Technology sourcing: an empirical analysis using firm-level patent data*

Rachel Griffith[†], Rupert Harrison[†‡] and John Van Reenen[§]

January 30, 2004

Abstract

This paper examines whether UK firms that locate their R&D activity in the USA benefit more than other UK firms from knowledge spillovers originating from US R&D. Patent data provides a firm-level measure of the location of innovative activity, which enables the identification of knowledge spillovers associated with "technology sourcing". Our identification strategy helps to address various problems usually encountered when measuring knowledge spillovers. We find evidence for knowledge spillovers associated with technology sourcing. Our main results suggest that the increase in the US R&D stock in manufacturing over 1990-2000 was associated with on average a 4% higher level of TFP for the UK firms in our sample. This compares with an average 6.5% higher level of TFP associated with the increase in their own R&D stocks over the same period.

JEL No. O32, O33, F23

Keywords: knowledge spillovers; technology sourcing; productivity

---

* **Acknowledgement:** The authors would like to thank Nick Bloom, Steve Bond and Michele Cincera for helpful comments. Financial support for this project was provided by the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the IFS. The data was developed with funding from the Leverhulme Trust.

[†]Institute for Fiscal Studies and University College London

[‡]Corresponding author (rupert.harrison@ifs.org.uk)

[§]Centre for Economic Performance and London School of Economics

1. Introduction

This paper examines whether UK firms that locate innovative activity in the USA benefit more than other UK firms from knowledge spillovers originating from US R&D. Several recent studies have found that gaining access to new technologies is an increasingly important reason for firms to locate R&D abroad, and that, as the technological leader in many industries, the USA is one of the principal recipients of this kind of R&D investment by subsidiaries of foreign firms.[1] Evidence that knowledge spillovers are partly geographical in scope provides a rationale for such 'technology sourcing' behaviour in order to overcome geographical barriers.[2] In this context the flow of knowledge from foreign R&D subsidiaries of domestic multinationals back to the domestic economy may play an important role in the diffusion of new technologies and productivity growth.

This has interesting implications for government policy. For example, governments commonly identify increasing the amount of R&D perfomed domestically as a policy goal, but a more relevant focus may be the amount of R&D performed by domestic firms, especially if this is located close to the world technological frontier. If so, a policy such as an R&D tax credit that encourages firms to repatriate R&D activity may be partly counterproductive.

This paper has two main advantages over most previous studies of international knowledge spillovers. First it uses a firm-level panel data set, which allows for better modelling of heterogeneity between firms than industry or country-level

[1] See for example von Zedtwitz and Gassman (2002) and Serapio and Dalton (1999)

[2] See for example Jaffe, Trajtenberg and Henderson (1993) and Keller (2002)

studies. Secondly, and more importantly, it uses information from patent data on the location of inventors and patent citations to create a geographical measure of firms' innovative activity. This provides a specific channel through which to identify international knowledge spillovers associated with technology sourcing.

The structure of this paper is as follows. Section 2 discusses two key motivations for our approach in the context of previous literature on knowledge spillovers. Section 3 presents the basic model and Section 4 describes the data. Section 5 explains our methodology and presents the empirical results, and a final section concludes.

2. Motivation

There are two key motivations behind our empirical approach, one empirical and one concerning the identification of knowledge spillovers. We discuss each in turn.

2.1. Empirical motivation

Several recent studies of the bahaviour of multinational firms have suggested that gaining access to new technologies is an increasingly important motivation for firms locating R&D activity abroad, and especially in the USA. Serapio and Dalton (1999) argue that much of the globalisation of innovative activity has involved foreign firms locating R&D activities in the USA in order to benefit from technology sourcing at the leading edge of technological innovation: "Foreign parent companies, particularly in the drugs/biotechnology and electronics industries, have established or acquired foreign R&D laboratories in the US in order to gain access to science and technology, and enhance their global capabilities for technology development and innovation." They document the fact that UK firms are a particularly significant part of this development, with the third highest R&D expenditures in the USA in 1996 of all foreign countries.

This interpretation of foreign R&D investment is in contrast with earlier interpretations which focussed on the importance of adapting technologies developed at home to the conditions of the foreign market.[3] Other research has found that firms location decisions differ between the "research" aspect of R&D and the "development" aspect. For example, von Zedtwitz and Gassmann (2002) identify four archetypes of R&D internationalisation based on whether research, development or both are internationally dispersed. They find that motivations for internationalising research are largely driven by the desire to access new technologies, while motivations for internationalising development are usually associated with adapting existing products and/or concepts. In this paper we attempt to use information from patent citations to capture firms' different motivations for locating R&D in the USA.

Much recent research, especially work using patent citations, suggests that technology sourcing may be a plausible mechanism for reducing the geographical localisation of knowledge spillovers. Jaffe et al (1993), and Jaffe and Trajtenberg (1998), find that even after controlling for other factors, patents whose inventors reside in the same country are typically 30% to 80% more likely to cite each other than inventors from other countries, and that these citations tend to come sooner.

[3] See Le Bas and Sierra (2002) for a discussion

They also find that localisation does fade over time, but only very slowly.

Singh (2003) uses patent citations to investigate the role of multinational subsidiaries in knowledge diffusion. He finds that greater MNC subsidiary activity increases cross-border knowledge flows between the host country and the the MNC home base, but the MNC home base gains more than domestic firms in the host country. Branstetter (2003) uses patent citations to measure the role of foriegn direct investment by Japanese firms in the USA in mediating flows of knowledge between the two countries. He finds that knowledge spillovers received by the investing Japanese firms tend to be strongest via R&D and product development facilities. Spillovers from from investing Japanese firms to the USA flow most strongly through greenfield affiliates in which Japanese firms are deploying superior technology or managerial practices.

However, we are aware of no studies that attempt to find empirical evidence for technology sourcing in terms of its effects on firm-level productivity. We believe that the information from firms' patents on inventor location and citations used in this study provides an ideal channel for identifying knowledge spillovers associated with technology sourcing.

2.2. Identification of knowledge spillovers

The second key motivation behind our empirical approach concerns the econometric identification of knowledge spillovers. Essentially, the dominant approach to estimating knowledge spillovers suffers from a serious identification problem that is not always sufficiently discussed, and is rarely addressed.

