NBER WORKING PAPER SERIES

PRODUCTIVITY GROWTH, KNOWLEDGE FLOWS, AND SPILLOVERS

Gustavo Crespi
Chiara Criscuolo
Jonathan E. Haskel
Matthew Slaughter

Working Paper 13959
http://www.nber.org/papers/w13959

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2008

Contact: Jonathan Haskel, Queen Mary, University of London, Economics Dept, London E1 4NS, j.e.haskel@qmul.ac.uk. 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 (www.ceriba.org.uk); 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. We thank for comments Beata Smarzynska Javorcik, Lee Branstetter and seminar participants at QMUL, Sussex and Kent. Errors are, of course, our own. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Productivity Growth, Knowledge Flows, and Spillovers
Gustavo Crespi, Chiara Criscuolo, Jonathan E. Haskel, and Matthew Slaughter
NBER Working Paper No. 13959
April 2008
JEL No. F23,O47,O57

ABSTRACT

This paper explores the role of knowledge flows and productivity growth by linking direct survey data on knowledge flows to firm-level data on TFP growth. Our data measure the information flows often considered important, especially by policy-makers, such as from within the firm and from suppliers, customers, and competitors. We examine (a) what are the empirically important sources of knowledge flows? (b) to what extent do such flows contribute to TFP growth? (c) do such flows constitute a spillover of free knowledge? (d) how do such flows correspond to suggested spillover sources, such as multinational or R&D presence? We find that: (a) the main sources of knowledge are competitors; suppliers; and plants that belong to the same business group ; (b) these three flows together account for about 50% of TFP growth; (c) the main "free" information flow spillover is from competitors; and (d) multinational presence contributes to this spillover.

Gustavo Crespi
The Freeman Centre
University of Sussex
Brighton, East Sussex
BN1 9QE
United Kingdom
g.a.crespi@sussex.ac.uk

Chiara Criscuolo
Centre for Economic Performance
London School of Economics
Houghton Street
London
WC2A 2AE
United Kingdom
c.criscuolo@lse.ac.uk

Jonathan E. Haskel
Economics Department
Queen Mary, University of London
Mile End Road,
London E1 4NS
j.e.haskel@qmul.ac.uk

Matthew Slaughter
Tuck School of Business
Dartmouth College
100 Tuck Hall
Hanover, NH 03755
and NBER
matthew.j.slaughter@dartmouth.edu

1 Introduction

Total factor productivity (TFP) growth is often modelled as increases in knowledge (Griliches, 1979). Increases in knowledge, in turn, are typically 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); and (iii) use of knowledge from outside the firm.

There are 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 (e.g., Branstetter, 1998, Jaffe and Trajtenberg, 2002). A substantial literature in international economics studies the possibility that information might come from nearby multinational enterprises (MNEs) or international trade (e.g., Smarzynska Javorcik, 2004, and references therein). 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 the major public-policy question of setting subsidies for R&D and multinational firms (Hanson, 2000).

Despite the importance of these three questions, it is well acknowledged that their answers are not fully known. The main problem is that it is very hard to measure knowledge flows across firms. To illustrate the problems surrounding knowledge flows and motivate what we do, consider the concrete example of Southwest Airlines, widely regarded as having introduced many frontier innovations such as ticketless boarding. In turn, many other airlines have copied them, including Europe's Ryanair. McGinn (2004) writes, "Ryanair CEO Michael O'Leary…visited Southwest's headquarters in Dallas … At the time, Southwest was already garnering accolades as the industry's big innovator … Flying back to Ireland after a few days with Southwest, O'Leary laid plans to replicate the strategy."[2] As we now explain, many conventional knowledge-flow measures would not capture this sort of innovation.

To measure knowledge flows, the two main methods in the literature are direct and indirect. The main *direct* method is to use information in patent citations. This method has a number of advantages. For patenting firms with a sufficient time series, one can generate a measure of knowledge stock. Using citations, one can trace knowledge flows between patentees and identify their industry and geography.

---

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

[2] In turn, at key stages Southwest itself developed key innovations by learning from others: "To improve turnaround of its aircraft at airports, Southwest sent observers to the Indianapolis 500 to watch pit crews fuel and service race cars. 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).

As is well-acknowledged, however, this method suffers from some difficulties. First, as for frontier innovations, not all frontier innovations are patentable, neither are all frontier patentable innovations chosen to be patented.[3] None of the innovations in the Southwest/Ryanair example above are patented. Bloom and Van Reenen (2002) report that in their sample of 59,919 U.K. firms, just 12 companies accounted for 72% of all patents. Second, recent research has indicated that many patent citations have been included by examining officers rather than inventors themselves: typical figures are 50% in the United States (Jaffe and Trajtenberg, 2002) and 91% in Europe (Criscuolo and Verspagen, 2005). This makes patent citations a noisy measure of information flows. Third, there is some interest in knowledge flows that patent citations either cannot measure or can only do so with difficulty. For example, policymakers are often interested in possible knowledge spillovers between MNEs (to whom they often pay a subsidy) and their local suppliers or customers. Finally, it is likely that many innovations in the economy are within, not on, the patentable frontier. In sum, knowledge flows as embodied in patents are likely to be key information flows and contribute to understanding frontier innovations. But it seems valuable to examine complementary data on knowledge flows in non-patenting firms, which constitute the vast majority of companies, particularly in the light of such problems above.

