_0^\infty} E \sum_{t=0}^{\infty} \beta^t U\left(c(l,t)\right)$$

subject to:

$$w(l,t) + \frac{\overline{R}(t)}{\overline{L}} = p(l,t)c(l,t) \forall l,t$$

where $U\left(c(l,t)\right)$ is the instantaneous utility of consumption $c(l,t)$ in period t, β is the discount factor, and E is the expectation operator. Consumption incurs a price $p(l,t)$. Income consists of the wage $w(l,t)$ and the share $1/\overline{L}$ of total land rent $\overline{R}(t)$ under the assumption that consumers hold a fully-diversified portfolio of land across locations. Due to free labour mobility, in each period t, utility is the same everywhere.

Production employs labour and land with technology:

$$x\left(L(l,t)\right) = Z(l,t)L(l,t)^{\mu}$$

where $x\left(L(l,t)\right)$ is output per unit of land, $Z(l,t)$ is total factor productivity and $L(l,t)$ is employment per unit of land with $\mu \in (0,1)$. The profit maximization problem of a firm can be stated as:

$$\max_{L(l,t)} \left(1 - \tau(l,t)\right) \left[p\left(l,t\right) Z(l,t)L(l,t)^{\mu} - w(l,t)L(l,t)\right]$$

where $\tau(l,t)$ is a tax on profits, levied by the local government of the county to which location l belongs, in order to finance investment in process innovation leading to an improved level of total factor productivity equal to $z_l Z(l,t)$.

In particular, the local government can buy a probability $\phi \in [0,1]$ of innovating at a cost $\psi(\phi)$ per unit of land proportional to wages, with $\psi'(\phi) > 0$ and $\psi''(\phi) > 0$. Successful innovation allows the government to draw z_l from a Pareto distribution with c.d.f. $F(z) = 1 - z^{-a}$ with $z \geq 1$. Under the assumption of risk neutrality, the local government of county G with land measure I then

solves:

$$\max_{\{\phi(l,t)\}_{l\in G}} \int_{l\in G} \frac{\phi(l,t)}{a-1} p(l,t)Z(l,t)L(l,t)^{\mu} dl - I\psi(\phi(l,t)) \tag{18}$$

where $\psi(\phi(l,t))$ is government investment in location l at time t, $\phi(l,t)$ is the probability that the government gets to draw from $F(z)$ in location l at time t, $1/(a-1)$ is the expected value of the total factor productivity gain for location l at time t conditional on the government getting to draw from $F(z)$ in that location. In other words, the local government spends on R&D to maximize the expected increase in the output value of its county net of the investment cost. The fact that the maximization problem (18) is static follows from a key assumption on the diffusion of innovation, which makes the best technology available to all neighbouring locations with a one-period delay with respect to the innovator. Matched with the assumption that counties are small, the one-period delay implies that a county's innovation decision today does not affect its expected level of technology tomorrow. Interestingly, (18) exhibits a scale effect as high-price, high-productivity and high-employment density locations will optimally innovate more.

As in the dynamic NEG framework presented earlier, Desmet and Rossi-Hansberg (2010) introduce geography through communication and transport frictions that hamper the geographical mobility of goods and ideas. For ideas, at time t, before the innovation decision, location l has access to the best spatially discounted technology available of the previous period, so *ex ante* $Z(l,t)$ equals:

$$Z^{-}(l,t) = \max_{r\in[0,1]} e^{-\delta|l-r|} Z(r,t-1)$$

where $\delta > 0$ measures the steepness of the spatial decay of diffusion. Based on this technology consumers costlessly relocate, which ensures that utility is the same across all locations, and wages are set. The fact that consumers hold fully-diversified portfolios of land in all locations implies that they need not be forward-looking when deciding where to locate. After consumers move, counties invest in innovation, and production takes place using the new technology $Z^{+}(l,t)$ so that *ex post* $Z(l,t)$ equals $Z^{+}(l,t)$. Due to land portfolio diversification, rents are redistributed from high-productivity to low-productivity locations, which therefore run trade surpluses and deficits respectively. Turning to the product market, transport costs again take the iceberg form: if one unit of the good is shipped from l to r, only $e^{\kappa|l-r|}$ units reach their destination. Hence, with perfect competition we have $p(r,t) = e^{-\kappa|l-r|} p(l,t)$.

In equilibrium labour and product markets clear. In the case of labour, at

each point in time the market clearing condition is:

$$\int_0^1 L(l,t)dl = \overline{L}$$

The market clearing condition in the product market is less straightforward. Following Rossi-Hansberg (2005), it is stated sequentially. In particular, one can start at one end of the one-dimensional space interval and accumulate production minus consumption in a given market (properly discounted by transport costs) until one reaches the other end of the interval. At the boundary, for markets to clear, 'excess supply' has to be equal to zero. Formally, let $H(l,t)$ define the stock of excess supply accumulated from location 0 to location l. By construction, $H(l,t)$ is defined by the initial condition $H(0,t) = 0$ and the differential equation:

$$\frac{\partial H(l,t)}{\partial l} = x(l,t) - c(l,t)L(l,t) - \kappa \left| H(l,t) \right|$$

where $x(l,t) = x(L(l,t) - \psi(l,t)/p(l,t),t)$ so that, at each location, we add to the stock of excess supply the amount of local output and subtract the amount of local consumption. We then need to adjust for the fact that if $|H(l,t)|$ is not zero and we increase l, we have to ship the stock of excess supply over a longer distance. This implies a per-unit cost in terms of the good equal to κ due to the iceberg transport costs. In the end, the good market clears if $H(1,t) = 0$.

At any period t the instantaneous equilibrium of this economy can be computed easily. Before innovation takes place, workers decide where to live. Although the realizations of innovation are random, counties are small, so that there is no aggregate uncertainty. This allows workers to anticipate prices correctly. In addition, workers observe wages and land rents. Once innovation is realized, one can compute actual production, actual distributed land rents and trade. The resulting prices should then be consistent with those used by workers when they decided where to live. Since decisions depend only on current outcomes, computing an equilibrium involves solving a functional fixed point each period. The dynamic growth process is determined by the sequence of those static points.

As usual, the spatial distribution of producers and workers results from the balance between agglomeration and dispersion forces. The diffusion of technology promotes agglomeration, as high levels of local employment raise the incentives to innovate. Due to spatial decay in the diffusion of innovation, pro-

ductivity is higher in locations close to high-employment clusters, which attracts employment and fosters more innovation. This agglomeration force is opposed by local congestion, as employment density reduces labour productivity. This arises because, with constant returns to labour and land, and given that land cannot be accumulated locally, there are local diminishing returns to labour. This form of local congestion tends to spread employment across locations given identical technology levels.

Growth is linked to geography because more uniform, but weaker, incentives to innovate are associated with dispersion, whereas agglomeration is associated with fewer, but more active, innovation centres. As a result, when activity is spatially dispersed, innovation relies more on the extensive margin (how many locations innovate) whereas the intensive margin (how much each location innovates) plays a key role when activity is agglomerated. Easier diffusion makes the extensive margin less important and aggregate growth is generally higher with agglomeration.

Growth is also higher for higher transport costs, as these lead to more concentrated production. In this respect, higher transport costs entail static losses but dynamic gains, through more agglomeration and thus innovation. This is different from the NEG framework discussed earlier, in which higher transport costs promote dispersion and slower growth. The difference is explained by the fact that Minerva and Ottaviano (2009) do not model locally non-reproducible land, so that no congestion arises from its use. This parallels the opposite predictions of the static models of Krugman (1991) and Helpman (1998) discussed earlier in the chapter.

The model of Desmet and Rossi-Hansberg (2010) implies that the concentration of employment in neighbouring locations leads to more innovation and faster growth. This effect is due to local density in a given location, and diffusion from locally dense neighbours. Desmet and Rossi-Hansberg (2012a) present a more general version of their framework in which two industries, manufacturing and services, interact because of trade. This extension reveals another channel through which agglomeration and growth are connected. Due to perfect competition, locations specialized in manufacturing exhibit higher producer ('mill') prices of services. This happens because low transport costs in serving local consumers in the manufacturing cluster allow service providers in those locations to remain competitive in terms of customer ('delivered') prices, despite higher producer prices. Manufacturing clusters will therefore have an incentive to import services from other locations. Their demand for imported services

will, however, fall with distance due to growing transport costs, so that locations closer to manufacturing clusters will tend to have higher employment, higher prices and greater innovation in services. Accordingly, the co-agglomeration of different industries is an additional source of local growth and innovation. This trade channel works on top of diffusion, and is reminiscent of the distinction between transport and communication costs drawn by Martin and Ottaviano (1999, 2001) in their dynamic NEG models.

In a quantitative exercise, Desmet and Rossi-Hansberg (2012a) show that their model can help to explain the evolution of the US economy over the last half-century. In particular, it can generate the reduction in the manufacturing employment share, the increased spatial concentration of services, the growth in service productivity starting in the mid-1990s, the rise in the dispersion of land rents in the same period, and several other spatial and temporal patterns.

In contrast to the model we presented above, where innovation is decided by local governments, Desmet and Rossi-Hansberg (2012a) explicitly model innovation as the outcome of firms making profit-maximizing choices. To produce, firms need to compete for non-replicable land. Since innovation can increase the productivity of that non-replicable land, firms realize they can enhance their bid for land by innovating. As a result, firms may optimally choose to innovate, in spite of the market being perfectly competitive and all profits being bidden away through land rents. The role of land in generating innovation in a perfectly competitive environment is discussed in further detail in Desmet and Rossi-Hansberg (2012b).

Moreover, Desmet and Rossi-Hansberg (2012a) show how the reallocation of employment toward services ultimately accelerates innovation in some locations specializing in services; from then onwards, service productivity increases together with manufacturing productivity, leading to a balanced growth path. Hence, their model is a full-fledged endogenous growth model with spatial heterogeneity, and one that can accommodate both structural transformation and a balanced growth path. The methods that Desmet and Rossi-Hansberg (2010, 2012a) use to deal with growth in an ordered geographical space are fairly straightforward to apply, relative to the underlying complexity of the problem. However, they can only be used in one-dimensional (or two-dimensional and symmetric) compact geographical spaces, and extending this approach to non-symmetric two-dimensional space would be a challenge.

6 Regional prosperity: data and methods

A common thread runs through many of the models we have considered: what happens at each location is a function of the outcomes and characteristics of all other locations. This raises a formidable identification problem for empirical researchers who want to isolate causal mechanisms, and the available empirical methods differ in how persuasively they achieve this. This section will first discuss the available data, and then some leading methods. Some of the most important studies are based on natural experiments, with estimates often obtained by difference-in-differences; since these methods are well known, we do not cover them in detail. For an extended discussion of the natural experiment approach in regional economics, see Holmes (2010). Some examples, and discussion, can be found in Diamond and Robinson (2010).

6.1 Data

Historically, one obstacle to work on regional growth has been the scattered nature of the available data. Researchers on national growth have long been able to draw on the Penn World Table and the World Development Indicators, but there is no close equivalent for sub-national data. Recently this has begun to change, in contributions by Gennaioli et al. (2013a,b), Lessmann (2011) and Mitton (2013). The regional data sets of Gennaioli et al. (2013a) and Mitton (2013) are especially comprehensive; the first covers 1569 regions from 110 countries, which together account for 74% of the world's land area and 97% of its GDP. Mitton's data set is broadly similar in coverage, but partially corrects for internal variation in the cost of living, using data on living costs compiled for a number of cities by the Economic Research Institute. More detailed data on output deflators and regional living costs are typically unavailable, however, as we discuss in the appendix.

These data sets are cross-sections; Lessmann (2011) has compiled a panel data set on regional inequality, but for a smaller number of countries. For a few countries, long-run data sets have been compiled going back to the nineteenth century, such as the work of Turner et al. (1997) and Mitchener and McLean (1999, 2003) on US data; the latter papers use some data on prices. For some countries, the populations of cities have been used to proxy regional development over centuries; see Acemoglu, Cantoni, Johnson and Robinson (2011) and Cantoni (2010) for examples and references.

The increased availability of establishment-level data for some countries can

be used to address some research questions. Another recent development is the ability to analyze data at smaller spatial scales even for developing countries. Harari and La Ferrara (2013) illustrate the potential of this approach: they study civil conflict in Africa at the sub-national level, based on areas that are 1 degree of latitude by 1 degree of longitude, and relating conflict to localized crop failures or climate shocks. Moving to a smaller scale requires a careful approach to spatial dependence and clustering; Barrios et al. (2012) is a recent treatment of this issue. One way to use data at small scales is to aggregate them up to a regional level, the origin of some variables in the Gennaioli et al. (2013a) and Mitton (2013) data sets. For discussion of the use of geographical information systems in regional economics, see Overman (2010).

One approach of particular interest, emerging from an interdisciplinary research effort, has been to use satellite data on light density at night to develop measures of income or population density at the sub-national level. As Chen and Nordhaus (2011) and Henderson et al. (2012) emphasize, this is especially attractive for measuring growth in countries where spatially-disaggregated statistics are unreliable or not available. One application would be to map changes in regional income for countries where hard-to-measure activity, like subsistence agriculture or an urban informal sector, is significant. Relative to the use of official data, the approach also allows population density and income to be estimated for smaller spatial scales. For example, using data on light density for 22,850 sub-national units in developing countries, Hodler and Raschky (2010) study whether foreign aid is disproportionately allocated to the home regions of national leaders. Michalopoulos and Papaioannou (2013) use light density to study whether regional outcomes in Africa are related to pre-colonial institutions, and the local traditions of political centralization in particular.

Measurement issues, especially for developing countries, require thought about how the data relate to the research questions of interest and the concepts used in theoretical models. For example, flows of remittances between regions, which can be significant, will influence regional income. The measured output of some regions can be heavily influenced by natural resource revenues, which will typically be transferred out of the region. To give a concrete example, the treatment of mining output for Indonesia influences findings about regional inequality and convergence (Hill et al. 2008). These points also suggest the importance of considering whether the data at hand correspond most closely to the regional equivalent of GDP ('output'), or GNP ('income'). The former is most relevant for productivity comparisons, the latter for studying regional differ-

ences in living standards. Whichever concept is adopted, measurement error in regional data is likely to be a significant problem, and its consequences remain under-explored by applied researchers.[26]

6.2 Spatial econometrics

If we recognize that regions are interdependent, statistical analysis has to proceed carefully. Outcomes at one location (for example, for productivity) will be closely linked to the outcomes and characteristics of other regions. This implies that the data-generating process will be characterized by spatial dependence; ignoring this dependence is risky, which is clear from a time-series analogy. A good econometrician knows that serial correlation is not solely an issue for inference, but often indicates that the empirical model has been mis-specified. This is why econometricians are wary of mechanical autocorrelation corrections, or exclusive reliance on clustering the standard errors. Related points apply to spatial data, and yet many economists continue to analyze regional data as if spatial dependence is a second-order problem. In fairness, it is true that spatial dependence is inherently harder to address than time-series dependence, because the one-dimensional ordering in time does not apply in the spatial case.

The field of spatial econometrics typically addresses this problem by prespecifying the *relative* strengths of interactions between regions, using the device of a spatial weight matrix. The entries in the matrix are often based on distances between locations or the existence of shared borders, although there is nothing in the approach which requires the interactions to be determined by physical geography. The literature is large and growing fast, and we highlight only the areas most relevant to the discussion later in the chapter. This brief introduction draws partly on Anselin (2001) and especially the longer survey by Anselin (2006).

For the case of N regions, a cross-section model with a spatial lag is conventionally expressed in matrix notation as:

$$y = \rho W y + X\beta + \varepsilon \tag{19}$$

where W is an $N \times N$ spatial weight matrix (typically normalized in some way) and ρ indexes the strength of the spatial spillovers. Given $\rho \neq 0$, the spatial lag

[26] Additional measurement issues are discussed in some of the contributions in Kanbur and Venables (2005). Measurement errors are likely even in the official data of developed countries. Cameron and Muellbauer (2000) examine this issue for the UK, by comparing the UK's Regional Accounts with alternative sources of information on earnings.

will be correlated with the disturbances ε, because the above model implies

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon \tag{20}$$

Expanding each inverse implies that y at each location is a function of X and ε at all locations, so that the effects of the explanatory variables and the errors are transmitted across space rather than confined to each region. A corollary is that Wy in (19) is necessarily endogenous and hence OLS estimates of that model will be inconsistent. The literature has developed alternative procedures for estimating such models, using either maximum likelihood or instrumental variables.

Alternatively, we could allow for spatial dependence in the errors rather than in the dependent variable, using the spatial error model:

$$y = X\beta + u \tag{21}$$

$$u = \lambda W u + \varepsilon \tag{22}$$

This is more closely related to a spatial lag model than it may seem. If we note that $u = (I - \lambda W)^{-1}\varepsilon$ we can use this in $y = X\beta + u$ and then have

$$y = X\beta + (I - \lambda W)^{-1}\varepsilon$$

which implies

$$y = \lambda W y + X\beta - \lambda W X\beta + \varepsilon$$

This model could be estimated with or without the implied parameter restrictions. This is usually referred to as the Spatial Durbin Model, by analogy to the derivation of common factor restrictions in time series models by Durbin (1960). One interpretation is that a spatial lag helps to address the issue of omitted variables that are spatially correlated, but this is only true if the spatial dependence corresponds to the relative interactions embedded in the weight matrix W.

There are two main interpretations of what spatial econometrics achieves. The first is that the spatial dependence is not itself of direct interest, but must be addressed to obtain reliable estimates of the parameters. In practice, some parameters become much harder to interpret when a model incorporates spatial spillovers. A common example would be attempts to link parameter estimates to the rate of convergence in a neoclassical growth model. When regional income levels are influenced by the income levels of neighbouring regions, the

theoretical counterpart of an estimated convergence rate is unclear, because the neoclassical growth model sits uneasily with the reality of interdependent regions.

An alternative interpretation is that spillovers are of direct interest. In that case, the estimate of ρ is seen as directly informative and not just a nuisance parameter. The problem here is that spatial econometric models are silent on mechanisms, and without a mechanism, adding a spatial lag of the dependent variable will often seem too ad hoc to be informative; Gibbons and Overman (2012) argue along these lines. They suggest that, for many applications, it would be more sensible to emphasize spatial lags of the explanatory variables. That approach is simpler to implement, and often easier to connect to theoretical models.

Another frequent criticism of the spatial approach is that the researcher's choice of weight matrix W is necessarily arbitrary, because there are many different possibilities. This criticism might sometimes go too far. There is a sense in which imposing $\rho = 0$ is an arbitrary choice too. Even a model with a misspecified weight matrix may have better properties than a model which does not acknowledge spatial dependence at all. Approaches based instead on structural models, such as the use of measures of market potential, also impose restrictions on the data that are best seen as maintained assumptions, and that are open to question. Given the inevitable uncertainty over the appropriate weight matrix, one way to make the analysis less arbitrary is to use Bayesian Model Averaging, as in Crespo Cuaresma and Feldkircher (2012) and LeSage and Fischer (2008). This allows a range of specifications to be considered, while formally acknowledging the researcher's uncertainty about the model and the nature of the spatial interactions.

As things stand, there are clear divisions in the literature about the usefulness of these methods. Corrado and Fingleton (2012), Gibbons and Overman (2012) and LeSage and Fischer (2008) provide extensive discussion, from a variety of perspectives. That opinion is divided can be seen from the different paths taken in the applied literature. The spatial econometric papers take care over dependence, but often adopt rather mechanical hypotheses about regional growth and the nature of spillovers. In contrast, many papers by growth economists and development economists put forward interesting hypotheses, but largely ignore the issue of spatial dependence, or adopt corrections such as spatially-clustered standard errors that do not address underlying problems with the regression specification. One improvement would be to adopt a spatial equivalent to HAC

estimators of standard errors, such as that developed by Kelejian and Prucha (2007); but this continues to emphasize the problems for inference rather than the structure of the estimated model. [27]

In recent panel data studies, a common approach to error dependence has been to interact time dummies with one or more regional characteristics. Versions of this are adopted in Acemoglu, Cantoni, Johnson and Robinson (2011), Burgess and Pande (2005), Burgess et al. (2005), and Cantoni (2010) among others. We call this an assumption of *proportional time effects*. It can be seen as a special case of the common factor structures studied in the macroeconometric literature, where the error term has a component $\phi_i f_t$. Here f_t is a vector of common factors, the effects of which are allowed to vary across the regions i by means of the (row) vector of factor loadings, ϕ_i. This could be a natural route to take for regional data. For example, an urban core of manufacturing and services might be strongly correlated with the national business cycle, while an agricultural periphery would be less correlated with the business cycle and more strongly correlated with climate variation and world food prices. In principle, a factor structure could account for much of the spatial dependence in the data. So far, there has not been much work analyzing regional data using these methods, but the techniques are developing rapidly and could be important for regional panel data models in particular. For surveys, see Eberhardt and Teal (2011) and Sarifidis and Wansbeek (2012).