The conventional approach follows Griliches (1979) by including a measure of some external knowledge pool in a production (or cost) function framework. The dependent variable is usually some measure of firm productivity, or it can be some measure of firms' innovative output, constituting a knowledge production function. Aside from many problems associated with the estimation of production functions, the most commonly-cited difficulty for identification of spillovers is that the "spillover pool" of outside knowledge available to a firm must be specified a priori. This problem is eloquently summed up by Griliches (1992): "To measure [spillovers] directly in some fashion, one has to assume either that their benefits are localised in a particular industry or range of products or that there are other ways of identifying the relevant channels of influence, that one can detect the path of the spillovers in the sands of the data".

Most studies address this problem by assuming that a firm is more likely to benefit from the R&D of other firms that are 'close' to it in some technological and/or geographical sense. In these models the 'spillover pool' available to firm i is equal to:

$$G_i = \Sigma_j w_{ij} R_j \tag{2.1}$$

where w_{ij} is some 'knowledge-weighting matrix' applied, for example, to the R&D expenditures of other firms R_j.

The literature contains many different approaches to constructing this matrix. Perhaps the simplest is to assume that w_{ij} is equal to one if the firm is the same industry and zero otherwise. Another method, suggested by Griliches (1979) and first used in Jaffe (1986), is to use firm-level data on patenting by class of patent, or sometimes the distribution of R&D spending across product fields, to locate firms in a multi-dimensional technology space. A weighting matrix is then constructed using the uncentered correlation coefficients between the location vectors of different firms.

However, these approaches to estimating spillovers suffer from another fundamental identification problem. This is that it is not easy to distinguish a spillovers interpretation from the possibility that any positive results are "just a reflection of spatially correlated technological opportunities" (Griliches, 1996). In other words, if new research opportunities arise exogenously in a firm's technological area, then it and its technological neighbours will do more R&D and may improve their productivity, an effect which may be erroneously picked up by a spillover measure.

This issue is discussed by Manski (1991) under the general title "the reflection problem". True knowledge spillovers correspond to an endogenous social effect, in the sense that an individual outcome (e.g. productivity) varies with the behaviour of the group (e.g. R&D spending). This can be differentiated from an exogenous social effect, whereby an individual outcome varies with the exogenous characteristics of the group, or a correlated effect whereby individuals in the same group tend to have similar outcomes beacuse they have similar characteristics or face similar environmental influences. Identification of endogenous effects is not possible unless prior information is available with which to specify the composition of reference groups. This is the role played by a knowledge weighting matrix, or even a simple industry-level measure of the spillover pool. However, even if this information is available, identification is not possible if the variables defining reference groups are functionally related to variables that directly affect outcomes. This is quite likely to be the case for many of the approaches found in the literature. For example, technological closeness is likely to be correlated with exogenous technological opportunity, and firms in the same industry are likely to be subject to similar supply or demand shocks. Thus the task for anybody trying to identify knowledge spillovers is to find a set of variables with which to define firms' reference groups that are not related to unobserved variables that directly affect the outcomes being measured.

3. The basic model

The basic approach follows Griliches (1979) and many subsequent papers by including measures of the external knowledge stock available to the firm in a firm-level production function. Thus we assume that the firm's value added can be written as follows

$$Y_{it} = Q(X_{it}, G_{it}) \tag{3.1}$$

where Y_{it} is real value added for firm i in year t, X_{it} is a vector of the firm's own inputs including labour, capital and the firm's own knowledge stock accumulated by doing R&D, and G_{it} is the external knowledge stock available to the firm.

As discussed above, a key assumption is how to define G_{it}. Because we want to identify geographical aspects of spillovers we assume that G_{it} is composed of a domestic and a foreign component, and do not restrict the response of the firm's value added to each component to be the same.

$$G_{it} = (D_{it}, F_{it}) \tag{3.2}$$

$$Y_{it} = Q(X_{it}, D_{it}, F_{it}) \tag{3.3}$$

The key innovation is that we allow the elasticity of value added with respect to the foreign and domestic external knowledge stocks to depend on a measure of the geographical location of the firm's innovative activity. So we have

$$\frac{\partial Y_{it}}{\partial D_{it}} \frac{D_{it}}{Y_{it}} = d(W_i^D) \tag{3.4}$$

$$\frac{\partial Y_{it}}{\partial F_{it}} \frac{F_{it}}{Y_{it}} = f(W_i^F) \tag{3.5}$$

where W_i^D and W_i^F are measures of the amount of the firm's innovative activity that is located at home or abroad respectively.

The most important aspect of our basic model is that the location measures allow identification of knowledge spillovers associated with technology sourcing in a way that should be less susceptible to the Manski-Griliches critique discussed earlier. While many studies claim identification of knowledge spillovers in this context from a positive response of value added to the external spillover pool, we only infer the existence of spillovers if the magnitude of that response depends positively on a direct proxy for a channel of knowledge transfer; in other words only if

$$\frac{\partial f(W_i^F)}{\partial W_i^F} > 0 \tag{3.6}$$

A positive response of value added to the spillover pool could be due to a spurious "correlated effect" if the variables used to define the spillover pool are related to unobserved variables that directly affect value added. Inferring the existence of knowledge spillovers simply from an observed positive response thus depends on the assumption that no such relationship between the two types of variables exists. In our approach identification depends only on the much weaker assumption that the nature of this relationship does not depend on our measure of the geographical location of innovative activity.

A concern remains that W_i^D and W_i^F are choice variables for the firm, and may thus be correlated with firm or industry-level technological shocks in a way that undermines our identification strategy. We have no exogenous instruments for the location of firms' innovative activity. However, we use pre-sample information to construct W_i^D and W_i^F. This ensures that they are not affected by technology shocks that also directly affect firm-level outcomes during the sample period.

4. Data

In order to implement our empirical strategy we need to measure three types of information: the location of firms' innovative activity, firms' productivity performance, and the domestic and foreign spillover pools available to firms. To do this we use three types of data source: data on patenting at the US Patent Office, firm accounts data, and OECD data on industry level R&D expenditure. We now describe the sources of these three types of data.