The *indirect* method of measuring knowledge flows typically regresses TFP growth on factors thought to be potentially causing information flows, such as the presence of multinational enterprises (MNEs) or international-trade status. This method does have the advantage of implicitly looking at non-patentable innovations and effects exemplified in case studies (such as information flows from suppliers) and it does try to speak to a number of prominent policy issues, such as subsidies to MNEs. However (quite apart of issues of measurement, simultaneity etc.) the problem with evidence based on TFP is that it is indirect and thus is consistent not just with information flows but with other likely effects as well.[4]

This paper brings new evidence to bear on these questions by linking direct data on knowledge flows with data on TFP growth. The TFP growth data are derived from firm-level business surveys conducted by the U.K. 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. These features minimize measurement problems of non-response and recall bias; that said, we shall explore robustness issues extensively.

The knowledge-flow data come from the Community Innovation Survey (CIS), an official EU-wide survey that asks business enterprises to report innovation outputs; innovation inputs; and, most

---

[3] 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.

importantly for our paper, sources of knowledge for innovation efforts. We use the second and third waves of the CIS.[5] For our purposes the major feature of the CIS is that it asks firms about their R&D and also their knowledge flows.[6] 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, and universities. The major advantages of these data are that they (a) directly measure many of the information flows that economists and policy-makers have identified as important, and they (b) cover all firms, not just patenting firms.

A feature of these data is that they are qualitative. As such, they may be thought to be inferior to hypothetical 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 hypothetical data would then have to be weighted by importance, which would introduce some qualitative element.[7] Following the patents literature, we shall try to correct for errors in measuring the "true" information flows. Perhaps most importantly, for each firm we re-scale each knowledge-flow variable to be the deviation from the average importance of all knowledge flows reported by the firm. With this transformation a firm that reports "very important" to all measures, simply because a respondent ticks all boxes in a column (e.g., because of hubris or true technological fecundity), scores zero for all flows. Thus we shall *not* correlate TFP growth with firms reporting high knowledge flows from suppliers. Rather, we shall correlate TFP growth with firms reporting that knowledge flows from suppliers are more important than the average of all their other knowledge flows.

In our analysis, we first 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 a first check on our data, which helps locate it in the literature, is to use patents as the measure of increased knowledge and so regress it on R&D and our measures of knowledge flows.[8] We find sensible

---

[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. We discuss this more below.

[6] We use too the data on patents applied for. We do not use the self-reported innovation question. A large number of papers have used the CIS self-reported innovation output data, see e.g. Hall and Mairesse (2006) for a survey. For our purposes here it is hard to interpret these data in terms of spillovers.

[7] For example, data collected by social network analysis (SNA) in management studies attempt to show informal relationships: who knows whom and who shares information and knowledge with whom. SNA typically gathers data about the relationships between a defined group/network of people with using questionnaires and/or interviews or software that tracks 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. In a recent paper on within-firm information flows, Cowgill, Wolfers and Zitzewitz (2008) use the location of trader's desks and a questionnaire on knowledge sources.

[8] Due to confidentiality we do not have information on the company names and so cannot link the data to patent citations data. Duguet and MacGarvie (2005) merge French CIS data with firms for whom they have patent citations and demonstrate a positive

results: patents are strongly associated with R&D (with a coefficient that is similar to other studies) and also with information flows from universities.[9] To the extent that university knowledge flows contribute to knowledge advances at the technological frontier, these seem sensible.

The heart of our analysis is to estimate the relation between TFP growth and both R&D and knowledge flows. This measures knowledge creation more broadly than patents and, under certain conditions, allows one to identify these information flows as spillovers. Our main result is a statistically significant association between TFP growth and information flows from three sources: other firms in an enterprise group, competitors, and suppliers. These effects are economically significant as well: taken together, these three information flows account for nearly 50% of observed variation in TFP growth. They are also robust to different measurement choices and different samples.

We then discuss if such knowledge flows are spillovers. We have no data on the prices that firms pay, if any, for information flows. We reason that information flows from competitors are likely to be uninternalised spillovers (Baldwin and Hamel, 2003, p.53 argue similarly); that information flows from within the company group are likely internalised; and that information flows from suppliers may or may not be spillovers.

Finally, we relate our direct measures of information flows with earlier indirect work. Specifically, we correlate our firm-level knowledge-flow data with industry-level R&D, MNE presence, competition, and TFP gaps. We think this is of interest since these proxies have often been related to TFP growth in the absence of data on the underlying information flows they purport to represent. A widely acknowledged potential problem of these indirect measures is that the relation between, say MNE presence and TFP growth could be due not just to information flows but also underlying technological factors that both boost TFP growth and cause MNEs to be present.[10] We find that our direct knowledge-flow measures from competitors have some correlation with MNE presence, but not with other indirect measures commonly used in the literature. This lends further support to the value of our direct measures of knowledge flows, and offers a general caveat to at least some earlier findings, which we discuss.

