CENTRE *for* ECONOMIC PERFORMANCE

CEP Discussion Paper No 785

April 2007

Productivity Growth, Knowledge Flows and Spillovers

Gustavo Crespi, Chiara Criscuolo, Jonathan E. Haskel and Matthew Slaughter

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

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

Abstract

This paper explores the role of knowledge flows and TFP growth by using direct survey data on knowledge flows linked to firm-level TFP growth data. Our knowledge flow data correspond to the kind of information flows often argued, especially by policy-makers, as important, such as within the firm, or from suppliers, purchasers, universities and competitors. We examine three questions (a) What is the source of knowledge flows? (b) To what extent do such flows contribute to productivity growth? (c) Do such flows constitute a spillover flow of free knowledge? Our evidence show that the main sources of knowledge are competitors; suppliers; plants that belong to the same group and universities. We conclude that the main "free" information flow spillover is from competitors and that multi-national presence may be a proximate source of this spillover.

Keywords: business services; structural change; economic growth; productivity
JEL Classifications: O11, M2,

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

Acknowledgements

Financial support for this research comes from the ESRC/EPSRC Advanced Institute of Management Research, grant number RES-331-25-0030 which is carried out at CeRiBA at the Business Data Linking Branch at the ONS; we are grateful to all institutions concerned for their support. This work contains statistical data from ONS which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. We thank the BDL team at ONS as usual for all their help with computing and facilitating research and Ray Lambert and Helen Simpson for help with CIS3. Errors are of course our own.

Gustavo Crespi is a Research Fellow at SPRU, University of Sussex and CeRiBA, Queen Mary College, University of London. Chiara Criscuolo is a Research Fellow at the Centre for Economic Performance, London School of Economics. Jonathan Haskel is Professor of Economics at Queen Mary's, University of London. Matthew Slaughter is an Associate Professor of Business Administration at Tuck Business School, Dartmouth College (USA).

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

ISBN 978 0 85328 160 3

1 Introduction

Productivity growth is generally modelled as due to increases in physical inputs and knowledge inputs. Increases in knowledge are then generally ascribed to three main sources (i) investment in new knowledge within the firm (e.g. R&D), (ii) use of existing knowledge from within the firm (e.g. from past discoveries or knowledge-sharing with other divisions of the firm) (iii) use of knowledge from outside the firms.

There are then at least three major questions concerning these outside knowledge flows.[1] First, what is their source? A substantial literature in IO studies the extent to which citations to other patents are "local", in terms of geographic or technical distance. A substantial literature in international economics studies the possibility that information might come from nearby multi-nationals or external trade. Second, to what extent do such flows contribute to productivity growth? Just as growth accounting seeks to account for productivity growth due to physical inputs, how much do knowledge flows raise productivity growth? Third, do such flows constitute a spillover flow of free knowledge? This final question is central for setting subsidies for R&D and multi-nationals both of which are major public policy issues in many countries.

Despite the importance of these questions it is well-acknowledged that their answers are not fully worked out. The main problem is that it is of course very hard to measure knowledge flows across firms. There are two main methods: direct and indirect. The main *direct* method is to use the information in patent citations. As is well-acknowledged, however, this literature suffers from some difficulties. One problem is that patents likely measure a selected form of knowledge increase, since not all innovations are patentable, neither are all patentable innovations chosen to be patented[2]. Bloom and Van Reenen (2002) report for example, that in their sample of 59,919 U.K. firms, just 12 companies accounted for 72% of all patents. Of course, knowledge flows as embodied in patents might be key information flows within the economy, and contribute to understanding frontier innovations. But it seems at least of some interest to assemble data on what knowledge flows look like in non-patenting firms, who constitute the vast majority of companies.[3]

The *indirect* method is typically to regress TFP growth on some factors thought to be potentially causing information flows such as the presence of MNEs or trading/exporting status. This method does have the advantage of attempting to test a number of effects that case studies suggest are important and it

---

[1] Similar questions apply to internal flows, but these are perhaps generally less emphasised.

[2] Jaffe and Trajtenberg (2002, p.3) say: "There are, of course, important limitations to the use of patent data, the most glaring being the fact that not all inventions are patented ….". Cohen, Nelson and Walsh (2000) report a survey of 1,478 R&D labs in US manufacturing that found that other methods of protecting intellectual property, such as secrecy, lead time and marketing were ranked higher than patents.

[3] Another problem is that recent research has indicated that as many as 50% of patent citations have been included by the examining officers rather than the inventor themselves. This makes patent citations a noisy measure of information flows.

does try to speak to a number of prominent policy issues. For example, many economies the world over give subsidies to multi-nationals and to exporting firms. In turn, a number of case studies have been used as evidence to suggest these might be justified by TFP–enhancing information flows from local multi-nationals or contact with other markets following exporting. However (quite apart of issues of measurement, simultaneity etc.) the problem with evidence based on TFP is that it is indirect and is consistent with information flows, but likely with other effects as well.[4]

This paper attempts to bring new evidence to bear on this question, by using direct data on knowledge flows linked to TFP growth data. The TFP growth data is derived from firm-level business surveys conducted by the UK *Office of National Statistics* in compiling the national accounts. The data have a high response rate, are collected annually and give data on labour, materials and capital inputs and relevant cost shares. This minimises at least the problems of measuring TFP growth with non-response and recall bias. Of course such measures are noisy and we shall explore robustness issues extensively.

The knowledge flow data is derived from linking these data to the *Community Innovation Survey*, an official EU-wide survey that asks business enterprises to report innovation outputs; innovation inputs; and, most importantly for our paper, sources of knowledge for innovation efforts. We use the second and third waves of the CIS and the panel contained therein.

For our purposes the major feature of the CIS is that it asks firms about their R&D and also their knowledge flows.[5] Firms are asked to rate the importance of knowledge flows for innovations from a number of sources such as suppliers, other firms in the firm group, customers, universities etc. The major advantage of these data is they have the advantage of trying to directly measure a number of the information flows that economists and policy-makers have identified as important, such as those from suppliers, universities etc.

The major disadvantage of these data is that they are qualitative. Some might therefore dismiss them on the basis that they are inferior to some quantitative knowledge flow measures e.g. the number of emails from suppliers, phone calls from clients or bytes of information from internal databases. Of course such data would then have to be weighted by their importance, which would almost inevitably introduce some qualitative element.[6] So, in the light of the gap in knowledge of non-patenting firms and the importance of the question, we shall proceed with investigating this measure. Following the patents

---

[4] Looking at the relation with TFP does have the advantage that one should be able to identify a spillover, provided of course the conditions for TFP to measure all priced inputs hold.

[5] It also asks about patents, data on which we use, and self-reported innovation outputs, which we do not use here. A large number of papers have used the CIS self-reported innovation output data. For our purposes here it is hard to interpret these data in terms of spillovers.

[6] For example, data collected by social network analysis (SNA) in management studies attempt to show informal relationships - who knows who and who shares information and knowledge with who. Also SNA typically gathers data about the relationships between a defined group/network of people with the use of questionnaires and/or interviews or with softwares that track directly e-mail messages or repository logs. The responses are then processed to create a network map of the knowledge flows within the group or network and to produce statistical analyses of the patterns in the data. Therefore even in SNA some qualitative information is introduced through the surveys

literature, we shall of course regard our variables as an error-ridden measure of the "true" information flows. We therefore attempt a number of corrections. We also have data on patents applied for by the firm and so use these data to place our information flow measures in the patents context.[7]

Our investigation and major findings are as follows. First, we re-scale each knowledge flow variable to be the deviation from the average importance of all knowledge flows in the firm. Thus for example a firm who reports "very important" to all measures, simply because they tick all boxes in a column, scores zero for each flow. Thus we shall *not* say that firms who report knowledge flows from suppliers also have high TFP growth. Rather we shall say that firms reporting that knowledge flows from suppliers are more important than the average of all their other knowledge flows have a positive association with TFP growth. Indeed, this transformation is sufficient, on our panel, to render insignificant any fixed effects in the relation between TFP growth and these transformed information flows.

Second, we posit a knowledge production function (Griliches, 1979) relating increases in knowledge to investment in knowledge, which we measure by R&D, and knowledge flows. Thus an immediate check on our data, which helps locate it in the literature, is to use patents as the measure if increased knowledge and so regress it on R&D and our measures of knowledge flows. We find sensible results: patents are strongly associated with R&D (with a coefficient that is exactly the same as other studies of this link such as Hausman, Hall and Griliches, 1984) and also with information flows from universities.[8] To the extent that university knowledge flows contribute to knowledge advances at the frontier, these seem sensible.

Third, we look at the relation between R&D, knowledge flows and TFP growth. We think this is of interest since it expands our knowledge change data beyond patents[9], and, under certain conditions, allows one to identify these information flows as spillovers. Our main results are a statistically significant association between TFP growth and above-firm average information flows from: other firms in the enterprise group, competitors and suppliers. The effects are economically significant as well, with such information flows "explaining" (in a growth accounting sense) about 50% of TFP growth. The effects are robust to different methods of measurement and different samples.

