The Impact of University Research on Corporate Patenting*

Christian Helmers
Wolfson College
University of Oxford

Mark Rogers
Harris Manchester College
University of Oxford

Preliminary Version: September 2009

ABSTRACT

This paper analyses the associations between the number of patenting manufacturing firms within a UK postcode area and (i) the number of universities in the same postcode area, (ii) the presence of engineering or biological sciences research departments, (iii) the number of researchers active in these departments, (iv) the 'quality' of research. The main contribution of this paper is to distinguish between the possible effect of universities on small firms, as opposed to large firms. The commonly held view is that location matters more for small firms than large firms. The results here confirms this view. In fact, in virtually all cases we only find a positive association for local universities on smaller firms. From a methodological point of view, we add to the existing literature by accounting for potential simultaneity between university research and patenting by local firms. Moreover, we also allow for the effects of the presence of universities in neighbouring postcode areas to influence firms' patenting activity by incorporating spatial effects.

KEYWORDS: Patents, universities, knowledge transfer, spillover

JEL Classification: L22, L26, O34

*Corresponding author: Mark Rogers, Harris Manchester College, Mansfield Road, OX1 3TD Oxford, United Kingdom, mark.rogers@hmc.ox.ac.uk.

1 Introduction

The capacity of university research to act as a catalyst for business sector innovations is of critical importance. Universities have a potential impact on firms' innovation in a variety of ways including: publication of fundamental research; university patenting and licensing; spin-offs and university incubators; joint research with firms; targeted knowledge transfer to firms; consultancy projects; training of students including continuous professional development and executive training programmes; and informal knowledge networks. From a policy perspective there are a number of key issues. Perhaps of paramount importance is how should money for university research be allocated? This involves choices over both how many universities should be funded and also the distribution of funding within the chosen universities. For example, is it important to ensure university research is spread across a wide range of regions? Equally, to what extent should research be focused on the 'best' universities and, of course, how should one determine the 'best'? In addition, there are a range of other issues surrounding university-business links including the incentives facing a university to a) patent and license, b) engage in joint research with business and c) actively pursue knowledge transfer programs.

The central research question is does a university's research have an association with the innovative activity of firms located close to the university? Innovative activity is notoriously difficult to measure but patents and R&D are frequently used proxies (see Greenhalgh and Rogers, 2009, for a full discussion). Since in the UK, as in other countries, data on the R&D activity of smaller firms is limited, this paper uses patenting as the proxy for innovation. Using patent data does direct the focus to specific types of innovations, but this may be appropriate when considering certain types of university research. The UK has developed an elaborate methodology to assess the quality of research called the Research Assessment Exercise (RAE), which is used as a metric to allocate research funds. We use data from the 2001 RAE in this paper. The data that allow us to analyse patenting come from the Oxford Firm Level Intellectual Property (OFLIP) database. This database contains the patenting activity of all registered firms in the UK over the period 2000 to 2007 along with information on firm-level performance and characteristics. This means that the analysis can consider separately micro firms or small and medium enterprises (SMEs), which might be thought to rely on local universities more than larger firms. We capture proximity between university research and private firms by focusing on the effect of university research on firms located within the same postcode area as universities.

In order to allow for neighbourhood effects from the presence of universities located in adjacent postcode areas, we incorporate also a spatially lagged measure for university presence. Moreover, we also address a critical issue in the assessment of the association between patenting and university research within a narrowly defined geographical area: simultaneity between university research and firms' innovative activity. We instrument the RAE research quality measure by a university's age. Universities in the UK are well suited for this identification strategy as there is large variation in university age across universities and postcode areas which we find to be strongly correlated with research quality. At the same time, we argue that a university's age is not associated with patenting of local companies other than through its effect on the quality of research undertaken at the corresponding university.

The paper is organised as follows. The next section reviews the existing empirical results on the role of universities in business innovation. Section 3 discusses some of the difficulties of analysing the link between university research and corporate patenting, together with outlining our approach. Section 4 discusses the data used. Section 5 explains the estimation method adopted. Section 6 describes our results and Section 7 offers some concluding thoughts.

2 Literature

This section reviews the existing empirical work on the broader impact of university research on corporate innovative activity. The literature essentially originates in Jaffe's (1989) seminal work in which he analyses spillovers from university research to corporate patent activity using the Griliches (1979) type knowledge production function framework. Corporate patenting in 29 US states is the basic dependent variable although this is broken down into different technology areas for the period 1972-77, 1979, and 1981.[1] The determinants of patenting are: the number of universities; university R&D (again broken down by technology area), total R&D performed by industry, a geographical coincidence indicator (an index that captures concentration of universities and firms) and a number of control variables such as population and public research laboratories. Jaffe allows for a simultaneous relationship between private sector R&D and university research by estimating a system of simultaneous equations. The results suggest that there are spillovers from universities to corporate patenting, most strongly in drugs but also in chemicals, electronics and mechanical arts. There is also an indication that university R&D causes private R&D (and not vice versa). Acs et al. (1992) complement Jaffe's study by using innovations as the dependent variable (from the 1982 US Small Business Administration innovation database). They also find positive associations between university research and innovation at the state level, although this is not the case for electronics-based industries. While Jaffe uses panel data, Acs et al. results are based only on a single year; however, they note that estimates from a single cross section and the pooled data using Jaffe's data are very similar.

More recently, Harhoff (1999) looks at the formation of new firms in 328 West-German counties over the period 1989 to 1993 and, specifically, how the existing industry structure and presence of publicly-funded research measured in 1987 and 1989 respectively affect rates of firm creation between 1989 and 1993.[2] Harhoff focuses his analysis on two important 2-digit industries, the electrical machinery and mechanical

[1] The technology areas are drugs/medical, chemical, electronics/optics/nuclear, mechanical arts and 'other'.

[2] The conditioning variables refer to the pre-reunification period, while start-ups are investigated also during the period directly after Germany's reunification. Harhoff does not account for re-unification effects and it is not clear to what degree the important structural shock induced by reunification has influenced start-up activity.

engineering sectors. These industries are further divided into high-technology and low-technology on the basis of R&D intensity at the 4- and 5-digit industry level. A Poisson Pseudo Maximum Likelihood model is then used, with number of new entrants as the dependent variable where the data is treated as a pooled cross-section. To model the effect of research on firm formation, Harhoff uses county-specific employment shares of R&D personnel, scientific personnel at universities as well as at extra-university research laboratories and institutes. Industry structure is captured by the share of the industry's employment in the county's total manufacturing employment, as well as through a Herfindahl concentration measure for the manufacturing industry. [3]. In addition, a wide range of other explanatory variables are included, which are similar to above studies. The results indicate that the employment share of scientists and engineers in universities and extra-university research institutions are positively associated with high-tech firm formation. Moreover, Harhoff finds new firm formation within the sectors studied to be persistent in highly specialised regions. Yet, high-tech start-ups are found to be much less persistent, i.e., high-tech start-ups are more likely to be found in counties with heterogeneous industry structures. The results also suggest a positive association between high-tech start-ups and the presence of business-oriented service providers.