How does our paper relate to other work using CIS data? Of those that study productivity, few link to production data as we do here and use TFP growth.[11] Of this work, most modelling uses the Crepon, Duguet and Mairesse (CDM, 1998) model; see also Loof and Heshmati (2002) and Hall and

---

correlation between citations and reported knowledge flows. The knowledge flow "analyzing competing products" is correlated with exporting and importing behaviour, see MacGarvie (2006).

[9] Strictly, information flows from universities that were rated as more important than the average of all information flows.

[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.

[11] The CIS questionnaire asks for sales and employment, but TFP growth cannot be computed. Thus many papers e.g. Mairesse and Mohen (20002) use the fraction of sales reported to be innovative as their output measure.

Mairesse (2006) for a survey. That model consists of a three equation system linking (i) productivity levels to self-reported innovation, (ii) self-reported innovations to inputs, such as R&D and information flows (iii) R&D to its conjectured determinants (such as size). A substantial part of the research effort is to account for the selection induced by the questionnaire design in many continental European countries, whereby non-innovators do not respond to most questions. Our work differs in a number of regards. First, U.K. questionnaire design means our U.K. data do not face this selection problem. Second, the CDM (1998) paper does not look at information flows, but rather the importance of "demand pull" and "technology push" (measured by asking firms to rate these factors' importance, see also Klomp and Van Leeuwen, 2001).[12] And third, we focus on TFP growth, not productivity levels.

To the best of our knowledge there are almost no papers that link TFP growth with information flows (and none with the transformation that we use).[13] Loof and Heshmati (2002) find a positive relation between labour productivity growth and self-reported innovation output and in turn self-reported innovation output is correlated positively with information from within the firm and from competitors, but negatively with information from customers. Loof and Heshmati (2006, Table VIII) find that, for services, innovation output is positively correlated with information from conferences, suppliers and via co-operation with others, but negatively correlated with within-firm sources. Van Leeuwen and Klomp (2006) study information flows indirectly by collapsing information from various sources. These variables along with others determine whether a firm innovates, which is then a determinant of its productivity growth (see their Appendix B).[14]

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 relations with multinational status, R&D presence and other common proxies for information sources. Section six concludes.

---

[12] Crepon et al (1998, p.142) use answers to the questions "Do you consider that in your firm innovation is determined through the impetus given by the market (relationships with customers, competitors" and "Do you consider in your firm innovation is determined by technology specific dynamics". Klomp and van Leeuwen (2001) use a similar question, see also Beneavente (2006) for Chile.

[13] Cassiman and Veugelers (1999, 2002) look at aggregated knowledge flows on and make/buy innovation decisions and co-operation in R&D ventures.

[14] Baldwin and Hanel (2003) also find information flows from competitors to be important self-reported sources of innovation, but do not use TFP growth data. Levin, Klevorick, Nelson, and Winter (1987) find, using a questionnaire, competitors inventing around patents to be an important reason for ineffectiveness of patents. This accords with our finding that learning from competitors matters.

2 Theory and 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)? 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) to regress $T\dot{F}P$ on MNE presence. The distance-to-frontier literature (e.g., Griffith, Redding and van Reenen, 1996) 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 (e.g., Jaffee, 1986) postulates that $A'_{\_i}$ can be measured by R&D or knowledge stocks outside the firm (e.g., in the industry) 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 production survey data on TFP growth. Second, with these data we begin by estimating (2) with patents as a dependent variable (the CIS 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 most work that has used the CIS. Due to data availability, almost no papers have matched the CIS with TFP growth 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 presence and other variables commonly used as indirect learning measures.

2.2 Data Description

We shall use two sets of data. To measure TFP growth we use production survey data from the official U.K. Annual Respondents Database from the Annual Businesses Inquiry. Data on knowledge and information flows come from the Community Innovations Survey.

2.2.1 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.[15]

There are a number of points to be noted. First, to 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 and machinery, buildings, and materials). To build capital stocks from investment we applied a perpetual-inventory method.

---

[15] 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).

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.[16] 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. [17] 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.

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 is our key variable of interest: information flows, for which the CIS question and form looks as follows.

"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				

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

[17] 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.

	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				

Column headers are N (not used) and L, M, H for, 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 above or in Criscuolo, et al, 2005). Second, modern theories assume that MNEs can better transfer knowledge within the enterprise than other like firms (Markusen, 2002). Thus a direct check on the data would be to see if knowledge flows among MNE plants in an enterprise group to be greater than flows among plants in a purely domestic enterprise group. This is indeed the case (Criscuolo, et al, 2005). Third, these data have a very high response rate on the CIS surveys.

One broad objection to these measures 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.[18] 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 (e.g., dues to a trade fair). We look at TFP rather than labour productivity and then review the conditions under which this should capture paid-for inputs.

Given these measurement concerns, 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

[18] One could try to collect employee time-use diaries to try to measure learning time from different sources. There are general time-use surveys such as 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 Mintzberg (1973) for the classic study (of nine managers).