Fourth, we ask if such knowledge flows are spillovers. We have no data on the prices that firms pay, if any, for information flows. *A priori* reasoning would suggest that information from other firms in the group is likely internalised, so this is unlikely to be a spillover. But it is hard to see how information flows from competitors is, so the latter would seem to be a spillover. Baldwin and Hanel (2003) argue similarly. Information flows from suppliers are not so amenable to *a priori* reasoning. Such flows have

---

[7] Due to confidentiality we do not have information on the company names and so cannot link the data to patent citations data.
[8] Strictly, information flows from universities that were rated as more important than the average of all information flows.
[9] Note we also show a positive and significant relation between patents and TFP growth.

been extensively documented in case studies, but it is not possible from case studies to know whether they have been capitalised or not. Since we use TFP, if the conditions for TFP are satisfied and our measurement is correct, then it is a spillover. However, neither of these might be the case. A supplier might for example tell a firm about a new machine, causing the firm to report such an information flow as important. But if the firm then pays the market rate for it, there would be no effect on TFP. More precisely, TFP growth, as we measure it, captures paid for inputs if the subinputs in each input aggregate are perfect substitutes (in efficiency units), the relative prices of sub-inputs reflect their relative marginal products and if the price of the input is not firm-specific (since we have no firm-specific input deflators). We focus on the last of these conditions and show some suggestion that information flows from suppliers could be associated with firms paying less for their inputs than others, which would not be a spillover. Thus we conclude that the main information flow spillover is from competitors.

Finally, having assembled some direct evidence, we re-examine the indirect work. We look at the relation between our knowledge flow data and R&D in the industry, MNE presence in the industry, competition and the TPF gap between the firm and the frontier. We think this is of interest since such proxies have often been related to TFP growth in the absence of data on the underlying information flow they are purported to represent. But, as is well-acknowledged, the relation between, say MNE presence and TFP growth could be due to both information flows and underlying technological factors that both boost TFP growth and cause MNEs to be present.[10] Thus the presence or absence of a relation with information flows should inform what is driving the correlations in this literature. We find positive relations between information flows (from competitors, relative to the mean for the firm) and the presence of R&D and MNEs. But, with respect to MNEs, the implied impact of MNE presence on TFP growth via information flows is much lower than that from the simple relation between MNE presence and TFP growth. This suggests that other studies that have relied on this relationship have overstated the spillover impact of MNEs. We find no relation between learning from suppliers and these variables, suggesting, that for these data at least, knowledge spillovers from competitors, of whom MNEs are important, are statistically the most important contribution to TFP growth. Finally, we find some relation between learning from clients and MNE presence in downstream industries (MNE presence weighted by the input/output flows). This supports the recent indirect work of, for example, Smarzynska Javorcik (2004), but since we find no link between learning from clients and TFP growth it suggests that this form of spillover does not operate, on our data set at least.

How does our paper relate to other work using innovation surveys? Using the Community Innovation survey, due to data availability, rather few studies have linked the CIS to TFP data. Of those papers most have concentrated on the direct impact of product and process innovations on productivity

---

[10] Also note that increase in MNE presence is accompanied by an increase in competition. This might give an additional spur to increasing the productivity of domestic plants.

(level and growth) rather than on the importance of knowledge spillovers (e.g.Crepon; Duguet and Mairesse, 1998; Loof and Heshmati 2002 and Klomp and Van Leeuwen, 2001 to cite a few).[11]

The plan of the rest of the paper is as follows. In the next section we set out the basic framework for analysis. Section three sets out the data and section four the equations to be estimated and the results. Section four looks at whether these are spillovers or not, section five at the relation with MNE and R&D presence and section six concludes.

2 Theory and first look at data

2.1 Theory outline

For firm *i*, measured total factor productivity growth, $T\dot{F}P$ is some combination of changes in the knowledge stock at firms, $\dot{A}$, demand shocks and other unobservables (ε_{1i}) and so can be written

$$T\dot{F}P_i = f(\dot{A}_i, \varepsilon_{1i}) \tag{1}$$

Following Griliches (1979) we may write changes in the firm knowledge stock, $\dot{A}_i$, as due to investment in new knowledge, such as R&D, and flows from the existing knowledge, which may be inside the firm (i) or outside (_i)) which we write as

$$\dot{A}_i = f(R_i, A'_i, A'_{\_i}, \varepsilon_{2i}) \tag{2}$$

where a prime indicates a flow from the inside and outside knowledge stocks A_i and $A_{\_i}$ and where ε_{2i} are the various other shocks, which might include elements of ε_{1i} (e.g. if ε_{1i} includes unmeasured changes in managerial ability that also affects knowledge production). In this framework, (at least) three questions arise. First, what are the relevant knowledge flows in (2) that determine productivity growth in (1)?[12]

[11] In fact most of these studies use principal factor analysis and group external knowledge sources in two categories: Science base and others.

[12] Consider for example some of the case study evidence e.g. post-Southwest Airlines low cost airlines: "In the 1990s other airlines around the world began to model their strategies around Southwest's, often after a visit by their managements to Dallas. The most successful of these included RyanAir, Easy Jet, and Go in Europe as well as Air Asia in the Far East" (Heskett, 2003, p. 5). McGinn (2004) writes, "In 1991, [RyanAir's now-CEO Michael] O'Leary, an accountant, visited Southwest's headquarters in Dallas … At the time, Southwest was already garnering accolades as the industry's big innovator … O'Leary liked what he saw … Flying back to Ireland after a few days with Southwest, O'Leary laid plans to replicate the strategy." As for Southwest's itself, "To improve turnaround of its aircraft at airports, Southwest sent observers to the Indianapolis 500 to watch pit crews fuel and service race cars. The airline recognized that pit crews performed, in a different industry and at much faster speeds, the same functions as airplane maintenance crews. New ideas about equipment fittings, materials management, teamwork, and speed subsequently contributed to a 50% reduction in the airline's turnaround time" (Frei, 2004, p. 2).

Second, to what extent are such flows spillovers? Third, can one distinguish such flows from other influences on productivity growth (here the εs)?

Very broadly, the patents literature offers direct evidence on knowledge flows in (2) by measuring $\dot{A}$ as patents and using citations as measures of knowledge flows $A'_{\_i}$. Other work is more indirect. The MNE spillovers literature postulates that the proximity of MNEs is a possible source of $A'_{\_i}$ and so combines (1) and (2) and regresses $T\dot{F}P$ on MNE presence. The distance-to-frontier literature postulates that $A'_{\_i}$ can be measured by TFP levels in a nominated frontier firm/set of firms, A'_i can be measured by TFP levels in the firm itself and so regresses $T\dot{F}P$ on the gap between frontier and own firm TFP. The R&D literature postulates that $A'_{\_i}$ can be measured by R&D or knowledge stocks outside the firm (i.e. in the industry for example) and so regresses $T\dot{F}P$ on own and outside R&D.

Our contribution, we believe, is the following. First, we match survey data on R&D and learning flows with administrative data on TFP growth. Second, with these data we begin by estimating (2) with patents as a dependent variable (the survey asks for the number of patents applied for) to see what this measure of $\dot{A}$ shows and therefore locate our work in line with previous work using different knowledge flow measures. Third, combining (1) and (2), we estimate the relation between $T\dot{F}P$ and these knowledge flows. We find robust statistical relations between $T\dot{F}P$ and various knowledge flows which, we argue, sheds light on what kind of knowledge flows are important for TFP growth in non-patenting firms. Fourth, we then collect data on R&D, MNE presence and distance-to-frontier measures and see if they are related to learning to better understand the indirect evidence on TFP growth.

As mentioned above, we believe that this work goes beyond work that has used the CIS. Due to data availability, few papers have matched the CIS with TFP data, which means that few papers can look at the spillovers issue. In addition, we are not aware of papers that have used our transformation of information flows, or explored their relation with MNE and other variables commonly used as indirect learning measures.

2.2 Data

We shall use two sets of data. To measure TFP growth, we use administrative data from the official UK businesses survey. Data on information flows and the like comes from the Community Innovations Survey.

2.2.1 The ARD data

We use the ARD (the Annual Respondents Database) which consist of successive cross-sections of input and output data reported by firms in response to the official business survey, the Annual Business Inquiry

(ABI). The ABI is an annual inquiry covering production, construction and some service sectors, but not public services, defence and agriculture.[13]

There are a number of points to be noted. First, reduce reporting burdens, multi-plant businesses are allowed to report on plants jointly. Such an amalgamation of plants is called a reporting unit and in practice most reporting units are firms. Reporting burdens are further reduced in some years by requiring only reporting units above a certain employment threshold to complete an ABI form every year (typically the threshold is 100 employees). So our data is best thought of as at the firm level, and mostly covers larger firms. Second, regarding data, firms report on turnover, employment (total headcount), wage bill, materials, and material costs and investment (in plant, building and materials). To build capital stock from the investment data we applied a perpetual inventory method, Martin (2002).