6.3 Regional growth regressions

A substantial fraction of the work on regional prosperity, especially that for developing countries, is based on cross-section or panel data growth regressions. Assessed as a whole, the literature inherits many of the issues of interpretation that have undermined the cross-country study of economic growth. As Durlauf et al. (2005, p.558) argue, the problem is not only that some regression-based studies are unreliable. A further problem arises on the consumption side: it can be hard, when presented with a particular study, to tell whether it has been executed well or badly. This means that even the best studies may be assigned relatively little weight.

Relative to the cross-country literature, the use of regional data may be much

[27] In Conley (1999), if an over-identified GMM approach is taken, spatial dependence is an issue for estimation as well as inference. More generally, a model which does not allow for spatial dependence is likely to be incomplete, again suggesting that spatial dependence matters for point estimates as well as standard errors.

less vulnerable to omitted variables. A common argument is that factors such as institutions and cultural norms vary greatly across countries, but less so within them. But the fact that regions are within the same polity can be a double-edged sword, because they may influence each other, and be subject to common shocks, to a much greater extent than countries. In some ways, the legacies of the cross-country literature have been unfortunate. Empirical studies often treat the units as essentially independent, or take the neoclassical growth model as the starting point, either explicitly or implicitly; this approach has problems at the country level, and seems all but untenable for regional data.

As already noted, many of the regression-based studies by growth economists and development economists fail to address spatial dependence. It is common for researchers to analyze variables or interventions which are highly correlated spatially, but the estimates may then be confounded by omitted spillovers, or spatially correlated variables such as aspects of physical geography or market access. At least some of these are known to be important features of the data, and there could be gains from combining the hypotheses of these studies with methods from spatial econometrics and macroeconometrics.

Another fundamental issue receives even less emphasis in the literature: the basic causal structure implicit in a regression sits uneasily with a spatial equilibrium. The most obvious and well-known problem is that many regional characteristics, such as average education or financial depth, are not fixed endowments, but endogenously determined outcomes. But this also hints at a deeper identification problem, and one that has been less widely noted. At first glance, regression-based methods give simple answers about the determinants of regional prosperity. But their interpretation is complicated by endogenous agglomeration. When a given variable changes, this could reconfigure spatial patterns of activity in ways that (for example) amplify the effects of minor differences, just as agglomeration can amplify minor differences in physical geography. This makes it hard to interpret estimated associations between regional growth and explanatory variables. To make this point concrete, consider the estimated growth effect β of a one-unit change in a given variable X for one region. If X increased by the same amount for all regions, would the growth effect be β for each region? This is rarely clear, but then it is hard to interpret the results from regression-based studies. Put differently, it is not clear what is being assumed about the simultaneous role of changes in the spatial distribution of activity. In the context of a spatial equilibrium, this attempted distinction is artificial and impossible to maintain, but that is precisely the point. It com-

plicates the interpretation of much empirical work, not least its consequences for policy. It also indicates the benefits of structural models, where meaningful counterfactuals are much easier to construct.

Another distinctive feature of regional data is that the cross-section dimension and the time dimension often have broadly similar magnitudes. This suggests that panel time series methods, such as those introduced by Pesaran and Smith (1995) and Pesaran et al. (1999), could be natural candidates for the analysis of regional data. For reasons that are not fully clear, few studies have applied these methods to regional questions; exceptions include Cameron and Muellbauer (2001) for Britain, and Trivedi (2006) for India. The first of these briefly explores time-series specifications in which the dependent and independent variables are formed as deviations from the average values of contiguous regions.

6.4 Structural models

We have repeatedly emphasized the dangers of analyzing spatial data without thinking in terms of a spatial equilibrium. That might suggest abandoning regressions in favour of calibrating or estimating structural models, often drawing heavily on the work we described in section 4 of the chapter. Examples of this approach include Donaldson (2010), Redding (2012), and Van Nieuwerburgh and Weill (2010). The quantitative use of structural models has many advantages: general equilibrium effects are accounted for, parameters should have a clear interpretation, and progress can be made even when some data (such as regional price levels or productivity levels) are lacking, by inferring these from other outcomes. Further, the use of a structural model allows counterfactual simulations and the quantification of welfare effects, both of which are attractive when policies are to be assessed. Holmes (2010) emphasizes that the approach can be used to evaluate policies that have never previously been implemented. He discusses the approach further, as does Combes (2011).

We review several of these studies below. If structural models have a weakness, it is the uncertainty over whether it is the model speaking or the data; the list of maintained assumptions is often extensive, and the data may know more than the model can say. There is a complementary role for reduced-form approaches, partly in drawing attention to interesting associations, and partly as a check on the maintained assumptions of any given structural model, as in Donaldson (2010). But it seems clear that structural models will have an important, even pre-eminent, place within the best future work on regional data.

6.5 Spatial discontinuity designs

We now discuss a method which has become a powerful way of identifying causal effects in the recent literature, and which drives some of the most important papers. This is to look for institutions or policies that sometimes stop (or change) at the borders of regions, and quantify their causal effects by comparing outcomes either side of the border. We illustrate the application of this method, and some of its pitfalls, using the classic paper by Holmes (1998).

Holmes was interested in whether state-level policies influence the location of manufacturing activity. It had long been known that manufacturing activity had grown slowly in the industrial north of the US and more rapidly in other regions, including in the 'right-to-work' states which weakened unions by state legislation that outlawed closed shops. But simple correlations between regional outcomes and a right-to-work indicator are not all that informative about causal effects, given that regions may differ in other ways, such as geography and climate. Holmes's answer to this problem was to identify sets of counties adjacent to borders, where right-to-work laws applied one side of the border and not the other. Using the presence of these laws as a proxy for more generally 'pro-business' policies, he found large effects: manufacturing's share of employment increased by about a third on crossing from an 'anti-business' state to a 'pro-business' state. As Holmes (1998, p. 671) explains, the power of this approach is that:

> ...at state borders, the geographic determinants of the distribution of manufacturing — for example, climate, soil fertility, access to transportation, and the level of agglomeration benefits — are approximately the same on both sides of the border. What differs at the border is policy.

In what follows, we call this approach a *spatial discontinuity design*. It has since been applied in other contexts, including to political institutions and financial reform. For now, we note that Holmes's paper not only demonstrates the power of this method, but also provides a careful account of its limitations. He notes that the effects of policy differences far from the border may be smaller than the effects close to the border. After all, a firm may be more influenced by policy differences between locations that are close to one another (and hence similar in terms of market potential) than between locations that are further distant. With this in mind, Holmes interprets his estimates as upper bounds on the effects of a statewide policy change. The same issue substantially complicates a welfare

analysis. A policy difference that shifts a firm from one location to another could have minor effects on welfare (for example, if a firm chooses to locate one side of a border rather than another) or major effects (for example, if policy differences compounded the decline of America's northern cities).

A second issue is that borders are not randomly generated. This means that an identifying assumption — geographic characteristics are the same either side of the border — will not always hold. The example in Holmes (1998) is that some state boundaries coincide with discontinuities in nature represented by mountain ranges and coal veins. As a result, he drops some observations, but acknowledges that there may be other unknown instances in the data. Although this is a limitation, spatial discontinuity designs are likely to hold various other characteristics constant, to an extent that is otherwise hard to achieve. They represent one of the most informative methods for learning about regional prosperity.

6.6 Synthetic controls

For some regional questions, an approach based on spatial discontinuity may be either infeasible or uninformative. This is especially likely when a researcher is interested in events or characteristics confined to a single region, or a small number of regions. As an example, consider a researcher interested in the effect of localized conflict on a single region's prosperity. There is not an obvious way to construct a counterfactual. Comparing outcomes with those of a neighbouring region may not work, because there is no guarantee that the two regions will share similar characteristics. The alternative is a less formal case study, but that has problems of its own (see Temple 1999).

Abadie and Gardeazabal (2003) considered this problem and introduced a method for constructing a 'synthetic control', which can be compared to the region of interest. In an application to Spanish regions, their specific aim was to quantify the economic effects of Basque terrorism on the Basque Country. To do this, they compared the evolution of Basque Country outcomes with a weighted average of other Spanish regions: the synthetic control. The weights were chosen so that the characteristics of the synthetic control resembled those of the Basque Country in the years before terrorism. The synthetic control can be seen as an approximation to the required counterfactual, the Basque Country without terrorism.

More formally, consider a case where there are J control regions available (in their case, the Spanish regions other than the Basque Country). The treated region has a set of K characteristics stored in a $(K \times 1)$ vector X_1. The J control

regions have corresponding pre-treatment characteristics stored in a $K \times J$ matrix X_0. Drawing on ideas in the statistical literature on matching, the suggestion of Abadie and Gardeazabal is to choose a $(J \times 1)$ column vector of weights $W = (w_1, ..., w_J)'$ in order to minimize

$$(X_1 - X_0 W)' V (X_1 - X_0 W)$$

subject to $w_i \geq 0$ and $\sum w_i = 1$, and where V is a diagonal matrix with non-negative components, which weight the different characteristics.

Hence, as well as choosing the set of K relevant characteristics, the researcher has to decide how to weight them. The diagonal elements of V could be based on subjective judgments about their relative importance. In their own application, Abadie and Gardeazabal use a more objective approach, and choose the elements of V so that GDP per capita in the synthetic control is close to that of the Basque country for the pre-treatment years. Once a researcher has chosen or estimated V, and obtained the weights W, outcomes can be compared between the region of interest and the synthetic control. For example, the GDP per capita of the control will just be a weighted average of the GDP per capita of the J regions, where the weights are the (possibly zero) individual elements of W.

The synthetic control method lends itself to graphical comparisons of outcomes, and robustness tests using the placebo approach familiar from the treatment effects literature. Applications to regional data are currently limited, but the method is especially likely to be useful when the number of regions is small, or the treatment of interest is confined to a small number of regions. It also provides a bridge between the regression-based methods favoured by economists, and the more qualitative, case-study approaches favoured in some other disciplines. An introductory overview of the method by Abadie et al. (2012) makes this point in relation to political science.

7 What determines regional prosperity?

We now turn to the empirical evidence on regional growth. Following the precedent of the cross-country literature, our use of the term 'growth' is deliberately elastic. We use the term to encompass the study of influences on levels (or relative levels) as well as influences on steady-state growth rates. In fact, most of what we have to say has more bearing on the former, and so 'prosperity' might be a better term.

There is a second ambiguity, to a far greater extent than in the cross-country

literature. Regional growth sometimes refers to an increase in population rather than productivity, as factors of production gravitate towards particular areas. In fact, some authors have argued that measures such as population density will better capture underlying differences in productivity and quality of life (for example, Rappaport and Sachs, 2003). Many spatial models predict that equalization of real incomes will be achieved through adjustment in nominal wages, price levels and local population sizes. Hence, the criterion for regional success, or the best interpretation of 'growth', varies across studies and research questions.

7.1 Physical geography

We start with physical geography, which can influence economic activity and population density through many channels; it is a more disparate topic than it might appear at first. Among the channels highlighted in the literature are access to the coast or transport networks, and physical transport costs more broadly; climate factors such as temperature or precipitation; and disease ecology.[28] Less obviously, as we also discuss below, physical geography can be a long-run influence on cultural and social norms, and local institutional development. And it may be especially important for developing countries, some of which are much larger and more heterogeneous internally than, say, the countries of Western Europe.

Geography can be thought of as influencing prices (partly through higher transport costs for remote regions), total factor productivity (in both agriculture and industry, see Dell, Jones and Olken, 2012) and also incentives for factor accumulation. One of the most well-known findings is that economic activity is disproportionately coastal: Gallup et al. (1999) report that the areas of the US, Western Europe, and northeast Asia that are within 100 kilometres of the coast contain just three percent of the world's inhabited land area, but thirteen percent of its population, and at least 32 percent of global GDP.

Coastal locations, by lowering the costs of external trade, can be seen as favouring high productivity. Rappaport and Sachs (2003) argue that the coastal concentration in the US derives primarily from a productivity effect. The direct benefit of coastal location will be amplified by effects on economic geography, as firms and populations form agglomerations in coastal areas. This can also introduce path dependence; Bleakley and Lin (2012) study this issue using the

[28] Relevant papers include Bloom and Sachs (1998), Dell, Jones and Olken (2012), and Sachs and Malaney (2002), respectively.

proximity of many American cities to historical obstacles to water navigation, where continued transport relied on overland hauling. They find that these obstacles continue to be associated with relatively high population densities, even though their direct relevance to transport costs has long since disappeared.

Some of the evidence that physical geography matters is based on studying the location of individual industries (Ellison and Glaeser, 1999; Davis and Weinstein, 2008). Much of this evidence is for developed countries, but Felkner and Townsend (2011) use detailed data for Thailand to show that enterprise locations are associated with various geographic characteristics.

When the data are analyzed at higher levels of aggregation, the interpretation is more difficult. The cross-section study by Gennaioli et al. (2013a) finds that average temperature has limited explanatory power for output per capita within countries. Mitton (2013) considers a wider range of geographic and climate variables; he finds that many are statistically significant, but collectively their explanatory power remains relatively modest. Dell, Jones and Olken (2009) and Nordhaus (2006) find some effects of temperature using variation at smaller spatial scales; in Nordaus (2006), there are opposing effects on output per capita and output per area. The latter is relevant because, given the nature of a spatial equilibrium, some climate effects are likely to be more readily observable in relative population density, rather than relative productivity. As we have emphasized throughout, in models with heterogeneous sectors and/or mobile workers, trying to infer fundamental influences on productivity from comparisons of average output per capita is not straightforward.[29]

Instead, it should be recognized that physical geography will operate partly through the spatial distribution of the population. Some aspects of physical geography may have a limited direct effect on production costs - for example, fewer cloudy days per year - but can still influence wages and incomes, through their impact on the location decisions of utility-maximizing mobile workers (Roback 1982). Using US data, Rappaport and Sachs (2003) point out that proximity to Great Lake or ocean coasts helps to explain population density in levels and changes; as well as a productivity explanation, there may also be a quality-of-life effect. Over the twentieth century, the US saw a large-scale movement of population towards areas with good weather. Many of the northern industrial cities have lost population over time, while cities in the Sun Belt have grown. Rappaport (2007) argues that, as US incomes have risen, an income effect on the

[29] In one of the first studies of these questions, Warner (2002) calls a version of this problem the 'mobility bias'.

demand for good weather has been an important driver of this adjustment.[30]

These ideas are supported by studies which consider factor incomes. There is a large literature which links regional variation in wages and rents to physical geography through amenities such as better weather. A general finding is that some of the regional variation in wages can be explained by differences in climate-related amenities (Roback, 1982; Beeson, 1991). And researchers primarily interested in the effects of economic geography sometimes find effects of measures of physical geography in regional wage regressions (for example, Amiti and Cameron, 2007).

Even this brief summary hints at the difficulties of studying physical geography in the context of a spatial equilibrium. It can influence productivity directly and via economic geography, and partly through the location decisions of workers based on amenities, while path dependence complicates this even further. Various researchers have sought to cut through these complexities by studying major shocks or perturbations. Much of this work points to the sustained importance of fixed regional characteristics, even in the face of other changes. For example, Hornbeck (2012b) studies agricultural land values in the Great Plains of the US from 1945 to 2002, and shows that long-run technological progress has not diminished the importance of local environmental advantages.

It is clear that, historically, some climate shocks have led to substantial population movements. Hornbeck (2012a) looks at the economic effects of the Dust Bowl, the severe drought and subsequent wind erosion of topsoil in sections of the American Plains in the 1930s. The erosion of topsoil greatly reduced agricultural productivity in the affected areas, leading to falls in the price of land, out-migration and diverted in-migration. As Hornbeck notes, adjustment was achieved mainly by population movements.

Other major shocks have also been studied, as in the work of Davis and Weinstein (2002). They show that the relative population densities of Japanese regions have been remarkably stable over the past 8,000 years, and that even large-scale shocks, such as the Allied bombing of Japanese cities during World War II, had only temporary effects on the Japanese city size distribution. The findings indicate the long-run importance of fixed characteristics of locations, including their physical geography.

Finally, a more complex set of arguments traces the influences of physical

[30]The argument is that, as consumption goes up, the marginal utility of consumption falls and hence individuals are more willing to forego income for the sake of better weather; they migrate to regions with better weather, forcing wages in those regions downwards and house prices upwards until a spatial equilibrium is restored. See also Desmet and Rossi-Hansberg (2013a).

geography on local institutions, cultural and social norms, the distribution of ethnic groups, and even political trajectories. Physical geography can sometimes manifest itself in profound differences of institutions and culture, with the semi-autonomous Federally Administered Tribal Areas (FATA) of northwestern Pakistan as a well-known example. More generally, as observers such as Scott (2009) have discussed, state-building sometimes founders in the mountains. Herbst (2000) argues that the large interiors of some African countries, with their low population densities and disconnected peripheries, have made it difficult for governments to maintain control over their territories, and have limited the development of effective states. Even more complex effects of geography are possible. China's ethnic geography and cultural differences, it is sometimes argued, partly reflect historical differences between areas suitable for arable farming (and hence permanent settlement) and the more nomadic cultures of the pastoral areas.[31] Moreover, geography can shape the response to historical events: looking at the specific issue of rugged terrain, Nunn and Puga (2012) argue that its direct costs have been offset, in Africa's case, by the protection it offered from the slave trade, with effects that have persisted to the present day.

Taken together, these points indicate a major challenge for empirical researchers: much remains to be done in understanding when and how physical geography influences regional prosperity. And aspects of this task seem increasingly urgent, given the scope for climate change to reshape productivity levels and specialization across the world, both across and within countries. Dell, Jones and Olken (2012) find that increases in temperature adversely affect output in poor countries, and may also have consequences for political stability. It should also be emphasized that, even if regional prosperity seems only modestly affected by temperature differentials between regions, the effects of climate change on national comparative advantage could be substantial. This in itself would be enough to drive new patterns of regional growth and decline for many of the world's countries, leaving aside other effects such as desertification.

Recent work by Krusell and Smith (2009), Hassler and Krusell (2012), and Desmet and Rossi-Hansberg (2013b) seeks to quantify the differential effects of climate change across distinct locations, an approach pioneered by Nordhaus in his development of the multi-region RICE model. Desmet and Rossi-Hansberg (2013b), in particular, emphasize the importance of a spatial dimension to the analysis: as the climate changes, welfare losses arise because of frictions in the

[31] See Kaplan (2013) for an overview of this argument.

movement of people and goods across locations. One consequence is that, in the presence of migration restrictions between the global South and the global North, the estimated welfare losses are much larger for the global South. Their work is based on the effects of temperature changes; for at least some countries, the uncertainties for future regional development are compounded by the possibilities of water stress, coastal flooding, changes in the incidence of extreme weather, and new risks to health.

7.2 Market access

Geographers have long pointed out that access to markets influences regional output levels. For example, Harris (1954) argued that the demand facing a given region depends on the distance-weighted GDP of all other regions. More recently, the empirical literature by economists has adopted measures of market access derived from structural models in the New Economic Geography tradition. We now provide a brief review of this literature; for more detailed surveys, see Combes (2011), Combes et al. (2008), Head and Mayer (2004), and Redding (2010). Assessed as a whole, the literature strongly supports the idea that market access or proximity influences regional prosperity.

In sections 4 and 5 of the chapter, we reviewed models in which firms in more central locations will face higher demand for their products and thus, initially, higher profits. The usual assumption of free entry will equalize rates of return across locations but, as Head and Mayer (2006) note, the adjustment could take place via local employment or production, or through changes in wages - or, more generally, the remuneration of immobile production factors. The bulk of the literature has focused on adjustment through wages, and much of our discussion will look at this mechanism. But there is also a literature which considers adjustment via employment and production changes, often drawing on models with a freely-tradable numéraire sector which makes wages invariant to demand. Key papers in this 'home market effect' literature include Davis and Weinstein (1999, 2003), Head and Ries (2001) and Hanson and Xiang (2004). Head and Mayer (2006) show how these papers relate to the literature on adjustment via wages. Empirically, it is often difficult to separate these two adjustment mechanisms cleanly.

Turning to adjustment via wage changes, a first strand in this literature builds on Redding and Venables (2004). Their influential paper takes the spatial distribution of production and expenditure as given, and considers the wages that firms in each location can afford to pay. Firms in more remote locations incur

higher trade costs when selling their products. This lowers the value added attributable to the factors of production; labour, as the relatively immobile factor, is affected most. Hence, the income and productivity levels of a region are influenced by its position relative to potential markets — in other words, by economic geography.

The empirical specification of Redding and Venables is ultimately based on the model of Krugman and Venables (1995). Symmetric, monopolistically-competitive firms from a given location i sell their tradable output in N different locations subject to trade costs. Demand is of the CES form and production takes place under increasing returns to scale. If labour is the only factor of production, free entry implies the following relation between (nominal) wages in location i and the demand and prices in all regions, which Redding and Venables call the 'wage equation':

$$w_i^{\sigma} = A \sum_{j=1}^{N} T_{ij}^{1-\sigma} E_j P_j^{\sigma-1} \tag{23}$$

where w_i denotes wages, E_j and P_j are respectively expenditure on traded goods and the CES price index in location j, T_{ij} are trade costs between locations i and j, σ the elasticity of substitution between product varieties produced by different firms, and A is a constant.[32] Trade costs take the familiar iceberg form: for every unit shipped only $1/T_{ij}$ units arrive, where $T_{ij} = 1$ would correspond to free trade.