4.1. Patent data

The IFS-Leverhulme database used in this paper is a combination of two datasets. Full details of the matching between the datasets can be found in Bloom and Van Reenen (2000), and the process is sketched in the Appendix at the end of this paper. The first dataset is the NBER patent citations data file which contains computerised records of over two million patents granted in the USA between 1901 and 1999. This is the largest electronic patent dataset in the world. The second dataset is the Datastream on-line service which contains accounts of firms listed on the London Stock Exchange over 1968-2000. The initial sample is all firms existing in 1985 with names starting with the letters A-L, plus any of the top 100 UK R&D performers not already included, in order to maximise the number of patents matched to firms. This gives 415 firms.

The intersection of the two datasets gave 266 firms who had taken out at least one patent between 1975 and 1998, categorised by date of application. The reason for restricting our attention to patents applied for after 1975 is that data on citations is only available for patents applied for after this date.

4.1.1. Inventor location

The main information that we use from the patent data is the country address of the inventor(s) listed on the patent application. Table 1 lists the primary inventor's country for the 63,733 patents matched to the 266 UK firms. For comparison, the final column lists the share of the primary inventor's country for the entire patent database of all patents registered in the USA between 1975 and 1998 (more than 2 million patents). As expected the share of UK inventors is much higher for the patents owned by the 266 UK firms (31.0% in column (2)) than for the whole sample of patents (3.0% in column (3)). Nevertheless, the US has the highest share of inventors even for the patents owned by the 266 UK firms (45.1%). The high share of patents owned by the 266 UK firms but invented in the USA is probably partly due to home-country bias from using a US dataset, but also reflects the county's strong innovative performance and the location of many UK firms in the USA. An overall bias towards US based patents should not be a problem as long as it is not different across firms in a way that is related to other firm characteristics.

4.1.2. Citations

We also use data on patent citations to refine our measures of the location of firms' innovative activity. We assume that a patent owned by a UK firm but invented by an inventor located in the USA is more likely to be associated with technology sourcing behaviour if it cites other patents whose inventors were located in the USA. In particular, if a patent owned by a UK firm but invented by an inventor located in the USA does not cite any other patents whose inventors were located in the USA, this suggests that the patent is unlikely to be associated with technology sourcing. Such a patent is more likely to be associated with other motivations for locating R&D abroad, such as adapting existing technologies to the local market.

The 63,733 patents matched to our 266 UK firms make 472,998 citations to other patents, an average of 7.4 citations made by each patent. Of these 472,998 citations, 405,788 have information on the country location of the cited inventor. 23.6% of the citations were made by inventors located in the UK, but only 6.5% of all the citations are to a patent whose inventor was located in the UK. In contrast, while 56.4% of the citations were made by inventors located in the USA, 64.3% of all the citations are to a patent whose inventor was located in the USA. Again, this probably illustrates both the fact that the data is from the US patent office, and the dominant global position of the USA in innovation.

Table 2 presents a cross-tab of the location of the citing and cited inventor for the 405,788 citations where this information is available. It is important to remember that all of these citations were made by patents that are owned by UK firms, even if the inventor was located in the US. Only 16.9% of citations made by UK inventors are made to another UK inventor, while 54.1% are made to a US inventor. In contrast, 74.0% of citations made by US inventors are made to other US inventors, while only 3.2% are made to UK inventors. This provides some very preliminary evidence that most patents owned by UK firms but invented by an inventor located in the US are building on other knowledge created in the USA.

4.1.3. Self-citations

We want to investigate whether firms are benefitting from external knowledge that has not been generated within the same firm. Because of this we want to control for self-citations, where a patent cites another patent that is owned by the same firm. 8.7% of all citations are made to patents owned by the same patenting subsidiary (or "assignee"), while a further 1.1% of all citations are made to a different assignee that is nevertheless part of the same parent firm. Table 3 shows a similar cross-tab to Table 2, except only for self-citations to a patent that is owned by the same parent firm. Unsurprisingly, the percentages in the diagonals (for example a UK inventor citing another UK inventor, or a US inventor citing another US inventor) are much higher than before. Interestingly, once we condition only on self-citations, patents owned by UK firms but invented in the US are not much more likely to cite UK inventors than was the case before (3.4% in Table 3 compared to 3.2% in Table 2). Thus, even within firms, the transfer of knowledge from the UK to the USA appears to be small compared to the transfer of knowledge within the USA.

Table 4 shows the same cross-tab as Table 2 once we have excluded these self-citations. These citations to patents outside of the same firm are the citations that we will use to refine our measure of technology sourcing behaviour. As before, the number of citations made by US inventors to UK inventors is small (3.2% of all citations made by US inventors), while the number of citations made by US inventors to other US inventors outside of their firm is large (71.8% of all citations made by US inventors).

4.1.4. Application dates

It is generally considered that physical proximity is more important for the flow of knowledge that is "tacit", in the sense that it is not easily codified or written down in manuals. The flow of tacit knowledge is more likely to be mediated through face-to-face meetings and personal interactions between scientists and/or engineers. It also seems likely that knowledge that has been created recently is more "tacit" than knowledge that was created longer ago. Thus firms that locate innovative activity in the US in order to gain access to pools of tacit knowledge are unlikely to be attempting to access knowledge that was created twenty or even ten years ago. For this reason we also use information on the application dates of each citing and cited patent in order to refine our measures of the location of firms' innovative activity. In particular we look at citations made to patents that were applied for within the last three years. For example, if a patent was applied for in 1989, we restrict our attention to the citations that it makes that are to patents that were applied for in 1986, 1987, 1988 or 1989. Table 5 shows the same cross-tab of the country of the citing and cited inventor for all non self-citations of this type. The proportions are similar to those in Table 4, although UK inventors are slightly more likely to cite other UK inventors than before, while US inventors are less likely than before to cite other US inventors. We will return to the patent data when we discuss how we calculate our measures of the location of firms' innovative activity.