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

which says that measured information flows are related to true information flows, plus some firm/time-varying variables Z , plus an error term.

A first source of measurement error that might drive a wedge between $A^{j\prime}_{it}$ and L^{j}_{it} is 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 against finding any significant effect of learning and biasing down included coefficients via attenuation bias. To the extent that it is fixed, we can control for this by using the two CIS cross-sections in a panel or by our firm-specific averaging we describe below. To the extent that it varies systematically with firm features (e.g., size), we can control for this with other observables.

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 replies to control for possible level differences in their opinions about what constitutes L, M, and H. There may also be the related problem of respondents tending to give similar answers across information flows: e.g., because of managerial hubris that a firm is an excellent learner.

In the empirical analysis we correct for this bias and for firm-specific effects using the following approach. We start by calculating the average of each firm's reported learning from all 17 information sources, $\overline{L}^{j}{}_{it}$ (converting the responses N, L, M, and H into 0, 1, 2, and 3). We then expressed each of the learning variables as its deviation 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*

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

where the average of all replies to the j questions by firm i is $\overline{L}^{j}{}_{it}$. With equation (4), a firm that always ticks the same box is allocated a zero on all information measures. A firm that ticks all box L, but on one information source ticks H, would have a 1 for that source and a zero for all others.

With the data transformation in equation (4) we aim to control for any unobservable firm effect in Z_{it} (and we will use $\overline{L}^{j}_{it}$ as a control as well). 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 I(L^j_{it}). Thus for there to be an omitted common variable that explained the relation between $T\dot{F}P$ and I(L^j_{it}), it would *not* be, e.g., that more aggressive managers both have better TFP growth and also report more aggressively on the CIS. Rather, it would have to be that an omitted variable causes 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 exclude this possibility, but it does seem unlikely to be a systematic feature of our data.

2.3 Summary Statistics and Sample Selection

The matched data set consists of 804 manufacturing observations that 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, which we exclude initially). These 804 firms consist of 752 firms that appear in either CIS2 and CIS3, plus 26 observed twice. The reason for this small panel element is that the CIS was not designed to be a panel across waves. 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, describes our full data.[19] Let us concentrate on the innovative activity figures: as in other data sets discussed earlier, 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. Turning to the learning data (transformations of the raw survey as in equation (4)), the median firm also does no learning, from any source. However, learning is less skewed than R&D and patents. 51% of our firms report learning from competitors more intensively than the average of learning from all sources (a share that, per equation (4), is smaller than fraction that 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) TFP growth firms (with growth measured 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. These split-sample differences suggest that TFP differences may indeed related to differences in information flows.

[19] We also analysed how the regression sample differs from the complete ARD and CIS samples. Relative to the ARD, our sample is bigger in terms of gross output, employment and productivity levels (but the standard deviation of these numbers is large). Relative to the CIS, our 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.

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.[20] First, it might be that marginal effects differ due to technology. If this is the case, then to the extent our sample is of large firms 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.[21]

3 Econometric Framework and Estimation Results

3.1 Using Patents to Estimate the Knowledge Production Function

Before turning to TFP growth at the heart of our analysis, we start by estimating the knowledge production function in equation (2) using patents as a measure of $\dot{A}$. Our purpose here is to (a) check our data against other studies 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, relative to other knowledge sources, to be important. Thus, we first implement (2) as follows

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

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; and we include fixed effects for region, two-digit industry, firm, and status (dummies measuring if they have recently merged and increased sales, merged and decreased sales, or are a start-up). We include status in all results below, in case adjustment costs might obscure the long-term relation between inputs and outputs. We show the L variables as lagged,

---

[20] 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.

[21] 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.

since they refer to learning over the past two years. The dependent variable is a count of patents, and so we used a negative binomial model with random effects (we rejected the Poisson model).

Table 2 reports our results. We start by replicating the regression of Griliches, Hall, and Hausman (1984), who on a sample of U.S firms performing R&D regressed patents applied for on (log) R&D expenditure, size and industry dummies. They obtained an elasticity of R&D expenditure of 0.45 (see their column 4, row 6 of Table 6). As column 1 shows, we obtained an elasticity of 0.33. This, then, seems like a good cross check on our data.

In column 2, we add the information-flow variables. We included information flows from other enterprises in the group $I(L_i^{GROUP})$, suppliers $I(L_i^{SUPPLIERS})$, competitors $I(L_i^{COMPET})$, clients $I(L_i^{CLIENTS})$, and universities $I(L_i^{UNIV})$. Learning from universities is positive and significant; to the extent 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 proxy for research investment with log number of persons employed in R&D at the firm, which, as discussed below, 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 expands the sample to all firms besides those doing R&D. Note that the coefficients 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 our data, and so we now turn to investigate $\dot{A}$ measured as TFP growth.