2.2.2 CIS data

The U.K. CIS is part of an EU-wide internationally agreed survey of businesses on innovation outputs; innovation inputs; and sources of knowledge for innovation efforts. There have been three waves of U.K. CIS surveys: CIS1 (covering 1991-3), CIS2 (1994-6) and CIS3 (1998-2000). We use CIS2 and CIS3 and the panel therein.[14] The CIS is an official survey of around 19,000 firms (CIS3) stratified using the IDBR with a 46% response rate. See Criscuolo (2005) for a further discussion of non-response bias etc. [15] Matching the CIS and ARD is simplified by the fact that the two surveys are carried out on the basis of the IDBR. Matching is thus by common survey identification number and not by address or postcode hopefully minimising matching induced measurement error.

We use the CIS data to measure a number of variables. First, for R&D, there are measures of persons engaged in R&D and R&D expenditure. In the results below we used the reported persons engaged in R&D (as a share of all employment) since this turns out to be the most reliably reported when comparing it to other R&D surveys (Haskel, 2005; results are not sensitive to other measures). One problem in using data for foreign firms is that they may undertake R&D elsewhere and so return zero employment in the CIS. This should however be picked up in the data on information flows.

---

[13] The ABI is based on the UK business register, the Inter-Departmental Business Register (IDBR), which contains the addresses of businesses, some information about their structure (e.g. domestic and foreign ownership) and some employment or turnover data (sometimes both), based on accounting and tax records. However, it does not contain enough data to calculate TFP since it does not have materials use or investment/capital (and much employment data is interpolated from turnover data).

[14] CIS1 is largely unusable due to a response rate of barely 10%.

[15] ONS selects survey recipients by creating a stratified sample of firms with more than 10 employees drawn from the IDBR SIC92 two-digit classes and eight employment-size bands. Production includes manufacturing; mining; electricity, gas and water; and construction. Services includes wholesale trade; transport, storage, and communication; and financial intermediation and real estate. Note that the survey, although voluntary is an official one and has a series of reminders to try to boost response. A response rate of 46% is very good compared with many voluntary (non-official) surveys. Response to the ABI is mandatory.

Second, the CIS data also asks firms to report the number of patents they have applied for. This is not a measures of patents granted, but a measuring of patenting activity that has been used in other studies, see e.g. Griliches, Hall and Hausmann (1984).

Third, perhaps the key measure, is information flows, for which the CIS asks the following question.

"Please indicate the sources of knowledge or information used in your technological innovation activities, and their importance during the period 1998-2000. (please tick one box in each row)

		N	L	M	H
Internal					
	Within the enterprise				
	Other enterprises within the enterprise group				
Market					
	Suppliers of equipment, materials, components or software				
	Clients or customers				
	Competitors				
Institutional					
	Universities or other higher education institutes				
	Government research organisations				
	Other public sector e.g. business links, Government Offices				
	Consultants				
	Commercial laboratories/ R&D enterprises				
	Private research institutes				
Specialised					
	Technical standards				
	Environmental standards and regulations				
Other					
	Professional conferences, meetings				
	Trade associations				
	Technical/trade press, computer databases				
	Fairs, exhibitions				
	Health and safety standards and regulations				

where the column answers to columns are N (not used) and L, M, H, respectively low, medium and high.

Before reviewing the disadvantages of this measure, it is perhaps worth noting some of the more favourable aspects of it. First, the information flow variables correspond with some of those flows identified as important in a number of studies e.g. flows from suppliers, from within the firm, from universities, see the discussion footnote 12 above or in Criscuolo, Haskel and Slaughter (2005). Second, modern theories of MNEs are built on the assumption that MNEs can better transfer knowledge within the enterprise than other like firms, see e.g. Markusen (2002). Thus a direct check on the data would be to see if knowledge flows between MNE plants in an enterprise group to be greater than flows between a like comparator (plants in a purely domestic enterprise group for example). This is indeed the case

(Criscuolo, Haskel and Slaughter, 2005). Third, these data have a very high response rate on the CIS surveys.

One broad objection to these measures however is that they are qualitative. It might be that one would prefer quantitative data on importance-weighted information flows. Thus, for example, if ideas flowed by email or phone calls or videoconferencing, one might try to collect data on the number of emails, phone calls or videoconferences.[16] As well as being a formidable task in of itself, such records would then have to be weighted by their importance, since it is unlikely that every email and call are of equal importance. In the absence of prices, such weighting would likely itself require qualitative surveys too. Thus we regard our measures as a (noisy) importance-weighted indicator. A second objection is that our survey data does not record if the information flow is paid for (a visit to a trade fair example). Thus we will look at TFP rather than labour productivity and review the conditions under which this should capture paid-for inputs. Given the doubts over measurement, following Bertrand and Mullainathan. (2001), it seems useful to think about these data in a measurement error framework. Denote the true learning flow as $A^{j\prime}{}_{it}$ and our measured flow from source j in time t for firm i as $L^{j}{}_{it}$ we can write

$$L^{j}_{\_it} = A'^{j}_{\_it} + Z_{it} + v_{it} \tag{3}$$

which says that measured information flows are related to true information flows plus some firm- time-varying variables Z plus an error term. The following sources of measurement error might drive a wedge between $A^{j\prime}{}_{it}$ and $L^{j}{}_{it}$: first, single-rater bias, i.e. the questionnaire is answered by one respondent in the firm who may give unrepresentative or inaccurate reports. To the extent that this is random, then this raises v_{it} and so biasing us against finding any significant effect of learning and biasing down included coefficients via attenuation bias. To the extent that it is fixed, we can explore this by using the two CIS cross-sections in a panel. To the extent that it varies systematically with, say size, we can enter size as a control

Second, by questionnaire construction, respondents reply on the basis of a Likert-type scale. Thus the scale has no meaningful cross-sectional variation due to the impossibility of comparing different respondent's options. There might also be a bias if some respondents tend, for example, to always tick the middle box for all answers.

In the empirical analysis we correct for this bias and for time varying firm specific effects using the approach described in more detail below.

---

[16] Or, for example, one could try to collect employee time-use diaries to try to measure learning time from different sources. There are of course general time-use surveys see e.g. the American Time Use Survey, but these do not cover time use at work. As for time use at work, there are some time use surveys of managers done by management researchers but on a small scale and not directed at learning, see e.g. Mintzberg (1973) for the classic study (of 9 managers).

We start by calculating the average of each firm's reported learning from all 17 information sources, L* (converting the responses into 0, 1, 2 and 3). We then expressed each learning variables as deviations from the average of each firm's learning. Finally, if that figure is positive, then we allocate a one to that (deviation) learning variable and zero otherwise. That is, we formed the following indicator function for source of knowledge *j*

$$\begin{aligned} &\mathrm{I}(L^j)_{it} = 1 \quad if \quad (L^j_{it} - \overline{L}^j_{it}) > 0, \quad \overline{L}^j_{it} = \sum_{j=1}^{J} L^j_{it} \\ &\mathrm{I}(L^j)_{it} = 0 \quad otherwise \end{aligned} \tag{4}$$

where the average of all replies to the j questions by firm i is $\overline{L}^j{}_{it}$. Thus a firm, who for example, always ticks box 3, is allocated a zero on all information measures, likewise a firm who always ticks box 1. A firm who ticks all box 1 but on one information source ticks 3 would have a 1 for that source and a zero for all others. In this way we hope to control for any unobservable firm effect in (3). Finally, we enter $\overline{L}^j_{it}$ as a control. In sum, we do *not* study the relation between $T\dot{F}P$ and $L^j{}_{\_it}$ but rather between $T\dot{F}P$ and $\mathrm{I}(L^j{}_{\_it} - \overline{L}^j{}_{it})$. Thus for there to be an omitted common variable that explained the relation between $T\dot{F}P$ and $\mathrm{I}(L^j{}_{\_it} - \overline{L}^j{}_{it})$, it would *not* be that say more aggressive managers have better TFP growth and report more aggressively. Rather, it would have to be that an omitted variable that causes both managers to have better TFP growth and report more aggressively on a particular learning variable relative to the others (and in the panel, that this over-reporting changed over time). We cannot of course exclude this possibility, but it does seem to reduce the possible biases from omitted common factors.

2.3 Averages

The matched data set consists of 804 observations who appear in the ARD and CIS2 or CIS3, and have complete TFP, R&D and information flow data and are UK firms (we have an extra 238 observations who are part of foreign MNEs who we exclude initially). This 804 firms consist in turn of 752 firms who appear in either CIS2 and CIS3 plus 26 observed twice). The reason for this small panel element is that the CIS panel is a coincidence from the two samples of around 8,000 observations in CIS3 and about 5,000 in CIS2. In addition, to match to TFP growth data on the ARD restricts the sample to manufacturing with full ARD data on inputs and outputs.