Another study using West-German data is by Fritsch and Slavtchev (2007). The authors use West-German data for NUTS-3 regions with the dependent variable being counts of patent applications between 1995 and 2000.[4] Fritsch and Slavtchev estimate a Griliches-type knowledge production function. The inputs are R&D (equal to the number of private sector employees in R&D),[5] and universities' regular as well as additional external funds. The data on university funding allows the authors to make inference with respect to the association between the presence of universities and patent applications within regions. Inter-regional spillovers are captured by allowing private R&D and university funds to affect patent counts in neighbouring regions (where the distance between regional centres is within 0-50km and 50-75km). Similar to the other studies, Fritsch and Slavtchev also include an industrial concentration index (in the form of a Gini coefficient). To account for the higher propensity to patent in the manufacturing sector compared to services, the authors also include a manufacturing specialisation index, which is the share of manufacturing employment relative to the national average. The authors estimate a negative binomial model for their panel that constrains any research impact on patenting to be after three years.[6] The results suggest that there is no evidence for university research measured as the universities' regular budget has any positive association with patenting. Only external funds are associated with increased

[3]This measure excludes the specific industry studied and captures a county's degree of diversification in the manufacturing sector

[4]Note that patents are allocated to regions using the address of the inventor, rather than the address of the firm. So an important assumption Fritsch and Slavtchev make is that firms' location and place of residence of inventors coincide within the same NUTS-3 region. Also, it is unclear whether only national patent applications are included or whether the data set contains also EPC and PCT patents.

[5]Fritsch and Slavtchev assume that every employee with a tertiary degree in engineering or natural sciences works in R&D.

[6]The authors argue that a random effects specification is more appropriate than fixed effects. Yet, since random effects require strict exogeneity of regressors, this may be problematic.

patent counts. While external funds also affect patenting within a 50km radius, the authors do not allow for spillover effects for regular funds. They also find private sector R&D to be positively associated with patent applications where the effect is weakened by distance across regions.

For the UK, Abramovsky et al. (2007) analyse the relation between university research on the location pattern of business R&D in six specific product groups at the establishment level. The data on R&D active firms comes from the ONS Business Entreprise Research and Development (BERD) data. The analysis uses 2-letter UK postcode areas as the unit of analysis. University research in a postcode area is proxied by the number of research departments that get 2001 RAE rankings of 5 or 5*, and those that get (lower) 1 to 4 rankings.[7] In addition, the log of the number of research students (also divided between universities ranked 4 and below and 5 and 5*) is included. Location of business R&D is the average number of establishments in a postcode area (i.e., no data on R&D spending is used) during the 2000-2003 period. Moreover, they also look at the number of new R&D performing entrants within postcode areas between 2000-2003 where they use the differences in RAE rankings between 1996 and 2001 as measures for university research quality in an attempt to account for unobserved heterogeneity of spatial units that is correlated with the level of research quality. Cross sectional regressions are run separately for R&D establishments in the different product groups,[8] as well as domestic/foreign establishments assuming that the dependent variables follow a negative binomial distribution. The different product group regressions include explanatory variables based on different university research departments (biology, chemistry medical, materials science, computer science, and electrical as well as mechanical engineering). A range of other control variables at the postcode area level are used including total manufacturing employment, diversification of manufacturing employment, skill levels in the population, and a dummy for science parks. The results indicate nearly no statistically significant correlation between the average number of R&D performing firms and the presence of universities, their number, or their overall research quality. Using the RAE rankings for specific university departments, there are some statistically significant associations between university research and the count R&D performers albeit the overall correlation pattern is weak. Nevertheless, the results suggest that 5 or 5* rankings have an influence in pharmaceuticals and chemicals, while RAE rankings 1 to 4 also have an effect in pharmaceuticals and additionally in machinery and communications equipment when the sample is restricted to foreign R&D active firms. With respect to the effect of the change in RAE rankings on the number of R&D entrants, the results confirm a statistically significantly positive association between the quality of chemistry research departments and entry of R&D performing firms in the pharmaceutical industry as well as of material science research departments on the chemical sector. All other coefficients capturing the change in research quality across research departments for the different sectors are not statistically significant, with the exception of the TV and radio equipment industry in which entry

[7]The RAE grade scale is 5* (highest), 5, 4, 3a, 3b, 2 or 1.

[8]The product groups are pharmaceuticals, chemicals, machinery, electrical, TV and radio, and motor vehicles.

is correlated with research quality of electrical and mechanical engineering departments.

Abramovsky and Simpson (2008) extend the analysis of Abramovsky et al. (2007) by analysing the determinants of the number of R&D conducting firms within postcode districts located in proximity to university research departments. In addition, they investigate whether firms that are located close to universities are more likely to engage in collaborative research. The dependent variable for the analysis of firms' location choice is a count of the average number of establishments reporting non-zero own R&D expenditure during 2000-2003 for a product group within postcode districts in the UK.[9] In order to carry out this research, Abramovsky and Simpson combine Business Entreprise Research and Development (BERD) data for 2000-2003 together with Community Innovation Survey data for the UK.[10] While the BERD data provides information on which firms conduct R&D, the CIS data are used to construct a measure for the existence of university-business links. As in Abramovsky et al. (2007), RAE 2001 data is used to construct measures for university presence and quality of research conducted at universities' research departments. In addition, in this paper the authors also consider the count of universities with a radius of 10km as well as a radius of 10 to 50km from the centre of each postcode district. As in all the studies discussed above, Abramovsky and Simpson use a large range of additional controls at the postcode district and area level, such as the number of employees, the percentage of employees with a tertiary degree in science or engineering, R&D intensity, and public funding for R&D, a density measure (count of postcodes at the district level), a measure of skill composition of work force at postcode area level, percentage of economically active population in postcode area, total manufacturing employment in postcode area, percentage of total manufacturing employment in relevant industry, and a measure indicating the presence of science parks. Similar to Harhoff (1999), the estimation of firms' location choice is carried out by a negative binomial model treating the data as a cross-section and running separate regressions for each product group. To estimate firms' propensity to conduct collaborative research with universities, the authors estimate a probit. The results suggest that pharmaceutical firms tend to locate close to world-class chemistry research departments. There is also some evidence for firms located close to universities to be more likely to engage in collaborative research in the fields of chemicals (with materials science departments) and vehicles (with mechanical engineering departments). At the same time, the authors also find that chemicals, vehicles and machinery industries tend to locate in areas with higher manufacturing employment and which are specialised in the respective industry.

While the paper offers additional insight with respect to business-university collaboration, the paper does not address the endogeneity problem inherent in this kind of analysis of firms' location choice. Moreover, the evidence for co-location may be

[9]The product groups considered are pharmaceuticals, chemicals, machinery, electrical machinery, TV and radio equipment, vehicles, precision instruments and aerospace. Note also that they restrict their sample to firms that report in the CIS to have introduced a product or process innovation or have ongoing or abandoned innovative activities or that have innovation-related expenditures over the past three years.

[10]They use both the CIS3 (1998-2000) and CIS4 (2002-2004) data.

confounded with more general unobserved agglomeration externalities. This problem becomes evident when considering the location of firms in London, where also the largest number of universities is found. Hence, what may be interpreted as co-location of firms and universities may be equally well be due to more general unobserved agglomeration effects than specific university-business spillovers. Note also that Abramovsky and Simpson do not consider the possibility for spatial autocorrelation in their analysis.

The endogeneity problem arising from agglomeration is specifically addressed by Kantor and Whalley (2009). In order to assess the effect of university spending on local private sector labour income, the authors take advantage of the fact that in the US a university's spending is a function of the market value of its endowment. This allows Kantor and Whalley to instrument university expenditures by the interaction of a university's initial endowment and time-varying stock market shocks. The results indicate that university expenditures have a minor albeit positive effect on labour income in large urban US counties. A 10 percent increase in university spending results in 0.5 percent higher private sector labour income. This effect intensifies for firms that are found to be technologically close to the research conducted at universities within the same county.