3.2 Using TFP Growth to Estimate the Knowledge Production Function

We next move to our TFP growth regressions, which form the core of our analysis. We start by estimating a simple Cobb-Douglas equation

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

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 indicator function for the deviation from the mean of

[22] As stated above, this variable is better measured than R&D expenditure. However, we found a few small firms that 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. We would like to adjust these but we only have R&D employment 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. Finally, the

learning from the j'th information sources. The following points regarding measurement and causality are worth noting. First on measurement, as is commonly the case we do not have firm-specific prices and hence the revenues deflated by industry prices is on the left hand side, relegating the log difference between firm and industry prices into ε (Klette and Griliches, 1996). We accordingly use product innovation as a proxy for this. We also look at possible input-specific prices, see below. Second, the specification in (6) implicitly constrains the αs to be the same, but we also experiment with TFP as a dependent variable, which controls for this.

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. Again, our data transformation in equation (4) reduces potential biases. The α coefficients could be biased too, but they are not our focus; we do experiment with ΔlnTFP as our regressand, which potentially removes much of this bias.[23]

Table 3 sets out our estimation results for equation (6). For convenience, we do not report the coefficients on the inputs or $I(L^{CLIENTS})$ and $I(L^{UNIV})$, as estimates for neither were significant in our data.[24] Looking at column 1, the R&D term is 0.13, in line with other studies, but not particularly precisely estimated. Turning to the information terms, there are positive and statistically significant effects from competitors $I(L^{COMPET})$, (significant at 10%), suppliers $I(L^{SUPPLIERS})$ and the enterprise group $I(L^{GROUP})$. This is the central finding of our paper: information flows from these three sources are associated with higher TFP growth.

We look at the economic significance of this finding below, but first we examine its robustness. Column 2 adds foreign MNE presence measured as the share in the two digit industry of foreign employment (recall our sample is only U.K. firms). This is positive and significant, at 0.13, without affecting the significance or coefficients of the main 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. We return to the meaning of this correlation below.

---

correlation between R&D employment levels in both periods is 0.85, suggesting relatively small changes in employment and so our results should not been too badly affected.

23 Because of just two cross-sections and a small panel element to our data, and because of the doubts summarised in Gorodnichenko (2006), we do not report results from IV-type estimators.

24 The coefficients (t-statistics) on the ΔlnK, ΔlnM, ΔlnL, $I(L^{CIENTS})$ and $I(L^{UNIV})$ terms were, respectively, 0.40 (5.22), 0.44 (5.64) and 0.11 (4.11), -0.007(0.87) and –0.011 (1.24). The t-statistics are clustered by firm to account for the (small number of) firms that appear in the two cross-sections; the unclustered t-statistics are very similar.

Columns 3 to 7 show robustness checks that do not much affect the information flow terms. Column 3 adds a pressure of competition measure, the change in market share, lagged two periods (Nickell, 1996) to test whether the competitors regressor is capturing information flows or some other pressure of competition. Change in market share is negative but insignificant, suggesting that falls in market share are (statistically) weakly associated with increased ΔlnTFP two periods later. Importantly, the I(L^{COMPET}) term is unaffected. Thus it would seem that I(L^{COMPET}) is not just proxying for firms facing more competitive pressure. Column 4 adds the product innovation term, but this again does not affect the learning terms. Column 5 adds a control for whether firms report joint ventures or other research cooperation with their competitors. Again, this does not affect the I(L^{COMPET}) term, suggesting that it is not just picking up research co-operation (which still might involve spillovers of information of course). Column 6 expands the sample to include foreign MNEs (with a dummy for these MNEs). This adds another 277 firms and the significance of the information-flow variables remains.

Column 7 uses Δln*TFP* as the dependent variable. Whilst the I($L^{SUPPLIERS}$) effect remains statistically significant, the I(L^{COMPET}) and I(L^{GROUP}) effects become less significant, although the magnitudes are hardly changed. Measurement error in firm-level cost shares , non-Dixit Stiglitz demand functions, or imperfect competition could mean that measured Δln*TFP* is more noisy than estimated output elasticities.

Table 4 splits our sample into firms who do and do not do the following: undertake R&D, apply for patents, and export. The most general form of the estimates would be to have multiple interactions but we have simply too small samples to have any precision here. Thus these results should be regarded as suggestive. There is some suggestion that the I(L^{COMPET}) effect is strongest for R&D performing firms, firm who do not patent, and firms who export: hinting that learning from competitors comes most from firms with the absorptive capacity to do so. The most significant I($L^{SUPPLIERS}$) and I(L^{GROUP}) effects come from non-R&D and non-exporters. More-detailed exploration of these effects awaits more data.

We conclude from Tables 3 and 4 that there exists a significant correlation between TFP growth and I(L^{COMPET}), I($L^{SUPPLIERS}$), and I(L^{GROUP}). We cannot completely rule out endogeneity bias, but as emphasised above, this bias would have to depend on some unobserved variable (e.g., management talent) that affects productivity growth (not its level), that is not captured in other regressors, and that also affects the deviation of learning from firm averages (not the average itself). We find this nature of bias unlikely, but do point out that it would likely imply that the true learning impacts of interest are smaller than the coefficient estimates in Tables 3 and 4.