Table 1 top panel shows our data.[17] Let us concentrate on the innovative activity figures: the median firm in the sample does no R&D and no patenting (indeed the top 12 firms in our data account for about 60% of all patenting activity; 92% of firms did not apply for a patent). Thus our data looks similar to the patenting data set out above. Turning to the learning figures, the median firm also does no learning, from any source. However, learning is less skewed than R&D and patents. The numbers in the table are the transformation as in (3) and indicate that 51% of the 804 CIS firms matched with the ARD report learning from competitors more intensively than the average of learning from all sources. Thus these numbers are smaller than the firms who report any learning from competitors. Learning from suppliers and clients, relative to the average is reported more often than learning from other enterprises in the enterprise group and from universities.

Finally, Table 1 divides the sample into low (middle panel) and high (lower panel) productivity growth firms (with growth relative to median three-digit industry productivity growth). The high productivity growth firms have similar R&D employment (with the median slightly higher) and apply for slightly more patents (2.93 against 2.45). They also make slightly more use of information flows in all cases.

Before moving to econometric analysis, we discuss what consequences our particular sample might have for inference about the marginal impact of these learning variables on TFP growth.[18] First, it might be that marginal effects differ due to technology e.g. they are different in larger firms. If this is the case then to the extent they are larger we will overstate the average marginal effect. Second, the effects might differ due to selection bias induced by our use of a matched sample. The obvious source of possible selection bias is that we use surviving firms, either that survive within the three years of each cross-section so that we can compute TFP numbers for them or that they survive between the three years of each cross-section. In each case this introduces selection to the extent that if low productivity growth firms only survive if they have had a positive shock, then the sample of surviving firms consists of high productivity growth firms plus low productivity growth firms who have had a beneficial shock in the early period. This flattens any positive relation between productivity growth and its drivers such as information flows, biasing co-efficients downwards and so causing us to understate actual co-efficients.[19]

---

[17] We also analysed how the regression sample differs from the complete ARD and CIS samples. Relative to the ARD, the sample is bigger in terms of gross output, employment and productivity levels, but the standard deviation of these numbers is large. Relative to the whole CIS the sample is again larger and but not more productive. It does more R&D and more patenting however, and learns from more sources than the whole CIS sample.

[18] Of course, the absolute levels of learning will likely be larger in our sample, since the sample is of larger firms and such firms do more learning. At issue here is however the marginal impact on productivity growth.

[19] Information flows were more prevalent in hi-tech industries. Note however that all our econometric work controls for industry dummies so that any unobserved factor that affects both industry TFP growth and industry-specific learning is controlled for.

3 Econometric framework and results

3.1 Estimating the knowledge production function, (2) using patents

Before turning to TFP growth, we start by estimating the knowledge production function using patents as a measure of $\dot{A}$. Our purpose here is to (a) check our data against other studies that have analysed the relation between patents (applied for) and R&D and (b) add in our information flow variables to locate our data alongside other studies. If we regard patents as reflecting changes in "frontier" knowledge, then we might expect information flows from universities to be important for example. Thus we implement (2) using the transformation of the learning terms as follows

$$\begin{aligned} PATENTS_{it} = \gamma_1 R_{it-1} + \Sigma \gamma_{2j} I(L^j)_{it-1} + \gamma_3 \overline{L}_{it-1} + \gamma_4 \ln SIZE_{it} \\ + \lambda_R + \lambda_I + \lambda_i + \lambda_i^{STATUS} + \varepsilon_{it} \end{aligned} \qquad (5)$$

where PATENTS are patents applied for over the past 3 years, R is R&D (measured as expenditure and employment shares, see below) we add log size since this is standard in many patenting equations and dummies for the region, the two digit industry, the firm and status[20]. The dependent variable is a count of patents and so we used a negative binomial model with random effects (we rejected the Poisson model).[21]

Table 2 reports our results. We start by duplicating, as far as possible the regression of Griliches, Hall and Hausmann (1984) who took a sample of US firms who performed R&D and regressed their patents applied for on (log) R&D, size and industry dummies, obtaining an elasticity of R&D expenditure of 0.33. As column 1 shows, our elasticity is 0.33. This then seems like a good cross check on our data.

In column 2, we add the information flow variables. In practice there were many of them and so we just included information flows (in deviation from the company-specific average L_i*) from other enterprises in the group I(L_i^{GROUP}-L_i*), suppliers I($L_i^{SUPPLIERS}$-L_i*), competitors I(L_i^{COMPET}-L_i*), clients I($L_i^{CLIENTS}$-L_i*) and universities I(L_i^{UNIV}-L_i*). Learning from universities is positive and significant; on the assumption that patenting reflects changes in frontier knowledge, this seems reasonable. Learning from other sources and the average level of learning of the firm are not significant. In the third column we use as a proxy for the research investment the log number of persons employed in R&D at the firm, which, as we discussed above, is better measured in our data: the elasticity of this is 0.546 again amongst the information flow variables the only significant coefficient is the one on university. Finally, column 4

[20] Firms are asked if they have recently merged and increased sales, merged and decreased sales or are a start-up. We include this in all our results in case (internal) adjustment costs are clouding the long-term relation between inputs and outputs in these cases. All our results are robust to their exclusion.

[21] We were unable to get the fixed effects version to converge.

expands the sample to all firms besides those doing R&D. It is notable here that the co-efficients on log R&D employees and the learning variables remain virtually unchanged.

In sum, these results are, we believe, interesting in themselves, suggesting a robust patents/R&D relation and a relation between patents and learning flows from universities. They also suggest some confidence in the data on patents and R&D at least and so we shall proceed to investigate other measures of $\dot{A}$.

3.2 Estimating the output and knowledge production function, (1) and (2), using TFP growth

We next move to the TFP/productivity growth regressions. We start by estimating a simple Cobb-Douglas equation

$$\begin{aligned}\Delta \ln Y_{it} = \alpha^K \Delta \ln x_{it}^K + \alpha^M \Delta \ln x_{it}^M + \alpha^L \Delta \ln x_{it}^L + \gamma_1 R_{it-1} + \Sigma \gamma_{2j} I(L_i^j)_{it-1} + \gamma_3 \overline{L}_{it-1} \\ + \lambda_R + \lambda_I + \lambda_i + \lambda_i^{STATUS} + \varepsilon_{it}\end{aligned} \tag{6}$$

where x^L, x^M and x^K are labour, material and capital inputs, *R* is measured here by the ratio of R&D employment to total employment[22], and $I(L)^j$ is the indictor function for the deviation from the mean of learning from the j'th information sources. The following points regarding measurement and causality are worth noting. First on measurement, we do not have firm-specific prices and hence the revenues deflated by industry prices is on the left hand side (see appendix for more details). Thus we use product innovation as a proxy for the gap between company and industry output prices. We also look at possible input-specific prices, see below. Second, the specification in (6) implicitly constrains the αs to be the same, but our use of TFP, which controls for this (under certain circumstances) indicates the biases here are not serious.

Regarding causality, the choice of inputs is of course endogenous and to the extent that firms choose what sources to learn from, the effects of I(L) is potentially biased. We cannot eradicate that bias in the absence of an experiment but the use of deviations from means and specification of growth means we hope to reduce potential biases as discussed above. Of course, the α coefficients as likely biased too,

[22] As stated above, this variable is better measured than R&D expenditure. However, we found a few small firms who had fractions of around 100% who we suspect are essentially R&D facilities. To guard against this we entered a dummy variable (not reported) for firms reporting fewer than 20 employees overall. Note that, following Schankerman (1981), the interpretation of γ_1 is complicated because R&D employees and their wage are included in x^L and s_L. We would like to adjust these but we only have R&D employent and expenditure measured at the end of each CIS period, whereas we calculate TFP over the whole CIS period (we could adjust by assuming the same proportion of employment over the period but with logs this would not affect Δx^L; it would affect s_L by a small amount though. In addition, R&D expenditure is measured very poorly in the CIS. Finally, the correlation between R&D employment levels in both periods is 0.85, suggesting relatively small changes in employment.

but they are not our focus of interest here: we do experiment with ΔlnTFP too which potentially removes a lot of the bias, see Appendix.[23]

Table 3 sets out our results of estimating (6). Column 1 shows the results using OLS. The terms on the inputs are precisely estimated but add up to below one, a common result on differenced data. The R&D term is 0.131, in line with other studies, but very imprecisely estimated. Turning to the information terms, there are positive and statistically significant effects from competitors (L^{COMPET}-L^*), , suppliers ($L^{SUPPLIERS}$-L^*) and the enterprise group (L^{GROUP}-L^*),. Information flows from clients (L^{CIENTS}-L^*) and universities (L^{UNIV}-L^*) do no appear to be significant on this sample at least . Recall that since we are using TFP growth, these latter findings are consistent with the idea that such flows *do not* affect productivity growth; or that they *do* affect productivity growth but are paid for by firms.