Equation (23) says that wages in region i depend on the sum of expenditure in all other regions, adjusted for price differences and discounted by bilateral trade costs. Redding and Venables call the summation term in (23) 'market access'. Other authors, including Head and Mayer (2006), prefer the term 'real market potential', to highlight the price component P_j absent from more traditional measures such as the Harris (1954) market potential. Redding and Venables estimate (23) for a cross-section of 101 developed and developing countries for the year 1994. They find that GDP per capita (used as a proxy for wages) is correlated with their measures of market access, even after controlling for characteristics such as institutions and resource endowments.

The findings suggest that relative prosperity has a spatial dimension, but the assumption that labour is immobile across locations is less attractive for regional

[32] Redding and Venables also allow for technological differences between firms in different locations, intermediate inputs and other internationally-mobile primary factors. This implies that wages will also depend on technology levels and the price of intermediate inputs in each location.

data than cross-country data. In contrast, Hanson (2005) allows for labour mobility. In order to obtain empirically-relevant spatial production patterns, with activity at each location, he follows Helpman (1998) and introduces a nontraded good (housing) to create an additional dispersion force. While real wages are equalized across regions in this model, nominal wages are still a function of market access, as well as housing stocks.[33] In more detail, we have:

$$w_i^{\sigma} = B \sum_{j=1}^{N} T_{ij}^{1-\sigma} E_j^{\frac{\sigma(\mu-1)+1}{\mu}} H_j^{\frac{(\sigma-1)(1-\mu)}{\mu}} w_j^{\frac{\sigma-1}{\mu}} \tag{24}$$

where B is a constant, H_j is the housing stock of region j (assumed to be in fixed supply) and μ is the expenditure share of the traded goods sector. Hanson refers to the summation term in (24) as the 'augmented market potential' of region i, again to distinguish it from simpler measures that do not correct for price variation. He estimates (24) on a sample of 3075 counties in the continental United States for the period 1970-1990, and finds a strong positive correlation between changes in augmented market potential and changes in nominal wages.

These two frameworks have been used to study the geographical variation in wages and output levels for a wide range of countries, regions and time periods. Breinlich (2006) and Head and Mayer (2006) use the Redding-Venables approach to explain the variation in output per capita and wages across European Union regions, arguing that labour mobility is relatively low. Both papers find that the measure of real market potential in (23) performs no better than the simpler Harris market potential, in terms of explanatory power as measured by the R^2. Using modifications of the Hanson approach, Brakman et al. (2004a) and Mion (2004) provide evidence for the importance of proximity to sources of demand for German and Italian regions respectively.

Recent research has extended the ideas to low-income and middle-income countries, often based on the Redding-Venables approach. Bosker and Garretsen (2012) find a positive correlation between market access and GDP per worker for sub-Saharan African countries. Fally et al. (2010) find a correlation between wages and market access for Brazilian states, Amiti and Cameron (2007) for Indonesian districts, and Hering and Poncet (2009, 2010) for Chinese provinces and cities.

A common finding is that, although the market access variables are significant, the magnitude of the estimated effect is substantially lower than in Red-

[33] See Hanson (2005, section 2) for a full derivation. Note that local expenditure is still taken to be exogenous despite full labour mobility.

ding and Venables (2004). One explanation is that (with the exception of Bosker and Garretsen) these newer studies work at a disaggregated level, using either firm- or worker-level data. This enables them to control for additional covariates which are likely to be correlated with market access, including human capital. Moreover, these papers study wage differences in a regional context, where labour mobility will promote the equalization of wages. The finding of Hering and Poncet (2010) that variation in market access has a stronger impact on wages of highly-skilled workers, whose mobility is more restricted in China, lends support to this explanation.

There are some ways in which the developed-country literature needs modification when applied to developing countries or regions. For sub-Saharan Africa, Bosker and Garretsen (2008) find that the correlation between GDP per worker and market access is relatively weak. Their preferred explanation is that manufacturing, the sector to which the wage equation applies most directly, is still underdeveloped in the African case. In their study of China, Hering and Poncet (2010) show that wages in private firms, and particularly in foreign firms, react strongly to variation in a city's market access, but wages in state-owned enterprises much less so. These findings suggest that structural and institutional conditions can influence estimated relationships between wages and market access.

Since the cross-section study of Redding and Venables (2004), a number of papers have shown that the correlation holds using variation in market access over time. As mentioned previously, Hanson (2005) correlates changes in nominal wages with changes in market potential. Head and Mayer (2010) apply the Redding and Venables approach to all countries in the world with available trade data over the period 1965-2003. Breinlich (2006) and Bosker and Garretsen (2012) estimate specifications which sometimes include region and country fixed effects, respectively. The general finding is that output or income remain correlated with market access, but the correlation is substantially reduced when using the within variation.

Can the correlation between income and market access be interpreted as causal? One issue is that market access might be correlated with other fundamental determinants of local income levels, such as institutions or endowments. This can work both ways, since some determinants may themselves be influenced by market access; Redding and Schott (2003) construct a model in which incentives to acquire human capital are lower in countries with weak market access. But the fact that market access effects are weaker in the within dimension

does suggest that market access may be correlated with time-invariant determinants of income levels omitted from cross-section regressions.

A further problem is that, in essence, wage equation estimates are based on regressions of 'own' income (w_i) on measures of income/expenditure levels in neighbouring cities, regions or countries (E_j). But as discussed in section 6.2, this leads to a correlation between regression disturbances and the market access variable, and inconsistent estimates. This is most evident from equations (19) and (20) once we realize that market access can be seen as a spatial lag of regional expenditure levels adjusted for price differences ($E_j P_j^{\sigma-1}$), where the $T_{ij}^{1-\sigma}$ are the elements of the spatial weight matrix.[34]

One approach has been to search for instrumental variables, but the exclusion restrictions are often questionable, and the scope for finding a time-varying instrument is limited. A more promising approach is to study quasi-natural experiments in which there is exogenous variation in market access. The pioneering work in this area is Hanson (1996, 1997), who uses the changes in market access generated by Mexico's trade liberalization in 1985. Focusing on the apparel sector, Hanson (1996) shows that the pre-liberalization period was characterized by a strong regional wage gradient, with wages declining with distance from Mexico City. The 1985 trade liberalization led to a partial breakdown of this gradient, which Hanson attributes to a relocation of apparel assembly production to regions bordering the United States. The evidence for other manufacturing sectors is weaker, although the earlier introduction of special enterprise zones near the border (the *maquiladoras* programme) led to a compression of regional wage differences (Hanson, 1997).

Another event, which seems even more likely to isolate an exogenous and sizeable change in market access, is the division of Germany in 1949 and its reunification in 1990. This is studied by Redding and Sturm (2008). They base their analysis on the model by Helpman (1998) but look at its predictions for equilibrium population sizes rather than nominal wages. They show that, consistent with the model's predictions, West German cities close to the border with East Germany experienced a substantial decline (and after reunification, recovery) in population growth relative to other West German cities.

In an extension of the Redding-Sturm approach, Brülhart, Carrère and Trionfetti (2012) use the end of the Cold War, and the fall of the communist regimes

[34] More precisely, we have wages or output per capita on the left-hand side of the market access equation (23), and regional expenditure levels on the right-hand side. In practice, however, regional wages and expenditure levels are highly correlated, and estimating a market access equation is conceptually similar to estimating equation (19)

in Eastern Europe, to isolate a change in market access. They study the differential impact on Austrian municipalities bordering former communist economies, relative to interior municipalities. In contrast to Redding and Sturm, they have both wage and employment data at their disposal, and can analyze adjustment through both channels. They find that wages and employment growth were both influenced positively by better market access, but the estimated impact on employment growth was about three times as large as the impact on wage growth. This again suggests that, in settings where labour mobility is important, studies focused on wages could miss important forms of adjustment.

Overall, the literature indicates that prosperity is strongly associated with market access, at a range of levels of aggregation. This association is consistent with formal models from the New Economic Geography literature, in which the link is causal. One qualification is that, in some circumstances, the effects of market access on wage and employment patterns will be observationally equivalent to the effects of technological spillovers and labour pooling (Duranton and Puga, 2004; Redding, 2010). Some papers seek to address these alternative explanations using control variables, but their treatment is often less sophisticated than the treatment of market access, and draws less heavily on structural models. Another remaining challenge, and one of particular relevance for this chapter, is to integrate the analysis of adjustments in wages and those in employment. In seeking to understand regional prosperity, it would be useful to know how labour mobility and various institutional constraints or frictions shape the relative importance of these two forms of adjustment.

7.3 Openness

Since the New Economic Geography borrows heavily from trade theory, a natural research topic has been the relationship between external trade, internal economic geography, and regional disparities. Fujita et al. (1999) analyze this issue in detail, suggesting that trade could work to disperse manufacturing industry as a whole, but also lead to the spatial clustering of specific industries. Given the empirical importance of market access effects, it seems inevitable that spatial patterns of activity will be influenced by the nature of external trade, as in the work of Hanson (1996) reviewed above.

Redding (2012) uses a structural model (reviewed in section 4 above) to examine the effects of a fall in trade costs between the US and Canada, leaving internal trade costs unchanged. Given its greater trade intensity with US states, Central Canada would gain more than Western Canada under population im-

mobility. But if the population is mobile across regions, the improved market access of Central Canada causes it to gain population, while Western Canada would see a decline in population. The endogenous reallocation of population continues until all Canadian regions gain equally from the fall in trade costs, in the absence of costs to mobility.

Empirical work on external trade and regional disparities has often taken a reduced-form approach. In a study of Latin America, Serra et al. (2006) argue that regional disparities modestly increased, at least temporarily, in the wake of trade liberalization; the effect seems especially marked for Mexico. A further issue, especially for developing countries, is the influence of FDI on regional prosperity. Brakman et al. (2009) review the literature on the relationship between international business, FDI and agglomeration. A small empirical literature studies the links between FDI and regional inequality directly, particularly for the Chinese case, where FDI has been heavily concentrated in the eastern provinces (Wei et al. 2009). Lessmann (2013) studies China, and a wider sample of 55 countries over thirty years, 1980-2009; his main result is that FDI inflows may increase regional inequality in developing countries, but there is no evidence of a similar effect in richer countries.

7.4 Transport and infrastructure

In his book *The Age of Capital*, Hobsbawm (1962) briefly recounts the story of teachers sent from Rome to Sicily in the 1860s with plans to standardize the school curriculum. Differences in regional idioms, and the extent of regional insularity, were so extreme that the Sicilians mistook the teachers for visitors from England. This story can stand in for others: differences in regional dialects and social norms testify to long spans of time in which regions were not closely integrated. What changed this insularity, in Italy as elsewhere, was in large part new technologies for transport and communication. It is a truism that lower transport costs have made regions, countries and the world smaller, and played a major role in reconfiguring the spatial distribution of economic activity. Williamson (2006) provides an account of the transport revolution of the nineteenth century, documenting substantial falls in transport costs, driven by canal-building, steamships and railways. In 1817, it took fifty-two days to ship a load of freight from Cincinnati to New York by wagon and riverboat; by 1852 this had fallen to six days (Williamson, 2006, p. 8).

As we saw in section 4, theoretical models differ in their predictions about the effects of lower internal transport costs, partly because lower costs make it

easier for the consumers of a rural periphery to be served from cities. The effect of transport costs has been central to the New Economic Geography, and detailed treatments can be found in Fujita et al. (1999) and Combes et al. (2008), among others. The ambiguity of the models makes empirical work especially important, but it is not easy to quantify the causal effect of infrastructure. Investment in transport and communications will sometimes respond to regional changes in activity or population that originated with other forces. When policy-makers are forward-looking, national and sub-national governments may invest in regions with good growth prospects, or that are politically important. The research questions are difficult, but also highly relevant: investments in infrastructure for depressed areas have often been central to regional policies in Europe, China and elsewhere.

One approach to identifying causal effects is to construct a structural model with a role for transport, such as a computable general equilibrium model. The huge advantage of this approach is that counterfactuals can be studied, by simulating the patterns of regional development under different assumptions about transport technology or infrastructure investment. Williamson (1974) is an early example of this approach, Herrendorf et al. (2012) a more recent one, both covering the effects of nineteenth-century transport changes in the USA. If this approach has a weakness, it lies in the ambiguity already noted: it is not clear how to choose between models, but conclusions about the effects of transport costs are sensitive to this choice.

The work of Donaldson (2010), briefly reviewed in section 4, develops a Ricardian trade model with many locations and commodities, and trade costs. He uses this to study the introduction of the railway network of pre-partition India, seen as reducing trade costs between districts. His reduced-form regression estimate is that access to the railway increases a district's real income by 16%, and he finds that lower trade costs account for the entirety of this reduced-form effect. Donaldson and Hornbeck (2012) study the nineteenth-century expansion of the US railway network, finding effects that are more than double those in the well-known 'social saving' approach of Fogel (1964).

Michaels (2008) studies the introduction of the US Interstate Highway System, which connected cities and border crossings, but also lowered trade costs for the rural counties crossed by new roads. He finds that these counties experienced significant increases in trade-related activity, but without major changes in specialization in the directions predicted by trade theory.[35] Banerjee et al.

[35]More precisely, motivated by trade theory, he finds small increases in the wage bill of skilled

(2012) study access to transport infrastructure in China, exploiting the historical importance of connections between the major cities of the nineteenth century and the Treaty Ports. They find that regions closer to historical transportation networks have significantly higher levels of GDP per capita and higher average firm profits, but there is no evidence that the advantaged regions grew more quickly over the period studied (1986-2006).

For some developing countries, regional prosperity may also be influenced by energy infrastructure, and particularly the extent of electrification. It has been estimated that around a quarter of the world's population lack access to electricity. Lipscomb et al. (2013) study the long-run effects of electrification in Brazil: using spatial variation in the scope for hydropower plants, they can isolate exogenous variation in the extension of the network. Their results suggest that electrification brings significant gains in educational attainment, employment rates and income per capita.

The effects of geographic characteristics on electrification programmes have also been exploited by Dinkelman (2011) and Rud (2012). Dinkelman studies the effects of a household electrification programme in South Africa, using land gradient to isolate exogenous variation in access. The results indicate significant effects on female employment, potentially due to time released from home production and increased small-scale labour demand. In a study of Indian states, Rud (2012) uses the uneven availability of groundwater for electric-pump-based irrigation schemes to instrument for the expansion of the electricity network. His panel data estimates indicate that an increase in rural connections of one standard deviation would increase a state's manufacturing output by almost 15%.

Massive investments in infrastructure often appeal to policy-makers seeking to accelerate development by concrete, visible means; Lenin once defined communism as Soviet power plus electrification. This political appeal might suggest a risk of over-investment, and we have already seen that the effects of transport investments can be ambiguous. Even for electrification, the analysis of welfare effects becomes more complicated in a spatial equilibrium. New infrastructure can induce population movements that increase the demands on locally-provided public goods. Dinkelman and Schulhofer-Wohl (2012) study the issue, again for household electrification in South Africa, and find that congestion effects can halve the estimated local welfare gains.

workers relative to unskilled workers in skill-abundant counties, and small reductions where skills were scarce; but there is no evidence for changes in the industrial composition towards industries intensive in the abundant factor.

7.5 Institutions and local political economy

The study of institutions has been central to recent work on comparative development. National institutions will be among the forces that shape patterns of regional specialization and relative incomes, partly because they will influence comparative advantage at the national level. But there may also be important local variation in institutions within countries, as Acemoglu and Dell (2010), Naritomi et al. (2012) and Tabellini (2010) all emphasize. Its consequences have now been investigated for countries in Africa, Latin America and South Asia. It also has implications for the study of national development: countries could share the same rating for institutional quality - effectively a weighted average across regions - but differ in their internal institutional variation, with consequences for agglomeration and overall activity.

The idea that institutions vary within countries needs a little justification. In federal countries in particular, such as Brazil, India and Mexico, some areas of law may be determined locally, and de jure institutions will then vary across regions. But even where de jure institutions are similar, there may be substantial inter-regional differences in how these institutions operate in practice, partly given the importance of informal institutions. Tabellini (2010) emphasizes that given institutions can function differently across locations. He suggests that the judicial system works differently in southern and northern Italy, even though the formal frameworks are similar. A further complication is that which institutions matter will depend on a region's specialization; the institutions most relevant to rural agriculture may differ from the institutions most relevant to urban firms, for example. Using surveys of public employees in Bolivia, Brazil, and Chile, Gingerich (2013) shows that perceived effectiveness varies across different government agencies within these countries.

Further, the nature of the political economy will vary across regions. This could include the extent to which local elections are free and fair, the extent of control exerted by local elites, and the effectiveness of the rule of law and the judiciary. There is now a large literature examining variation in political economy at the sub-national level, with Besley and Burgess (2002) and Baland and Robinson (2008) as just two examples. Many instances of 'sub-national authoritarianism' have been documented for democracies in developing countries and transition economies; for example, the dominance of sub-national government by single parties was a feature of the US South until the later part of the twentieth century (see Gibson 2005). Through these mechanisms, there could be significant variation across regions in the quality of government, the provision

of local public goods, and in the rule of law and contract enforcement.

Local institutions will be a determinant of the comparative advantage of regions, in the same way that national institutions appear to shape the comparative advantage of countries (for example, Nunn 2007). One way in which the regional context differs is that individuals have considerable scope to relocate. Hence, when local public goods and amenities are better in some areas than others, and valued by individuals, migration across regions will take place until these advantages are offset by congestion — more intensive use of amenities — or higher living costs, such as housing costs. Similarly, firms that use local public goods intensively will tend to relocate to regions that provide these goods effectively. Hence, variation in local institutions will influence regional prosperity and population density.

In the past, the study of these effects has been hampered by the lack of data on institutional variation at the local level. The leading approach has been to study natural experiments. The well-known study by Banerjee and Iyer (2005) look at a variety of outcomes across Indian districts, notably agricultural investments and productivity, and relates them to historical variation in land rights under British rule in the nineteenth century. Banerjee and Iyer find that outcomes have diverged: those districts where land rights were given to landlords rather than cultivators have significantly worse outcomes in the post-independence period. Some of the divergence takes place relatively late, in 1965-80, which they attribute to the varying political trajectories of (historically) landlord and non-landlord districts. Their leading explanation is that, in districts where land rights were given to landlords, this led to a class-based and antagonistic politics, with consequences for policy priorities and public investment that have persisted for many decades. The results indicate that regional variation in public investment and development expenditure can make a material difference to outcomes at the sub-national level. It also seems clear that variation in local institutions and political trajectories have long-lived effects on regional outcomes. Further work on India by Iyer (2010), exploiting exogenous variation in direct British colonial rule versus indirect rule, reaches similar conclusions: areas that were under direct rule continued to have higher levels of poverty and infant mortality well into the post-colonial period.

Along similar lines, Naritomi et al. (2012) study local institutions in Brazil, finding that institutional quality and the distribution of land have been influenced by the distinct colonial histories of different regions. Acemoglu, García-Jimeno and Robinson (2012) study Colombia, identifying persistent effects of

slavery on various outcomes, exploiting spatial variation in slavery associated with the presence of gold mines during the seventeenth and eighteenth centuries. Dell (2010) finds similarly long-lived effects of a forced labour scheme in Peru, the 'mita'. The identification strategy is a spatial discontinuity design, based on comparing outcomes either side of a section of the geographic boundary of the affected area. Although the scheme was abolished in 1812, Dell establishes that its effects can still be seen today in substantially lower consumption levels, a greater incidence of child stunting, and greater prevalence of subsistence farming in the affected districts. She argues that these effects arose because the mita districts followed a different political trajectory, based on communal land tenure, compared to non-mita districts. The latter provided a more stable land tenure system that encouraged public goods provision, including education. The main results show how local institutions can have long-lived effects on spatial disparities. Moreover, these disparities seem to correspond to differences in life chances and opportunities that have not been eliminated by the possibility of migration between regions, even over many decades.

Further evidence of the long reach of history comes from the innovative study of Michalopoulos and Papaioannou (2013), briefly mentioned earlier. They study the relationship between contemporary sub-national development in Africa and measures of pre-colonial political centralization, where the latter reflect the extent of levels of political jurisdictions above the local (village) level. The initial results are based on a sample of roughly 500-700 geographic units. The measure of contemporary development is derived from satellite data on light density at night; this also allows a higher-resolution analysis of around 66,000 geographic units. Further, they also compare light density across contiguous ethnic homelands of groups that differ in their traditions of political centralization. The results consistently suggest that differences in light density, and hence in the density of contemporary economic activity, are related to long-standing differences in institutional traditions: areas with traditions of political hierarchy have higher development levels. As they note, the data on light density open up many further research possibilities.