Overall, the empirical literature finds some evidence that university research may have a positive association on surrounding firms' R&D and patenting activity, as well as local labor income. At the same time, the literature focusing on R&D and patenting is rather descriptive in investigating the co-location of university research and innovative activity of private companies. Yet, the evidence by Kantor and Whalley (2009) suggests that the relationship between firms' innovative activity and the presence/quality of university research is simultaneous since both universities and firms are likely to benefit from collaboration. Also the issue of confounding university-business links with unobserved agglomeration externalities, most evident in the case of London, demands attention. Finally, while Harhoff (1999) and Fritsch and Slavtchev consider the issue of spatial autocorrelation, the potential implications of spatial spillovers across spatial units deserve more in-depth analysis. Spatial units, fixed either at some postcode or county level, draw their boundaries in rather arbitrary ways, making it likely for innovation as well university research to spill over to neighbouring spatial units.

3 Identification issues

As has been indicated above, there remain a series of challenges in analyzing the relationship between university research and business innovation. In this paper we take corporate patenting as a proxy for innovation, something that is commonly done, but nevertheless an assumption we should openly discuss. It is well known that patents are a noisy measure of innovation due to different propensities to patent across firms and industries. These differences can be due to differences in firm-level strategies, as well as large heterogeneity in underlying innovations. Moreover, not all innovations are patentable, such as for example innovations in managerial practices or in the creative industries. Patent data do, however, have certain advantages. In our case the main advantage is we can identify the patenting activity of micro firms and SMEs. R&D data are almost always only available for the largest firms.[11] Hence, it is only by using patent data that we can test the hypothesis that the impact of universities varies across firm size. It is clear, however, the propensity to patent is much higher in some industries, we therefore restrict our attention to research that tends to generate patentable innovations, namely engineering related departments as well as medicine/biology/chemistry related department.

A second challenge is that there are no direct measures of the links between universities and firms (i.e., the precise channels of knowledge transmission remain unobserved and unspecified). As put by Jaffe (1989: 957) 'If the mechanism is primarily journal publications, then geographic location is probably unimportant in capturing the benefits of spillovers. If [...] the mechanism is informal conversations, then geographic proximity to the spillover source may be helpful or even necessary in capturing the spillover benefits.' Many argue that knowledge transfer between universities and private firms occurs through channels that operate within a certain geographical distance, such as frequent (informal) face-to-face interaction or are directly influenced by geographical proximity, such as personal networks, seminars and workshops etc. Another important factor may be the location choice of recent university graduates. If graduates tend to choose a location to work or establish their own business in proximity to their university, geographical distance plays a role. Our initial assumption is that university knowledge transmission is restricted to a postcode area (and, implicitly, is uniform within that postcode area). This type of assumption is widespread in the literature, but is clearly unsatisfactory. We do relax this assumption by allowing universities in neighbouring postcodes to generate knowledge flows. However, this 'local' assumption is still unappealing in a global world with rapid travel and communication. In fact, we could hypothesise that geographical location should not matter, especially for large, sophisticated firms with access to wide networks, both nationally and internationally. Hence the hypothesis is that the impact of universities on smaller firms within a postcode region should be greater that the impact on larger firms. The view that local universities can have important impacts on smaller firms is familiar, but statistical tests of this are rare. We therefore choose to split the sample into patenting micro firms & SMEs and patenting large firm.

A further concern, which is rarely discussed, concerns simultaneity. The presence and quality of university research and the patenting outcomes of private firms is likely to be simultaneous. This means that not only firms gain from knowledge transfer from universities, but that also university research benefits from private firms' innovative activity. For example, successful local private firms may support the university with research grants or consultancy contracts, or the successful innovation of private firms may stimulate and direct new (basic) research. This problem is closely linked with endogeneity arising from agglomeration of economic activity. If universities and in-

[11] In particular, in the UK even the Office for National Statistics does not have comprehensive data on smaller firms hence, for example, they cannot produce statistics on the geographical spread of micro and SME R&D activity.

novative firms co-locate in economically dense areas where unobserved agglomeration externalities exist, a positive correlation between university presence and patenting by firms may be observed without there being an actual link. Even including variables accounting for agglomeration will not suffice to avoid endogeneity if these externalities remain unobserved. In order to control for such simultaneity and agglomeration, we adopt an instrumental variable approach. We argue that universities' age is an informative and valid instrument for the quality of university research. It is informative as it correlates highly with our measures of research quality. It is a valid instrument since we argue that the age of a university should have no effect on current patenting other than through research quality.

Modeling firms' patenting decisions

How should one attempt to model any relationship between university research and its impact on firm-level patenting? The previous literature, including those that use R&D rather than patents, has considered both quantity and quality. There has also been recognition that different types of research will have differential impacts across industries and technical areas. For example Jaffe (1989) found the effect of university research on patenting to be more visible within technical areas. Relatively little attention has been focused on lag time, although the work of Adams (1990) found that university research (proxied by journal publications) had an impact on productivity after two decades. In addition, there has also been an inability to differentiate between impacts on small and large firms, since most databases do not have firm-level patent data.[12] However, Acs et al. (1994) and Audretsch (1998) suggest that universities are particularly important as a source of innovative knowledge for small firms. Also, previous research points to the importance of location in the relationship between innovative firms and university research. Fritsch and Slavtchev (2007), for example, find substantial clustering of patent applications in regions, particularly in urban agglomeration areas.

We assume the following relationship. Patenting P_i by companies in postcode area i is some function of university research U_i within area i and a vector $\mathbf{X}$ of k covariates with dimension $k \times 1$.

$$P_i = f(U_i, \mathbf{X}_i) \tag{1}$$

Patenting by companies is further broken down by large firm versus small firm (which is micro firm and SME combined) where we estimate separate regressions by size category. We employ different measures for university research including (i) the number of universities within a postcode area, (ii) the presence of engineering or biological sciences research departments, (iii) the number of researchers active at these departments at universities located within a given postcode area, and (iv) 'quality' of research conducted at these departments as assessed by the RAE 2001.[13] Measures (i) and (iii) aim to capture the amount of research undertaken within a postcode area.

[12] Although Acs et al (1992) do repeat the Jaffe analysis using small business innovation data.

[13] For more information on the RAE 2001 see www.rae.ac.uk/2001.

The intuition is that the larger the amount of research conducted within a postcode area, the greater the potential for knowledge transfer is and hence the higher the level of patenting should be. Only measure (iv) is a true measure of research quality. To measure quality, we split the total number of researchers into the number of researchers that received grades 1-4 versus those that received top grades 5 or 5* in the RAE 2001. Moreover, we also use the single overall RAE grade received by a university, which can be regarded as a global measure for the quality of research conducted at a university.[14]

Covariates included in $\mathbf{X}$ are (i) the population density (i.e., the number of people per hectare as indicated in the Census, 2001) within postcode area i, (ii) the log of the total number of people employed in the manufacturing sector (Census, 2001), (iii) the diversification of the industrial production within a postcode area where the measure varies between 0, indicating no diversification, and 1 indicating complete diversification diversification,[15] (iv) the ratio of unskilled to skilled labour where the information comes from the labour Census 2001, and (v) log R&D by region as reported by the ONS.

Covariates (i) and (ii) are included to control for agglomeration of economic activity in order to avoid endogeneity caused by co-location as discussed above. The inclusion of (iii) above relates to the debate about so called Marshall-Arrow-Romer (MAR) and Jacob externalities (Glaeser et al., 1992). MAR externalities arise when industries are concentrated within a location. In contrast, Jacobs externalities (Jacobs, 1969) emerge as a result of the diversity of industries within a location. Hence, MAR externalities can be regarded as intra-industry spillovers while Jacobs externalities are inter-industry spillovers.[16] Covariates (iv) and (v) capture the level of technology and research within postcode areas.