The other columns examine robustness. Column 2 uses random effects which improves the precision of the estimates. Columns 3 and 4 repeat the exercise but using ΔlnTFP as the dependent variable. The coefficients on learning are hardly changed, with the coefficient on R&D rising slightly to 0.167. The standard errors around the competitors and other enterprise variables rise somewhat.

Table 4 shows further robustness checks, with ΔlnY as the dependent variable (checks with ΔlnTFP as dependent are set out in Appendix 6). In the first column we add foreign MNE presence measured as the share in the two digit industry of foreign employment (recall our sample are only UK firms). This is positive and significant, at 0.137, without affecting the significance or coefficients of the main learning variables. Thus MNE presence has an effect over and above the learning variables. This could be consistent with MNE presence having no impact on firm learning but simply being associated with the same unobserved technological progressivity conditions that drive ΔlnTFP. Alternatively, it could be consistent with leaning from MNEs that is not measured by our variables.

Columns 2 and 3 report regressions for firms who undertake positive R&D and none respectively and similarly column 4 and 5 for patenters and non-patenters. The samples fall and the precision of the estimates falls somewhat. However, comparing column 2 with 3 and column 4 with 5 we might see that for firms who do not do R&D TFP growth is positively associated from learning relatively more from suppliers and from other firms within the group; while for firms which do R&D and for firms that do not patent there is a significant positive association of TFP growth and learning relatively more from competitors.

The remaining columns add further measures without much affecting the information flow terms. Column 6 adds a pressure of competition measure, the change in market share, lagged two periods, to test whether the information flow from competitors variable is picking up information flows or some other

---

[23] With just two cross-sections of data and a small panel element, plus the doubts summarised in Gorodnichenko (2006) we did

pressure of competition. The change in market share term is negative but insignificant, suggesting that falls in market share are (statistically) weakly associated with increased ΔlnTFP two periods later and the $I(L^{COMPET})$ term rises in sign and significance. Thus it would seem that the information flow variables have an effect that is stronger and over and above the competitive pressure variables.

Column 7 adds a dummy if the firm has reported introducing a new product innovation to control for the change in demand terms that appear since we do not observe plant-specific prices. None of the learning effects are much affected. Column 8 adds log (1+ patents applied for) and finds patents positive in their effect on TFP, and column 9 expands the sample to include foreign MNEs (with a dummy if the firm is a foreign MNE). The significance of the information flow variables remains.

The appendix sets out the results using ΔlnTFP as a dependent variable. The patterns of correlation are very similar indeed but the information flow variables are a little less well-determined. Whilst the ΔlnTFP implicitly allows for separate output elasticities, it could be more noisy if the cost shares are mismeasured or if the demand function is not Dixit-Stiglitz. Overall, then we believe there is evidence of a significant correlation between TFP growth and R&D and (L^{GROUP}-L^*), ($L^{SUPPLIERS}$-L^*), and (L^{COMPET}-L^*). We cannot of course rule out endogeneity bias, but as emphasised above, bias would have to depend on some variable, say changing management ability, that affects TFP growth (not its level), enjoys excess returns (i.e. is not captured in TFP) and also affects the deviation of learning from the average (not the average). All this suggests, to us, that we would hope to minimise the effects from endogeneity bias. Without an actual experiment we cannot remove effects of course, but on the assumption that such a variable causes both more TFP growth and more learning relative to the average, endogeneity bias would suggest our co-efficients are an upper bound on the true effect.

3.3 Economic significance

The above work has documented the statistical significance of these results. What is the economic significance? The coefficients on the relatified learning variables enable us to read off the TFP growth gains from learning which are approximately 1.5% from (L^{COMPET}-L^*)L, 1.5% from ($L^{SUPPLIERS}$-L^*), and 1.7% (L^{GROUP}-L^*), totalling 4.7% for a firm who uses all three sources of information (relative to the average).

What is the yardstick against which to judge this? The most straightforward way to judge this is against the interquartile range (IQR) of ΔlnTFP in the sample, which is 9.9% (in other words the firm at the 75% percentile of the TFP growth rate distribution has annual TFP growth which is 9.9 percentage

---

not experiment with IV type estimators.

points higher per year than the firm at the 25th percentile of the TFP growth distribution).[24] The sum of our main learning effects, 4.7% is thus "accounting for" about 50% of TFP growth. Thus if the "measure of our ignorance" is 9.9%, these learning variables reduce our ignorance by 50%.

Note in passing that the assumption that the productivity gain from a firm who, for example, has all three sources of learning, can be measured by summing the coefficients from the regression assumes that there is no productivity disadvantage to learning from all three sources together.[25]

4 Are these spillovers?

We believe we have set out some evidence of correlations in the data with the significant information flows being I(L^{GROUP}-L^*), I($L^{SUPPLIERS}$-L^*), and I(L^{COMPET}-L^*). Can these information flows be considered spillovers? If all is well-measured and competitive conditions hold, then the relation of these flows to TFP would suggest they are spillovers. However, these conditons may not hold and we have no direct data on whether the information has been paid for or not. But we can make some progress initially by *a priori* reasoning.

Consider first I(L^{COMPET}-L^*). It seems highly unlikely that competitor information would be paid for. Unless there is a joint venture occurring it seems hard to think of a mechanism by which companies would pay competitors for information.[26] Thus it seems reasonable to conclude that this is indeed a spillover of information.[27]

Consider second, I(L^{GROUP}) i.e. knowledge flows from other enterprises in the enterprise group. That such flows affect TFP growth are consistent with Klette (1996) for example, who finds that R&D performed in other plants in a group of plants influences TFP growth over and above that performed at the particular plant. He argues this is evidence consistent with within-firm spillovers. The question is then whether such information flows are internalised by firms. Standard theory would assume so but there may be imperfections in control within firms such that they constitute spillovers. Thus, it seems safest to conclude that these spillovers are internalised within a firm (the individual plant has a return above its plant-specific return to R&D because of information sharing across plants within the group), but the extent of the excess returns to other firms depend on knowledge flows via other mechanisms (such as knowledge sharing with competition or suppliers).

---

[24] We run a regression of ΔlnY on ΔlnK, ΔlnL and ΔlnM, plus industry and other dummies (excluding the L variables) and calculated the IQR of the residual. We feel that expressing the fraction in terms of the fraction of the IQR is more appropriate in the context of our data which is essentially a cross-section all in terms of the deviation from the industry mean. Growth accounting typically expresses the coefficient times the inputs as a fraction of total productivity growth, but our work here is in terms of deviation from the industry mean and hence average ΔlnTFP in the sample is, aside from rounding error, zero.

[25] We would test this from our data by interacting the learning sources in the regression, but we suspect we have too few degrees of freedom to do this effectively.

[26] In unreported analysis, available from the authors we control for the participation of the firm to joint ventures with competitors as this information is available from the innovation survey. The inclusion of the joint-venture variable in our specification did not change the conclusions reached in Table 2 and Table 3.

Third, consider the information flows from suppliers. This case is less amenable to *a priori* reasoning and so let us review some of the other issues involved. First, in terms of supporting evidence, the case study evidence provides evidence of both forward and backward linkages between MNEs and domestic firms suggestive of learning from suppliers (and customers), see e.g. Rodriguez-Clare (1996), or the survey in Hanson (2000). Second, econometric evidence typically has looked at presence of MNEs in the same industry, which is consistent with information flows from suppliers depending how wide the industry is (the two digit vehicles industry would include suppliers to the vehicle industry for example). Smarzynska (2004) uses MNE presence in Lithuania weighted by an input/output table and finds a positive relation between domestic firm productivity and the downstream presence of MNEs i.e. backward linkages. On our data, we would look for domestic firms reporting learning from customers rather then suppliers. Forward linkages are found to be important in Romania (Merlede and Schoors, 2006).

In terms of measurement we might wrongly infer the presence of spillovers if the conditions upon which TFP measures spillovers do not hold. One condition would be where the efficiency units of the mix of inputs in the firm are not measured by its nominal value. In turn this is when the market power in the purchasing of one input relative to the industry drives a wedge between the relative price of the input and its relative marginal product, causing TFP to mismeasure the efficiency units of the input. [28] In turn this wedge would have to be correlated with information flows. The sign of this correlation is hard to determine. The other condition arises since we have no input-specific deflators and so our measured real inputs of factor X, $\ln X^{MEAS}_{it}$, is derived from the value $W^X_i X_i$, divided by an industry-wide input price index, W^X_I giving $\ln X_{it}^{MEAS} = \ln X_{it} + \ln W_{it}^X - \ln W_{It}^X$. If then the information allows the particular firm to obtain the good more cheaply than the industry, then such a firm has an apparent rise in measured TFP growth due to the mismeasurement of firm-specific inputs (this is an analogous argument to the biases in TFP with mismeaurement of firm-specific output prices, see e.g. Klette and Griliches, 1996). Without further data on company-specific input prices, which is almost never available in a Production Census dataset, this has to remain a caveat over our results.