In the literature to date, the leading sceptics are Gennaioli et al. (2013a) and Mitton (2013). To explore the question, Gennaioli et al. run a simple regression of regional GDP per capita on country dummies and a proxy for economic institutions constructed from sub-national data extracted from the Enterprise Survey and *Doing Business* reports; there are 496 regional-level observations, across 79 countries. They find that, although there is significant regional variation in their

institutions measure, it explains very little of the regional variation in GDP per capita. Mitton (2013) obtains similar results. It is not clear how to reconcile their findings with the studies based on natural experiments, although the latter give more emphasis to political institutions (as opposed to economic institutions) and exploit exogenous variation. One possible story is that perceptions of institutions depend on a region's specialization. It might also be that measured variation in economic institutions is endogenous and linked to the scope for corruption and the politicization of economic activity, which could be greater in richer or more industrialized regions. Gennaioli et al. (2013a, p. 128) note that, on average, economic institutions are perceived as weaker in a country's richest region than in its poorest. Nevertheless, these explanations are speculative, and a good instrument for sub-national variation in economic institutions seems unlikely to emerge.

7.6 Culture and social norms

The sub-national variation in cultural and social norms, and social capital, has also been studied. The concept of social capital often lacks well-defined boundaries, but partly because of this, it is a useful umbrella term for social norms such as trust and civic engagement. A contemporary economist might also consider the density of social networks, and the quality of the links within them. Much of the recent interest in social capital can be attributed to the work of Putnam (1993), who contrasted the levels of trust and civic participation between regions of Italy, and argued that these differences in social norms had far-reaching consequences, partly acting through political outcomes.

Using regional data for Europe, Tabellini (2010) analyzes the relationship between regional incomes (and growth rates) and measures of cultural norms, such as trust, respect for others, and respect for individual independence and autonomy. His study includes country fixed effects and instruments the cultural variables using long-run historical data, nineteenth-century literacy rates and early (1600-1850) political institutions, both measured at the regional level.[36] His results suggest that, although the regions within each country have long shared the same formal institutions, historical data help to explain contemporary outcomes, with the effects mediated by cultural and social norms. The tenor of these findings is consistent with Banerjee and Iyer (2005) and Dell (2010), suggesting that regional outcomes can often be traced back many decades.

[36] It should be noted, however, that the political institutions measure has only limited measured variation within some countries.

Within Russia, Acemoglu, Hassan and Robinson (2011) find long-run effects of the persecution, displacement and mass murder of Jews by the Nazis during World War II; the cities where this was most intense have grown relatively slowly, and show greater support for communist politicians. The administrative districts (oblasts) most affected have lower average wages and income per capita. Acemoglu et al. attribute these effects to the changes in social structure brought by the Holocaust.

Other work on social norms uses variation at borders. For example, Becker and Boeckh (2011) study communities in Eastern Europe either side of the border of the former Habsburg Empire. Using survey data for 2006, they find that historical affiliation of an area with the Empire is associated with higher trust and less corruption in courts and the police, even though the Empire was broken up almost a century ago, in 1918. Along similar lines, Grosfeld and Zhuravskaya (2012) use the historical partition of Poland among three Empires - Russia, Austria-Hungary, and Prussia - and find effects either side of former borders on religious beliefs, voting patterns, and political beliefs, including support for democracy. Again, it is natural to think that some of these differences would have implications for the spatial pattern of activity, and also interesting that spatial differences in culture and social norms can persist for decades.

One natural question is whether religious differences influence regional outcomes. Using data on the counties of Prussia in the late nineteenth century, Becker and Woessmann (2009) find that Protestant counties are relatively prosperous, and attribute this to higher levels of literacy prompted by Luther's emphasis on schooling. In contrast, Cantoni (2010) exploits religious variation across the German lands of the Holy Roman Empire; using data on populations for 272 cities over 1300-1900, he finds that the paths taken by Catholic and Protestant cities and regions are virtually indistinguishable.

The literature we have reviewed emphasizes differences across regions in cultural and social norms. A less obvious argument is that differences across countries could influence the agglomeration process. For example, patterns of labour mobility may differ between societies that are relatively atomistic and individualistic, and those where close family ties are especially valued. Duranton et al. (2009) establish some interesting associations between historical family types, classified for medieval Europe, and variation in outcomes within countries. Investigating such hypotheses across countries is not straightforward, however: a cross-section analysis is limited by the small number of countries in the world, and a panel data analysis by the lack of time-series variation in cultural norms.

Nevertheless, the importance of social ties is worth stressing. Economists often emphasize the benefits of mobility, and fluid economic arrangements are seen as important for efficiency. But a highly mobile labour force is also one in which networks of family, friends and neighbourhood connections are repeatedly disrupted. Most of the theoretical models present a society that is atomized by construction, and there are no frictions that arise from ties to family and friends. It is possible that such a world could exist, but few of us would want to live there.

Recent empirical work by Belot and Ermisch (2009) and Dahl and Sorenson (2010) indicates the potential importance of social ties to location decisions. This raises the possibility of nonpecuniary externalities to mobility that are not always acknowledged. The social capital perspective complicates the picture still further: Sennett (1998) raised concerns that a modern, highly-educated workforce may be mobile, but also rootless and rarely socially engaged. More broadly, economic adjustments across regions and cities are likely to have cultural and social consequences, changing the character of areas in unpredictable ways.[37] These considerations are hard to define, and might seem a topic for sociologists rather than economists, but for the inconvenient fact that welfare effects sometimes involve non-economic mechanisms.

7.7 Entrepreneurship, skills and ideas

The role of entrepreneurship in prosperity is one of the most vexed questions in regional economics. Glaeser et al. (2010) open their discussion with the following questions: can the economic history of Detroit be told without Henry Ford and Alfred Sloan? Would Ford have achieved the same success if he had worked in Houston? Would Silicon Valley have experienced its remarkable growth without Frederick Terman and William Shockley? These questions hint at some degree of indeterminacy in the evolution of regional specialization and prosperity. They seem to open the way to a 'Great Man' approach in which, to misquote Thomas Carlyle, the history of regions is nothing but the biography of great men and women.

As in the more general study of history, this idea is unsettling. Taken to its extreme, it radically undermines attempts to generalize about regional growth. But equally clearly, there are limits on the extent to which individuals (and individual companies) can be decisive; Silicon Valley was more likely to take shape

[37] As an example, Solnit (2013) provides a brief account of contemporary San Francisco which emphasizes the losses that can accrue as employment patterns and living costs change.

in California than Alaska. These considerations suggest that the idea of entrepreneurship should be invoked by historians and economists in rather different ways. The historian of a region might want to draw heavily on what Klepper (2011) terms 'nano-economics', the study of specific companies and entrepreneurs, their spin-off companies, and other legacies. This endeavour would have lessons for the study of regional prosperity, but is not coterminous with it. Economists will typically want to think about models in which entrepreneurship is an outcome or mechanism within a much larger process. Put differently, explaining what happened in retrospect (to Detroit, or Silicon Valley) is not the same exercise as understanding how entrepreneurship shapes regional prosperity in general, even if there is some overlap.

This hints at some difficulty in framing the relevant research questions. Glaeser et al. (2010) argue that entrepreneurs will play a crucial role in the extent to which cities and regions are economically dynamic, and they survey some of the literature in this area. They emphasize that many of the forces which could drive agglomeration - spatial differences in input availability, access to ideas, and local culture or institutions - will also influence the extent of entrepreneurship. We note a corollary: seeking to quantify 'the' effect of entrepreneurship on regional prosperity makes little more sense than seeking to quantify the effect of agglomeration. Serious empirical work on entrepreneurship has to contend with its endogeneity to a range of economic and social forces, and this helps to explain why there is not more work on the topic.

Progress might depend on ingenious use of natural experiments, with Glaeser et al. (2012) as a leading example. They argue that, in the US, areas close to mines were more likely to specialize in industries like steel, with significant scale economies and dominated by large firms; as a result, the conditions for entrepreneurship were less likely to arise. They find that proximity to historical mining deposits (in 1900) is indeed associated with larger firms and fewer start-ups decades later, and use this proximity as an instrument for entrepreneurship. Across cities, entrepreneurship is strongly associated with faster employment growth even in IV estimates.

Entrepreneurship plays a central role in the structural model of Gennaioli et al. (2013a), discussed in section 4.3 above. Individuals can choose between employment and entrepreneurship; more able individuals self-select into entrepreneurship and, as in Lucas (1978), especially able entrepreneurs run larger and more productive firms. The most important empirical consequence is that human capital formation is placed centre-stage, as a source of highly-able indi-

viduals; and traditional Mincerian wage regressions, or development accounting exercises, risk understating the effect of schooling on regional prosperity, because some of the returns to schooling are reflected in capital income rather than wages. Felkner and Townsend (2011) also construct an occupational choice model with a role for entrepreneurship, but emphasizing the role of local access to finance; they simulate the model on detailed data for Thailand, and compare the paths taken by spatial enterprise concentration with those seen in the data.

The model in Gennaioli et al. (2013a) is essentially static, and designed to explain outcomes in a cross-section of regions. But one reason for being interested in entrepreneurship is that it may help to explain why some regions successfully reinvent themselves, while others lose dynamism. It is also related to ideas about information transmission and the extent to which individuals and firms are linked through various networks; this suggests the benefits of integrating the analysis of regions with ideas from urban economics. Glaeser and Gottlieb (2009) argue that, to be successful, modern cities increasingly depend on the links between urban density and the transmission of ideas. For some countries, the spatial concentration of the highly-skilled is likely to be increasing; Moretti (2012) argues this for the US.[38]

Particular instances of entrepreneurship, firm entry, and the concentration of skilled workers in particular locations, are often thought to be associated with universities, as in Stanford's influence on Silicon Valley, or the cluster of technology companies around Cambridge in the UK. Traditionally, the study of some of the effects of universities has drawn on the local multipliers and impact assessments developed in the regional science literature.[39] But quantifying the wider benefits of universities for the local transmission of ideas, innovation, firm entry, or the 'creative classes' of Florida (2002) is even harder.

Universities are not randomly assigned across locations, and natural experiments are hard to find. Moretti (2004) studies the social returns to education in the US, partly by using the land-grant colleges of 1862 and 1890 to instrument for differences in the share of college graduates across cities. He finds that a higher college share not only increases the wages of less-educated workers, but also those of the well-educated, consistent with a role for human capital externalities. An alternative approach is to use more detailed data, perhaps at the

[38] See also Ganong and Shoag (2013) for further discussion and references.

[39] Armstrong and Taylor (2000) includes an introduction to these approaches, acknowledging some important objections. Using, instead, an econometric approach to local multipliers, Moretti (2010) finds that one additional skilled job in the traded sector will generate 2.5 jobs in providing goods and services.

establishment level, to study particular mechanisms; for example, Abramovsky et al. (2007) find that business R&D in the UK is sometimes located close to highly-ranked university research departments in related disciplines.

7.8 Local financial development

Does local financial development matter? In a well-known paper, Jayaratne and Strahan (1996) used data on US states to study the effects of bank deregulation on economic growth. Most US states began the 1970s with restrictions on the expansion of bank branches within and across state borders; over the following twenty-five years, the majority of these states eliminated or loosened the controls. Using an empirical model with state fixed effects, Jayaratne and Strahan find that branching deregulation significantly increased the growth rate of state personal income per capita and gross state product per capita. The evidence that this was achieved by a greater volume of commercial lending is not strong. Instead, the lifting of branching restrictions seems to have resulted in a lower share of non-performing loans in the state total, and a lower share of loans being written off each year. This evidence on the quality of lending is not conclusive, since the loan portfolios of banks may have changed in their size composition and in the riskiness of borrowers. Nevertheless, the paper suggests that local financial intermediation can influence regional prosperity.

More recently, the same policy reform has been revisited by Huang (2008) using a spatial discontinuity design. He compares performance across pairs of contiguous counties either side of a state border, where one county is affected by deregulation earlier than the other. This approach allows for heterogeneity in treatment effects over time and across states. The evidence that deregulation had economic benefits seems noticeably weaker than in the Jayaratne and Strahan study, although it is not clear whether the alternative approach to identification has led to more reliable estimates, or just to greater imprecision. The results do not appear to be driven by spillovers of deregulation across borders, since Huang also compares outcomes using 'hinterland' counties within the still-regulated states, further from the border. He emphasizes that the instances of significant growth accelerations in his study all occur relatively late in the reform process, after 1985, and interprets this in terms of 'learning by observing', so that states which liberalized later tended to have better outcomes on average.

For those developed countries without restrictions on inter-regional lending or spatial variation in regulation, a sceptic might argue that local financial development cannot matter. Comparing regions of Italy, Guiso et al. (2004) find

evidence that it does: for example, the ratio of new firms to population is 25 percent higher in the most financially developed Italian region, compared to the least. Their study exploits a 1936 banking law, which had persistent effects on the number of bank branches, and can be used as an instrument for the exogenous supply of credit. Natural experiments have also been found in other countries. In a study based on Russian data, Berkowitz et al. (2012) use regional variation in banking that arose at the end of the Soviet era, and its establishment of specialized banks ('spetsbanks'). They find that the presence of the spetsbanks increased within-region lending to firms and individuals, but had no discernible effect on income per capita. Regions with spetsbanks are associated with increased employment rates, however.

Chen et al. (2010) study venture capital in the US, noting that venture capital firms, and venture-capital-financed companies, are heavily concentrated in just three metropolitan areas (Boston, New York and San Francisco). They associate this with localized knowledge spillovers in sectors especially likely to draw on venture capital; with localized knowledge spillovers across venture capital firms; and with entrepreneurs that seek finance from previously-successful venture capital firms. These features could lead to a virtuous circle as entrepreneurs locate businesses close to funding sources, and other venture capital firms enter at the same location; conversely, other regions may experience a vicious circle. Chen et al. suggest that policies which increase the number of venture-backed investments in a region will increase the chances of venture capital firms establishing offices in that region.

For developing countries, there are complicating factors, not least the close connections between the banking sector and the state that are found in some countries. China is an important example: Démurger et al. (2002) argue informally that the monopoly state banking system has contributed to regional inequality, by limiting access to external finance in the interior provinces and by assigning priority for lending to the state-owned enterprises in the coastal and north-eastern regions. For India, Burgess and Pande (2005) investigate the effects of a large state-led expansion in bank branches in rural areas; they find that it increased deposit mobilization and lending, and lowered rural poverty. Fafchamps and Schündeln (2013) study local financial development in Morocco, at a lower level of aggregation, corresponding roughly to a city or county: they find that access to a bank increases firm entry, raises firm growth, and lowers the likelihood of firm exit.

7.9 Other policies and regulations

We have already discussed some well-known papers on specific policies and regulations. One approach to local policy variation uses spatial discontinuity designs, as in the paper by Holmes (1998) reviewed in section 6.5 above. A more recent example is the work of Duranton et al. (2011) on local taxation, based on pairing establishments across borders in the UK. But the literature on local and regional policies in developed countries is sufficiently extensive to require a dedicated survey of its own, which space does not permit.

Instead, we briefly review some evidence for developing countries, much of it based on the states of India. The case of India is interesting because policies and regulations have varied across states and over time. Aghion et al. (2008) study the effects of dismantling the License Raj, a 1951 system of controls that regulated entry and production in the formal manufacturing sector. The elimination of these barriers to investment and entry affected states differently: those industries in states with pro-employer labour market institutions grew faster than those in states with pro-worker institutions. Since pro-worker institutions seem to be directly associated with weaker industrial performance, the overall effect of de-licensing was to increase the disadvantages of states with pro-worker labour market institutions. Earlier, the panel data study of Besley and Burgess (2004) had already found that pro-worker labour market regulation was associated with lower output, employment, investment and productivity in the formal manufacturing sector; higher output in the informal sector; and higher rates of urban poverty.

Besley and Burgess remark that, in this case, specific attempts to redress the balance of power between capital and labour seem to have worked against the interests of the poor. Their earlier panel data study, Besley and Burgess (2000), examined an alternative redistributive policy, land reform. They find that reforms which changed the terms of land contracts lowered poverty and raised agricultural wages, although this may have been accompanied by lower average income. Implementing land reform had a poverty-reducing effect equivalent to growth in income per capita of around 10 percent. Since the estimated effects vary with the exact type of land reform, a further lesson of their study is that the specific details of a policy intervention can matter a great deal. A remaining question raised by these papers, not straightforward to answer, is the effect of policy variation on regional disparities when states are linked in a spatial equilibrium. Although labour mobility across Indian states is likely to be low, entrepreneurs and firms must still decide where to locate, and agglomeration

and growth can be determined jointly even when labour is immobile.

Policies can also influence regional TFP through their effect on factor misallocation. For example, Brandt et al. (2013) study distortions within China, and find that most of their estimated within-province distortions are due to the misallocation of capital between the state and the non-state sectors; this misallocation lowers province-level TFP. For China, there has also been some work on policy-induced barriers to trade between provinces; see Young (2000) and Holz (2009) for alternative views on their importance.

7.10 Conflict

For some countries, localized conflict can influence the relative prosperity of regions, both in the short run and in ways that unfold over time. India's class conflicts take their most extreme form in the 'Naxalite' or Maoist peasant uprisings, which affect the Red Corridor within eastern states.[40] The affected states are among the poorest in the country, but since economic and political outcomes are jointly determined, identifying the causal effect of conflict is difficult. Comparing Indonesian provinces, Hill et al. (2008) suggest that conflict has been a factor in the slow growth of Maluku and, to a lesser extent, resource-rich Aceh; relevant to the latter, Morelli and Rohner (2010) examine the relationship between the spatial distribution of natural resources and the risk of conflict, including the rise of secessionist movements.

One approach to recovering a causal effect is that of Abadie and Gardeazabal (2003). They use their synthetic control method, reviewed in section 6.6 above, to study the effect of Basque terrorism. They find that it reduced GDP per capita in the Basque Country, relative to a synthetic control region without terrorism, by ten percentage points.[41]

Another branch of the literature studies the effect of wartime destruction on the spatial distribution of population, or the relative outcomes of affected regions. The aim is not to investigate the overall humanitarian or economic costs of conflict or war, but to see whether past (localized) destruction influences later regional outcomes. Two well-known studies consider the effects of World War II bombing, by Davis and Weinstein (2002) and Brakman et al. (2004b); the former was briefly discussed in section 7.1 above, and both are reviewed in detail in

[40] Banerjee and Iyer (2005) note that the regions most associated with this conflict are areas where landlord-based systems were implemented under British rule.

[41] As well as the synthetic control method, they also used an event study of the stock prices of firms significantly exposed to the Basque Country, to show that these stock prices outperformed when the 1998-99 ceasefire became credible, and underperformed at the end of the ceasefire.

Brakman et al. (2009). Their review concludes that some shocks had permanent effects, consistent with models of agglomeration in which there are multiple equilibria and path dependence.

Miguel and Roland (2011) study the long-term regional effects of the US bombing of Vietnam, noting that it was heavily concentrated in a subset of their 584 sample districts, and hence with scope for differential effects across regions. They find that districts heavily bombed between 1965 and 1975 had moderately lower consumption (compared to other districts) in 1992-3, but this effect had disappeared by 2002; nor do they find significant long-run effects on the relative poverty rates, electricity infrastructure, literacy, or population density of the affected areas. A complicating factor is that the Vietnamese government undertook major reconstruction efforts; but otherwise, the findings indicate that long-run patterns of spatial activity are largely independent even of damaging bombing campaigns. Miguel and Roland interpret this as evidence against simple models of regional poverty traps.

8 Regional Decline

One of the simplest points about regional economics is also one of the most fundamental. The invisible hand is more active at some times and places than others, and once distinct points in space are introduced into economic theory, the conventional arguments that markets can be Pareto-efficient no longer apply. Markets, left to themselves, can establish patterns of regional growth and decline that involve many economic and social costs. One of the best reasons to study regional growth might be to learn how to forestall or reverse regional decline.

In this section, we discuss some of the processes involved in decline. Its analysis is partly the obverse of regional growth; for example, the results of Holmes (1998) tell us not only about the growth of US states with pro-business policies, but also the relative decline of states without them. Similarly, the evolution of location-specific advantages, such as market access, can explain decline as well as growth. Yoon (2013) argues that reductions in local advantages help to explain the decline of the US Rust Belt, compounded by a reversal of agglomeration and a decline in the quality of local public goods. This perhaps hints that decline raises specific issues of its own, which have been under-researched. Our treatment will be relatively discursive and speculative, emphasizing areas for future research rather than drawing heavily on existing work.

What do we mean by decline, and does it matter? Regions could be declining

in terms of absolute or (more often) relative living standards and welfare indicators, but also in terms of absolute or relative population, since one response to economic decline will be out-migration and diverted in-migration. This second kind of decline is often a symptom of the first, but has interest in itself, as a distinct process. In the US, a country usually judged to have high labour mobility, it is perhaps not surprising that the primary response to the Dust Bowl was out-migration (Hornbeck 2012). Similarly, the Rust Belt has seen its share of the US population decline.

With few exceptions, economists generally take a benign view of factor mobility, and see it as a powerful equilibrating force. It is true that out-migration will sometimes benefit both migrants and those who remain behind, but this is not inevitable. In other ways it has the potential to compound the problems of a declining region, as Myrdal (1957) discussed. One complicating factor is selective migration; those who leave a declining region will often be the young and well-educated. Even in the absence of conventional human capital externalities, this form of out-migration could be self-reinforcing, and have social and political consequences for the declining region.[42] This is not to deny, as Myrdal seems to have done, that the logic of a spatial equilibrium will reassert itself. But the process of reaching it may involve significant costs, especially where the decline is absolute rather than relative.