4.2. Accounts data

The initial sample of 415 firms was cleaned for estimation. This included ensuring that employment observations were available, deleting firms with less than five consecutive observations over 1990 - 2000, and excluding firms for which there were jumps greater than 150% in any of the key variables (capital, labour, sales). Capital stock was constructed by a perpetual inventory method as in Bloom and Van Reenen (2000). The data does not include intermediate inputs, so value added was constructed as the sum of total employment costs, operating profit, depreciation and total interest charges. Because of UK accounting regulations, most of the firms did not report R&D expenditure before 1989, and so the analysis is restricted to the years 1990-2000.[4] An R&D capital stock was constructed using a perpetual inventory method and an assumed 15% rate of obsolescence. The results are robust to different rates. R&D activity is also included in the main labour and capital variables so any estimated returns to R&D are "excess" returns.[5]

Although these are "UK firms" in the sense that they are listed on the London Stock Exchange, a key feature of the data is that it relates to the firm's global

[4] Even after 1989 when a firm reports zero R&D it is not clear that this corresponds to a true zero, although it is unlikely to perform a large amount of R&D. In the results presented in this paper, a dummy variable was used to denote reported zero R&D expenditure, but the results are not sensitive to the exact treatment of reported zeros.

[5] See Griliches (1979)

activities. As discussed later this has potentially important consequences for the interpretation of our results. For now we maintain the assumption that, while a firm's innovative activity may be located anywhere in the world, its production activity is located in the UK. We examine the validity of this assumption and the consequences of any violations later on.

4.3. Spillover pool data

The domestic and foreign spillover pools were constructed using the OECD's "Analytical Business Expenditure on R&D" dataset (ANBERD, 2002) on R&D spending by two-digit manufacturing industry (ISIC Revision 3) in the UK and the USA. A stock measure was constructed using a perpetual inventory method and an assumed 15% rate of obsolescence[6], with a starting year of 1987. Although there are various problems with using industry-level measures as discussed above, this data has the crucial advantage for our purposes that it contains R&D expenditures by geographical location of the R&D activity. This would be extremely hard if not impossible to recreate using a weighted sum of other firms' R&D. Our measure also has the advantage of including all R&D carried out in each industry in each country, and not just the R&D of the other sampled firms.

Because the source of identification in our model comes from the way the response of value added to the spillover pool depends on the geographical location of innovative activity, the possibility of spurious "correlation effects" due to a spillover pool constructed at the industry-level should not be a serious problem.

[6] We experimented with other depreciation rates but the results were not significantly changed.

However, in order to at least partly control for industry level cyclical effects and shocks not associated with knowledge spillovers, we also include as a robustness check two-digit industry-level value added in the UK and USA. This was taken from the OECD's "Structural Analysis" database (STAN, 2003). It turns out that none of our results is affected by including these value-added terms.

After cleaning as described above and limiting the sample to manufacturing firms we are left with 1794 observations on 188 firms, 141 of which are matched to at least one patent. Table 6 reports summary statistics.

5. Methodology and Results

5.1. Functional form

We consider a Cobb-Douglas production function with constant returns in labour and capital inputs

$$Y_{it} = A_{it} L_{it}^{\alpha} K_{it}^{1-\alpha} R_{it}^{\beta} D_{jt}^{\gamma_1} F_{jt}^{\gamma_2} \tag{5.1}$$

where i indexes a firm, j indexes the firm's two digit industry, and t indexes the year. Y_{it} is real value added, L_{it} is observed labour inputs, K_{it} is a measure of the firm's capital stock, R_{it} is a measure of the firm's own R&D stock, and D_{jt} and F_{jt} are the R&D stock in the firm's two-digit industry in the UK and the USA respectively. We assume that the elasticities of value added with respect to the external knowledge stocks are a linear function of firm-specific measures of the location of innovative activity

$$\gamma_1 = \theta_1 + \theta_2 W_i^{UK} \tag{5.2}$$

$$\gamma_2 = \phi_1 + \phi_2 W_i^{US} \tag{5.3}$$

where a positive estimate of ϕ_2 would provide evidence of knowledge spillovers associated with technology sourcing from the USA.

5.2. Location measures

We use several measures of W_i^{UK} and W_i^{US} . The basic measure is constructed as the proportion of the firm's total patents applied for between 1975 and 1989 where the inventor is located in the UK or the USA respectively. They are both equal to zero if the firm has no patents. Becuase our firm panel runs from 1990 to 2000 the location measures are based purely on pre-sample information. As discussed above, this ensures that the location measurea are not affected by technology shocks that also directly affect firm-level outcomes during the sample period.

This form for the measure of the geographical location of innovative activity discards two types of information in the patent data. The first is variation over time, so that the measure represents an average of the location of the firm's innovative activity over the period 1975-1989. The second type of information is the total number of the firm's patents. While this may be relevant information, normalising the location measures to a proportion between zero and one helps to deal with difficulties associated with firm size and differences in propensity to patent across industries.

As mentioned above we also use information on patent citations to refine our measure of W_i^{UK} and W_i^{US} . A key theme in the literature is that technology sourcing is not the only motivation for firms to locate innovative activity abroad. In particular, firms may do R&D abroad in order to adapt existing technologies to new markets. Our empirical approach to this issue is to use data on citations to eliminate patents that are unlikely to represent technology sourcing behaviour. Consider two extreme cases for a patent that is owned by a UK firm but that was invented in the US: if the patent only cites patents owned by the same firm and whose inventors were located in the UK then the patent is more likely to represent activity associated with adapting an existing technology to the US market; on the other hand, if the patent cites many patents that are not owned by the firm and whose inventors were located in the US then the patent is more likely to represent technology sourcing behaviour. If we want to investigate whether there is evidence for technology sourcing behaviour in productivity outcomes, then we wish not to use the first type of patent when constructing our location measures.

To implement this approach, our second measure of W_i^{UK} and W_i^{US} excludes patents that do not cite any other patents whose inventors were located in the same country. We also exclude patents that do cite inventors from the same country, but only inventors within the same parent firm. The measure of W_i^{US} is thus equal to the proportion of the firm's total patents where: (1) the inventor is located in the USA and (2) the patent cites at least one other patent whose inventor was both located in the US and did not work for the same parent firm.

Our third and final measure of W_i^{UK} and W_i^{US} is the same as the second measure, except that it also uses information on the time-lag between the citing and cited patent. As discussed above, technology sourcing behaviour is likely to be associated with gaining access to pools of "tacit" knowledge. Given that knowledge that was created recently is more likely to have tacit characteristics, we include only citations to patents whose application date is no more than three years prior to that of the citing patent. The third measure of W_i^{US} is thus equal to the proportion of the firm's total patents where: (1) the inventor is located in the USA and (2) the patent cites at least one other patent that was applied for within the last three years and whose inventor was both located in the US and did not work for the same parent firm.