4 Data

The data for the analysis comes from two main sources. Data on firm-level patenting comes from the Oxford Firm Level Intellectual Property (OFLIP) database. The database draws on the Financial Analysis Made Easy (FAME) data that covers the entire population of registered UK firms (FAME downloads data from Companies House records).[17] OFLIP contains additional information on the IP activity of firms in the form of patents and trade marks. In this paper we use publications of both UK and EPO patents (in 2001) as our measure of patenting. OFLIP has been constructed by

[14] We use the maximum grade received by all universities within a postcode area.

[15] The manufacturing diversification measure is constructed as the sum of squares of the share of 4-digit SIC within postcode areas using OFLIP. It takes a value of 0 if a single 4-digit SIC produces all the output in the postcode, and tends to 1 as diversification increases. Note that OFLIP is appropriate for this since it contains data on all two million UK registered firms.

[16] The question of whether diversity or concentration promotes innovation has been analysed by Feldman and Audretsch (1999) who find for their cross-section of data that innovative activity in complementary industries which share a common fundamental science base to cluster both in terms of production activity and innovation.

[17] In this paper we use firms to mean registered firms. Hence firm refers to the legal entity that organizes production, in contrast to census-type data that uses the plant or production unit.

matching the FAME database and a number of firm-level IP datasets.[18]

The second source of data is the RAE for 2001, which is collated by the Higher Education Funding Council of England and Wales (HEFCE). The HEFCE data provide a range of indicators from which those listed in Table 1 are selected.[19] Note that even though the RAE is collated in 2001 it relates to research activity over the period 1996 to 2001.[20] This allows for a considerable time lag in the effect of university research on patenting even bearing in mind the usual 18 month lag between a patent's application and publication date.[21] Population density and skills data come from the England and Wales Census 2001 (and also Scottish Census 2001). The R&D data is downloaded direct from ONS website (database rdbd7) and is at the reginal level since more disaggregated data is not collected. The age of universities used as an instrument for research quality has been collected directly from universities' websites. Table 1 shows the summary statistics for the variables.

Table 1: Summary statistics, 2001

	Mean	Std. Dev.	Min	Max
Count of patentees	12.10	11.34	0	85
Count of large patentees	2.53	3.31	0	17
Count of small patentees	9.57	8.75	0	68
Number of universities (0, 1, 2, 3 or more)	0.91	0.96	0	3
Number of engineering departments	1.27	1.83	0	9
Number of biological science dept.	1.45	2.42	0	14
Number of engineering dept. $\leq$ grade 4	0.81	1.21	0	7
Number of engineering dept. grade 5 or 5*	0.50	1.10	0	5
Number of bio. science dept. $\leq$ grade 4	0.68	1.12	0	6
Number of bio. science dept. grade 5 or 5*	0.79	1.75	0	10
University age	70.58	143.94	0	834
Population density	10.97	18.06	0.12	93.77
Log manufacturing employment	10.15	0.84	6.58	11.87
Diversification of industry	84.69	13.61	19.46	97.14
Ratio of skilled to unskilled	0.58	0.37	0.26	2.31
Log R&D	6.75	0.89	4.83	8.07

[18] For details on the matching process and further details on the database see Rogers and Helmers (2009).

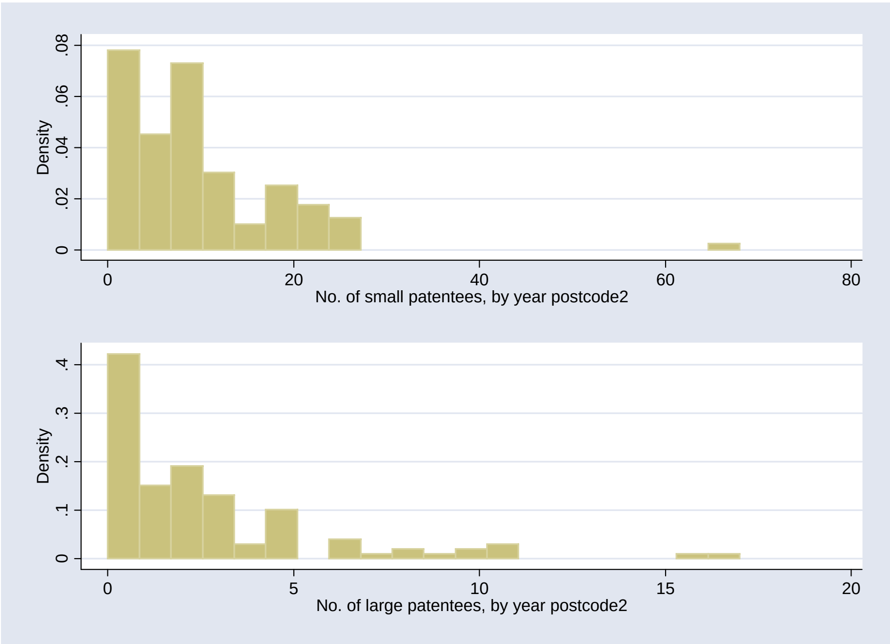
[19] Our definition of 'engineering' department includes General Engineering, Chemical Engineering, Electrical and Electronic Engineering, Mechanical, Aeronautical and Manufacturing Engineering and Mineral and Mining Engineering. Our definition of 'biological sciences' department includes Clinical Laboratory Sciences, Pre-Clinical Studies, Anatomy, Physiology, Pharmacology, Pharmacy, Biological Sciences, and Chemistry.

[20] For humanities the period is 1994-2000 but we do not use RAE data on humanities in this paper.

[21] Note that the correlation between patenting in 2001 and subsequent years is very high (when aggregated to the postcode level). The correlation coefficient between 2001 and 2002 is 0.95 and only falls to 0.9 for between 2001 and 2005. This implies that an analysis of the lag structure of any impacts from university research is problematic.

Figure 1 shows histograms for two of the dependent variables: the count, by postcode, for large firms that patent and also for small firms (which are both micro firms and SMEs). Large firms are defined as those with less than £39 million in total assets (the EU definition is Euro 43 million).[22] For large firms, in 42 of the 117 postcodes there are no patentees at all, and the postcode area (Birmingham) with the largest number of large-firm patentees has 17 large firms that patented in 2001. Restricting attention to smaller firms, there are only four postcodes that have no patenting firms. The postcode with the greatest number of smaller patentees is, again, Birmingham with 68.[23]

Figure 1: Histograms of large and small patentees, by postcode area



5 Estimation

Our outcome variable is a count variable, i.e., it assumes non-negative discrete values and has no natural ceiling. We begin with estimating the model ignoring the fact that the dependent variable is not normally distributed. The problems with applying OLS

[22]Total assets are used since this variable has the best coverage in the OFLIP database. This is due to the fact that in the UK all firms have a legal requirement to report total assets, but not total revenue or employment.

[23]The influence of Birmingham for the regressions on smaller firms has been checked and it does not affect the qualitative results, although the magnitude of coefficients does change.

in a setting with a discrete outcome variable are well known from the binary response case. Most importantly, $E(y|\mathbf{x})$ should be non-negative while OLS will usually still result in $\mathbf{x}\beta' < 0$ (i.e. predicted values are negative). Non-linearity of $E(y|\mathbf{x})$ is another characteristic of count data ignored by OLS. An obvious alternative is to estimate the model assuming a Poisson distribution. The problem with the assumption of a Possion distribution is its equidispersion property, i.e., it assumes that the mean and variance are the same. However, in our data, we find overdispersion to be present, which means the mean is not equal to the variance. To account for overdispersion, we use the negative binomial model as proposed by Cameron and Trivedi (1986).