The New Economic Geography literature has investigated decline in terms of the combined effects of changes in inter-regional trade costs and assumptions on labour mobility. In Puga (1999), when trade costs are high, industrial activity is dispersed. If trade costs fall, this promotes the agglomeration of activities with increasing returns. This is compounded by migration, implying the relative (and perhaps absolute) decline of some regions. But if workers do not move across regions, then as trade costs fall further, firms become increasingly sensitive to cost differentials across regions, and industry will spread out once more. Nocco (2005) considers a variant of this model with a role for knowledge spillovers across regions. A natural question is whether agglomeration is optimal; to consider this, Ottaviano and Thisse (2002) study a two-region economy with skilled workers that are mobile, and unskilled workers that are not. Market forces lead to the optimal outcome when trade costs are high or low, but for intermedi-

[42] In some ways, points such as these — the limitations and constraints on migration as an equilibrating force — have been better understood in the literature on developing countries. Lipton (1980, p.15) writes that it is 'perfectly consistent to claim, as I do, that the migrant on average gains from migration, but the village he leaves behind loses'. See also Kanbur and Rapoport (2005).

ate levels of trade costs, agglomeration takes place when dispersion is socially desirable.

When regional decline is discussed, an idea often heard is that policy-makers should seek to protect people rather than places. Some economists seem to take the view that, if out-migration is taking place, so be it. But this view risks leaving too much out. Some consequences of a region 'emptying out' are inefficient, and involve multiple externalities. Infrastructure and social overhead capital will be written off or less well utilized, and the local tax base eroded. Movements of population to other areas will require new investment and increase demands on local public goods. Declining regions are likely to become low-trust, high-crime regions. Most of these outcomes will not be internalized by migrants, and it is often hard to see anything creative in the destruction of social capital.

As a practical matter, there is a large literature on regional policy, understood as a response to decline, whether in terms of relative economic position, or sustained out-migration. But one constraint on this literature is that some of the mechanisms underlying decline, such as crime, social unrest and local political consequences, are complex. There is also a risk that regional policy could be too reactive. Once decline is under way, disadvantages can accumulate, and may be hard to reverse. Anticipation of regional decline could even be self-fulfilling, which again suggests the need to consider regional problems in dynamic terms. On the basis that prevention is better than cure, one issue for policy will be a given region's extent of diversification, and hence its robustness to shocks. But, at the risk of labouring the obvious, there is no good reason to expect that decentralized markets will lead to the optimal degree of diversification, not least given the many externalities involved. This in itself could justify some degree of intervention.

To emphasize the lack of diversification of, say, 1960s Detroit could seem a little too easy, a form of retrospective wisdom. After all, not many are currently calling for Silicon Valley to diversify. But as Glaeser (2011) emphasizes, Detroit's problem was that its fortunes were closely tied not just to a small number of sectors but to a small number of firms, the 'Big Three' of Chrysler, Ford and General Motors; to an uncomfortable extent, Detroit was a three-company town. The quantitative exercise of Alder et al. (2013) attributes much of the Rust Belt's wider decline to a lack of competition and powerful unions.[43] A related perspective could draw on Gabaix (2011), who argues that idiosyncratic shocks to

[43]See also Desmet and Rossi-Hansberg (2013a) for a quantitative account of the declining populations of Rust Belt cities, partly in terms of relatively large local frictions.

large firms can account for aggregate business cycle fluctuations to a significant degree. This same idea of the 'granularity' of economic activity could also be applied to regional growth and regional decline, and may be especially important for relatively small countries.

Another, closely related, lesson of Detroit might be the potential for path dependence, or regional 'lock in', a theme of some recent work by geographers. Specialization is an endogenous outcome, the consequences of which unfold over time, and that interact with later shocks. Klepper (2010) argues that the post-war development of Detroit - and that of Silicon Valley - was partly driven by successful spinoffs from high achieving firms, and 'organizational reproduction'. A natural corollary is that, in the long term, success in narrowly-defined areas could crowd out other entrepreneurial activities (Glaeser et al. 2010). A city or region could become locked in to particular sectors or lines of activity, bringing the risk of future decline. Martin and Sunley (2006) discuss work on path dependence in more detail, emphasizing that not much is know about why some regional economies lose dynamism, while others evolve and continually reinvent themselves.

Some argue that the solution to regional decline is to promote clusters of firms in particular sectors. The practical importance of clusters, as a source of higher productivity or a response to regional decline, continues to divide opinion.[44] The work of Klepper (2010) implies that understanding specific industrial clusters requires detailed attention to their genealogy. The most famous examples appear to have developed in a largely organic way, rather than through external intervention; there is room for debate over whether pro-cluster policies would be effective, even if desirable.

It might be easier to achieve consensus when the analysis of regional decline focuses on the labour market. Kline and Moretti (2013) discuss the possibility of hiring subsidies that vary across locations, as a candidate place-based policy. An alternative approach emphasizes the potential benefits of local ownership: the argument is partly that locally-owned firms are less likely to reduce employment in the face of negative shocks. Kolko and Neumark (2010) investigate this hypothesis for the US, finding that the greatest benefits come not from small independent businesses but corporate headquarters, followed by locally-owned chains. A different place-based policy adopted by some countries, including the UK, is to locate public sector offices in depressed regions. Again, this may help to promote stability, although a general equilibrium analysis is needed.

[44] See Duranton (2011) for an especially sceptical view.

These observations point to the importance of studying regional decline in more depth. In the meantime, reading Adam Smith, or for that matter most contemporary textbooks in economics, would provide little assistance to the citizens of cities and regions that confront decline. Their problems deserve more attention from economists. Until this happens, the mechanisms and costs of decline will be comprehended deeply only by those directly involved, and with much to lose. What the invisible hand gives, it can also take away.

9 Conclusions

The study of regional growth is often thought to be simpler and more straightforward than national growth. We have emphasized, instead, various ways in which it is harder. Regional outcomes are best seen in terms of a spatial equilibrium. Regions are interdependent, and their locations matter. For example, theoretical models predict that market access influences relative prosperity and population density, and these predictions are supported by a variety of evidence. Meanwhile, labour mobility implies that incomes, populations and living costs are all endogenous, and must be considered jointly. The days when a textbook on regional economics could legitimately base most of its discussion on the neoclassical growth model are gone.

Some of the other inheritances from the cross-country literature have been problematic. Many empirical studies treat the observations on regional units as if they derived from independent entities. But with regional outcomes tightly linked in various ways, it is rarely straightforward to identify causal effects from regional data, or to relate the estimated effects to underlying quantities of interest. For example, in the regression-based studies, it is rarely clear how to interpret the estimated effects of a given variable on productivity or growth. Do these estimates hold constant the spatial distribution of population and economic activity, or do they partly reflect endogenous changes in agglomeration?

This distinction becomes especially important whenever a researcher seeks to draw lessons for national growth, or regional policies. The problem can be seen with an extreme example. If one region gains from a specific policy only by expanding at the expense of another, any analysis which implicitly holds fixed the spatial distributions of population and activity will be misleading about welfare effects. The arithmetic of regional policy is complicated, and what is an addition for one region may be subtraction from another, even when the policy aimed at multiplication. As elsewhere in the study of regional data, the quantitative

application of structural models seems the most promising response.

We have also tried to highlight some areas where additional research seems especially needed. When economists consider the possibility of a regional problem, they typically examine disparities in average living standards, and their evolution over time. The empirical literature on this topic is vast. Yet some of the most important regional problems are likely to be those where areas are persistently losing population. We have emphasized that regional decline, conceived in these terms, is likely to be a distinct process, and one that has rarely been studied by economists.

In passing, we have also drawn attention to some of the burdens of adjustment, such as the non-pecuniary externalities that can arise through mobility. Formal models sometimes indicate that even modest economic changes, with similarly modest welfare effects, involve substantial redistributions of population and economic activity. One possible conjecture is that a market economy may sometimes involve 'too much' ongoing relocation. Limiting mobility is rarely attractive, however. A more promising avenue would be to investigate forms of economic arrangements that lessen the need for mobility in the first place.

We will not attempt even a short summary of the forces that influence regional growth. Too much remains uncertain, and it is a sign of the current health of the literature that any such survey would quickly become dated. Instead, we have emphasized recent developments in the study of regional growth, both theoretical and empirical. As well as the impetus from the New Economic Geography literature, the increasing availability of (and interest in) regional data sets mean that the field is evolving at a great pace. Methods such as spatial discontinuity designs and the use of natural experiments have shed new light on causal effects, while the quantitative application of structural models is likely to be highly informative. Combined, these developments suggest that regional growth has become a particularly exciting area of economics, rich in data, interesting research questions, new methods, and increasingly sophisticated models. The study of growth has belatedly entered its own Space Age, and there is no going back.

10 Appendix: data and methods

This appendix discusses regional price deflators; criticisms of the beta convergence approach; and some alternative methods for studying regional growth.

10.1 Regional price deflators

Within countries, prices for identical goods often differ across locations. Ideally, it would be possible to compare output and income across regions in real terms, in the same way that the Penn World Table allows comparisons of real output across countries. Accurate comparisons of output or productivity across regions require PPP deflators or measures of regional output that aggregate goods and services at a common set of prices (such as producer or 'mill' prices). Similarly, for the study of differences in the standard of living across regions, it would be useful to have cost-of-living deflators, partly based on housing costs.

In practice, real comparisons of regional output are rarely possible over long spans of time; only a few countries, including Canada and China, release data which allow for price differences across provinces. Sometimes, deflators may be available for just a few points in time. Aten (2008) and Aten and D'Souza (2008) have undertaken this for the US. The cross-country, cross-section data set of Mitton (2013) adjusts for some differences across regions in the cost of living, by linking regions to data on living costs for particular cities. More generally, measures of regional inequality can adjust for price differences at particular dates, but in principle, a researcher studying regional growth and convergence needs deflators for each date.

This makes it important to consider the main influences on regional price levels, and their variation over time. If labour is homogeneous and mobile, a spatial equilibrium requires real incomes to be equalized across locations; in that case, regions with higher nominal wages must have higher price levels for goods and services, and/or higher housing costs. This result emerges from general equilibrium models, such as those developed in Redding (2012). In his analysis, market access also matters: well-connected (less remote) regions will tend to have relatively low consumer prices for tradable goods. For a migration equilibrium, this must be offset by a higher population that drives up land prices and living costs, and hence equalizes real wages.

It seems likely, at least for developed countries, that national statistical agencies already have some of the raw price data needed to construct regional-level deflators. Deaton and Dupriez (2011) note that the agencies are 'strangely reticent' on this topic.[45] For prices to be representative of a region, data on the spatial distribution of population are also needed; but if prices differ across locations, such data are needed in any case, to derive national-level deflators that

[45]Nevertheless, work is likely to emerge using disaggregated data on purchases and prices, from other sources; see Handbury and Weinstein (2011).

are representative. For cost-of-living deflators, a major component is likely to be housing costs.

In the absence of official data, an open question is whether empirical researchers could make progress by imputing price levels. One approach to living costs assumes that households with the same budget share of food, but in different locations, have the same level of welfare; a comparison of their nominal expenditure levels then reveals the relative price levels at the different locations. But the key assumption, that households with the same budget share of food have the same welfare, is strong; see Deaton and Dupriez (2011).

An alternative approach might use simple assumptions about the sensitivity of price levels to development levels or measures of market access, perhaps drawing on theoretical models. These relationships could then be used to map between observable variables and the unobserved true deflators, at least for the purpose of a sensitivity analysis. One question, which could be studied using the currently available data, is whether regional deflators are sometimes stable enough (relative to one another) that growth and convergence studies can give reliable answers even in their absence. A related question is the extent to which cost-of-living deflators can proxy for the price levels of output needed for productivity comparisons. The results of Redding (2012) suggest that this could be risky, not least if market access varies widely across locations.

10.2 Beta convergence

In their empirical work on convergence, Barro and Sala-i-Martin (1991, 2004) assume that steady-states are similar across regions. This assumption does a lot of work. It means that an explicit theory of steady-state positions is not required. From an econometric point of view, it provides a justification for studies of absolute convergence across regions, of the kind they and other authors have carried out. But in their work, the similarity of steady-states is assumed rather than established. It seems unattractive on theoretical grounds: in a market economy with labour mobility, the average product of labour will necessarily vary across regions, due to composition effects among other forces.

We could still ask whether their approach is informative about the extent of long-run disparities. One perspective on this is to look at the R^2 of an absolute convergence regression. A typical model would have the form

$$(y_{it} - y_{it-\tau})/\tau = \eta + \beta y_{it-\tau} + \phi_t + \varepsilon_{it} \tag{25}$$

where y_{it} is the logarithm of output per capita for region i at time t. Sala-i-Martin (1996) and Barro and Sala-i-Martin (2004) argue that estimates of β often correspond to a convergence rate of around 2% a year. The regressions that Barro and Sala-i-Martin present for the US states and Japanese prefectures (their Tables 11.1 and 11.2) often have a relatively low R^2 for short subperiods. But over longer spans of time (1880-2000 for the US, 1930-90 for Japan) their simple regression has an R^2 of 0.92 for both countries. At first glance, this indicates that steady-state positions are similar. But this is misleading: their regression omits fixed effects, which could proxy for time-invariant determinants of relative income levels. In the absence of these fixed effects, it is likely that the parameter estimates are biased, and the high R^2 is misleading.

When the regression (25) is discussed in the literature, β is typically regarded as the main parameter of interest. For the study of regional growth, we should be interested in the more general model

$$(y_{it} - y_{it-\tau})/\tau = \eta_i + \beta y_{it-\tau} + \phi_t + \varepsilon_{it} \tag{26}$$

The variance of the region-specific effects (the η_i) will be denoted σ_η^2, and should be seen as a key parameter of interest. After taking out the common time effects ϕ_t, the model in (26) implies that region i is mean-reverting with long-run mean $\mu_i = \eta_i/\beta$. The cross-section variance of μ_i therefore depends on β^2 and also on σ_η^2. But also note that, given continued shocks, each region's output will continue to vary over time.

For the US, using a fixed effects estimator on ten-year subperiods doubles the estimated rate of convergence (results not reported). For a large sample of countries with sub-national data, Gennaioli et al. (2013b) find that including regional fixed effects greatly increases the estimated rate of convergence. The assumption in Barro and Sala-i-Martin that $\sigma_\eta^2 = 0$ seems hard to defend. Barro and Sala-i-Martin provide an alternative justification, which is that $y_{it-\tau}$ may be uncorrelated with η_i. But if the process has been running for any length of time, this alternative assumption is also unattractive, because a mean-reverting process such as (26) will necessarily generate a correlation between output per capita and the fixed effects.

There is another reason for querying this approach. In the cross-country literature, the process in (26) has a structural justification: it approximates transitional dynamics in the vicinity of a balanced growth path. But for regions, the neoclassical growth model should not be expected to apply, given inter-regional flows of capital and labour. Hence, for regional data, (26) is not structural, but

only a way of capturing the time-series dependence in the data. Instead, Gennaioli et al. (2013b) suggest the use of a specification in which each region's factor input (perhaps some broad notion of capital) is a Cobb-Douglas function of its endowment of that factor - based on past investment - and the level that would obtain under full mobility. As they acknowledge, this assumption is ad hoc, but it leads to a simple specification which generalizes the standard conditional convergence regression. The extent of the barriers to factor mobility can be estimated from the data, although their estimates indicate higher barriers than might have been expected.

Whichever model is adopted, using a single lag may give a misleading picture. Regional living standards could be influenced by omitted variables which are themselves autocorrelated, and so e_{it} will be serially correlated. For regional data, a natural generalization of 26 is:

$$(y_{it} - y_{it-\tau})/\tau = \eta_i + \beta y_{it-\tau} + \phi_t + u_{it} + \varepsilon_{it} \tag{27}$$

$$u_{it} = \rho u_{it-\tau} + v_{it} \tag{28}$$

which implies

$$(y_{it} - y_{it-\tau})/\tau = \eta_i' + (\beta + \rho/\tau)y_{it-\tau} - (\beta\rho + \rho/\tau)y_{it-2\tau} + \phi_t' + v_{it} + \varepsilon_{it} - \rho\varepsilon_{it-\tau}$$

and makes clear the likely inadequacy of a model with just one lag. There are further reasons that serial correlation is likely. In the cross-country literature, the neoclassical growth model can be used to argue that cross-section and time-series variation in the β parameter should be limited. This seems less plausible for regions, and the heterogeneity will lead to serially correlated errors. Measurement error, partly due to time-varying regional price levels, could also lead to serial correlation. These points suggest that beta-convergence regressions, with or without fixed effects, have significant weaknesses. The remainder of this appendix considers some alternatives.

10.3 Time series approaches

Recent studies draw heavily on the implications of convergence for the time series properties of regional data. Bernard and Durlauf (1996) showed how to relate different concepts of convergence to time series properties. To fix ideas,

we will initially consider how a researcher should proceed in the case of two regions. The choice of the null hypothesis needs thought, and should depend on the claim that a researcher is interested in seeking to falsify. If the two regions are believed to be on parallel growth paths, and a researcher wants to see if this claim can be falsified, a natural approach is to look at their (log) output gap and apply stationarity tests.

When the regions are genuinely following parallel growth paths, a stationarity test such as that of Kwiatkowski et al. (1992) should not reject the null of stationarity. Alternatively, if a researcher wants to examine a claim of divergence, the natural approach is to test whether the output gap contains either a stochastic trend (using a unit root test) or a deterministic linear time trend (as when log incomes in the two regions are trend-stationary processes with different trend growth rates).[46] Note, however, the maintained assumption that long-run steady-states are time-invariant, a point we return to shortly.

Extending these ideas to N regions, output gaps could be defined relative to a particular benchmark region, or a weighted average, as in early work such as Carlino and Mills (1993, 1996). But this approach becomes problematic if one or more regions are diverging from the others. The results will vary with the choice of benchmark, and using a weighted average will indicate non-convergence even when a subset of regions is moving together. In principle, a more attractive approach is to allow each region's growth to be a function of the $N-1$ output gaps with other regions (Carvalho and Harvey 2005). But a flexible version of this, with separate catch-up coefficients for each ordered pair, implies $N(N-1)$ parameters and hence becomes difficult or impossible to estimate when the number of regions is large.[47]

An attractive alternative is that of Pesaran (2007), who develops a test based on all $N(N-1)/2$ pairwise output gaps. Taking the null of interest to be 'non-convergence', Pesaran shows that under this null, the fraction of pairwise output gaps for which a unit root is rejected should be close to the size of the unit root test that has been applied (e.g., 5%). The fraction of pairwise gaps for which a unit root is rejected can be taken as a measure of the extent of convergence.

[46] One potential complication here is that convergence could be present but slow, so that the log output gap is a fractionally integrated process. For a study that includes an application of this idea to data on the contiguous US states, see Mello (2011).

[47] A somewhat related approach is to apply multivariate tests for cointegration, such as Johansen's method, as in the early study of cross-country convergence by Bernard and Durlauf (1995). This approach can provide evidence on the number of common stochastic trends likely to be driving the output movements of the N regions. But again, it becomes infeasible when the number of regions is large.

The detailed data on rejections will divide the sample into groups for which non-convergence is rejected, and regions that are evolving independently. This is more informative than collapsing the issue to a binary opposition between convergence and divergence.

Pesaran's approach has been applied to the contiguous US states by Mello (2011) and to European regions by Le Pen (2011). The fraction of pairwise gaps for which a unit root is rejected is typically low, suggesting non-convergence for the majority of regions, even though other tests provide clear evidence of mean reversion. In response to these findings, Le Pen (2011) argues for the importance of structural breaks — that is, mean shifts in the output gaps. But this highlights a fundamental dilemma for time series approaches. For the tests to have some power, long spans of data are needed, but then it is harder to maintain the assumption that relative steady-states are time-invariant. If steady-states are evolving over time, this breaks the direct connection between time-series properties and convergence concepts.[48]

The dilemma arises partly from taking a univariate approach to a process influenced by a wider range of variables. In the literature on national growth, the steady-state positions are typically modelled as stable functions of a few variables, as in Mankiw et al. (1992). This is harder to implement for regional data, partly because data on potential control variables are often lacking, and partly because interesting models of regional disparities may not yield simple expressions for steady-states. At least as a way of describing the data, an alternative approach uses the behavior of the cross-section dispersion (or inequality) in regional income to draw conclusions about the underlying statistical processes. Evans (1996) showed that if the units such as regions follow independent random walks, then the (cross-section) log variance will be integrated of order one around an upward quadratic trend. If the regions are instead believed to have converged and to be driven by a common trend, the log variance will be stationary and fluctuate around a constant mean.[49] This notion of convergence does not require the cross-section variance to decline monotonically to zero, an outcome that is unlikely for a collection of stochastic processes. Evans (2000) includes an application of this idea to the contiguous US states.