3.3 Economic Significance of Our Key Estimates

Tables 3 and 4 have documented the statistical significance of knowledge flows on productivity growth. What is the economic significance? The coefficients on our key learning variables enable us to read directly the TFP-growth gains: approximately 1.5% from I(L^{COMPET}), 1.5% from I($L^{SUPPLIERS}$), and 1.7% I(L^{GROUP}). This means that a firm learning from all these three sources (relative to the average) enjoys TFP growth that is 4.7 percentage points faster than a firm without such learning, ceteris paribus.

A straightforward yardstick against which to judge this magnitude is the inter-quartile range of ΔlnTFP in our sample. This range is 9.9%: the firm at the 75th percentile of the distribution of TFP growth has annual TFP growth that is 9.9 percentage points higher than the firm at the 25th percentile of this distribution.[25] The sum of our key learning effects, 4.7 percentage points, can thus account for nearly 50% of the observed raw variation in TFP growth. So if the ex ante measure of our ignorance about what drives TFP growth is 9.9 percentage points, our key learning variables reduce this ignorance by nearly half. In light of the long-standing effort in many literatures to explain TFP performance, we regard our results here as potentially very important.

4 Are These Knowledge Flows Spillovers?

Can the information flows I(L^{GROUP}), I($L^{SUPPLIERS}$), and I(L^{COMPET})be considered knowledge 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 conditions 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}). It seems highly unlikely that competitor information would be paid for. Unless there is a joint venture occurring (which we have controlled for) it seems hard to think of a mechanism by which companies would pay competitors for information. Thus it seems reasonable to conclude that this is indeed a spillover of information.[26] For a similar argument, see Baldwin and Hanel (2003).

Consider second I(L^{GROUP}), knowledge flows from other enterprises in the enterprise group. That such flows affect TFP growth is 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

---

[25] 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.

[26] Whether this is Pareto-improving in a general equilibrium sense depends upon the mechanism by which information from competition affects, e.g., managerial effort (see, e.g., Vickers, 1996).

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).

Third, consider $I(L^{SUPPLIERS})$. 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 than 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.[27] 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 mismeasurement of firm-specific output prices (e.g., Klette and Griliches, 1996). Without further data on company-specific input prices, which are almost never available in micro data like ours, this has to remain a caveat over our (and perhaps others') results.

---

[27] 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

5 Relating Our Direct Knowledge Measures to Indirect Alternatives

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 use proxies such as industry R&D or MNE presence. 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 such that a correlation between MNE presence and productivity growth is not 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 that, of course, 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.[28] 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 then our 804-firm 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

---

$M^* = P^M M / P_1^{\ M}$. 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).

column. 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, via competition.

The next columns of Table 5 report results with I($L^{SUPPLIER}$) as regressand. 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 vertical spillovers, columns 5 and 6 report results using MNE and R&D presence weighted by the input/output table, in this case to measure learning via the supply chain.[29] Neither effect is significant, suggesting either no statistically significant influence on these data over this time period, or perhaps poor measurement of the vertical "distance" to these learning flows.

What can we conclude? 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. Second, we also find support for the idea that R&D in the industry is correlated with more information flows from competitors, at least in the larger sample. Third, both these results suggest 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. There is some suggestion of greater competition in the industry raising learning from suppliers, but the statistical significance of this effect is confined to the larger sample.

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 magnitude? The implied magnitude 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. This magnitude is quite small in absolute terms, and also is smaller than the indirect estimates in the literature (e.g., Haskel, Pereira, and Slaughter, 2007). As earlier studies acknowledged, this magnitude suggests that earlier studies with indirect knowledge-flow measures may have mis-estimated learning effects because of unobserved factors (e.g., technological innovation) driving both ΔlnTFP and the propensity of MNEs to situate in a particular industry.[30]

---

[28] We also included a dummy for firms with TFP above the 90th percentile.
[29] 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. That fraction is calculated using the input/output tables which are collapsed from product to industries using the conversion concordance supplied with the tables. The tables are the 2000 tables, consistent with the 2005 Blue Book, downloaded from < http://www.statistics.gov.uk/about/methodology_by_theme/inputoutput/archive_data.asp>.
[30] 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

6 Conclusions

This paper has tried relate TFP growth with knowledge flows, with our main innovation being direct rather than indirect measures of these knowledge flows. Our analysis addressed

(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 our direct measures relate to the many indirect measures in the literature?

One of the main contributions of the paper was to match census-of-production data on outputs and inputs of firms with questionnaire data on knowledge flows. We have argued that the knowledge-flow data extend the patenting literature by offering an important step towards understanding knowledge flows for non-patenting firms, which in reality constitute nearly all firms. We have transformed our raw questionnaire responses to address respondent bias and measurement error.

Our main findings on the four questions above are as follows. First, we estimate a statistically significant association between TFP growth and above-firm average information flows from three sources: other firms in the enterprise group, competitors, and suppliers. There is less effect from universities and clients, though the patent results suggest universities are statistically significant sources of information flows for patenting. Second, these information flows are economically significant: they account for nearly 50% of TFP growth. The effects are robust to many different methods of measurement and different samples. Third, we argued that flows from competitors are spillovers, whilst flows from suppliers remain uncertain. Fourth, we found that of our direct knowledge-flow measures, competition is most robustly correlated with one of the indirect measures commonly used in the literature, namely MNE presence. This lends further support to the value of our direct measures of knowledge flows, and offers a general caveat to at least some earlier findings.