Assume a conditional mean function

$$E(y_i|\mathbf{x}_i) = \exp(\mathbf{x}_i\beta) \tag{2}$$

The corresponding variance is

$$Var(y_i|\mathbf{x}_i) = \exp(\mathbf{x}_i\beta)[1 + \alpha_i^2 \exp(\mathbf{x}_i\beta)] \tag{3}$$

where α_i^2 denotes the variance of α_i which can be interpreted as a measure of unobserved heterogeneity. Parameters β and α are jointly obtained from maximizing the following log-likelihood function

$$l_i(\beta, \alpha^2) = \alpha^2 \log\left[\frac{\alpha^{-2}}{\alpha^{-2} + \exp(\mathbf{x}_i\beta)}\right] + y_i \log\left[\frac{\exp(\mathbf{x}_i\beta)}{\alpha^{-2} + \exp(\mathbf{x}_i\beta)}\right] + \tag{4}$$

$$+ \log[\Gamma(y_i + \alpha^{-2})/\Gamma(\alpha^{-2})] \tag{5}$$

where $\Gamma(\cdot)$ is the gamma function defined for $r > 0$ by $\Gamma(r) = \int_0^\infty z^{r-1} \exp(-z)dz$.

IV Approach

The relationship between the quality of research undertaken at a university and patenting activity of private firms may be characterized by a simultaneous relationship. This means that (x denoting university research) is, in part, determined by y (the number of patentees) and hence x is correlated with the error term ϵ. This means that in the DGP, the outcome variable and the regressor are simultaneously generated. In order to address the simultaneity concern, we employ an IV approach. The instrument has to be informative and valid. An instrumental variable is informative if it is correlated with the endogenous variable, which can easily be verified. Validity requires that the variable is uncorrelated with the disturbance term. This is achieved if the instrument affects the outcome variable exclusively through the endogenous variable conditional on all other exogenous covariates.

As before, we first estimate a linear probability model assuming normality using a standard IV estimator. Denoting the instrumental variable as z, we assume that $E(z, \epsilon) = 0$. Hence, the IV estimator is

$$\beta_{IV} = (z'x)^{-1}z'y \tag{6}$$

with $\hat{Var}(\beta_{IV}) = \hat{\sigma}^2(z'x)^{-1}(z'z)(x'z)^{-1}$. Since the linear model has the drawbacks discussed above, we employ a control function approach to account for endogeneity assuming a negative binomial distribution. As suggested in Wooldridge (2001), we assume that our structural model is

$$E(y_{1i}|y_{2i}, \mathbf{x}_i, e_i) = \exp(\beta_1 y_{2i} + \mathbf{x}_i'\beta_2 + e_i) \tag{7}$$

where y_2 denotes endogenous research quality, $\mathbf{x}$ is a vector of covariates, and e is an unobserved latent variable. If research quality is endogenous, we have $E(y_2, e) \neq 0$ while we still assume that $E(\mathbf{x}, e) = 0$. We assume that y_2 is given by a linear reduced form

$$y_{2i} = \mathbf{x}_{1i}'\phi_1 + \phi_2 z_i + \varepsilon_i \tag{8}$$

Crucially, z_i is a variable that is correlated with endogeneous research quality while it does not affect y_1 other than through y_2 conditional on $\mathbf{x}$. We assume that $E(z, \varepsilon) = 0$. Since we have only a single exclusion restriction, the model is exactly identified. While this does not pose any problem for identification, it limits our ability to test for the validity of the instrument. In order to estimate Equation (7) using the structure imposed on the endogenous variable in Equation (8), we have to make a rather restrictive assumption on the error terms. We assume that

$$e_i = \gamma_i \varepsilon_i + \xi_i \tag{9}$$

where ξ is white noise and independent of ε. The assumption made in Equation (9) that e and ε are linearly related always holds if e and ε are jointly normal distributed. Clearly, if $\gamma \neq 0$, y_{2i} is endogenous. Using (8) and (9), we can rewrite (7) as

$$E(y_{1i}|y_{2i}, \mathbf{x}_i, \varepsilon_i) = \exp(\beta_1 y_{2i} + \mathbf{x}_i'\beta_2 + \gamma_i \varepsilon_i) \tag{10}$$

in order to obtain unbiased and consistent estimators for β_1, β_2, and γ_i. In practice, we estimate Equation (10) using a two-step procedure, estimating first (8) which allows recovering an estimate for $\hat{\varepsilon}_i$ which is plugged into (10) where inference is based on bootstrapping in the second stage.

Spatial Approach

Some of the previous studies discussed in Section 2 tested (e.g., Harhoff, 1999)[24] and accounted for spatial autocorrelation (Fritsch and Slavtchev, 2007).[25] Spatial dependence may arise in our setting because borders of postcode areas do not necessarily

[24] Note that Harhoff uses a Lagrange Multiplier test to test for the presence of spatial autocorrelation in the error term as well as to test for a spatial lag term.

[25] Fritsch and Slavtchev (2007) included the average residual of adjacent regions, which is problematic because these residuals have been obtained based on the assumption of no spatial autocorrelation in the error term.

coincide with geographical coverage of economic activity. To capture geographic proximity, we construct a spatial weight matrix which assumes the value of one if postcode areas are adjacent and otherwise zero. Hence, weights are binary. In order to capture spatial spillovers of university research, we construct the weighted sum of our university research measures in adjacent postcode areas, i.e., $\sum_{j=1}^{N} w_{ij}x_j$ where $w_{ij} = 1$ if postcode areas i and j are adjacent and $w_{ij} = 0$ otherwise. If there is no spatial dependence in the error term and no spatially lagged dependent variable is added to the basic specification, OLS and ML still yield consistent estimates. From a theoretical point of view there is no reason to include a spatially lagged dependent variable and we therefore concentrate on testing for the absence of spatial autocorrelation in the residuals using a Lagrange Multiplier test.[26]

6 Results

Table 3 shows a set of OLS regressions as a baseline. The first two columns indicate support for the basic hypothesis that smaller firms benefit from close proximity to universities, but large firms do not. The other variables that are significant for both large and smaller firms are: log of manufacturing employment, the ratio of skilled to unskilled workers, and log R&D (although note that the coefficients are of different magnitude). For the regression with smaller firms, population density appears to have a negative effect. The last four columns use the number of engineering and biological sciences departments, instead of number of universities. A similar pattern of results is found. While a larger number of engineering and biological sciences departments is associated with larger numbers of patenting small firms within postcode areas, no such statistically significant association is found for larger firms. As discussed above, OLS is not well suited for count data. Table 4 repeats the specifications in Table 3 using a negative binomial model. The qualitative results are unchanged (except for the number of engineering departments which is now marginally statistically significant for large firms). In summary, when considering associations between the presence of local universities and patenting, in general only smaller firms are found to exhibit positive associations.

Table 5 looks at associations between the number of researchers in engineering departments and also the number in biological sciences and the number of patentees using both OLS (Columns (1)-(4)) and the negative binomial model (Columns (5)-(8)). Again, only the number of small firms is associated with the number of researchers located at universities within postcode areas (although this result is strongest for engineering).

Table 6 investigates this further by considering the quality of research conducted at the different departments (as assessed by the RAE in 2001). The RAE grades departments from 1 (lowest) to 5* (highest), but in these regressions we use a count measure based on those with grades 1-4 versus those 5 or 5*. For example, the central Birmingham postcode has five engineering departments, with two of these graded 1-4 and

[26] For a detailed description of the test in the context of spatial econometrics, see Anselin (1988).

three with grades 5 or 5*. These regressions indicate that quality does matter. In both engineering and the biological sciences only the number of 5 or 5* ranked departments has a significant, positive association with the number of firms patenting. Notably, in these regressions, both small and large firm patenting is associated with the number of 5 or 5* ranked engineering departments. This is not the case for the number of 5 or 5* ranked biological sciences departments where again only the regressions for smaller firms show a statistically significant coefficient associated with the number of 5 or 5* ranked departments.