Table 7 reports summary statistics of the three location measures for the 141 firms that are matched to at least one patent. The mean and median values of the weights become smaller as the requirements become more restrictive, in other words as we first condition only on the location of the inventor, then on location and citation characteristics, and then finally on location, citation and time-lag characteristics. The measures for the UK become smaller more rapidly as we condition on citations. This reflects the smaller number of citations that are made to UK inventors than to US inventors.

We estimate the basic functional form described above in logs

$$\begin{aligned}(y_{it} - k_{it}) &= \alpha(l_{it} - k_{it}) + \beta r_{it} + \theta_1 d_{jt} + \phi_1 f_{jt} + \theta_2 W_i^{UK} d_{jt} + \phi_2 W_i^{US} f_{jt} \\ &\quad + \theta_3 W_i^{UK} + \phi_3 W_i^{US} + a_{it} \qquad (5.4)\end{aligned}$$

5.3. Estimation

We assume that the residual productivity term takes the form

$$a_{it} = t_t + \eta_i + u_{it}. \tag{5.5}$$

where the year dummies control for common macro effects and the firm effect and stochastic productivity shock may be correlated with the regressors. We allow for arbitrary heteroskedasticity and possible serial correlation in the stochastic productivity shock. We include industry dummies in all regressions. We estimate using Systems-GMM, where the information from the levels equation helps to alleviate the weak instruments problem associated with first-difference GMM when series are persistent.[7] The additional moment conditions take the form

$$E[\Delta x_{i,t-s}(\eta_i + u_{it})]. = 0 \tag{5.6}$$

for $s = 1$ when $u_{it} \sim AR(0)$ and for $s = 2$ when $u_{it} \sim AR(1)$, where x_{it} indicates the regressors being instrumented. This requires the first moments of x_{it} to be time-invariant, conditional on common year dummies. We test the vaidity of the additional moment conditions using a Sargan difference test.

We assume that all firm-level variables are endogenous, while in our final specification all industry-level variables are treated as strictly exogenous. We examine specifications where the industry-level R&D stocks are treated as endogenous and the results are not significantly affected. The results are also robust to lagging the industry-level variables by one period, in which case they can be treated as

[7] See Blundell and Bond (1999) for an exposition and a production function example

pre-determined. We instrument firm-level variables in the differenced equation with their levels lagged from two to five times inclusive, and in the levels equation by their first-differences lagged once, as well as by all time and industry dummies and all exogenous variables.

5.4. Empirical Results

Table 8 presents results for the basic production function and the basic spillover and value added terms. Column (1) is OLS without imposing constant returns to scale in labour and capital, while column (2) does impose constant returns. The hypothesis of constant returns to scale is not rejected at the 5% level. Column (3) is the basic production function using Systems-GMM. The coefficient on capital is very similar to the OLS case. The estimated elasticity with respect to own R&D corresponds to a median private excess rate of return to R&D of about 15%, which is similar to that found in other studies.[8] Tests are presented for first and second order serial correlation in the first-differenced residuals, with robust p values in brackets. Neither test ever rejects the hypothesis of no serial correlation. This justifies the use of twice lagged instruments in the difference equation and once lagged instruments in the levels equation. A Sargan test of overidentifying restrictions is not significant, and neither is a Sargan difference test of the extra moment conditions implied by the levels equation.

Columns (4) and (5) introduce the main industry level spillover terms. The spillover terms are treated as strictly exogenous in column (4) and as endogenous

[8] See Griliches (1992)

in column (5). The coefficients are not significantly affected. The main spillover terms do not enter significantly in either case, and the coefficient on the UK term becomes close to zero when value added is included to control for industry-level shocks in column (6). Thus we find no evidence for either domestic or international knowledge spillovers in a conventional specification. Neither value added terms are significant, althought they are both positive.

Table 9 presents the key interaction results. Column (1) is OLS, with the UK and US location interactions included. The coefficients on the basic industry level R&D variables are insignificant as before, as is the coefficient on the UK location interaction. However, the coefficient on the US interaction is positive and significant at the 1% level, suggesting the existence of knowledge spillovers associated with technology sourcing from the USA. The fact that the UK interaction is not significant is not very surprising for a sample of UK firms, in that the marginal effect of locating innovative activity in the UK on the firm's ability to benefit from spillovers from UK R&D is likely to be smaller than in the US case. The significant negative effect of the US location measure W_i^{US} itself is only observed conditional on the inclusion of the interaction terms, and it enters positively when the interactions are not included. The median marginal effect of W_i^{US} on value added remains positive.

Column (2) is the same specification estimated by Systems-GMM. The coefficient on the US location interaction is very similar although it is less precisely estimated. Nevertheless, it remains significant, although not quite at the 5% level. All the other coefficients are similar to the OLS case. Column (3) uses the second type of location weight that is refined using data on citations. The coefficient on the US interaction term is now both larger and more precisely estimated than in column (2), and is significant at the 1% level. This suggests that the citation information does indeed provide a more refined measure of location, and provides further support for the existence of international knowledge spillovers associated with technology sourcing. Column (4) uses the third type of location weight that is further refined using data on the time-lag between citing and cited patents. The coefficient on the US interaction term is significantly larger and remains significant at the 1% level, providing further evidence of the technology sourcing hypothesis.

5.5. Location of production activity

A further issue relates to the fact that the data represents firms' global activity. Although we have been assuming that production activity is located in the UK, this is not completely true in practice. It is possible that the location measure W_i^{US} is not only proxying for the location of innovative activity, but also for the location of production. In other words, firms with innovative activity in the USA are likely also to have productive activity located there. If this is the case, then we may be picking up not only international spillovers but also domestic spillovers within the USA, with all the ensuing identification issues that were discussed earlier.

We attempt to control for this by using the separate reporting of domestic employment to total employment. 117 out of 188 firms report domestic employment separately to total employment at least once during 1990-2000. For those that do not report separately we assume that all employment is domestic. Of those 117 firms, 53 report total employment greater than domestic employment at least once. We drop these firms from the sample and re-estimate our model on the remaining 135 firms, which we expect to have little or no foreign production activity. Column (5) presents the same specification as column (4) except now only for the 135 firms. The results are very similar, although the UK interaction becomes negative but insignificant. These results suggest that the initial results were not primarily driven by the location of firms' production activities.

