Choose the Neighbor Before the House: Agglomeration Externalities in UK Science Parks*

Christian Helmers
Universidad Carlos III de Madrid

February 2011

ABSTRACT

This paper investigates the presence of agglomeration externalities and their effect on innovative activity of companies located in two geographically adjacent science parks in the UK. The analysis investigates whether technological proximity of firms located in science parks increases firm-level innovative activity measured as patenting. The effect of agglomeration is identified from heterogeneity in firms' relative positions in a network constructed by using firms' multiple industry classifications. The results provide evidence for the presence of positive intra-industry knowledge spillovers at the SIC 2-digit level among firms located in science parks and thus suggest that tenant firms benefit from positive agglomeration externalities in science parks that concentrate on specific, while still relatively broad, business areas.

KEYWORDS: Science park, agglomeration externality, patent, network centrality

JEL Classification: D21, D24, L25

*I thank Ian Cockburn, Markus Eberhardt, Marcel Fafchamps, Bronwyn Hall, Jacques Mairesse, Manasa Patnam, and Mark Rogers for helpful discussions and comments. The paper was presented at the Knowledge in Organizations Conference at Monte Verità, EPIP 2010 in Maastricht, and at a seminar at LSE. Useful comments by conference and seminar participants are gratefully acknowledged. Email: christian.helmers@uc3m.es

1 Introduction

Clusters of firms are an old phenomenon; during the first Industrial Revolution in the UK, for example, the cotton industry was heavily concentrated in Lancashire within the Oldham-Bolton-Manchester triangle, and Leeds was the center for the woollen industry. In more recent times, clusters of high-tech firms, such as Silicon Valley in California or Boston's Route 128 in Massachusetts have gained fame and are often referred to as role models for other locations to promote the emergence of a larger number of innovative companies.

The underlying intuition is that clusters of high-tech firms generate some form of agglomeration externalities, be it knowledge spillovers or thick labor markets, which make firms within the cluster more productive and innovative than they would have been had they located anywhere else outside of the cluster. Firms might even never have been founded outside of clusters. Policymakers in developed and developing countries alike have latched onto this simple concept and thus promote the clustering of high-tech firms. In the EU, the European Commission has even elevated the promotion of clusters to one of its nine strategic priorities for successfully promoting innovation within the Lisbon Agenda (EUCOM, 2008).

There is evidence that the agglomeration of high-tech firms occurred naturally in some places, such as Boston's Route 128, from a combination of the availability of physical resources, such as high-skilled labor and world-class research institutions, and agglomeration externalities (Dorfman, 1983). However, evidence for the UK (Devereux et al., 2004) suggests that it is low-tech manufacturing industries, such as textiles, ceramics, and cutlery, rather than high-tech industries that are geographically concentrated. Devereux et al. (2004) suggest that knowledge spillovers are not sufficient to drive high-tech industry agglomeration in the UK. Yet, using UK data at the NUTS3-level, Rice et al. (2006) find a positive correlation between productivity and agglomeration of economic activity. To promote the clustering of high-tech industries and achieve productivity gains resulting from agglomeration, policy makers in the UK have put in place various incentive schemes. One of the favored tools to induce the emergence of clusters of high-tech firms are science parks. Science parks are usually linked directly to universities or public research institutions, and are therefore seen as a channel for university-industry transfer further benefitting the emergence and success of highly innovative companies.

There has been a tremendous increase in the number of science parks in the UK and worldwide. In the UK, the first two science parks were opened in 1973. In 1983, there were seven parks in operation, by 1989 there were already 38 (Westhead and Batstone, 1998) and 46 by 1999 (Siegel et al., 2003b). Link and Scott report the existence of 81 university research parks in 2002, and I counted a total of 85 parks in operation in 2010 out of which 72 (85 percent) are directly linked to a university or public research institution. Westhead and Batstone (1998) report that a total of 3,800 people were employed in science parks in 1985, rising to 16,587 in 1992. According to the latest available data from the UK Science Park Association (UKSPA), a total of 76,603 individuals was employed in science parks in the UK in 2008.[1] In other countries, including developing countries, notably China (e.g. Tan, 2006) and Brazil (Cabral and Dahab, 1998), science parks have also been widely adopted as a means to promoting high-tech companies.

[1]Some science parks in the UK are not part of UKSPA and therefore not included in its annual statistics. Hence, total employment in all science parks in the UK is even higher.