Another route is to develop methods for describing growth paths of eco-

[48]Note that a pairwise output gap process which is stationary, but with mean shifts, is compatible with either catching-up or divergence, depending on whether the mean of the output gap shifts downwards or upwards, respectively. The time series approach then becomes harder to implement, and interpret.

[49]For some related discussion, see Ng (2008) and Pesaran (2007).

nomies that are converging, while separating out long-run effects from cyclical components, by using unobserved component models. For studies of US convergence that adopt this approach, see Carvalho and Harvey (2005) and Carvalho, Harvey and Trimbur (2007).

10.4 Distribution dynamics

A popular approach has been 'distribution dynamics', developed for cross-country data by Quah (1993) and applied to regional data by Quah (1996). This approach characterizes transitions of income per capita between income classes, using a transition matrix whose elements are the probabilities of moving from one income class to another. There are important ways in which this is more flexible than a panel data model, and more informative about the underlying process. It provides direct information about mobility between income classes, and the stationary distribution implied by a given transition matrix will reveal tendencies latent in observed realizations of income levels (Quah 1993). Under the strong assumption that the transition probabilities remain stable over time, the stationary distribution provides a long-run forecast of the shape of the distribution of regional income levels.

Kremer et al. (2001) make the useful observation that, when considering income levels for aggregate economic units, banded into wide classes, it is likely that the only non-zero transition probabilities are those between adjacent income classes. More dramatic relative movements are unlikely for countries or regions, at least over short spans of time. In this case, the ratios of the individual elements of the stationary distribution can be derived as ratios of transition probabilities. But since a ratio can be sensitive to a small change in its denominator, the estimated stationary distribution may be sensitive to small changes in the estimated transition probabilities. Hence, at least when the stationary distribution is the main result of interest, one drawback of this approach is a lack of robustness. Kremer et al. suggest an alternative method, which is to iterate the estimated process over a limited number of future periods and study the outcome, rather than emphasizing the stationary distribution.

A further problem arises from the discretization that is often used to construct the transition matrix. An alternative is to treat the state space as continuous and model the joint distribution of outcomes at t and $t + \tau$, as in the cross-country work of Quah (1997) and Johnson (2005), for example. But given the number of regions typically available to a researcher, there is not a great deal of information from which to estimate something as complex as a joint

distribution, again implying a lack of robustness.

Despite its problems, an attractive aspect of the distribution dynamics approach is that it can be used to investigate regional polarization. Quah's work on national growth is strongly associated with his 'twin peaks' result, the finding that the stationary distribution is bimodal. In the regional growth literature, his methods have been the most popular approach to the study of polarization. Some other methods are available, using various ways of defining polarization (see Anderson et al. 2012 and Zhang and Kanbur 2001).

10.5 Multimodality and mixture densities

The hypothesis of polarization is a special case of a more general idea, that of convergence clubs. A common intuition is that, over time, regions might sort into distinct groups or clubs, such as rich and poor. Their respective positions could then reflect disparate steady-states, or even the possibility of multiple equilibria. These ideas have been discussed repeatedly in the literature. Using Quah's methods can provide some insight, but approaches have emerged which provide more direct information on the existence of clubs and their membership.

One approach is 'bump hunting', or the use of formal statistical tests for multimodality, as in Pittau and Zelli (2006). But for many purposes, a more informative approach is to model the cross-section distribution of a regional variable as a mixture distribution. To give an example from regional economics, if regions tend to belong either to an industrialized and services-oriented urban core or to a rural, agricultural periphery, the data might be generated by a mixture distribution with two components. The data for a given region are then drawn from one component distribution with some probability, and the other component with the complementary probability; the idea generalizes readily to mixtures with more than two components. Methods for finite mixtures can be adopted to estimate characteristics of the components, such as means and variances, and also provide a probabilistic classification that can be used to assign ('fuzzily') any given region to one of the component distributions.

For investigating convergence clubs, alternatives to the mixture density approach address parameter heterogeneity in various ways. Canova (2004) is one of the first contributions along these lines. Other methods for sample-splitting include a regression tree approach as in Johnson and Takeyama (2001), or the methods for inference for threshold estimation developed by Hansen (2000). These approaches typically invoke simple parametric models estimated on subsamples, indicating the extent of parameter heterogeneity. This seems most use-

ful when the hypotheses of interest can be captured by simple regression specifications, but as we have emphasized, general equilibrium models often rule this out.

References

[1] Abadie, Alberto, Diamond, Alexis and Hainmueller, Jens (2012). Comparative Politics and the Synthetic Control Method. Manuscript, Harvard, June.

[2] Abadie, Alberto and Gardeazabal, Javier (2003). The economic costs of conflict: a case study of the Basque Country. *American Economic Review*, 93(1), 113-132.

[3] Abramovsky, Laura, Harrison, Rupert and Simpson, Helen (2007). University Research and the Location of Business R&D. *Economic Journal*, 117(519), C114-C141.

[4] Acemoglu, Daron, Cantoni, Davide, Johnson, Simon and Robinson, James A. (2011). The consequences of radical reform: the French Revolution. *American Economic Review*, 101(7), 3286-3307.

[5] Acemoglu, Daron and Dell, Melissa (2010). Productivity differences between and within countries. *American Economic Journal: Macroeconomics*, 2(1), 169-188.

[6] Acemoglu, Daron, García-Jimeno, Camilo and Robinson, James A. (2012). Finding Eldorado: Slavery and long-run development in Colombia. *Journal of Comparative Economics*, 40(4), 534-564.

[7] Acemoglu, Daron, Hassan, Tarek A. and Robinson, James A. (2011). Social Structure and Development: A Legacy of the Holocaust in Russia. *Quarterly Journal of Economics*, 126(2), 895-946.

[8] Aghion, Philippe, Burgess, Robin, Redding, Stephen J. and Zilibotti, Fabrizio (2008). The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India. *American Economic Review*, 98(4), 1397-1412.

[9] Aiello, Francesco and Scoppa, Vincenzo (2000). Uneven Regional Development in Italy: Explaining Differences in Productivity Levels. *Giornale degli Economisti e Annali di Economia*, 59(2), 270-298.

[10] Alder, Simeon, Lagakos, David and Ohanian, Lee (2013). The decline of the US Rust Belt: a macroeconomic analysis. Manuscript, January.

[11] Amiti, Mary and Cameron, Lisa (2007). Economic Geography and Wages. *Review of Economics and Statistics*, 89(1), 15-29.

[12] Anas, Alex, Arnott, Richard and Small, Kenneth A. (1998). Urban spatial structure. *Journal of Economic Literature*, 36(3), 1426-1464.

[13] Anderson, Gordon, Linton, Gordon and Leo, Teng (2012). A polarization-cohesion perspective on cross-country convergence. *Journal of Economic Growth*, 17(1), 49-69.

[14] Anselin, Luc (2001). Spatial econometrics. In Baltagi, Badi H. (ed.) *A companion to theoretical econometrics*. Blackwell, Oxford.

[15] Anselin, Luc (2006). Spatial econometrics. In Mills, Terence C. and Patterson, Kerry (eds.) *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*. Palgrave Macmillan, Basingstoke.

[16] Arkolakis, Costas, Costinot, Arnod, and Rodríguez-Clare, Andrés (2012). New Trade Models, Same Old Gains? *American Economic Review*, 102(1), 94-130.

[17] Armstrong, Harvey and Taylor, Jim (2000). *Regional economics and policy* (third edition). Blackwell, Oxford.

[18] Aten, Bettina H. (2008). Estimates of State and Metropolitan Price Parities for Consumption Goods and Services in the United States, 2005. BEA paper.

[19] Aten, Bettina H. and D'Souza, Roger J. (2008). Regional price parities: comparing price level differences across geographic areas. *Survey of Current Business*, 88, 64-74.

[20] Azzoni, Carlos R. (2001). Economic growth and regional income inequality in Brazil. *Annals of Regional Science*, 35(1), 133-152.

[21] Bairoch, Paul (1993). *Economics and world history*. Harvester Wheatsheaf, Hemel Hempstead.

[22] Baland, Jean-Marie and Robinson, James A. (2008). Land and Power: Theory and Evidence from Chile. *American Economic Review*, 98(5), 1737-65.

[23] Baldwin, Richard E. and Forslid, Rikard (2000). The core-periphery model and endogenous growth: stabilizing and destabilizing integration. *Economica*, 67(267), 307-324.

[24] Baldwin, Richard E., Forslid, Rikard, Martin, Philippe, Ottaviano, Gianmarco I. P. and Robert-Nicoud, Frederic (2003) *Economic Geography and Public Policy*. Princeton University Press, Princeton.

[25] Baldwin, Richard E. and Martin, Philippe (2004). Agglomeration and regional growth, in Henderson J. V. and Thisse J. F. (eds.), *Handbook of Regional and Urban Economics*, vol. 4, Elsevier, Amsterdam, 2671-2711.

[26] Baldwin, Richard E., Martin, Philippe and Ottaviano, Gianmarco I. P. (2001). Global income divergence, trade, and industrialization: The geography of growth take-offs. *Journal of Economic Growth*, 6(1), 5-37.

[27] Baldwin, Richard E. and Okubo, Toshihiro (2006). Heterogeneous firms, agglomeration and economic geography: Spatial selection and sorting. *Journal of Economic Geography*, 6(3), 323-346.

[28] Banerjee, Abhijit, Duflo, Esther and Qian, Nancy (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. NBER working paper no. 17897.

[29] Banerjee, Abhijit and Iyer, Lakshmi (2005). History, institutions, and economic performance: the legacy of colonial land tenure. *American Economic Review*, 95(4), 1190-1213.

[30] Barrios, Salvador and Strobl, Eric (2009). The dynamics of regional inequalities. *Regional Science and Urban Economics*, 39(5), 575-591.

[31] Barrios, Thomas, Diamond, Rebecca, Imbens, Guido W., and Kolesár, Michal (2012). Clustering, spatial correlations, and randomization inference. *Journal of the American Statistical Association*, 107:498, 578-591.

[32] Barro, Robert J. and Sala-i-Martin, Xavier (1991). Convergence across States and Regions. *Brookings Papers on Economic Activity*, 22(1), 107-182.

[33] Barro, Robert J. and Sala-i-Martin, Xavier (2004). *Economic Growth* (second edition). MIT Press, Cambridge, MA.

[34] Becker, Sascha O. and Boeckh, Katrin (2011). The Empire is dead, long live the Empire! Long-run persistence of trust and corruption in the bureaucracy. CEPR discussion paper no. 8288.

[35] Becker, Sascha O. and Woessmann, Ludger (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *Quarterly Journal of Economics*, 124(2), 531-596.

[36] Beeson, Patricia E. (1991). Amenities and Regional Differences in Returns to Worker Characteristics. *Journal of Urban Economics*, 30(2), 224-241.

[37] Behrens, Kristian and Robert-Nicoud, Frederic (2012). Survival of the fittest in cities: Agglomeration, selection and polarisation. CIRPÉE Discussion Paper no. 09-19, revised February 2012.

[38] Behrens, Kristian, Duranton, Gilles and Robert-Nicoud, Frederic (2010). Productive cities: Sorting, selection and agglomeration. CEPR Discussion Paper n. 7922.

[39] Behrens, Kristian, Mion, Giordano, Murata, Yasu and Südekum, Jens (2011). Spatial frictions. CEPR Discussion Paper no. 8572.

[40] Belot, Michèle and Ermisch, John (2009). Friendship ties and geographical mobility: evidence from Great Britain. *Journal of the Royal Statistical Society: Series A*, 172(2), 427-442.

[41] Bencivenga, Valerie R. and Smith, Bruce D. (1997). Unemployment, Migration, and Growth. *Journal of Political Economy*, 105(3), 582-608.

[42] Berkowitz, Daniel, Hoekstra, Mark and Schoors, Koen (2012). Does finance cause growth? Evidence from the origins of banking in Russia. NBER working paper no. 18139.

[43] Bernard, Andrew B. and Durlauf, Steven N. (1995). Convergence in international output. *Journal of Applied Econometrics*, 10(2), 97-108.

[44] Bernard, Andrew B. and Durlauf, Steven N. (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71(1-2), 161-173.

[45] Bernard, Andrew, Eaton, Jonathan, Jensen, J. Bradford and Kortum, Samuel (2003). Plants and productivity in international trade. *American Economic Review*, 93(4), 1268-1290.

[46] Bernard, Andrew B., Redding, Stephen J. and Schott, Peter K. (2013). Testing for Factor Price Equality with Unobserved Differences in Factor Quality or Productivity. *American Economic Journal: Microeconomics*, forthcoming.

[47] Besley, Timothy and Burgess, Robin (2000). Land Reform, Poverty Reduction, And Growth: Evidence From India. *Quarterly Journal of Economics*, 115(2), 389-430.

[48] Besley, Timothy and Burgess, Robin (2002). The Political Economy Of Government Responsiveness: Theory And Evidence From India. *Quarterly Journal of Economics*, 117(4), 1415-1451.

[49] Besley, Timothy and Burgess, Robin (2004). Can Labor Regulation Hinder Economic Performance? Evidence from India. *Quarterly Journal of Economics*, 119(1), 91-134.

[50] Black, Duncan, and Henderson, J. Vernon (1999). A theory of urban growth. *Journal of Political Economy*, 107(2), 252-284.

[51] Bleakley, Hoyt and Lin, Jeffrey (2012). Portage and Path Dependence. *Quarterly Journal of Economics*, 127(2), 587-644.

[52] Bloom David E. and Sachs, Jeffrey D. (1998). Geography, Demography, and Economic Growth in Africa. *Brookings Papers on Economic Activity*, 29(2), 207-296.

[53] Bosker, Maarten and Garretsen, Harry (2012). Economic Geography and Economic Development in Sub-Saharan Africa. *The World Bank Economic Review*, 26(3), 93-136.

[54] Boucekkine, Raouf, Camacho, Carmen and Zou, Benteng (2009). Bridging the gap between growth theory and the new economic geography: The spatial Ramsey model. *Macroeconomic Dynamics*, 13(1), 20-45.

[55] Brakman, Steven, Garretsen, Harry, and Schramm, Marc (2004a). The Spatial Distribution of Wages: Estimating the Helpman-Hanson Model for Germany. *Journal of Regional Science*, 44(3), 437-466.

[56] Brakman, Steven, Garretsen, Harry, and Schramm, Marc (2004b). The strategic bombing of German cities during World War II and its impact on city growth. *Journal of Economic Geography*, 4(2), 201-218.

[57] Brakman, Steven, Garretsen, Harry and van Marrewijk, Charles (2009). *The new introduction to geographical economics*. Cambridge University Press, Cambridge.

[58] Brandt, Loren, Tombe, Trevor, and Zhu, Xiaodong (2013). Factor market distortions across time, space and sectors in China. *Review of Economic Dynamics*, 16(1), 39-58.

[59] Breinlich, Holger (2006). The Spatial Income Structure in the European Union — What Role for Economic Geography? *Journal of Economic Geography*, 6(5), 593-617.

[60] Brito, Paulo (2004). The dynamics of growth and distribution in a spatially heterogeneous world. Working Papers, Department of Economics, ISEG, WP13/2004/DE/UECE.

[61] Brock, William and Xepapadeas, Anastasios (2008). Diffusion-induced instability and pattern formation in infinite horizon recursive optimal control. *Journal of Economic Dynamics and Control*, 32(9), 2745-2787.

[62] Brock, William and Xepapadeas, Anastasios (2009). General pattern formation in recursive dynamical systems models in economics. Fondazione Eni Enrico Mattei working paper no. 2009.49.

[63] Brülhart, Marius, Carrère, Céline and Trionfetti, Federico (2012). How wages and employment adjust to trade liberalization: Quasi-experimental evidence from Austria. *Journal of International Economics*, 86(1), 68-81.

[64] Burgess, Robin and Pande, Rohini (2005). Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment. *American Economic Review*, 95(3), 780-795.

[65] Burgess, Robin, Pande, Rohini and Wong, Grace (2005). Banking for the Poor: Evidence From India. *Journal of the European Economic Association*, 3(2-3), 268-278.

[66] Cameron, Gavin and Muellbauer, John (2000). Earnings biases in the United Kingdom Regional Accounts: some economic policy and research implications. *Economic Journal*, 110, F412-F429.

[67] Cameron, Gavin and Muellbauer, John (2001). Earnings, unemployment, and housing in Britain. *Journal of Applied Econometrics*, 16(3), 203-220.

[68] Canova, Fabio (2004). Testing for convergence clubs in income per capita: a predictive density approach. *International Economic Review*, 45(1), 49-77.

[69] Cantoni, Davide (2010). The economic effects of the Protestant Reformation: testing the Weber hypothesis in the German lands. Manuscript, Universitat Pompeu Fabra, December.

[70] Carlino, Gerald and Mills, Leonard O. (1993). Are US regional incomes converging? A time series analysis. *Journal of Monetary Economics*, 32(2), 335-346.

[71] Carlino, Gerald and Mills, Leonard O. (1996). Are US regional incomes converging? Reply. *Journal of Monetary Economics*, 38(3), 599-601.

[72] Carvalho, Vasco M. and Harvey, Andrew C. (2005). Growth, cycles and convergence in US regional time series. *International Journal of Forecasting*, 21(4), 667-686.

[73] Carvalho, Vasco M., Harvey, Andrew C. and Trimbur, Thomas (2007). A note on common cycles, common trends, and convergence. *Journal of Business & Economic Statistics*, 25(1), 12-20.

[74] Caselli, Francesco (2005). Accounting for Cross-Country Income Differences. In Aghion, Philippe and Durlauf, Steven N. (eds.) *Handbook of Economic Growth*, Vol. 1A. North-Holland, New York.

[75] Chen, Henry, Gompers, Paul, Kovner, Anna and Lerner, Josh (2010). Buy local? The geography of venture capital. *Journal of Urban Economics*, 67(1), 90-102.

[76] Chen, Xi and Nordhaus, William (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences (US)*, May 24, 108(21), 8589-8594.

[77] Clark, Gordon L., Gertler, Meric S. and Feldman, Maryann P. (eds.) (2000). *The Oxford Handbook of Economic Geography*. Oxford University Press, Oxford.

[78] Combes, Pierre-Philippe (2011). The empirics of economic geography: how to draw policy implications? *Review of World Economics*, 147(3), 567–592.

[79] Combes, Pierre-Philippe, Duranton, Gilles and Gobillon, Laurent (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2), 723-742.

[80] Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, Puga, Diego and Roux, Sébastien (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6), 2543-2594.

[81] Combes, Pierre-Philippe, Mayer, Thierry, and Thisse, Jacques-François (2008). *Economic Geography*. Princeton University Press, Princeton NJ.

[82] Conley, Timothy G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1), 1-45.

[83] Corden, W. Max and Findlay, Ronald (1975). Urban Unemployment, Intersectoral Capital Mobility and Development Policy. *Economica*, 42(165), 59-78.

[84] Córdoba, Juan Carlos (2008). On the distribution of city sizes. *Journal of Urban Economics*, 63(1), 177-197.

[85] Corrado, Luisa and Fingleton, Bernard (2012). Where is the economics in spatial econometrics? *Journal of Regional Science*, 52(2), 210-239.

[86] Cowell, Frank (2011). *Measuring inequality*. Oxford University Press, Oxford.

[87] Crespo Cuaresma, Jesús and Feldkircher, Martin (2012). Spatial filtering, model uncertainty and the speed of income convergence in Europe. *Journal of Applied Econometrics*, forthcoming.

[88] Cuñat, Alejandro, and Maffezzoli, Marco (2007). Can Comparative Advantage Explain the Growth of US Trade? *Economic Journal*, 117(520), 583-602.

[89] Dahl, Michael S. and Sorenson, Olav (2010). The migration of technical workers. *Journal of Urban Economics*, 67(1), 33-45.

[90] Davis, Donald R., and Dingel, Jonathan I. (2012). A spatial knowledge economy. NBER Working Paper No. 18188.

[91] Davis, Donald and Weinstein, David (1999). Economic Geography and Regional Production Structure: An Empirical Investigation. *European Economic Review*, 43(2), 379–407.

[92] Davis, Donald and Weinstein, David (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review*, 92(5), 1269–1289.

[93] Davis, Donald and Weinstein, David (2003). Market Access, Economic Geography and Comparative Advantage: An Empirical Assessment. *Journal of International Economics*, 59(1), 1–23.

[94] Davis, Donald and Weinstein, David (2008). A Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science*, 48(1), 29–65.

[95] Deaton, Angus and Dupriez, Olivier (2011). Spatial price differences within large countries. Manuscript, Princeton University, July.

[96] Dell, Melissa (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica*, 78(6), 1863-1903.

[97] Dell, Melissa, Jones, Benjamin and Olken, Benjamin (2009). Temperature and incpme: reconciling new cross-sectional and panel estimates. *American Economic Review*, 99(2), 198-204.

[98] Dell, Melissa, Jones, Benjamin and Olken, Benjamin (2012). Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics*, 4(3), 66-95.

[99] Démurger, Sylvie, Sachs, Jeffrey D., Woo, Wing Thye, Bao, Shuming, Chang, Gene and Mellinger, Andrew (2002). Geography, Economic Policy, and Regional Development in China, *Asian Economic Papers*, 1(1), 146-197.