5.6. Robustness

We consider several robustness checks to the results in Table 9. First we include further interactions of the industry level R&D measures with a zero-one dummy that indicates whether the firm has any patents at all or none. This is to check that the results on the location interactions are not driven by patenting firms having higher "absorptive capacity" than non-patenting firms, since non-patenting firms by definitions have values of W_i^{UK} and W_i^{US} equal to zero. Neither of the interactions with the patenting dummy is ever signficant, and the positive significant interaction with W_i^{US} remains, suggesting that the results are not driven by absorptive capacity.

Secondly we replace the industry level R&D measures with industry level value added, in order to check that the results are not driven by industry level shocks unrelated to R&D. None of the value added terms is significant, and when we include value added and R&D terms together the coefficients on the R&D terms are similar to before, suggesting that it is indeed the R&D stocks that are driving the results.

We also lagged all the industry level R&D terms by one period, so that they could be considered pre-determined. Again the main results are not affected. Finally we relaxed the assumption of constant returns to scale in labour and capital, which did not affect the main results.

6. Summary and Conclusions

The results presented in this paper provide some evidence for the existence of knowledge spillovers associated with technology sourcing. The idea that firms might invest in R&D activity in a technologically advanced country such as the US in order to gain access to spillovers of new "tacit" knowledge has been suggested in the literature, as discussed above, but we know of no studies that have attempted to find evidence for this in observed productivity outcomes.

Our main results suggest that the increase in the US R&D stock in manufacturing over 1990-2000 was associated with on average a 4% higher level of TFP for the UK firms in our sample. This compares with an average 6.5% higher level of TFP associated with the increase in their own R&D stocks over the same period. Thus spillovers from the US contributed about two-thirds of the effect of firms' own R&D. Our results also suggest that for a UK firm, shifting 10% of its innovative activity (as measured by patent applications) to the US from the UK while keeping its overall level of R&D stock the same (e.g. changing W_i^{US} from 0.30 to 0.40 and W_i^{UK} from 0.70 to 0.80 while keeping R_{it} the same) is associated with an increase in its TFP level by between 3% and 7%. This effect is the same order of magnitude as that of a doubling in its R&D stock.

Our result has interesting implications for policy. Governments are generally keen to promote higher levels of domestic R&D activity, and the countries of the EU have recently expressed an aspiration to raise the level of R&D spending within the EU to 3% of GDP. However a question arises as to whether a country should be concerned with the total amount of R&D expenditure located domestically, or the total amount of R&D performed by domestic firms anywhere in the world, especially if R&D performed abroad is more productive and provides access to new technologies. As an example, the UK has recently introduced R&D tax credits that apply only to R&D located in the UK. This might have partially counterproductive effects if it encourages firms to repatriate R&D activity from abroad.

7. References

Aghion, P., Bloom, N., Blundell, R., Griffith, R. and Howitt, P. (2002), "Competition and innovation: An inverted U relationship", NBER Working Paper No. 9269

Bernstein, J. I. and Nadiri, M. I. (1989), "Research and development and intra-industry spillovers: An empirical application of dynamic duality", Review of Economic Studies 56, pp 249-269

Bloom, N. and Griffith, R. (2001), "The internationalisation of R&D", Fiscal Studies 22 (3), pp 337-355

Bloom, N. and Van Reenen, J. (2000), "Real options, patents, productivity and market value: Evidence from a panel of British firms", Institute for Fiscal Studies Working Paper No. W00/21

Blundell, R. and Bond, S. (1999), “GMM estimation with persistent panel data: An application to production functions”, Institute for Fiscal Studies Working Paper No. W99/4

Branstetter, L. (1996), “Are knowledge spillovers international or intranational in scope? Microeconometric evidence from the U.S. and Japan”, NBER Working Paper No. 5800

Branstetter, L. (2003), “Is foreign direct investment a channel of knowledge spillovers? Evidence from Japan’s FDI in the United States”, Columbia Business School and NBER, mimeo

Cincera, M. and Van Pottelsberghe da la Potterie, B. (2001), "International R&D spillovers: a survey", Cahiers Economiques de Bruxelles, 169, pp 3-32

Cohen, W. M. and Levinthal, D. A. (1990), “The implications of spillovers for R&D investment and welfare: A new perspective”, Administrative Science Quarterly 35, pp 128-152

Criscuolo, P. (2002), "Reverse technology transfer", Unpublished Paper, University of Sussex

Griffith, R., Redding, S. and Van Reenen, J. (2001), “Mapping the two faces of R&D: Productivity growth in a panel of OECD industries”, Institute For Fiscal Studies Working Paper W02/00

Griliches, Z. (1979), “Issues in assessing the contribution of research and development to productivity growth”, Bell Journal of Economics 10 (1), pp 92-1167

Griliches, Z. (1992), “The search for R&D spillovers”, Scandinavian Journal of Economics 94, supplement, pp S29-S47

Griliches, Z. (1994), “Productivity, R&D, and the data constraint”, American Economic Review 84, (1), pp 1-23

Griliches, Z. (1996), “R&D and productivity: The unfinished business”, in Z.Griliches (1998), R&D and Productivity: The Econometric Evidence, The University of Chicago Press, Chicago and London

Griliches (1998), R&D and Productivity: The Econometric Evidence, The University of Chicago Press, Chicago and London

Griliches, Z. and Mairesse, J. (1995), “Production functions: The search for identification”, NBER Working Paper No. 5067

Grossman, G. M. and Helpman, E. (1991), Innovation and Growth in the Global Economy, MIT Press, Cambridge, Mass.