Taken together, we think our results help illuminate productivity growth and the information flows that are widely thought to help drive it. Future work will hopefully build on our results with more and better data—e.g., on the absorptive capacity of firms.

---

etc. Note too, that the presence of MNEs has an independent effect on company TFP even when controlling for information flows and if this is another spillover, perhaps via information flows on non-technological factors, then this would boost the economic effect of MNEs.

References

Aghion, Philippe, Nick Bloom, Richard Blundell. Rachel Griffith and Peter Howitt, "Competition and innovation: an inverted U relationship." 2005. *Quarterly Journal of Economics*, Vol. 120, No. 2, pp. 701-728.

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

Benavente, Jose Miguel. 2006. "The role of research and innovation in promoting productivity in Chile." *Economics of Innovation and New Technology*. 15(4), pp. 301-315.

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.

Branstetter, Lee. 1998. "Looking for International Knowledge Spillovers: A Review of the Literature with Suggestions for New Approaches," *Annales d'Economie et de Statistique* No. 49/50, 1998, pp. 517-540.

Cassiman, B., and Veugelers, R. 2002. "R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium," *American Economic Review*, vol. 92(4), pp. 1169-1184.

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

Cowgill, B, Wolfers, J and E Zitzewitz. 2008. "Using Prediction Markets to Track Information Flows: Evidence from Google," working paper, available from http://bocowgill.com/GooglePredictionMarketPaper.pdf.

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.

Criscuolo, Chiara, Haskel, Jonathan E., and Matthew J. Slaughter. 2005. "Global Engagement and the Innovation Activities of Firms." NBER Working Paper #11479.

Criscuolo, P. and Verspagen, B. 2005. "Does it matter where patent citations come from? Inventor versus examiner citations in European patents." Research Memoranda 017, Maastricht : MERIT, Maastricht Economic Research Institute on Innovation and Technology.

Duguet, E., and M. MacGarvie. 2005. "How Well Do Patent Citations Measure Flows of Technology?" *Economics of Innovation and New Technology*, vol. 14:5.

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/

Griffith, Rachel; Redding, Stephen, and Van Reenen, John. 2004. "Mapping the two faces of R&D: productivity growth in a panel of OECD industries." *Review of Economics and Statistics*.

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

Hall, Bronwyn H. and Mairesse, Jacques. 2006. "Empirical Studies of Innovation in the Knowledge-Driven Economy." *Economics of Innovation and New Technology*. 15(4):289-299.

Haskel, Jonathan E. 2005. "Matching the CIS and BERD", working paper.

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. 2007. "Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?" *Review of Economics and Statistics*, 89(2), 2007, pp. 482-496.

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. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value." *American Economic Review*, 76(5), pp. 984–1001.

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*, 11(4), pp. 343-61.

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

Klomp, L., and van Leeuwen,G. 2001. "Linking Innovation and Firm Performance: A New Approach." *International Journal of the Economics of Business*, 8:343-364.

Levin, R.C., Klevorick A., Nelson, R.R., and Winter, S.G. 1987. "Appropriating the Returns from Industrial R&D," *Brookings Papers on Economic Activity*, 3, pp.783-820.

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

Lööf, Hans and Almas Heshmati. 2006. "On the Relationship between Innovation and Performance: A Sensitivity Analysis." *Economics of Innovation and New Technology*. 15(4), pp. 317-344.

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

MacGarvie, Megan. 2006. "Do Firms Learn from International Trade?" *The Review of Economics and Statistics*, 88(1).

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 Consequences", 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), pp. 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.

Van Leeuwen, George and Luuk Klomp. 2006. "On the Contribution of Innovation to Multi-Factor Productivity Growth." *Economics of Innovation and New Technology*, 15(4), pp. 367-390.

Veugelers, R., and Cassiman, B. 1999. "Make and Buy in Innovation Strategies: Evidence from Belgian Manufacturing Firms," *Research Policy*, Elsevier, 28(1), pp. 63-80.

Vickers, John. 1995. "Concepts of Competition." *Oxford Economic Papers*, 47(1), pp. 1-23.

Table 1
Characteristics of Firms:
Total Sample, Above-Median TFP Growth, and Below-Median TFP Growth

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
Explaining Patents Applied For

	(1)	(2)	(3)	(4)
	RD_exp>0	RD_exp>0	RD_pers>0	All
Ln RD 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: This table reports estimates of equation (5), where the dependent variable is patents applied for. Estimates are reported for 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 reported in brackets. * significant at 10%, ** significant at 5%; *** significant at 1%.