---

[27] Vickers (1996) sets out a series of models where competition is Pareto-improving due to information spillovers.

[28] To see this, suppose that $Y=F(M^*)$ where M^* is effective materials, which is unobserved. In turn suppose that $M^*=M_1+(1+\phi)M_2$ where M_1 and M_2 are different material volumes, which are also unobserved and φ is the relative marginal product of M_2 to M_1 i.e. φ converts the quantity of M_2 into efficiency units. φ is of course unobservable, but the first order conditions for a firm give that $\phi=(P_2^M - P_1^M)/ P_1^M$ where the subscript is the price of the particular factor, so that we can write $M^*= P^M M/ P_1^{\ M}$. Hence, under these conditions, TFP measures the mix of different inputs acquired (and it would not, if, for example, the M^* equation functional form does not hold, or ϕ does not equal relative prices).

5 Learning flows and R&D, MNE presence, competition and the productivity gap

In this final section we try to relate our results to the indirect literature on spillovers. The regressions above have used explicit measures of learning. Many studies do not have explicit learning measures and hence other proxies are used (such as R&D in the industry, presence of MNEs etc.). This section explores whether there is a relation between these proxies and our learning measures. We think this of interest since it is sometimes argued that these proxies capture other effects besides knowledge flows. For example, it is argued that MNEs are likely to situate in more technologically progressive industries which would also have faster growing firms and hence there is a correlation between MNE presence and productivity growth but not one driven by knowledge. Thus, if there were no relation between our knowledge measures and these proxies this would cast doubt on the TFP/MNE correlation being driven by learning (or on the learning measure of course). Note of course that we just have three years of UK data and such correlations as we obtain may not shed light on many of the findings that are based on data for other countries.

To do this we regress

$$I(L_i^j)_{it} = \beta^j Z_{it} + \lambda_R + \lambda_I + \lambda_i + \lambda_i^{STATUS} + \varepsilon_{it} \qquad (7)$$

where the left hand side is the indicator function for the j'th learning source and Z are a number of candidate variables suggested by the indirect spillovers literature such as: MNE presence, R&D in the industry, competition and distance of the firm to the productivity frontier. This regression is estimated by probit with marginal effects reported. Since many of the learning sources are measured at the three-digit industry level, e.g. R&D, foreign MNE employment share and industry price-cost margin, we enter the industry dummies at a two-digit level. The distance to the frontier measure is the TFP of the firm at the 90th percentile less the productivity of the firm under consideration and thus is positive, with a higher measure corresponding to a greater distance from the frontier firm.[29] Finally, we can estimate (7) for all firms for whom we have complete information flow and industry data and for our sample of 804 firms only; we show both for completeness.

Table 5 reports our results. Column 1 and 2 reports results for I(L^{COMPET}), using 3,528 firms for whom we have complete information and our 804 sample above. R&D and MNE presence in the three-digit industry are both positively correlated with learning from competitors, statistically significant at 1% levels in the first column, but with only MNE presence significant at 10% in the second column.

[29] We also included a dummy for firms with TFP above the 90th percentile.

I(L^{COMPET}) is not however significantly correlated with the TFP gap or Price-cost margins in either column. All this suggests some support for the interpretation that MNE presence conveys spillovers to domestic firms, but via competition.

Columns 3, 4, 5 and 6 of Table 5 report results with I($L^{SUPPLLIER}$) as dependent variable. In column 3, on the full sample, there is a negative and statistically significant effect of our inverse competition measure on this learning source, suggesting that more competition (lower PCMs) is correlated with more learning. This seems to be the only effect of any statistical strength in these columns however. Given the interest in forward and backward learning, column 7 and 8 reports results using MNE presence weighted by the input/output table, in this case to measure backward learning.[30] As column 5 shows, the point estimate on the large sample is positive and significant, but on the small sample, negative with t=1.63.

Finally, column 7 and 8 look at the effects of I($L^{CLIENTS}$) learning from clients on forward linkages and finds them to be positively and significantly associated.

What then can we conclude from this? First, the correlations here on the bigger sample at least support the idea that MNE presence, both horizontally and vertically, are correlated with information flows from competitors, supplier and clients. Second, we also find support for the idea that R&D in the industry is correlated with more information flows from competitors. Third, linked with our findings above, this suggests positive spillovers from industry R&D and MNE presence, via competition, to TFP growth. Fourth, the path of possible spillovers via suppliers is not quite so clear. The TFP growth sample gives a positive correlation between knowledge flows and TFP growth, but the knowledge flows are only positively and significantly correlated with forward linkages from MNEs in the fuller sample. Thus we cannot be sure of the source of these spillovers if they are spillovers.

Finally whilst the general pattern of results supports the idea that MNE presence is statistically significantly linked with more learning from competitors which in turn is linked with more productivity growth, what is the economic effect? The implied effect of MNEs on ΔlnTFP is the coefficient on learning in the ΔlnTFP equation, 0.015 times the coefficient on MNE presence in this equation, 0.359 which equals 0.0054. In Haskel, Pereira and Slaughter (2002) where we did not have data on information flows, we obtained an effect of the fraction of MNE presence on ΔlnTFP of 0.055 (Table 3, lowest panel). The results here suggest, as these papers have discussed, this figure overstates the effect due to information flows, perhaps because there is also a technological effect driving both ΔlnTFP and the propensity of MNEs to situate in a particular industry. In Haskel, Pereira and Slaughter we also suggested that on that coefficient, many schemes to attract MNEs had overpaid relative to their spillover benefits.

---

[30] That is, we calculated MNE presence by industry and weight it, for a firm in industry J, by the fraction of output in the industry supplied from the other industries. The forward linkage measure is weighted by the fraction supplied to other industries.

The data here suggest spillover benefits that are almost exactly 1/10th of those in that paper, re-inforcing the conclusion that many of these schemes have overpaid MNEs.[31]

6 Conclusion

This paper has tried to model TFP growth as being due to knowledge flows as in many approaches to the knowledge production function. Using direct evidence on TFP growth and knowledge flows, it has tried to estimate

(a) which knowledge flows are the source of TFP growth
(b) what is the impact on TFP growth
(c) do such knowledge flows constitute spillovers?
(d) how do such direct measures relate to the many indirect measures in the literature (such as R&D or MNE presence in the industry).

We have done this by matching census of production data on outputs and inputs of firms and questionnaire data on knowledge flows. We have argued that the knowledge flow data complement the patenting literature and are a first step in understanding knowledge flows for non-patenting firms. In addition, the questionnaire data seems to accord with the patents literature, suggesting that R&D and information from universities is particularly strongly correlated with patenting activity. Finally, we have transformed the questionnaire response to try to better cope with respondent bias and Likert-scale type measurement error.

Our main findings on the four questions above are as follows. First, our main results are a statistically significant association between TFP growth and above-firm average information flows from: other firms in the enterprise group, competitors and suppliers. Second, such flows are economically significant as well, with such information flows "explaining" (in a growth accounting sense) about 50% of TFP growth. The effects are robust to different methods of measurement and different samples. Third, we believe that flows from competitors are spillovers, whilst flows from suppliers remain uncertain. Fourth, we find positive relations between information flows (from competitors, relative to the mean for the firm) and the industry R&D and MNEs. We find some relations between learning from clients and MNE presence in downstream industries and learning from suppliers and MNE presence in upstream industries. This supports the use of indirect measures to proxy knowledge flows but the implied effect of MNEs on productivity growth via learning is rather smaller than that estimated by indirect methods in the literature.

[31] Strictly speaking, the Community Innovation Survey only reports on learning about technologies so our estimates might be a lower bound to the true knowledge spillovers from MNEs if domestic firms also learn on management; organisation; marketing etc.

Of course, it would be preferable to have more and better data on information flows and TFP growth, and sufficient data to study absorptive capacity at the firm. However, we think this paper is at least a first step in better understanding the TFP growth residual and the associated IO and international literature on the knowledge flows that might drive it.

References

Baldwin, John, and Petr Hanel. 2003. *Innovation and Knowledge Creation in an Open Economy*: Canadian Industry and International Implications, Cambridge (MA): Cambridge University Press.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review*, 91(2), pp. 67-72.

Bloom, Nicholas and John M. Van Reenen. 2002. "Patents, Real Options, and Firm Performance." *Economic Journal*, 112, pp. C97-C116.

Cohen, W., Nelson, R. and Walsh, J., (2000), "Protecting Their Intellectual Assets: Appropriabilty Conditions and Why US Manufacturing Firms Pantent (Or Not)", NBER working paper 7552.

Crépon, B., E. Duguet and J. Mairesse (1998), "Research and Development, Innovation and Productivity: An Econometric Analysis at the Firm Level", *Economics of Innovation and New Technology*, 7(2), 115-158.

Frei, Frances X. 2004. *Rapid Rewards at Southwest Airlines*. Harvard Business School Case #9-602-065.