Hall, B. H. (1996), "The private and social returns to R&D", in Technology, R&D and the Economy, ed. B. Smith and C. Barfield, Brookings Institution and AEI, Washington D.C., pp 140-162

Hall, B. and Mairesse, J. (1996), "Estimating the productivity of research and development: An exploration of GMM methods using data on French and United States manufacturing firms", NBER Working Paper No. 5501

Harhoff, D. (2000), "R&D spillovers, technological proximity, and productivity growth – Evidence from German panel data", Schmalenbach Business Review 52, pp 238-260

Jaffe, A. (1986), "Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value", American Economic Review 76, pp 984-1001

Jaffe, A. and Trajtenberg, M. (1998), "International knowledge flows: Evidence from patent citations", NBER Working Paper No. 6507

Jaffe, A., Trajtenberg, M. and Henderson, R. (1993), "Geographic localisation of knowledge spillovers as evidenced by patent citations", Quarterly Journal of Economics 108 (3), pp 577-598

Keller, W. (1996), "Are international R&D spillovers trade-related? Analyzing spillovers among randomly-matched trade partners", SSRI Working Paper No. 9607, University of Wisconsin, Madison

Keller, W. (2002), "Geographic localization of international technology diffusion", American Economic Review, 92, pp 120-142

Kuemmerle, W. (1999), "The drivers of foreign direct invetsment into research and development: an empirical investigation", Research Policy, 30 (1), 1-24

LeBas, C. and Sierra, C. (2002), "Location versus home country advantages in R&D activities", Research Policy, 31, pp 589-609

Mairesse, J. (1995), "R&D productivity: A survey of the econometric literature", Paper presented at the CEPR workshop on R&D spillovers, Lausanne, Jan. 27/28 1995

Manski, C. (1991), "Identification of endogenous social effects: The reflection problem", Review of Economic Studies 60 (3), pp 531-42

Nickell, S. (1996), "Competition and corporate performance", Journal of Political Economy 104

OECD (2003), "Structural Analysis Data Base", Paris

OECD (2002), "Analytical Business Enterprise Research and Development", Paris

Reinganum, J. (1989), "The timing of innovation: Research, Development and Diffusion", in R. Schmalensee and R. D. Willig (eds), Handbook of Industrial Organization, North-Holland, Amsterdam, Vol. 1, pp 849-908

Scherer, F. M. (1982), "Interindustry technology flows and productivity growth", Review of Economics and Statistics LXIV, pp 627-634

Serapio, M. G. Jr. and Dalton, D. H. (1999), "Globalization of industrial R&D: an examination of foreign direct investments in R&D in the United States", Research Policy 28, pp 303-316

Singh, J. (2003), "Knowledge diffusion and the role of multinational subsidiaries", Harvard Business School and Department of Economics, mimeo

Spence, A. M. (1984), "Cost reduction, competition and industry performance", Econometrica 52, pp 101-121

Trajtenberg, M. (1990), Economic Analysis of Product Innovation: The Case of CT Scanners, Harvard University Press, Cambridge, Mass.

von Zedtwitz, M. and Gassmann, O. (2002), "Market versus technology drive in R&D internationalisation: four different patterns of managing research and development", Research Policy 31, pp 569-588

8. Data Appendix

The full data matching process can be found in Bloom and Van Reenen (2000), but the main aspects are sketched here. From the population of public firms quoted on the London Stock Exchange, a random sample of all companies whose names began with the letters 'A' through 'L' were selected. Also selected were the top 100 R&D performing firms in the UK in order to maximise the number of patents that could be matched. For all of these 415 firms Who Owns Whom 1985 was used to manually match each patenting subsidiary to their parent companies. This process was subsequently checked for all large subsidiaries and outliers using the Internet. Being a manual matching process, the matching accuracy appears to be quite good, and is certainly substantially greater than a computerised flexible string search. In direct comparisons this uncovered only about 10% of the matches found manually.

Table 1: Country of inventor

Country of Inventor	(1) Number of Patents matched to our UK firms	(2) % Share of patents matched to our UK firms	(3) % Share of all USPTO patents
UK	19,745	31.0	3.0
USA	28,731	45.1	55.7
Japan	4,411	6.9	18.8
Germany	2,481	3.9	7.9
France	1,457	2.3	3.0
Other	6,908	10.8	11.6
Total	63,733	100	100

Notes: 63,733 patents matched to 266 UK firms; final column refers to all patents registered at the US Patent Office between 1975 and 1998

Table 2: Location of citing and cited inventors: all patents matched to our sample of UK firms

Cited country:	UK	USA	Other	Total
Citing country:				
UK	16,233	52,024	27,889	96,146
	(16.9%)	(54.1%)	(29.0%)	(100%)
USA	7,298	167,912	51,790	227,000
	(3.2%)	(74.0%)	(22.8%)	(100%)
Other	3,014	40,784	38,844	82,642
	(3.6%)	(49.4%)	(47.0%)	(100%)
Total	26,545	260,720	118,523	405,788
	(6.5%)	(64.3%)	(29.2%)	(100%)

Notes: 63,733 patents making a total of 472,998 citations; 405,788 citations have data on the location of the cited inventor

Table 3: Location of citing and cited inventors : only self-citations by our sample of UK firms

Cited country:	UK	USA	Other	Total
Citing country:				
UK	10,391 (90.3%)	654 (5.7%)	462 (4.0%)	11,507 (100%)
USA	853 (3.4%)	22,732 (91.5%)	1,261 (5.1%)	24,846 (100%)
Other	627 (6.2%)	1,100 (10.8%)	8,443 (83.0%)	10,170 (100%)
Total	11,871 (25.5%)	24,486 (52.6%)	10,166 (21.8%)	46,523 (100%)

Notes: 63,733 patents making a total of 46,523 self-citations

Table 4: Location of citing and cited inventors : only non self-citations by our sample of UK firms

Cited country:	UK	USA	Other	Total
Citing country:				
UK	5,842 (6.9%)	51,370 (60.7%)	27,427 (32.4%)	84,639 (100%)
USA	6,445 (3.2%)	145,180 (71.8%)	50,529 (25.0%)	202,154 (100%)
Other	2,387 (3.3%)	39,684 (54.8%)	30,401 (42.0%)	72,472 (100%)
Total	14,674 (4.1%)	236,234 (65.8%)	108,357 (30.2%)	359,265 (100%)

Notes: 63,733 patents making a total of 426,475 non self-citations; 359,265 of these non self-citations have data on the location of the cited inventor