[100] Desmet, Klaus, Ghani, Ejaz, O'Connell, Stephen D. and Rossi-Hansberg, Esteban (2012). The spatial development of India. World Bank policy research working paper no. 6060.

[101] Desmet, Klaus and Rossi-Hansberg, Esteban (2009). Spatial growth and industry age. *Journal of Economic Theory*, 144(6), 2477-2502.

[102] Desmet, Klaus and Rossi-Hansberg, Esteban (2010). On Spatial Dynamics. *Journal of Regional Science*, 50(1), 43-63.

[103] Desmet, Klaus and Rossi-Hansberg, Esteban (2012a). Spatial development. Manuscript; earlier version NBER Working Paper No. 15349.

[104] Desmet, Klaus and Rossi-Hansberg, Esteban (2012b). Innovation in space. *American Economic Review, Papers and Proceedings*, 102(3), 447-52.

[105] Desmet, Klaus and Rossi-Hansberg, Esteban (2013a). Urban Accounting and Welfare. *American Economic Review*, forthcoming.

[106] Desmet, Klaus and Rossi-Hansberg, Esteban (2013b). On the spatial economic impact of global warming. Manuscript, Princeton.

[107] Diamond, Jared and Robinson, James A. (2010). *Natural experiments of history*. Belknap, Harvard.

[108] Dinkelman, Taryn (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review*, 101(7), 3078-3108.

[109] Dinkelman, Taryn and Schulhofer-Wohl, Sam (2012). Migration, congestion externalities, and the evaluation of spatial investments. CEPR discussion paper no. 9126.

[110] Dixit, Avinash K., and Stiglitz, Joseph E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3), 297-308.

[111] Donaldson, Dave (2010). Railroads of the Raj: estimating the impact of transportation infrastructure. NBER working paper no. 16487. *American Economic Review*, forthcoming.

[112] Donaldson, Dave and Hornbeck, Richard (2012). Railroads and American economic growth: a 'market access' approach. Manuscript, March.

[113] Dorling, Daniel (2011). *Injustice*. Policy Press, Bristol.

[114] Drèze, Jean and Sen, Amartya (1997). *Indian Development: Selected Regional Perspectives*, WIDER Studies in Development Economics, Clarendon Press, Oxford.

[115] Duclos, Jean-Yves, Esteban, Joan and Ray, Debraj (2004). Polarization: Concepts, Measurement, Estimation. *Econometrica*, 72(6), 1737-1772.

[116] Duranton, Gilles (2007). Urban evolutions: The fast, the slow, and the still. *American Economic Review*, 97(1), 197-221.

[117] Duranton, Gilles (2011). 'California Dreamin': The Feeble Case for Cluster Policies. *Review of Economic Analysis*, 3(1), 3-45.

[118] Duranton, Gilles, Gobillon, Laurent and Overman, Henry G. (2011). Assessing the Effects of Local Taxation using Microgeographic Data. *Economic Journal*, 121(555), 1017-1046.

[119] Duranton, Gilles and Monastiriotis, Vassilis (2002). Mind the gaps: The evolution of regional earnings inequalities in the UK 1982-1997. *Journal of Regional Science*, 42(2), 219–256.

[120] Duranton, Gilles and Puga, Diego (2004). Micro-foundations of urban agglomeration economies. In Henderson, Vernon and Thisse, Jacques-François. *Handbook of Regional and Urban Economics*, Vol. 4, North-Holland, Amsterdam.

[121] Duranton, Gilles, Rodríguez-Pose, Andrés and Sandall, Richard (2009). Family types and the persistence of regional disparities in Europe. *Economic Geography*, 85(1), 23-47.

[122] Durbin, James (1960). The Fitting of Time-Series Models. *Review of the International Statistical Institute*, 28(3), 233-244.

[123] Durlauf, Steven N. (2012). Poverty traps and Appalachia. In James Ziliak (ed.) *Appalachian Legacy: Economic Opportunity After the War on Poverty*, Brookings Institution Press, Washington DC.

[124] Durlauf, Steven N., Johnson, Paul A. and Temple, Jonathan R. W. (2005). Growth Econometrics. In P. Aghion and S. N. Durlauf (eds.) *Handbook of Economic Growth, Volume 1A*, North-Holland: Amsterdam, pp. 555-677.

[125] Durlauf, Steven N., Johnson, Paul A. and Temple, Jonathan R. W. (2009). The Econometrics of Convergence. In Terence C. Mills and Kerry Patterson (eds.) *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*. Palgrave Macmillan.

[126] Eaton, Jonathan, and Eckstein, Zvi (1997). Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics*, 27(4-5), 443-474.

[127] Eaton, Jonathan and Kortum, Samuel (1999). International technology diffusion: Theory and measurement. *International Economic Review*, 40(3), 537-570.

[128] Eaton, Jonathan and Kortum. Samuel (2002). Technology, geography, and trade. *Econometrica*, 70(5), 1741-1779.

[129] Eberhardt, Markus and Teal, Francis (2011). Econometrics For Grumblers: A New Look At The Literature On Cross-Country Growth Empirics. *Journal of Economic Surveys*, 25(1), 109-155.

[130] Eeckhout, Jan (2004). Gibrat's Law for (all) cities. *American Economic Review*, 94(5), 1429-1451.

[131] Ellison, Glenn and Glaeser, Edward (1999). The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *American Economic Review: Papers and Proceedings*, 89(3), 311-316.

[132] Enflo, Kerstin Sofia (2010). Productivity and employment — is there a trade-off? Comparing Western European regions and American states 1950-2000. *Annals of Regional Science*, 45(2), 401-421.

[133] Evans, Paul (1996). Using cross-country variances to evaluate growth theories. *Journal of Economic Dynamics and Control*, 20(6-7), 1027-1049.

[134] Evans, Paul (2000). Income dynamics in regions and countries. In Hess, Gregory D. and van Wincoop, Eric (eds.) *Intranational macroeconomics*. Cambridge University Press, Cambridge.

[135] Evans, Paul and Karras, Georgios (1996). Do economies converge? Evidence from a panel of US states. *Review of Economics and Statistics*, 78(3), 384-388.

[136] Fafchamps, Marcel and Schündeln, Matthias (2013). Local financial development and firm performance: evidence from Morocco. *Journal of Development Economics*, 103, 15-28.

[137] Fally, Thibault, Paillacar, Rodrigo, and Terra, Cristina (2010). Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics*, 91(2), 155-168.

[138] Fedorov, Leonid (2002). Regional inequality and regional polarization in Russia, 1990-99. *World Development*, 30(3), 443-456.

[139] Feldman, Maryann P. (1994). *The Geography of Innovation*. Kluwer Academic Publishers, Dordrecht.

[140] Felkner, John S. and Townsend, Robert M. (2011). The geographic concentration of enterprise in developing countries. *Quarterly Journal of Economics*, 126(4), 2005-2061.

[141] Florida, Richard (2002). *The rise of the creative class*. Basic Books, New York.

[142] Fogel, Robert W. (1964). *Railroads and American economic growth: essays in economic history*. Johns Hopkins University Press, Baltimore.

[143] Fujita, Masahisa, Krugman, Paul R. and Venables, Anthony J. (1999). *The Spatial Economy: Cities, Regions and International Trade*. MIT Press, Cambridge MA.

[144] Fujita, Masahisa and Ogawa, Hideaki (1982). Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2), 161-96.

[145] Fujita, Masahisa and Thisse, Jacques-François (2002). *Economics of Agglomeration*. Cambridge University Press: Cambridge.

[146] Fujita, Masahisa and Thisse, Jacques-François (2003). Does geographical agglomeration foster economic growth? And who gains and loses from it? *Japanese Economic Review*, 54(2), 121-145.

[147] Gabaix, Xavier (1999). Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, 114(3), 739-767.

[148] Gabaix, Xavier (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3), 733-772.

[149] Gabriel, Stuart A., Mattey, Joe P. and Wascher, William L. (2003). Compensating differentials and evolution in the quality-of-life among U.S. states. *Regional Science and Urban Economics*, 33(5), 619-649.

[150] Gajwani, Kiran, Kanbur, Ravi and Zhang, Xiaobo (2006). Comparing the evolution of spatial inequality in China and India: a fifty-year perspective. DSGD discussion papers 44, International Food Policy Research Institute (IFPRI).

[151] Gallup, John Luke, Sachs, Jeffrey D. and Mellinger, Andrew D. (1999). Geography and Economic Development. *International Regional Science Review*, 22(2), 179-232.

[152] Ganong, Peter and Shoag, Daniel (2013). Why has regional income convergence in the US declined? Manuscript, Harvard.

[153] Gennaioli, Nicola, La Porta, Rafael, Lopez-de-Silanes, Florencio, and Shleifer, Andrei (2013a). Human Capital and Regional Development. *Quarterly Journal of Economics*, 128(1), 105-164.

[154] Gennaioli, Nicola, La Porta, Rafael, Lopez-de-Silanes, Florencio, and Shleifer, Andrei (2013b). Growth in regions. NBER working paper no. 18937.

[155] Gibbons, Stephen and Overman, Henry (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2), 172-191.

[156] Gibson, Edward L. (2005). Boundary control: subnational authoritarianism in democratic countries. *World Politics*, 58(1), 101-132.

[157] Gingerich, Daniel W. (2013). Governance indicators and the level of analysis problem: empirical findings from South America. *British Journal of Political Science*, 2, 1-38.

[158] Glaeser, Edward L. (2011). *Triumph of the city*. Macmillan, London.

[159] Glaeser, Edward L. and Gottlieb, Joshua D. (2009). The wealth of cities: agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, 47(4), 983-1028.

[160] Glaeser, Edward L., Kerr, Sari Pekkala, and Kerr, William R. (2012). Entrepreneurship and urban growth: an empirical assessment with historical mines. NBER working paper no. 18333.

[161] Glaeser, Edward L., Rosenthal, Stuart S. and Strange, William C. (2010). Urban economics and entrepreneurship. *Journal of Urban Economics*, 67(1), 1-14.

[162] Glaeser, Edward L. and Saks, Raven E. (2006). Corruption in America. *Journal of Public Economics*, 90(6-7), 1053-1072.

[163] Graham, Bryan S. and Temple, Jonathan R. W. (2006). Rich nations, poor nations: how much can multiple equilibria explain? *Journal of Economic Growth*, 11(1), 5-41.

[164] Grosfeld, Irena and Zhuravskaya, Ekaterina (2012). Persistent effects of empires: evidence from the partitions of Poland. Discussion paper, SSRN.

[165] Grossman, Gene and Helpman, Elhanan (1991). *Innovation and Growth in the Global Economy*. MIT Press, Cambridge MA.

[166] Guiso, Luigi, Sapienza, Paola, and Zingales, Luigi (2004). Does local financial development matter? *Quarterly Journal of Economics*, 119(3), 929-969.

[167] Guriev, Sergei and Vakulenko, Elena (2012). Convergence between Russian regions. CEFIR/NES working paper no. 180.

[168] Handbury, Jessie and Weinstein, David E. (2011). Is new economic geography right? Evidence from price data. NBER working paper no. 17067, May.

[169] Hanson, Gordon (1996). Localization Economies, Vertical Organization, and Trade. *American Economic Review*, 86(5), 1266-1278.

[170] Hanson, Gordon (1997). Increasing Returns, Trade, and the Regional Structure of Wages. *Economic Journal*, 107(440), 113-133.

[171] Hanson, Gordon (2005). Market Potential, Increasing Returns, and Geographic Concentration. *Journal of International Economics*, 67(1), 1-24.

[172] Hanson, Gordon and Xiang, Chong (2004). The Home Market Effect and Bilateral Trade Patterns. *American Economic Review*, 94(4), 1108-1129.

[173] Harari, Mariaflavia and La Ferrara, Eliana (2013). Conflict, climate and cells: a disaggregated analysis. CEPR discussion paper no. 9277.

[174] Harris, Chauncy (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 44(2), 315-348.

[175] Harris, Richard (2010). Models of regional growth: past, present and future. *Journal of Economic Surveys*, 25(5), 913-951.

[176] Hassler, John and Krusell, Per (2012). Economics and Climate Change: Integrated Assessment in a Multi-Region World. Manuscript, IIES, Stockholm.

[177] Head, Keith and Mayer, Thierry (2004). Empirics of Agglomeration and Trade. In Henderson, Vernon and Thisse, Jacques-François (eds.) *Handbook of Regional and Urban Economics*, Vol. 4, North-Holland, Amsterdam.

[178] Head, Keith and Mayer, Thierry (2006). Regional Wage and Employment Responses to Market Potential in the EU. *Regional Science and Urban Economics*, 36(5), 573–595.

[179] Head, Keith and Mayer, Thierry (2010). Gravity, market potential and economic development. *Journal of Economic Geography*, 10(1), 1–14.

[180] Head, Keith and Ries, John (2001). Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade. *American Economic Review*, 91(4), 858-876.

[181] Helpman, Elhanan (1998). The size of regions. In D. Pines, E. Sadka and I. Zilcha (eds.) *Topics in Public Economics: Theoretical and Applied Analysis*, Cambridge University Press, Cambridge, 33-54.

[182] Helpman, Elhanan, and Krugman, Paul R. (1985). *Market Structure and International Trade*. MIT Press, Cambridge MA.

[183] Henderson, J. Vernon, Storeygard, Adam and Weil, David N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2), 994-1028.

[184] Henderson, J. Vernon and Wang, Hyoung Gun (2005). Aspects of the rural-urban transformation of countries. *Journal of Economic Geography*, 5(1), 23-42.

[185] Herbst, Jeffrey (2000). *States and power in Africa.* Princeton University Press, Princeton.

[186] Hering, Laura and Poncet, Sandra (2009). The impact of economic geography on wages: Disentangling the channels of influence. *China Economic Review*, 20(1), 1-14.

[187] Hering, Laura and Poncet, Sandra (2010). Market Access Impact on Individual Wages: Evidence from China. *Review of Economics and Statistics*, 92(1), 145-159.

[188] Herrendorf, Berthold, Schmitz Jr., James A. and Teixeira, Arilton (2012). The Role of Transportation in U.S. Economic Development: 1840-1860. *International Economic Review*, 53(3), 693-715.

[189] Hill, Hal, Resosudarmo, Budy P. and Vidyattama, Yogi (2008). Indonesia's changing economic geography. *Bulletin of Indonesian Economic Studies*, 44(3), 407-435.

[190] Hirschman, Albert O. (1958). *The strategy of economic development*. Yale University Press, New Haven.

[191] Hobsbawm, Eric J. (1962). *The Age of Capital*. Weidenfeld and Nicolson, London.

[192] Hodler, Roland and Raschky, Paul A. (2010). Foreign aid and enlightened leaders. Monash discussion paper no. 54/10.

[193] Holmes, Thomas J. (1998). The effect of state policies on the location of manufacturing: evidence from state borders. *Journal of Political Economy*, 106(4), 667-705.

[194] Holmes, Thomas J. (2010). Structural, experimentalist, and descriptive approaches to empirical work in regional economics. *Journal of Regional Science*, 50(1), 5-22.

[195] Holz, Carsten A. (2009). No Razor's Edge: Reexamining Alwyn Young's evidence for increasing interprovincial trade barriers in China. *Review of Economics and Statistics*, 91(3), 599-616.

[196] Hornbeck, Richard (2012a). The enduring impact of the American Dust Bowl: short- and long-run adjustments to the environmental catastrophe. *American Economic Review*, 102(4), 1477-1507.

[197] Hornbeck, Richard (2012b). Nature versus Nurture: The Environment's Persistent Influence through the Modernization of American Agriculture. *American Economic Review: Papers & Proceedings*, 102(3), 245-249.

[198] Hsieh, Chang-Tai and Klenow, Peter (2010). Development Accounting. *American Economic Journal: Macroeconomics*, 2(1), 207-223.

[199] Huang, Rocco R. (2008). Evaluating the real effect of bank branching deregulation: comparing contiguous counties across US state borders. *Journal of Financial Economics*, 87(3), 678-705.

[200] Isard, Walter, Azis, Iwan J., Drennan, Matthew P., Miller, Ronald E., Saltzman, Sidney and Thorbecke, Erik (1998). *Methods of interregional and regional analysis*. Ashgate, Aldershot.

[201] Iyer, Lakshmi (2010). Direct versus indirect colonial rule in India: long-term consequences. *Review of Economics and Statistics*, 92(4), 693-713.

[202] Jayaratne, Jith and Strahan, Philip E. (1996). The finance-growth nexus: evidence from bank branch deregulation. *Quarterly Journal of Economics*, 111(3), 639-670.

[203] Johnson, Paul A. (2000). A nonparametric analysis of income convergence across the US states. *Economics Letters*, 69(2), 219-223.

[204] Johnson, Paul A. (2005). A continuous state space approach to 'Convergence by Parts'. *Economics Letters*, 86(3), 317-321.

[205] Johnson, Paul A. and Takeyama, Lisa N. (2001). Initial conditions and economic growth in the US states. *European Economic Review*, 45(4-6), 919-927.

[206] Judt, Tony (1996). *A Grand Illusion? An essay on Europe*. Hill and Wang.

[207] Kaldor, Nicholas (1970). The case for regional policies. *Scottish Journal of Political Economy*, 17, 337-348. Reprinted in F. Targetti and A. P. Thirlwall (eds.) (1989). *The Essential Kaldor*. Duckworth, London.

[208] Kanbur, Ravi and Rapoport, Hillel (2005). Migration selectivity and the evolution of spatial inequality. *Journal of Economic Geography*, 5(1), 43-57.

[209] Kanbur, Ravi and Venables, Anthony J. (2005). *Spatial inequality and development*. UNU-WIDER and Oxford University Press, Oxford.

[210] Kaplan, Robert D. (2012). *The revenge of geography*. Random House, New York.

[211] Kelejian, Harry H. and Prucha, Ingmar R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1), 131-154.

[212] Klepper, Steven (2010). The origin and growth of industry clusters: The making of Silicon Valley and Detroit. *Journal of Urban Economics*, 67(1), 15-32.

[213] Klepper, Steven (2011). Nano-economics, spinoffs, and the wealth of regions. *Small Business Economics*, 37(2), 141-154.

[214] Kline, Patrick and Moretti, Enrico (2013). Place based policies with unemployment. *American Economic Review*, 103(3), 238-243.

[215] Knight, John B. and Gunatilaka, Ramani (2011). Does economic growth raise happiness in China? *Oxford Development Studies*, 39(1), 1-24.

[216] Kolko, Jed and Neumark, David (2010). Does local business ownership insulate cities from economic shocks? *Journal of Urban Economics*, 67(1), 103-115.

[217] Kongsamut, Piyabha, Rebelo, Sergio and Xie, Danyang (2001). Beyond Balanced Growth. *Review of Economic Studies*, 68(4), 869-882.

[218] Kremer, Michael, Alexei Onatski and James Stock (2001). Searching for prosperity, *Carnegie-Rochester Conference Series on Public Policy*, 55(1), 275-303.

[219] Krugman, Paul (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3), 483-499.

[220] Krugman, Paul (1995). *Development, Geography, and Economic Theory*. MIT Press, Cambridge, MA.

[221] Krugman, Paul R. and Venables, Anthony J. (1995). Globalization and the inequality of nations. *Quarterly Journal of Economics*, 110(4), 857-880.

[222] Krusell, Per and Smith, Anthony A. (2009). Macroeconomics and Global Climate Change: Transition for a Many-Region Economy. Manuscript, IIES, Stockholm.

[223] Kwiatkowski, Denis, Phillips, Peter C. B., Schmidt, Peter and Shin, Yongcheol (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3), 159-178.

[224] Lagakos, David and Waugh, Michael (2013). Selection, Agriculture, and Cross-Country Productivity Differences. *American Economic Review*, 103(2), 948-980.

[225] Le Pen, Yannick (2011). A pair-wise approach to output convergence between European regions. *Economic Modelling*, 28(3), 955-964.

[226] LeSage, James P. and Fischer, Manfred (2008). Spatial Growth Regressions: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis*, 3(3), 275-304.

[227] Lessmann, Christian (2011). Spatial inequality and development - is there an inverted-U relationship? CESifo working paper no. 3622.

[228] Lessmann, Christian (2013). Foreign direct investment and regional inequality: a panel data analysis. *China Economic Review*, 24, 129-149.

[229] Lipscomb, Molly, Mobarak, A. Mushfiq, and Barham, Tania (2013). Development effects of electrification: evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics*, 5(2), 200-231.

[230] Lipton, Michael (1980). Migration from rural areas of poor countries: The impact on rural productivity and income distribution. *World Development*, 8(1), 1-24.

[231] Lucas, Robert E., Jr., (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics*, 9(2), 508-523.

[232] Lucas, Robert E. Jr. and Rossi-Hansberg, Esteban (2002). On the internal structure of cities. *Econometrica*, 70(4), 1445-1476.

[233] Magrini, Stefano (2004). Regional (di)convergence. In Henderson, J. Vernon and Thisse, Jacques-François (eds.) *Handbook of Regional and Urban Economics*, volume 4, North-Holland, Amsterdam, 2741-2796.