Table 3
TFP Growth and Knowledge Flows,
Baseline Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	basic	MNE share	competit	prod inn	co-op	MNE in sample	TFP
R&D/Emp	0.13	0.12	0.14	0.11	0.12	-0.01	0.17
	(1.39)	(1.34)	(1.85)*	(1.25)	(1.36)	(0.16)	(1.66)*
$I(L^{COMPET})$	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(1.88)*	(1.93)*	(2.13)**	(1.83)*	(1.81)*	(1.69)*	(1.36)
$I(L^{SUPPLIER})$	0.02	0.02	0.01	0.02	0.02	0.02	0.02
	(2.01)**	(2.02)**	(1.83)*	(2.01)**	(2.05)**	(2.37)**	(2.24)**
$I(L^{GROUP})$	0.02	0.02	0.01	0.02	0.02	0.01	0.01
	(1.97)**	(1.90)*	(1.83)*	(1.94)*	(2.02)**	(1.87)*	(1.48)
MNE share		0.13					
		(1.67)*					
Competition			-0.40				
			(0.81)				
Product innov				0.02			
				(1.15)			
Co-operation					-0.00		
					(0.02)		
MNE dummy						0.01	
						(1.44)	
Observations	804	804	614	804	804	1081	804
R-squared	0.58	0.58	0.66	0.58	0.58	0.58	0.06

Notes: This table reports estimates of equation (6), where the dependent variable is change in log output. Robust t statistics in parentheses, clustered. * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions include, not reported, ΔlnK, ΔlnM, ΔlnL, I(LCLIENTS), year dummy, a dummy for firms under 20 employees, the mean information response and a constant. The sample excludes foreign MNEs, except for column 6. The specification in column (7) differs in that its regressand is calculated TFP growth (see text for details).

Table 4
TFP Growth and Knowledge Flows, Additional Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	basic	R&D=0	R&D>0	Pat=0	Pat>0	EXP=0	EXP>0
R&D/Emp	0.13		0.11	0.04	0.18	-0.23	0.08
	(1.39)		(1.15)	(0.35)	(1.02)	(0.63)	(0.78)
$I(L^{COMPET})$	0.01	0.01	0.02	0.02	0.01	0.01	0.02
	(1.88)*	(0.58)	(2.30)**	(1.87)*	(0.83)	(0.34)	(2.24)**
$I(L^{SUPPLIER})$	0.02	0.03	0.01	0.01	0.02	0.05	0.00
	(2.01)**	(2.32)**	(0.66)	(1.54)	(0.97)	(2.96)***	(0.36)
$I(L^{GROUP})$	0.02	0.03	0.01	0.01	0.01	0.05	0.01
	(1.97)**	(2.24)**	(0.66)	(1.43)	(0.36)	(2.23)**	(0.69)
Observations	804	337	467	641	163	223	581
R-squared	0.58	0.71	0.55	0.61	0.53	0.61	0.64

Notes: This table reports additional estimates of equation (6), where the dependent variable is change in log output. Robust t statistics in parentheses, clustered. * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions include, not reported, ΔlnK, ΔlnM, ΔlnL, I(LCLIENTS), year dummy, a dummy for firms under 20 employees, the mean information response and a constant.

Table 5
Correlates with Knowledge Flows

	(1)	(2)	(3)	(4)	(5)	(6)
	COMPETITORS	COMPET.	SUPPLIERS	SUPPL.	SUPPLIERS	SUPPL.
(R&D Inten)(I)	0.25	0.56	-0.01	0.15		
	(2.77)***	(1.49)	(0.06)	(0.44)		
MNE share (I)	0.31	0.36	0.10	-0.15		
	(3.10)***	(1.92)*	(0.98)	(0.91)		
Gap (i)	-0.01	-0.03	-0.03	-0.13		
	(0.22)	(0.30)	(1.01)	(1.02)		
Price-cost marg (I)	-0.15	-0.02	-0.37	-0.23		
	(0.98)	(0.05)	(2.06)**	(0.53)		
MNE vertical (I)					-0.76	-0.04
					(1.44)	(0.05)
R&D/Emp vertical (I)					4.50	-3.30
					(1.48)	(0.73)
Observations	3,528	804	3,252	804	1,872	804

Notes: This table reports estimation results for equation (7), where the dependent variable is learning from the source indicated in the column heading. Marginal effects are reported for probit estimates, with 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. Estimation by probit, marginal effects reported. R&DI is the ratio of R&D expenditure to turnover in the 3 digit industry calculated from the BERD survey. GAPi is the TFP gap between firm i and the 90th percentile 4 digit firm zero using industry shares as cost shares (we also included a dummy for firms above the 90th percentile, not reported). Price-costI is unweighted 3 digit industry price cost margin. MNE share of employment in foreign MNEs in the three digit industry. Price-cost marg (I) is the industry level unweighted price cost margin, an inverse measure of competition used by Aghion et al (2005). The “vertical” terms are the industry measures of MNE shares and R&D intensity weighted by the input/output relation between the industry the firm is in and MNE and R&D presence in all other industries, see text. 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, sample falls in column 5 since only manufacturing data available. Column 1-4 all include two digit industry effects., columns 5 and 6 do not since the vertical terms are only available at two digit level due to the difficulties of making the IO table compatible with three-digit industries and the severe inaccuracies in the IO coefficients at the level of disaggregation.