Table 7 uses the maximum RAE grade achieved by engineering departments within a postcode area. Similarly, Table 8 shows the results for the maximum RAE grade achieved by biological sciences departments within postcode areas. The first four columns in both tables show again results using OLS and a negative binomial model. It is clear that in both cases the coefficients associated with research quality as assessed by the RAE are statistically significant only in the regressions for smaller firms. This means that there is a positive correlation between small firm patentees within postcode areas and better RAE grades both in engineering and biological sciences. The following four columns (Columns (5)-(8)) report results when using university age as an instrument for overall research quality. Columns (5) and (6) report standard IV OLS results whereas Columns (7) and (8) report the results from using a control function approach. Since the control function approach attempts to account for endogeneity using a negative binomial model, it is our preferred specification. The negative coefficient of the control term indicates that the latent factor captured by university age, which is positively correlated with research quality, is negatively correlated with the number of patentees within postcode areas. Hence, simultaneity appears to cause a downward bias in the coefficients associated with with research quality. Also, we find the magnitude of the bias to be considerable as for example for biological sciences the coefficients more than doubles when using the control function approach. As before, we only find coefficients of instrumented research quality to be statistically significant for small firms.

Finally, Table 9 shows results when a spatially lagged university variable is included in the specification to account for neighbourhood effects. The variable represents the average of adjacent postcode areas' university variables. Column (2) therefore suggests that not only the number of universities within a postcode area but also the number of universities in adjacent postcode areas is positively correlated with the number of small firm patentees. When we look at the number of engineering and biological sciences departments in Columns (3)-(6), we only find this positive neighbourhood effect for the number of engineering departments. Note that we also test for spatial autocorrelation in the error term using a Lagrange Multiplier test. We find that the null hypothesis of absence of spatial autocorrelation is never rejected for the sample of small firms while it is not rejected for large firms at the 5-percent level.

7 Conclusion

This paper has analysed the link between university research and the patenting activity of firms located close to universities. There are a number of major challenges in such an analysis. First, university research has many different dimensions, including subject area, quantity and quality. Second, the channels through which this research may impact on firms are unclear, as are the lag times involved. Following from this is the uncertainty of how to define 'local' (i.e. why it may be that 'local' firms benefit more from university research). Third, there are important and difficult issues surrounding simultaneity and endogenous location choices. This paper has approached challenges in the following way. The analysis is based upon the 117 two-digit postcode areas in the UK. For each of these postcodes we construct the following measures of university research activity (i) the number of universities in the same postcode area, (ii) the presence of engineering or biological sciences research departments, (iii) the number of researchers active in these departments, (iv) the 'quality' of research (as assessed by the RAE 2001). Similarly, for each postcode we construct a count of the number of patentees, broken down by large firm patentees and small firm patentees. Our hypothesis is that any impact of university research is more likely to occur for small firms. In contrast, large firms are likely to have access to many universities' research from around the UK, if not globally, hence it is less likely to find a specific impact from local universities.

We do not claim that even when instrumenting university research quality, we uncover causal effects between university research and corporate patenting. The analysis merely shows that, in almost all specifications, it is only the number of small firm patentees that show a significant, positive correlation with university research. This is not to say that university research has no impact on large firms, only that any such effects cannot be detected with the two-digit postcode specification used here (or even when allowing for neighbouring postcodes as in Table 9).

The analysis uncovers a number of other findings. First, when considering research from engineering and biological sciences departments separately, the regression analysis indicates that the number and scale of engineering departments has a stronger association with the number of small firm patentees. Second, using the RAE 2001 grading of engineering and biological sciences research departments, the analysis finds that research quality matters: only the departments with the highest RAE grades exhibit positive and significant associations with the number of small firm patentees. Third, in order to try to remove potential simultaneity, we use the age of the university as an instrument for RAE grade. The results indicate that the main findings are supported, in fact the magnitude of the coefficients rise. Lastly, the paper also relaxes the assumption that university research can only impact on firms within a two-digit postcode by incorporating neighbourhood effects (from adjacent postcode areas). This specification still indicates no impact on large firms, but does indicate that small firms benefit from neighbourhood effects.

We view this research as only a first step in a fuller understanding of the impact of university research. Further research should attempt to refine the various measures of university research. We have also not investigated the possible lag structure of the impact of university research. Further, it is well known that universities have increasingly engaged in technology transfer activities, including spin-off companies and the creation of science parks. These activities need to be incorporated into the analysis. The simple measure of patenting used here (a count of patentees) can also be refined by looking at the volume and nature of such patenting (e.g. whether international or national patent, and also by using IPC categories). Alternative methods of identification should also be pursued, as should further analysis of spatial effects.

References

[1] Abramovsky Laura and Simpson Helen (2008): 'Geographic Proximity and Firm-University Innovation Linkages: Evidence from Great Britain', CMPO Working Paper No. 08/200.

[2] Abramovsky Laura, Harrison Rupert, Simpson Helen (2007): 'University research and the location of business RandD', *Economic Journal*, Vol. 117(519), pp. C114-C141.

[3] Adams J. (1990). 'Fundamental Stocks of Knowledge and Productivity Growth.' *Journal of Political Economy* 98(4): 673-702.

[4] Anselin Luc (1988): 'Spatial Econometrics: Methods and Models', *Kluwer Academic Publishers.*

[5] Audretsch David (1998): 'Agglomeration and the Location of Innovative Activity', *Oxford Review of Economic Policy*, Vol. 14, No. 2, pp. 18-29.

[6] Cameron Colin and Trivedi Pravin (1986): 'Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests', *Journal of Applied Econometrics*, Vol. 1(1), pp. 29-53.

[7] Feldman Maryann and Audretsch David (1999): 'Innovation in cities: Science-based diversity, specialization, and localized competition', *European Economic Review*, Vol. 43, pp. 409-429.

[8] Fritsch Michael and Slavtchev Viktor (2007): 'Universities and Innovation in Space', *Industry and Innovation*, Vol. 14, No. 2, pp. 201-218.

[9] Greenhalgh Christine and Rogers Mark (2009): *Innovation, Intellectual Property and Economic Growth.* Princeton, Princeton University Press.

[10] Grilliches Zvi (1979): 'Issues in Assessing the Contribution of Research and Development to Productivity Growth', *The Bell Journal of Economics*, Vol. 10, No. 1, pp. 92-116.

[11] Rogers Mark and Helmers Christian (2009): 'An Analysis of the Innovative Activity of Entrepreneurial and Young Firms', mimeo.

[12] Harhoff Dietmar (1999): 'Firm Formation and Regional Spillovers - Evidence from Germany', *Economics of Innovation and New Technology*, Vol. 8, pp. 27-55.

[13] HM Treasury (2003): 'Lambert Review of University-Business Collaboration - Final Report'.

[14] Jaffe Adam (1989): 'Real Effects of Academic Research', *American Economic Review*, Vol. 79, No. 5, pp. 957-970.

[15] Kantor Shawn and Whalley Alexander (2009): 'Do Universities Generate Agglomeration Spillovers? Evidence from Endowment Value Shocks', *NBER Working Paper No. 15299.*

[16] Wooldridge Jeffrey (2001): 'Econometric Analysis of Cross Section and Panel Data', *MIT Press*.