Gorodnichenko, Y., (2006), "Using firm optimization to evaluate and estimate returns to scale," (previous title: "Estimating Returns to Scale: Critique of Popular Estimators and New Solutions to Old Problems"), working paper available at http://www-personal.umich.edu/~ygorodni/

Griliches, Zvi (1979). Issues in Assessing the Contribution of Research and Development to Productivity Growth.' *The Bell Journal of Economics* 10: 92-116."

Heskett, James L. 2003. *Southwest Airlines 2002: An Industry Under Seige*. Harvard Business School

Case #9-803-133.

Hanson, Gordon H. 2000. "Should Countries Promote Foreign Direct Investment?" G24 paper.

Haskel, Jonathan E., Sonia C. Pereira, and Matthew J. Slaughter. 2002. "Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?" National Bureau of Economic Research Working Paper #8724.

Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the R&D-Patent Relationship." *Econometrica*, 52, pp. 909-938.

Jaffe, Adam, and Manuel Trajtenberg. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge, MA: MIT Press.

Klette, Tor Jakob and Zvi, Griliches (1996). "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," *Journal of Applied Econometrics*, vol. 11(4), pages 343-61, July-Aug.

Klette, Tor Jakob. 1996. "R&D, Scope Economies, and Plant Performance." *RAND Journal of Economics*, 27(3), pp. 502-522.

Klomp,L., and vanLeeuwen,G. (2001). "Linking innovation and firm performance: A new approach." *International Journal of the Economics of Business*, 8:343-364.

Lööf, H. and A. Heshmati (2002), "Knowledge Capital and Performance Heterogeneity: A Firm Level Innovation Study," *International Journal of Production Economics*, 76(1), 61-85.

Markusen, James R. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press.

McGinn, Daniel. 2004. "Is This Any Way To Run an Airline?" *Newsweek*, October 4 Issue.

Mintzberg, H. , (1973), The Nature of Managerial Work, Harper and Row

Merlede, B., and Schoors, K., (2006), "FDI and the Consquences", University of Gent Working Paper 2006/372.

Rodriguez-Clare, Andres. (1996). "Multinationals, Linkages, and Economic Development." *American Economic Review* 86 (4), September, pp. 852-873.

Schankerman, Mark (1981) "The effects of double-counting and expensing on the measured returns to R&D", *The Review of Economics and Statistics*, 63(3):454{458.

Smarzynska Javorcik, Beata. (2004). "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages." *American Economic Review*, 94(3), pp. 605-627.

Table 1: Characteristics of firms with TFP growth above and below the median

Growth		Lgo	Lemp	YL_ard	rd_emp	patapply	I(L^{COMPET})	I($L^{SUPPLIER}$)	I(L^{CIENTS})	I(L^{GROUP})	I(L^{UNIV})
Total	Median	9.85	5.63	4.23	0.00	0.00	1.00	1.00	1.00	0.00	0.00
	Mean	9.81	5.52	4.3	0.03	2.69	0.51	0.65	0.68	0.41	0.19
	SD	1.4	1.14	0.7	0.06	17.92	0.50	0.48	0.47	0.49	0.40
	N	804	804	804	804	804	804	804	804	804	804
Below	Median	9.68	5.56	4.13	0.00	0.00	0.00	1.00	1.00	0.00	0.00
	Mean	9.59	5.39	4.2	0.03	2.45	0.49	0.62	0.65	0.38	0.19
	SD	1.46	1.17	0.77	0.06	21.72	0.50	0.49	0.48	0.49	0.39
	N	409	409	409	409	409	409	409	409	409	409
Above	Median	10.03	5.67	4.33	0.01	0.00	1.00	1.00	1.00	0.00	0.00
	Mean	10.05	5.65	4.4	0.03	2.93	0.54	0.68	0.71	0.44	0.20
	SD	1.29	1.1	0.61	0.06	12.87	0.50	0.47	0.45	0.50	0.40
	N	395	395	395	395	395	395	395	395	395	395

Notes: Firms are allocated to TFP growth below and above the median in their three-digit industry.

Table 2: Knoweldge production functions using patents to measure new knowledge

	(1)	(2)	(3)	(4)
	RD_exp>0	RD_exp>0	RD_pers>0	All
LnRD Exp	0.330	0.307		
	(4.83)***	(4.31)***		
Ln RD Emp			0.546	0.528
			(7.43)***	(7.46)***
$I(L^{COMPET})$		-0.142	-0.121	-0.097
		(0.69)	(0.65)	(0.54)
$I(L^{SUPPLIER})$		0.009	0.224	0.204
		(0.04)	(1.22)	(1.17)
$I(L^{CLIENTS})$		0.080	-0.107	-0.119
		(0.31)	(0.50)	(0.59)
$I(L^{GROUP})$		0.135	-0.026	-0.064
		(0.69)	(0.15)	(0.39)
$I(L^{UNIV})$		0.599	0.581	0.626
		(3.00)***	(3.36)***	(3.80)***
MEAN		0.215	0.215	0.344
		(1.05)	(1.21)	(2.09)***
Observations	355	355	467	804
Number of firms	344	344	445	752

Notes: Estimates by negative binominal allowing for random effects. Equations include (not reported): dummy for firms under 20 employees, log size, industry, regional and start-up dummies. T statistics in brackets. * significant at 10%, ** significant at 5%; *** significant at 1%.

Table 3: Estimates of production functions with information flow variables

	(1)	(2)	(4)	(5)
	CDoug, OLS	CDoug, RE	TFPG, OLS	TFPG, RE
ΔlnK	0.399***	0.400***	0.017	0.017
	(0.077)	(0.051)	(0.073)	(0.046)
ΔlnM	0.440***	0.440***		
	(0.079)	(0.018)		
ΔlnL	0.106***	0.106***		
	(0.026)	(0.021)		
R&D Emp	0.131	0.131*	0.167	0.167***
	(0.094)	(0.067)	(0.103)	(0.062)
$I(L^{COMPET})$	0.015*	0.015*	0.010	0.010
	(0.008)	(0.008)	(0.008)	(0.008)
$I(L^{SUPPLIER})$	0.017**	0.017**	0.017**	0.017**
	(0.008)	(0.008)	(0.007)	(0.008)
$I(L^{CLIENTS})$	-0.007	-0.008	-0.008	-0.008
	(0.008)	(0.010)	(0.009)	(0.009)
$I(L^{GROUP})$	0.017**	0.017**	0.011	0.011
	(0.008)	(0.008)	(0.008)	(0.007)
$I(L^{UNIV})$	-0.011	-0.011	-0.002	-0.002
	(0.009)	(0.010)	(0.008)	(0.009)
MEAN	-0.008	-0.008	-0.003	-0.003
	(0.007)	(0.007)	(0.006)	(0.007)
Observations	804	804	804	804
R-squared	0.58		0.07	

Notes, * significant at 10%, ** significant at 5%; *** significant at 1%. regressions include, not reported a year dummy, a dummy for firms under 20 employees and a constant.

Table 4: Estimates of production functions with information flow variables: robustness tests (dependent variable ΔlnY)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	SH_MNE	*RD>0*	*RD==0*	*PAT>0*	*PAT==0*	*DMShare*	*Prod in*	*Patents*	*Add MNEs*
R&D Emp	0.125	0.115	0.000	0.175	0.042	0.208*	0.114		-0.005
	(0.093)	(0.099)	(0.000)	(0.173)	(0.107)	(0.111)	(0.091)		(0.075)
I(L^{COMPET})	0.015*	0.022**	0.009	0.014	0.018**	0.019**	0.014*		0.012*
	(0.008)	(0.009)	(0.014)	(0.017)	(0.009)	(0.008)	(0.008)		(0.007)
I($L^{SUPPLIER}$)	0.017**	0.007	0.031**	0.019	0.014	0.014*	0.017**		0.017**
	(0.008)	(0.010)	(0.013)	(0.020)	(0.009)	(0.008)	(0.008)		(0.007)
I(L^{GROUP})	0.016*	0.007	0.032**	0.008	0.013	0.013*	0.016*		0.013*
	(0.008)	(0.011)	(0.014)	(0.020)	(0.010)	(0.008)	(0.008)		(0.007)
Share FOR_I	0.137*								
	(0.078)								
D2Lmshare						-0.402			
						(0.497)			
MNE dummy									0.011
									(0.008)
Ln(1+Patents)								0.07	
								(0.05)	
Product in							0.021		
							(0.019)		
Observations	804	467	331	163	635	614	804	804	1081
R-squared	0.58	0.55	0.70	0.53	0.61	0.63	0.58	0.57	0.58

Notes, * significant at 10%, ** significant at 5%; *** significant at 1%. regressions include, not reported, ΔlnK, ΔlnM, ΔlnL, I(LCLIENTS), I(LUNIV), year dummy, a dummy for firms under 20 employees, the mean information response and a constant. Neither I(LCLIENTS) or I(LUNIV) were remotely significant.