[234] Mankiw, N. Gregory, Romer, David and Weil, David N. (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics*, 107(2), 407-37.

[235] Marshall, Alfred (1920). *Principles of Economics*, 8th edition. Macmillan, London.

[236] Martin, Philippe and Ottaviano, Gianmarco I. P. (1999). Growing locations: Industry location in a model of endogenous growth. *European Economic Review*, 43(2), 281-302.

[237] Martin, Philippe and Ottaviano, Gianmarco I. P. (2001). Growth and agglomeration. *International Economic Review*, 42(4), 947-968.

[238] Martin, Ron and Sunley, Peter (2006). Path dependence and regional economic evolution. *Journal of Economic Geography*, 6(4), 395-437.

[239] Melitz, Marc (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695-1725.

[240] Melitz, Marc and Ottaviano, Gianmarco I. P. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75(1), 295-316.

[241] Melitz, Marc, and Redding, Stephen (2013). Firm Heterogeneity and Aggregate Welfare. NBER working paper no. 18919.

[242] Mello, Marcelo (2011). Stochastic convergence across US states. *Macroeconomic Dynamics*, 15(2), 160-183.

[243] Michaels, Guy, Rauch, Ferdinand and Redding, Stephen J. (2012). Urbanization and Structural Transformation. *Quarterly Journal of Economics*, 127(2), 535-586.

[244] Michaels, Guy (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *Review of Economics and Statistics*, 90(4), 683-701.

[245] Michalopouklos, Stelios and Papaioannou, Elias (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica*, 81(1), 113-152.

[246] Miguel, Edward and Roland, Gérard (2011). The long-run impact of bombing Vietnam. *Journal of Development Economics*, 96(1), 1-15.

[247] Milanovic, Branko (2005a). *Worlds Apart*. Princeton University Press, Princeton.

[248] Milanovic, Branko (2005b). Half a world: regional inequality in five great federations. *Journal of the Asia Pacific Economy*, 10(4), 408-445.

[249] Minerva, G. Alfredo and Ottaviano, Gianmarco I. P. (2009). Endogenous growth theories: Agglomeration benefits and transportation costs. In Capello, Roberta and Nijkamp, Peter (eds.) *Handbook of Regional Growth and Development Theories*. Edward Elgar, Cheltenham.

[250] Mitchener, Kris James and McLean, Ian W. (1999). U.S. Regional Growth And Convergence, 1880-1980. *Journal of Economic History*, 59(4), 1016-1042.

[251] Mitchener, Kris James and McLean, Ian W. (2003). The productivity of US states since 1880. *Journal of Economic Growth*, 8(1), 73-114.

[252] Mion, Giordano (2004). Spatial externalities and empirical analysis: the case of Italy. *Journal of Urban Economics*, 56(1), 97-118.

[253] Mitton, Todd (2013). The wealth of subnations: geography, institutions and within-country development. Manuscript, Brigham Young University.

[254] Moene, Karl Ove (1988). A reformulation of the Harris-Todaro mechanism with endogenous wages. *Economics Letters*, 27(4), 387-390.

[255] Morelli, Massimo and Rohner, Dominic (2010). Natural resource distribution and multiple forms of civil war. University of Zurich working paper no. 498.

[256] Moretti, Enrico (2004). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1-2), 175-212.

[257] Moretti, Enrico (2010). Local multipliers. *American Economic Review: Papers and Proceedings*, 100(2), 373-377.

[258] Moretti, Enrico (2012). *The new geography of jobs*. Houghton Mifflin Harcourt, New York.

[259] Murata, Yasusada (2008). Engel's law, Petty's law and agglomeration. *Journal of Development Economics*, 87(1), 161-177.

[260] Myrdal, Gunnar (1957). *Economic theory and under-developed regions*. Duckworth, London.

[261] Naritomi, Joana, Soares, Rodrigo R. and Assunção, Juliano J. (2012). Institutional Development and Colonial Heritage within Brazil. *Journal of Economic History*, 72(2), 393-422.

[262] Neary, J. Peter (2001). Of Hype and Hyperbolas: Introducing the New Economic Geography. *Journal of Economic Literature*, 39(2), 536-561.

[263] Ng, Serena (2008). A simple test for nonstationarity in mixed panels. *Journal of Business and Economic Statistics*, 26(1), 113-127.

[264] Ngai, L. Rachel and Pissarides, Christopher A. (2007). Structural Change in a Multisector Model of Growth. *American Economic Review*, 97(1), 429-443.

[265] Nguyen, Binh T., Albrecht, James W., Vroman, Susan B., and Westbrook, M. Daniel (2007). A quantile regression decomposition of urban-rural inequality in Vietnam. *Journal of Development Economics*, 83(2), 466-490.

[266] Nocco, Antonella (2005). The rise and fall of regional inequalities with technological differences and knowledge spillovers. *Regional Science and Urban Economics*, 35(5), 542-569.

[267] Nocke, Volker (2006). A gap for me: Entrepreneurs and entry. *Journal of the European Economic Association*, 4(5), 929-956.

[268] Nordhaus, William (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences*, 103(10), 3510-3517.

[269] Nunn, Nathan (2007). Relationship-Specificity, Incomplete Contracts and the Pattern of Trade. *Quarterly Journal of Economics*, 122(2), 569-600.

[270] Nunn, Nathan and Puga, Diego (2012). Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics*, 94(1), 20-36.

[271] Okubo, Toshihiro, Picard, Pierre and Thisse, Jacques-François (2010). The spatial selection of heterogeneous firms. *Journal of International Economics*, 82(2), 230-237.

[272] Olivetti, Claudia and Paserman, M. Daniele (2013). In the name of the son (and the daughter): intergenerational mobility in the United States, 1850-1930. CEPR discussion paper no. 9372.

[273] Oswald, Andrew J. and Wu, Stephen (2010). Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A, *Science*, 29 January, 327 (5965), 576-579.

[274] Oswald, Andrew J. and Wu, Stephen (2011). Well-Being across America. *Review of Economics and Statistics*, 93(4), 1118-1134.

[275] Ottaviano, Gianmarco I. P. (2012). Agglomeration, trade and selection. *Regional Science and Urban Economics*, 42(6), 905-1068.

[276] Ottaviano, Gianmarco I. P. and Thisse, Jacques-François (2001). On economic geography in economic theory: increasing returns and pecuniary externalities. *Journal of Economic Geography*, 1(2), 153-179.

[277] Ottaviano, Gianmarco I. P. and Thisse, Jacques-François (2002). Integration, agglomeration and the political economics of factor mobility. *Journal of Public Economics*, 83(3), 429-456.

[278] Ottaviano, Gianmarco I. P. and Thisse, Jacques-François (2004). Agglomeration and economic geography. In Henderson, J. Vernon and Thisse Jacques-Françpos (eds.) *Handbook of Regional and Urban Economics*, Vol. 4, North-Holland, Amsterdam.

[279] Ottaviano, Gianmarco I. P. and Thisse, Jacques-François (2005). New Economic Geography: what about the N? *Environment and Planning A*, 37(10), 1707-1725.

[280] Ottaviano, Gianmarco I. P., Tabuchi, Takatoshi and Thisse, Jacques-François (2002) Agglomeration and trade revisited, *International Economic Review*, 43(2), 409-435.

[281] Overman, Henry G. (2010). 'Gis a job': what use geographical information systems in spatial economics? *Journal of Regional Science*, 50(1), 165-180.

[282] Overman, Henry G. and Puga, Diego (2002). Regional unemployment clusters. *Economic Policy*, 34, 115-143.

[283] Overman, Henry G., Rice, Patricia and Venables, Anthony J. (2010). Economic linkages across space. *Regional Studies*, 44(1), 17-33.

[284] Pesaran, M. Hashem, Shin, Yongcheol and Smith, Ron P. (1999). Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association*, 94(446), 621-634.

[285] Pesaran, M. Hashem, and Smith, Ron (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1), 79-113.

[286] Phan, Diep and Coxhead, Ian (2010). Inter-provincial migration and inequality during Vietnam's transition. *Journal of Development Economics*, 91(1), 100-112.

[287] Picard, Pierre M. and Okubo, Toshihiro (2012). Firms' locations under demand heterogeneity. *Regional Science and Urban Economics*, 42(6), 961-974.

[288] Pittau, Maria Grazia (2005). Fitting regional income distributions in the European Union. *Oxford Bulletin of Economics and Statistics*, 67(2), 135-161.

[289] Pittau, Maria Grazia and Zelli, Roberto (2006). Empirical evidence of income dynamics across EU regions. *Journal of Applied Econometrics*, 21(5), 605-628.

[290] Pittau, Maria Grazia, Zelli, Roberto and Gelman, Andrew (2010). Economic disparities and life satisfaction in European regions. *Social Indicators Research*, 96(2), 339-361.

[291] Prager, Jean-Claude and Thisse, Jacques-François (2012). *Economic geography and the unequal development of regions*. Routledge, London.

[292] Puga, Diego (1999). The rise and fall of regional inequalities. *European Economic Review*, 43(2), 303-334.

[293] Putnam, Robert D. and others (1993). *Making democracy work: civic traditions in modern Italy*. Princeton University Press.

[294] Pyke, Frank, Becattini, Bruno and Sengenberger, Werner (1990). *Industrial Districts and Inter-firm Cooperation in Italy*. International Institute for Labour Studies, Geneva.

[295] Quah, Danny (1993). Empirical cross-section dynamics in economic growth. *European Economic Review*, 37(2-3), 426-434.

[296] Quah, Danny T. (1996). Regional convergence clusters across Europe. *European Economic Review*, 40(3-5), 951-958.

[297] Quah, Danny T. (1997). Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs. *Journal of Economic Growth*, 2(1), 27-59.

[298] Rappaport, Jordan (2007). Moving to Nice Weather. *Regional Science and Urban Economics*, 37(3), 375-398.

[299] Rappaport, Jordan and Sachs, Jeffrey (2003). The United States as a coastal nation. *Journal of Economic Growth*, 8(1), 5-46.

[300] Redding, Stephen (2010). The Empirics of New Economic Geography. *Journal of Regional Science*, 50(1), 297-311.

[301] Redding, Stephen J. (2012). Goods trade, factor mobility and welfare. NBER working paper no. 18008.

[302] Redding, Stephen and Schott, Peter (2003). Distance, skill deepening and development: will peripheral countries ever get rich? *Journal of Development Economics*, 72(2), 515-541.

[303] Redding, Stephen and Sturm, Daniel (2008). The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review*, 98(5), 1766-1797.

[304] Redding, Stephen and Venables, Anthony (2004). Economic Geography and International Inequality. *Journal of International Economics*, 62(1), 53-82.

[305] Rice, Patricia and Venables, Anthony J. (2003). Equilibrium regional disparities: theory and British evidence. *Regional Studies*, 37(6), 675-686.

[306] Rice, Patricia, Venables, Anthony J. and Pattachini, Eleonora (2006). Spatial determinants of productivity: analysis for the regions of Great Britain. *Regional Science and Urban Economics*, 36(6), 727-752.

[307] Roback, Jennifer (1982). Wages, Rents, and the Quality of Life. *Journal of Political Economy*, 90(6), 1257-1278.

[308] Rosenthal, Stuart S., and Strange, William C. (2004) Evidence on the nature and sources of agglomeration economies. In Henderson, J. Vernon and Thisse, Jacques-François (eds.) *Handbook of Regional and Urban Economics*, Volume 4, North-Holland, Amsterdam, 2119-2171.

[309] Rossi-Hansberg, Esteban (2005). A spatial theory of trade. *American Economic Review*, 95(5), 1464-1491.

[310] Rossi-Hansberg, Esteban and Wright, Mark (2007). Urban structure and growth, *Review of Economic Studies*, 74(2), 597-624.

[311] Rud, Juan Pablo (2012). Electricity provision and industrial development: evidence from India. *Journal of Development Economics*, 97(2), 352-367.

[312] Sachs, Jeffrey D. and Malaney, Pia (2002). The economic and social burden of malaria. *Nature*, 415(6872), 680-685.

[313] Sala-i-Martin, Xavier (1996). Regional cohesion: evidence and theories of regional growth and convergence. *European Economic Review*, 40(6), 1325-1352.

[314] Sarafidis, Vasilis, and Wansbeek, Tom (2012) Cross-sectional Dependence in Panel Data Analysis. *Econometric Reviews*, 31(5), 483-531.

[315] Satchi, Mathan and Temple, Jonathan R. W. (2009). Labor markets and productivity in developing countries. *Review of Economic Dynamics*, January, 12(1), 183-204.

[316] Scoppa, Vincenzo (2007). Quality of Human and Physical Capital and Technological Gaps across Italian Regions. *Regional Studies*, 41(5), 585–599.

[317] Scotchmer, Suzanne, and Thisse, Jacques-François (1992). Space and competition: a puzzle. *Annals of Regional Science*, 26(3), 269-286.

[318] Scott, James (2009). *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. Yale University Press, New Haven.

[319] Sen, Amartya (1980). Equality of What? In S. McMurrin (ed.) *Tanner Lectures on Human Values, Volume 1*. Cambridge University Press, Cambridge.

[320] Sennett, Richard (1998). *The corrosion of character*. New York, Norton.

[321] Serra, Maria Isabel, Fernanda Pazmino, Maria, Lindow, Genevieve, Sutton, Bennett, and Ramirez, Gustavo (2006). Regional convergence in Latin America. IMF working paper no. 06/125.

[322] Seya, Hajime, Tsutsumi, Morito, and Yamagata, Yoshiki (2012). Income convergence in Japan: a Bayesian spatial Durbin model approach. *Economic Modelling*, 29(1), 60-71.

[323] Shioji, Etsuro (2001). Public capital and economic growth: a convergence approach. *Journal of Economic Growth*, 6(3), 205-227.

[324] Skoufias, Emmanuel and Katayama, Roy S. (2011). Sources of welfare disparities between and within regions of Brazil: evidence from the 2002-2003 household budget survey (POF). *Journal of Economic Geography*, 11(5), 897-918.

[325] Solnit, Rebecca (2013). Diary. *London Review of Books*, 35(3), 7 February, 34-35.

[326] Starrett, David (1978). Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, 17(1), 21-37.

[327] Tabellini, Guido (2010). Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association*, 8(4), 677-716.

[328] Tamura, Robert (2012). Development Accounting and Convergence for US States. Manuscript, Clemson University.

[329] Temple, Jonathan (1999). The New Growth Evidence. *Journal of Economic Literature*, 37(1), 112-156.

[330] Temple, Jonathan R. W. (2005). Dual economy models: a primer for growth economists. *The Manchester School*, 73(4), 435-478.

[331] Temple, Jonathan and Woessmann, Ludger (2006). Dualism and cross-country growth regressions. *Journal of Economic Growth*, 11(3), 187-228.

[332] Trivedi, Kamakshya (2006). Educational human capital and levels of income: Evidence from states in India, 1965-92. *Journal of Development Studies*, 42(8), 1350-1378.

[333] Turner, Chad, Tamura, Robert, Mulholland, Sean E., and Baier, Scott (2007). Education and income of the states of the United States: 1840Ű2000. *Journal of Economic Growth*, 12(2), 101-158.

[334] Van Nieuwerburgh, Stijn and Weill, Pierre-Olivier (2010). Why has house price dispersion gone up? *Review of Economic Studies*, 77(4), 1567-1606.

[335] Venables, Anthony J. (1996). Equilibrium locations of vertically linked industries. *International Economic Review*, 37(2), 341-359.

[336] Venables, Anthony J. (2005). Spatial disparities in developing countries: cities, regions, and international trade. *Journal of Economic Geography*, 5(1), 3-21.

[337] Ventura, Jaume (1997). Growth and interdependence. *Quarterly Journal of Economics*, 112(1), 57-84.

[338] Walz, Uwe (1996). Transport costs, intermediate goods, and localized growth, *Regional Science and Urban Economics*, 26(6), 671-695.

[339] Warner, Andrew (2002). Institutions, geography, regions, countries and the mobility bias. CID working paper no. 91, Harvard.

[340] Weeks, Melvyn and Yao, James Yudong (2003). Provincial Conditional Income Convergence in China, 1953-1997: a Panel Data Approach. *Econometric Reviews*, 22(1), 59-77.

[341] Wei, Kailei, Yao, Shujie, and Liu, Aying (2009). Foreign direct investment and regional inequality in China. *Review of Development Economics*, 13(4), 778-791.

[342] Williamson, Jeffrey G. (1965). Regional inequality and the process of national development: a description of the patterns. *Economic Development and Cultural Change*, 13(4), 1-84.

[343] Williamson, Jeffrey G. (1974). *Late nineteenth-century American development: a general equilibrium history*. Cambridge University Press, Cambridge.

[344] Williamson, Jeffrey G. (2006). *Globalization and the poor periphery before 1950*. MIT Press, Cambridge, MA.

[345] Yang, Dennis Tao and Zhu, Xiaodong (2013). Modernization of agriculture and long-term growth. *Journal of Monetary Economics*, forthcoming.

[346] Yoon, Chamna (2013). The decline of the Rust Belt: a dynamic spatial equilibrium analysis. Manuscript, University of Pennsylvania.

[347] Young, Alwyn (1991). Learning by doing and the dynamic effects of international trade. *Quarterly Journal of Economics*, 106(2), 369-405.

[348] Young, Alwyn (2000). The razor's edge: distortions and incremental reform in the People's Republic of China. *Quarterly Journal of Economics*, 115(4), 1091-1135.

[349] Young, Alwyn (2013). Inequality, the Urban-Rural Gap and Migration. Manuscript, LSE.

[350] Zhang, Xiaobo and Kanbur, Ravi (2001). What Difference Do Polarisation Measures Make? An Application to China. *Journal of Development Studies*, 37(3), 85-98.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1231	Michalis Drouvelis Nattavudh Powdthavee	Are Happier People Less Judgmental of Other People's Selfish Behaviors? Laboratory Evidence from Trust and Gift Exchange Games
1230	Dan Anderberg Helmut Rainer Jonathan Wadsworth Tanya Wilson	Unemployment and Domestic Violence: Theory and Evidence
1229	Hannes Schwandt	Unmet Aspirations as an Explanation for the Age U-Shape in Human Wellbeing
1228	Bénédicte Apouey Andrew E. Clark	Winning Big But Feeling No Better? The Effect of Lottery Prizes on Physical and Mental Health
1227	Alex Gyani Roz Shafran Richard Layard David M Clark	Enhancing Recovery Rates: Lessons from Year One of the English Improving Access to Psychological Therapies Programme
1226	Stephen Gibbons Sandra McNally	The Effects of Resources Across School Phases: A Summary of Recent Evidence
1225	Cornelius A. Rietveld David Cesarini Daniel J. Benjamin Philipp D. Koellinger Jan-Emmanuel De Neve Henning Tiemeier Magnus Johannesson Patrik K.E. Magnusson Nancy L. Pedersen Robert F. Krueger Meike Bartels	Molecular Genetics and Subjective Well-Being
1224	Peter Arcidiacono Esteban Aucejo Patrick Coate V. Joseph Hotz	Affirmative Action and University Fit: Evidence from Proposition 209
1223	Peter Arcidiacono Esteban Aucejo V. Joseph Hotz	University Differences in the Graduation of Minorities in STEM Fields: Evidence from California
1222	Paul Dolan Robert Metcalfe	Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation

1221	Andy Feng Georg Graetz	A Question of Degree: The Effects of Degree Class on Labor Market Outcomes
1220	Esteban Aucejo	Explaining Cross-Racial Differences in the Educational Gender Gap
1219	Peter Arcidiacono Esteban Aucejo Andrew Hussey Kenneth Spenner	Racial Segregation Patterns in Selective Universities
1218	Silvana Tenreyro Gregory Thwaites	Pushing On a String: US Monetary Policy is Less Powerful in Recessions
1217	Gianluca Benigno Luca Fornaro	The Financial Resource Curse
1216	Daron Acemoglu Ufuk Akcigit Nicholas Bloom William R. Kerr	Innovation, Reallocation and Growth
1215	Michael J. Boehm	Has Job Polarization Squeezed the Middle Class? Evidence from the Allocation of Talents
1214	Nattavudh Powdthavee Warn N. Lekfuangfu Mark Wooden	The Marginal Income Effect of Education on Happiness: Estimating the Direct and Indirect Effects of Compulsory Schooling on Well-Being in Australia
1213	Richard Layard	Mental Health: The New Frontier for Labour Economics
1212	Francesco Caselli Massimo Morelli Dominic Rohner	The Geography of Inter-State Resource Wars
1211	Stephen Hansen Michael McMahon	Estimating Bayesian Decision Problems with Heterogeneous Priors
1210	Christopher A. Pissarides	Unemployment in the Great Recession
1209	Kevin D. Sheedy	Debt and Incomplete Financial Markets: A Case for Nominal GDP Targeting
1208	Jordi Blanes i Vidal Marc Möller	Decision-Making and Implementation in Teams

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7404 0612
Email info@cep.lse.ac.uk Web site http://cep.lse.ac.uk