Table 2: Correlation Matrices

	Engineering Sciences						
	Patenting	# University	# Engineering Department	# Researchers	Quality	# Researchers 5*	# Researchers 4*
Patenting	1.000						
# University	0.442	1.000					
# Engineering	0.476	0.878	1.000				
# Researchers	0.476	0.847	0.962	1.000			
Quality	0.433	0.825	0.931	0.966	1.000		
# Researchers	0.532	0.664	0.776	0.749	0.679	1.000	
# Researchers	0.331	0.786	0.872	0.808	0.794	0.428	1.000

	Biological Sciences						
	Patenting	# University	# Biology Department	# Researchers	Quality	# Researchers 5*	# Researchers 4*
Patenting	1.000						
# University	0.442	1.000					
# Biological	0.432	0.803	1.000				
# Researchers	0.396	0.788	0.970	1.000			
Quality	0.358	0.756	0.921	0.977	1.000		
# Researchers	0.470	0.646	0.835	0.785	0.709	1.000	
# Researchers	0.263	0.698	0.830	0.785	0.752	0.434	1.000

Table 3: Regressions: OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	Large firms	Small firms	Large firms	Large firms	Small firms	Small firms
No. of universities	0.378	2.105**				
	(0.362)	(0.961)				
No. engineering dept.			0.270		1.396**	
			(0.180)		(0.512)	
No. bio science dept.				0.079		1.035**
				(0.150)		(0.374)
Population per hectare	-0.055	-0.162**	-0.054	-0.055	-0.161**	-0.165**
	(0.030)	(0.051)	(0.030)	(0.030)	(0.048)	(0.051)
Log manufacturing employment	1.566***	5.562***	1.499***	1.699***	5.299***	5.724***
	(0.400)	(1.290)	(0.424)	(0.419)	(1.298)	(1.330)
Ind. diversification	-0.011	0.035	-0.011	-0.011	0.034	0.041
	(0.018)	(0.031)	(0.018)	(0.018)	(0.031)	(0.032)
Skilled/unskilled labour	6.392***	15.006***	6.300***	6.630***	14.733***	14.245***
	(1.594)	(2.686)	(1.674)	(1.838)	(2.627)	(3.541)
Log R&D	0.727*	1.578**	0.740**	0.682*	1.616***	1.474**
	(0.290)	(0.515)	(0.276)	(0.283)	(0.454)	(0.465)
Constant	-20.782***	-69.246***	-20.133***	-21.760***	-66.541***	-69.824***
	(4.370)	(14.188)	(4.490)	(4.551)	(14.009)	(14.175)
R-squared	0.318	0.470	0.326	0.311	0.498	0.495
N	117	117	117	117	117	117

Note: * indicates significance at 10%; ** at 5%; *** at 1%.

Table 4: Regressions: Negative Binomial

	(1)	(2)	(3)	(4)	(5)	(6)
	Large firms	Small firms	Large firms	Large firms	Small firms	Small firms
No. of universities	0.130	0.167***				
	(0.110)	(0.057)				
No. engineering dept.			0.080*		0.094***	
			(0.047)		(0.030)	
No. bio science dept.				0.016		0.062***
				(0.044)		(0.023)
Population per hectare	-0.018	-0.022***	-0.019	-0.018	-0.022***	-0.022***
	(0.012)	(0.005)	(0.012)	(0.013)	(0.005)	(0.005)
Log manufacturing employment	0.936***	0.743***	0.919***	1.010***	0.723***	0.760***
	(0.203)	(0.109)	(0.193)	(0.198)	(0.112)	(0.104)
Ind. diversification	0.000	0.007	-0.000	0.000	0.007	0.007
	(0.007)	(0.004)	(0.007)	(0.007)	(0.005)	(0.004)
Log manufacturing employment	2.671***	2.028***	2.698***	2.862***	2.032***	2.048***
	(0.677)	(0.263)	(0.652)	(0.663)	(0.278)	(0.354)
Log R&D	0.342***	0.194***	0.349***	0.324**	0.199***	0.181***
	(0.101)	(0.049)	(0.101)	(0.100)	(0.043)	(0.045)
Constant	-12.744***	-8.523***	-12.547***	-13.390***	-8.299***	-8.577***
	(2.286)	(1.223)	(2.198)	(2.303)	(1.248)	(1.184)
N	117	117	117	117	117	117
Observations	117	117	117	117	117	117

Note: * indicates significance at 10%; ** at 5%; *** at 1%.

Table 5: Regressions: OLS and Negative Binomial

	OLS				Negative Binomial			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Large firms	Large firms	Small firms	Small firms	Large firms	Large firms	Small firms	Small firms
No. engineering researchers	0.011		0.062***		0.002		0.003***	
	(0.007)		(0.019)		(0.002)		(0.001)	
No. bio science researchers		0.001		0.021		0.0002		0.001*
		(0.003)		(0.007)		(0.374)		(0.000)
Population per hectare	-0.049	-0.055*	-0.136***	-0.154***	-0.018	-0.017	-0.020***	-0.021***
	(0.031)	(0.031)	(0.051)	(0.054)	(0.011)	(0.013)	(0.005)	(0.005)
Log manufacturing empl.	1.479***	5.751***	5.139***	6.059***	0.938***	1.025***	0.733***	0.794***
	(0.420)	(0.442)	(1.227)	(1.451)	(0.195)	(0.197)	(0.112)	(0.110)
Ind. diversification	-0.010	-0.011	0.039	0.039	-0.0003	0.0004	0.007	0.007*
	(0.018)	(0.018)	(0.030)	(0.030)	(0.006)	(0.007)	(0.004)	(0.004)
Skilled/unskilled labour	6.066***	6.776***	13.355***	14.231***	2.737***	2.874***	2.005***	2.075***
	(1.836)	(1.859)	(3.041)	(3.563)	(0.663)	(0.665)	(0.312)	(0.365)
Log R&D	0.660**	0.665**	1.204**	1.290**	0.319***	0.321***	0.161***	0.167***
	(0.286)	(0.282)	(0.492)	(0.504)	(0.099)	(0.098)	(0.046)	(0.048)
Constant	-19.366***	-22.161***	-61.958***	-71.524***	-12.539***	-13.531***	-8.122***	-8.808***
	(4.469)	(4.741)	(12.965)	(15.109)	(2.285)	(2.237)	(1.271)	(1.257)
R-squared	0.366	0.344	0.548	0.505				
N	117	117	117	117	117	117	117	117

Note: * indicates significance at 10%; ** at 5%; *** at 1%.

Table 6: Regressions: OLS and Negative Binomial

	OLS				Negative Binomial			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Large firms	Large firms	Small firms	Small firms	Large firms	Large firms	Small firms	Small firms
No. engineering dept: <4	-0.121		0.184		-0.034		0.038	
	(0.244)		(0.556)		(0.085)		(0.049)	
No. engineering dept:5 or 5*	0.749*		2.608***		0.205**		0.139**	
	(0.405)		(0.979)		(0.104)		(0.059)	
No. bio science dept: <4		-0.256		0.011		-0.083		0.022
		(0.281)		(0.635)		(0.096)		(0.044)
No. bio science dept:5 or 5*		0.251		1.562**		0.063		0.079**
		(0.271)		(0.761)		(0.068)		(0.038)
Population per hectare	-0.060**	-0.053	-0.178***	-0.156***	-0.024	-0.018	-0.022***	-0.022***
	(0.028)	(0.032)	(0.042)	(0.055)	(0.011)	(0.013)	(0.005)	(0.005)
Log manufacturing empl.	1.513***	1.738***	5.438***	5.823***	0.899***	1.004***	0.721***	0.760***
	(0.464)	(0.433)	(1.359)	(1.384)	(0.180)	(0.190)	(0.107)	(0.102)
Ind. diversification	-0.013	-0.016	0.027	0.024	-0.001	-0.002	0.006	0.006
	(0.018)	(0.017)	(0.027)	(0.028)	(0.006)	(0.007)	(0.004)	(0.004)
Skilled/unskilled labour	6.135***	6.473***	14.533***	13.681***	2.813***	2.842***	2.042***	2.040***
	(1.815)	(1.969)	(3.141)	(3.921)	(0.651)	(0.671)	(0.305)	(0.360)
Log R&D	0.689**	0.654**	1.436**	1.394***	0.332***	0.330***	0.188***	0.178***
	(0.273)	(0.301)	(0.462)	(0.497)	(0.099)	(0.098)	(0.044)	(0.046)
Constant	-19.519***	-21.370***	-65.559***	-68.429***	-12.128***	-13.122***	-8.154***	-8.477***
	(4.561)	(4.542)	(13.680)	(13.562)	(2.024)	(2.206)	(1.170)	(1.130)
R-squared	0.389	0.357	0.550	0.535				
N	117	117	117	117	117	117	117	117

Note: * indicates significance at 10%; ** at 5%; *** at 1%.