Table 6: Summary statistics

	Mean	Median	Standard Deviation	Min	Max
Observations	9.5	10	1.8	5	11
Employees	10,711	1,750	27,564	34	288,000
Value added (£m)	372	49	928	1.5	8,244
Capital stock (£m)	515	52	1,415	1.3	11,110
R&D stock (£m)	144	1.8	597	0	4,860

Notes: 188 firms, 1990-2000; all monetary amounts are in 1995 currency, deflated using OECD manufacturing sector deflator; value added is constructed as the sum of total employment costs, operating profit, depreciation and interest payments; capital stock and R&D stock are constructed using a perpetual inventory method as described in the text

Table 7: Summary statistics for patenting firms

	Mean	Median	Standard Deviation	Min	Max
Total patent applications	240	40.5	657	1	5820
UK Location Weight	0.354	0.274	0.363	0	1
UK Location + Citation Weight	0.082	0.017	0.145	0	1
UK Location + Citation Within 3 Years	0.019	0.000	0.054	0	0.5
USA Location Weight	0.462	0.425	0.379	0	1
USA Location + Citation Weight	0.417	0.368	0.349	0	1
USA Location + Citation Within 3 Years	0.162	0.134	0.184	0	1

Notes: 141 firms matched to at least one patent; location weights are constructed as described in the text

Table 8 : Basic production function results

	(1) OLS	(2) OLS	(3) GMM	(4) GMM	(5) GMM	(6) GMM
ln (L) $_{it}$	0.635	0.669	0.672	0.675	0.676	0.679
	(0.060)	(0.048)	(0.072)	(0.075)	(0.070)	(0.071)
ln (K) $_{it}$	0.329	0.331	0.328	0.325	0.324	0.321
	(0.046)					
ln (R&D) $_{it}$	0.028	0.012	0.022	0.021	0.021	0.021
	(0.008)	(0.007)	(0.010)	(0.010)	(0.009)	(0.009)
ln (UK R&D) $_{jt}$				0.046	0.088	0.016
				(0.093)	(0.131)	(0.134)
ln (US R&D) $_{jt}$				0.002	-0.013	0.001
				(0.061)	(0.072)	(0.074)
ln (UK Value Added) $_{jt}$						0.105
						(0.080)
ln (US Value Added) $_{jt}$						0.084
						(0.065)
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Firms	188	188	188	188	188	188
Observations	1794	1794	1794	1794	1794	1794
1st order serial correlation	-	-	-1.217 (0.223)	-1.221 (0.222)	-1.224 (0.221)	-1.226 (0.220)
2nd order serial correlation	-	-	-0.981 (0.327)	-0.972 (0.331)	-0.977 (0.328)	-1.033 (0.302)
Sargan	-	-	77.69 (0.457)	78.74 (0.423)	152.34 (0.500)	149.23 (0.571)
Sargan difference	-	-	19.73 (0.411)	18.94 (0.438)	35.68 (0.427)	31.70 (0.516)

Notes: Dependent variable is the log of value added; the time period is 1990-2000; columns (1) and (2) are OLS with robust standard errors in brackets, clustered on industry; columns (2) to (6) impose constant returns to scale in labour and capital; the hypothesis of constant returns to scale in labour and capital never rejected at the 5% level; columns (3) to (6) are systems-GMM, with one-step robust standard errors in brackets, except for tests where p values in brackets; Labour and firm R&D stocks are assumed endogenous; industry R&D stocks are assumed strictly exogenous in column (4) and endogenous in columns (5) and (6); industry value added is assumed strictly exogenous in column (6); endogenous variables are instrumented by levels lagged from two to five times in the differences equation and differences lagged once in the levels equation, as well as by all exogenous variables and year and industry dummies

Table 9 : Interactions results

	(1) OLS	(2) GMM	(3) GMM	(4) GMM	(5) GMM
Location weight:	Location	Location	Location & Citation	Location & Citation within 3 years	Location & Citation within 3 years
ln (L/K) $_{it}$	0.669	0.679	0.677	0.671	0.619
	(0.048)	(0.069)	(0.070)	(0.071)	(0.074)
ln (R&D) $_{it}$	0.012	0.020	0.019	0.017	0.023
	(0.008)	(0.010)	(0.009)	(0.009)	(0.012)
ln (UK R&D) $_{jt}$	0.044	0.036	0.037	0.031	0.094
	(0.094)	(0.092)	(0.092)	(0.089)	(0.106)
ln (US R&D) $_{jt}$	-0.046	-0.025	-0.029	-0.021	0.033
	(0.075)	(0.063)	(0.061)	(0.061)	(0.083)
W_i^{UK} *** ln (UK R&D)** $_{jt}$	0.033	0.029	0.101	0.392	-0.586
	(0.023)	(0.029)	(0.093)	(0.259)	(0.394)
W_i^{US} *** ln (US R&D)** $_{jt}$	0.070	0.065	0.082	0.160	0.201
	(0.020)	(0.033)	(0.029)	(0.052)	(0.060)
W_i^{UK}	-0.285	-0.262	-0.793	-3.124	3.642
	(0.160)	(0.194)	(0.659)	(2.251)	(2.615)
W_i^{US}	-0.635	-0.592	-0.742	-1.515	-1.900
	(0.226)	(0.322)	(0.299)	(0.519)	(0.609)
Firms	188	188	188	188	135
Observations	1794	1794	1794	1794	1267
1st order serial correlation	-	-1.223	-1.223	-1.222	-1.208
		(0.221)	(0.221)	(0.222)	(0.227)
2nd order serial correlation	-	-0.972	-0.961	-0.939	-1.011
		(0.331)	(0.337)	(0.348)	(0.312)
Sargan	-	81.83	81.02	80.22	78.68
		(0.332)	(0.355)	(0.378)	(0.425)

Notes: Dependent variable is the log of value added divided by capital stock; the time period is 1990-2000; column (1) is OLS with robust standard errors in brackets, clustered on industry; columns (2) to (7) are systems-GMM; one-step robust standard errors in brackets, except for tests where p-values in brackets; firm-level variables assumed endogenous and industry level variables assumed strictly exogenous; endogenous variables are instrumented by levels lagged from two to five times in the differences equation and differences lagged once in the levels equation, as well as by all exogenous variables and year and industry dummies; column (7) restricts the sample to "domestic" firms, i.e. firms that never report domestic employment to be less than total employment