Table 5: Regressions to explain deviation of information from competitiors from learning average

(dependent variable 1/0 if learning rated higher than average of all learning types. Estimates by probit, marginal effects reported)

	(1)	(2)	(3)	(4)	(5	(6)	(7)	(8)
Dependent:	$I(L^{COMPET})$	$I(L^{COMPET})$	$I(L^{SUPPL})$	$I(L^{SUPPL})$	$I(L^{SUPPL})$	$I(L^{SUPPL})$	$I(L^{CLIENT})$	$I(L^{CLIENT})$
$R\&D_I$	0.252	0.580	-0.011	0.145				
	(2.77)***	(1.60)	(0.06)	(0.45)				
MNE $share_I$	0.308	0.359	0.097	-0.139				
	(3.10)***	(1.88)*	(0.98)	(0.85)				
GAP_I	-0.006	-0.029	-0.032	-0.123				
	(0.22)	(0.29)	(1.01)	(0.98)				
Price-$cost_I$	-0.153	-0.014	-0.370	-0.224				
	(0.98)	(0.04)	(2.06)**	(0.51)				
MNE back					-0.286	0.506		
					(1.63)	(7.40)***		
MNE forw							0.344	0.788
							(1.93)*	(10.02)***
Observations	3528	804	3532	804	804	3631	804	3631

Notes: Robust z statistics in parentheses. * significant at 10%, ** significant at 5%; *** significant at 1%. Regressions include two digit industry dummies, year dummy, status change dummies and a included a dummy for firms with TFP above the 90th percentile. Estimation by probit, marginal effects reported. $R\&D_I$ is the ratio of R&D expenditure to turnover in the 3 digit industry calculated from the BERD survey. GAP_I is the TFP gap with the 90th percentile 4 digit firm zero using industry shares as cost shares and including a dummy (gapdum not reported for negative gaps). Price-$cost_I$ is unweighted 3 digit industry price cost margin. MNE share of employment in foreign MNEs in the three digit industry. All big sample regressions were on 3,631 observations but some observations were unique to the industry and were absorbed, thus the number of observations shown varies to reflect this.

Appendix 6: Estimates of production functions with information flow variables (dependent variable: ΔlnTFP)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	TFPG, SH_MNE	TFPG, RD>0	TFPG, RD==0	TFPG, PAT>0	TFPG, PAT==0	TFPG, DMShare	TFPG, PRODINOV==0	TFPG, PRODINOV	TFPG, Patents only	TFPG, Patents	TFPG,MNE
ΔlnK	0.019	0.016	0.006	-0.060	0.034	0.030	0.018	0.017	0.026	0.017	-0.019
	(0.073)	(0.077)	(0.133)	(0.168)	(0.080)	(0.054)	(0.088)	(0.074)	(0.36)	(0.073)	(0.063)
R&D Emp	0.160	0.154	0.000	0.228	0.031	0.222**	-0.123	0.168	0.137	0.138	0.011
	(0.102)	(0.110)	(0.000)	(0.167)	(0.117)	(0.110)	(0.099)	(0.103)	(1.35)	(0.103)	(0.078)
dummy	-0.052*	-0.060	-0.045	0.067	-0.053*	-0.132	-0.054*	-0.053*	-0.054	-0.051	-0.027
	(0.032)	(0.064)	(0.031)	(0.060)	(0.031)	(0.162)	(0.030)	(0.032)	(1.69)	(0.032)	(0.030)
$I(L^{COMPET})$	0.011	0.011	0.012	0.006	0.014	0.014	0.032***	0.010		0.011	0.008
	(0.008)	(0.010)	(0.014)	(0.017)	(0.009)	(0.009)	(0.012)	(0.008)		(0.008)	(0.007)
$I(L^{SUPPLIER})$	0.017**	0.014	0.022*	0.029	0.011	0.016**	0.011	0.017**		0.017**	0.017***
	(0.007)	(0.009)	(0.013)	(0.018)	(0.008)	(0.008)	(0.011)	(0.007)		(0.007)	(0.007)
$I(L^{CLIENTS})$	-0.008	-0.009	-0.010	-0.003	-0.014	-0.009	-0.021	-0.008		-0.007	0.004
	(0.009)	(0.011)	(0.015)	(0.020)	(0.010)	(0.009)	(0.013)	(0.009)		(0.009)	(0.007)
$I(L^{GROUP})$	0.010	0.002	0.023	0.017	0.006	0.015*	0.002	0.011		0.011	0.010
	(0.008)	(0.009)	(0.015)	(0.019)	(0.009)	(0.008)	(0.012)	(0.008)		(0.007)	(0.007)
$I(L^{UNIV})$	-0.003	0.004	-0.012	-0.000	-0.003	-0.005	0.003	-0.002		-0.004	-0.004
	(0.008)	(0.009)	(0.019)	(0.017)	(0.010)	(0.009)	(0.015)	(0.008)		(0.008)	(0.007)
Share FOR_I	0.176**										
	(0.072)										
MEAN	-0.003	-0.008	-0.005	-0.036*	0.007	-0.002	0.001	-0.003	0.001	-0.005	-0.008
	(0.006)	(0.009)	(0.011)	(0.019)	(0.007)	(0.007)	(0.009)	(0.006)	(0.22)	(0.007)	(0.006)
D2Lmshare						-0.435					
						(0.464)					
ΔlnM											
ΔlnL											
MNE											0.008
											(0.008)
Patents									0.006	0.006	
									(1.21)	(0.005)	
Product in							0.000	-0.001			
							(0.000)	(0.019)			
Observations	804	467	331	163	635	614	387	804	804	804	1081
R-squared	0.07	0.11	0.11	0.20	0.06	0.11	0.14	0.07	0.06	0.07	0.06

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

784	Richard Layard Guy Mayraz Stephen Nickell	The Marginal Utility of Income
783	Gustavo Crespi Chiara Criscuolo Jonathan Haskel	Information Technology, Organisational Change and Productivity Growth: Evidence from UK Firms
782	Paul Castillo Carlos Montoro Vicente Tuesta	Inflation Premium and Oil Price Volatility
781	David Metcalf	Why Has the British National Minimum Wage Had Little or No Impact on Employment?
780	Carlos Montoro	Monetary Policy Committees and Interest Rate Smoothing
779	Sharon Belenzon Mark Schankerman	Harnessing Success: Determinants of University Technology Licensing Performance
778	Henry G. Overman Diego Puga Matthew A. Turner	Decomposing the Growth in Residential Land in the United States
777	Florence Kondylis	Conflict-Induced Displacement and Labour Market Outcomes: Evidence from Post-War Bosnia and Herzegovina
776	Willem H. Buiter	Is Numérairology the Future of Monetary Economics? *Unbundling numéraire and medium of exchange through a virtual currency and a shadow exchange rate*
775	Francesco Caselli Nicola Gennaioli	Economics and Politics of Alternative Institutional Reforms
774	Paul Willman Alex Bryson	Union Organization in Great Britain *Prepared for symposium for the Journal of Labor Research on "The State of Unions: A Global Perspective"*
773	Alan Manning	The Plant Size-Effect: Agglomeration and Monopsony in Labour Markets
772	Guy Michaels	The Effect of Trade on the Demand for Skill – Evidence from the Interstate Highway System
771	Gianluca Benigno Christoph Thoenissen	Consumption and Real Exchange Rates with Incomplete Markets and Non-Traded Goods
770	Michael Smart Daniel M. Sturm	Term Limits and Electoral Accountability

769	Andrew B. Bernard Stephen J. Redding Peter K. Schott	Multi-Product Firms and Trade Liberalization
768	Paul Willman Alex Bryson	Accounting for Collective Action: Resource Acquisition and Mobilization in British Unions
767	Anthony J. Venables	Shifts in Economic Geography and their Causes
766	Guy Michaels	The Long-Term Consequences of Regional Specialization
765	Fabrice Murtin	American Economic Development Since the Civil War or the Virtue of Education
764	Carlo Rosa Giovanni Verga	The Impact of Central Bank Announcements on Asset Prices in Real Time: Testing the Efficiency of the Euribor Futures Market
763	Benjamin Aleman-Castilla	The Effect of Trade Liberalization on Informality and Wages: Evidence from Mexico
762	L. Rachel Ngai Roberto M. Samaniego	An R&D-Based Model of Multi-Sector Growth
761	Mariano Bosch	Job Creation and Job Destruction in the Presence of Informal Labour Markets
760	Christian Hilber Frédéric Robert-Nicoud	Owners of Developed Land Versus Owners of Undeveloped Land: Why Land Use is More Constrained in the Bay Area than in Pittsburgh
759	William Nickell	The CEP-OECD Institutions Data Set (1060-2004)
758	Jean Eid Henry G. Overman Diego Puga Matthew Turner	Fat City: the Relationship Between Urban Sprawl and Obesity
757	Christopher Pissarides	Unemployment and Hours of Work: the North Atlantic Divide Revisited

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7955 7595 Email info@cep.lse.ac.uk
Web site http://cep.lse.ac.uk