Table 7: Regressions: OLS and Negative Binomial

	OLS		Negative Binomial		IV OLS		Control Function	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Large firms	Small firms	Large firms	Small firms	Large firms	Small firms	Large firms	Small firms
RAE grade engineering	0.217	0.868***	0.060	0.058***	0.022	0.288	0.114	0.089***
	(0.142)	(0.322)	(0.050)	(0.022)	(0.150)	(0.281)	(0.079)	(0.031)
Population per hectare	-0.049	-0.142***	-0.016	-0.019***	-0.054*	-0.157***	-0.016	-0.020***
	(0.029)	(0.051)	(0.012)	(0.005)	(0.031)	(0.052)	(0.015)	(0.006)
Log manufacturing employment	1.497***	5.618***	0.933***	0.758***	1.747***	6.358***	1.019***	0.808***
	(0.422)	(1.390)	(0.212)	(0.117)	(0.482)	(1.587)	(0.255)	(0.131)
Ind. diversification	-0.009	0.038	0.001	0.007*	-0.011	0.035	0.0002	0.007
	(0.017)	(0.031)	(0.007)	(0.004)	(0.018)	(0.029)	(0.008)	(0.004)
Skilled/unskilled labour	6.022***	14.337***	2.614***	1.985***	6.816***	16.691***	2.776***	2.133***
	(1.628)	(2.795)	(0.719)	(0.317)	(1.792)	(2.856)	(0.841)	(0.367)
Log R&D	0.714**	1.424**	0.335***	0.178***	0.668**	1.287**	0.325***	0.175***
	(0.286)	(0.508)	(0.098)	(0.047)	(0.281)	(0.525)	(0.106)	(0.051)
Constant	-20.035***	-68.819***	-12.681***	-8.565***	-22.161***	-75.125***	-13.462***	-9.065***
	(4.497)	(14.842)	(2.401)	(1.319)	(5.067)	(16.537)	(2.775)	(1.532)
Control Term							-0.097	-0.053*
							(0.077)	(0.030)
1st Stage Partial R^2					0.536	0.536		
F-Test					109.814	109.814		
					(0.000)	(0.000)		
R-squared	0.362	0.499			0.633	0.453		
N	117	117	117	117	117	117	117	117

Note: * indicates significance at 10%; ** at 5%; *** at 1%.

Table 8: Regressions: OLS and Negative Binomial

	OLS		Negative Binomial		IV OLS		Control Function	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Large firms	Small firms	Large firms	Small firms	Large firms	Small firms	Large firms	Small firms
RAE grade bio science	0.033	0.662**	0.001	0.038*	0.020	0.264	0.019	0.077**
	(0.131)	(0.323)	(0.045)	(0.023)	(0.138)	(0.261)	(0.078)	(0.038)
Population per hectare	-0.054	-0.145***	-0.017	-0.020***	-0.054*	-0.157***	-0.017***	-0.021***
	(0.031)	(0.054)	(0.013)	(0.005)	(0.031)	(0.052)	(0.016)	(0.006)
Log manufacturing employment	1.735***	5.910***	1.035***	0.789***	1.750***	6.400***	1.074***	0.812***
	(0.427)	(1.430)	(0.222)	(0.121)	(0.478)	(1.593)	(0.259)	(0.136)
Ind. diversification	-0.011	0.031	0.0004	0.007*	-0.011	0.032	0.0001	0.007
	(0.017)	(0.029)	(0.007)	(0.004)	(0.018)	(0.028)	(0.008)	(0.004)
Skilled/unskilled labour	6.781***	15.351***	2.892***	2.114***	6.828***	16.859***	2.919***	2.186***
	(1.740)	(3.138)	(0.704)	(0.348)	(1.755)	(2.866)	(0.867)	(0.388)
Log R&D	0.676**	1.493**	0.320***	0.180***	0.671**	1.328**	0.316***	0.177***
	(0.290)	(0.495)	(0.100)	(0.046)	(0.283)	(0.525)	(0.110)	(0.046)
Constant	-22.076***	-71.687***	-13.629***	-8.860***	-22.200***	-75.636***	-13.795***	-9.092***
	(4.635)	(15.278)	(2.509)	(1.354)	(5.025)	(16.599)	(2.798)	(1.494)
Control Term							-0.027	-0.055
							(0.087)	(0.036)
1st Stage Partial R^2					0.595	0.595		
F-Test					161.621	161.621		
					(0.000)	(0.000)		
R-squared	0.344	0.484			0.309	0.446		
N	117	117	117	117	117	117	117	117

Note: * indicates significance at 10%; ** at 5%; *** at 1%.

Table 9: Regressions: Negative Binomial Spatial Lag of University Research

	(1)	(2)	(3)	(4)	(5)	(6)
	Large firms	Small firms	Large firms	Small firms	Large firms	Small firms
No. of universities	0.167	0.206***				
	(0.117)	(0.056)				
Spatial Lag No. of universities	0.035	0.041***				
	(0.033)	(0.015)				
No. engineer dept.			0.077	0.109***		
			(0.055)	(0.030)		
Spatial Lag No. engineer dept.			-0.002	0.017*		
			(0.021)	(0.009)		
No. biological sciences dept.					0.019	0.107***
					(0.051)	(0.031)
Spatial Lag No. bio science dept.					0.002	0.009
					(0.017)	(0.006)
Population per hectare	-0.016	-0.021***	-0.018	-0.022	-0.018	-0.022***
	(0.012)	(0.004)	(0.011)	(0.004)	(0.013)	(0.004)
Log manufacturing employment	0.879***	0.686***	0.927***	0.683***	0.999***	0.688***
	(0.207)	(0.104)	(0.209)	(0.114)	(0.216)	(0.118)
Ind. diversification	0.0004	0.007	-0.0004	0.006	0.0003	0.006
	(0.007)	(0.004)	(0.007)	(0.004)	(0.007)	(0.004)
Skilled/unskilled labour	2.426***	1.806***	2.271***	1.931***	2.825***	1.920***
	(0.702)	(0.249)	(0.668)	(0.269)	(0.711)	(0.271)
Log R&D	0.343***	0.198***	0.348***	0.207***	0.324***	0.204***
	(0.103)	(0.050)	(0.101)	(0.043)	(0.101)	(0.043)
Constant	-12.242***	-8.078***	-12.615***	-7.969***	-13.286***	-7.968***
	(2.275)	(1.154)	(2.329)	(1.248)	(2.484)	(1.299)
LM test spatial autocorr	3.862	0.020	5.854	0.919	4.948	0.932
	(0.049)	(0.889)	(0.016)	(0.338)	(0.026)	(0.334)
N	117	117	117	117	117	117

Note: * indicates significance at 10%; ** at 5%; *** at 1%.