While science parks are commonly believed to generate ample positive agglomeration externalities in the form of knowledge spillovers for their tenant firms, there is to date little quantitative evidence for this effect and even less is known with regard to the determinants of these externalities. In a recent UKSPA good practice note which serves member science parks as advice on how to manage their parks, the authors acknowledge that *it is essential that [the] network [of tenant firms] has the right balance to the ingredients, otherwise it will not enable maximum value to be generated through the networking interactions. For example, a network that comprises of too many SMEs will often lack the necessary expertise to protect and exploit IP, to leverage funding and investment opportunities and secure routes to market* (Leake and Treloar, 2010). Hence, the right composition of different types of tenant firms is a condition for maximizing potential gains for tenant firms from co-locating in a science park. This, however, begs the questions what the ideal composition of firms is in a park, i.e., does a set of homogeneous firms, for example bio-tech, generate more inter-firm spillovers than a set of heterogeneous firms, for example bio-tech and IT companies? This question is related to a broader empirical debate: do agglomeration externalities arise for firms within the same industry or rather across industries? Hence, finding an answer to this question for the particular environment of science parks also informs the broader debate on agglomeration externalities which has not yet settled on a commonly-agreed conclusion.[2]

The relevant debate in the economic geography literature revolves around the so-called Marshall (1920), *intra*-industry, and Jacobs (1969), *inter*-industry, externalities. Agglomeration externalities of the Marshal type have been labeled MAR externalities (Glaeser et al., 1992) after Marshall (1920), Arrow (1962), and Romer (1990). MAR externalities occur between firms within the same industry and are also referred to as localization externalities. Externalities as discussed by Jacobs (1969) occur across industries and are often referred to as urbanization externalities. According to her view, variety of local industries spurs growth rather than agglomeration of firms within the same industry. One may interpret the different implications of MAR and Jacobs externalities as the former leading to spatial economies of scale, while the latter leads to spatial economies of scope (Döring and Schnellenbach, 2005).

While broad in terms of the empirical approaches and data used to investigate inter- and intra-industry externalities,[3] the literature so far has struggled with a number of empirical challenges: first, the choice of the unit of analysis usually is somewhat arbitrary. From a theoretical as well as policy point of view, it is unclear whether for example postcode areas are more suitable for analysis than regional-level data — this is the well-known modifiable areal unit problem in spatial analysis (Openshaw, 1984). Second, the problems in identifying agglomeration effects are similar to the well-known issues regarding the identification of peer effects. Most notably, Manski's (1993) reflection problem posits that within a linear-in-means framework, it is impossible to infer from the observed mean distribution of an outcome variable whether average behavior within a group affects the individual outcome of members of that group because of the self-referential nature of the peer relationship. This implies that any attempt to capture the effect of agglomeration externalities on the outcome variable of interest by regressing an individual firm's outcome on the average outcome of all firms located in the chosen unit of analysis is not

[2]For a review of the literature on agglomeration externalities see Fujita and Thisse (1996) and more recently Puga (2010).

[3]See for example Ciccone and Hall (1996), Henderson (1986, 2003). This literature is discussed in more detail in Section 2.

identified. Third, firms do not choose their location randomly; location choice is the result of firms' optimally balancing a range of determining factors potentially including agglomeration externalities. If some of these factors are unobserved to the researcher and correlated with a firm's behavior of interest, any estimate associated with empirical measures of agglomeration externalities will be biased and inconsistent.

The present analysis focuses on firms located in two directly adjacent science parks in the UK: the Cambridge Science Park and St. John's Innovation Centre. I construct an eight-year panel containing *all* firms that located in these two, geographically adjacent science parks during the period 2000-2007 to analyze the question of whether similarity of firms within science parks affects firm-level innovative performance, measured as patenting. Analyzing firms within science parks provides a natural definition of the unit of analysis. The reflection problem is addressed by assuming a specific structure of peer interaction. Agglomeration effects are defined as effects other firms within the science park have on a firm's innovative activity. In equilibrium, the impact of these 'peer effects' on a firm's innovative activity varies as a function of a firm's relative technological distance from any other firm in the science park. Technological distances are defined by SIC codes reported by firms. Given that about a fifth of firms in the sample reports more than a single SIC code, individual firms in the science parks differ according to their relative location within the SIC-code network spanned by all firms in the science park. This network location-specific firm-level heterogeneity allows identification of endogenous peer effects. Moreover, the peer effects model allows to distinguish between direct and 'global' effects where direct effects result from firms within the same industry whereas 'global' externalities spread across firms in different industries throughout the park. Finally, concerns regarding endogenous sorting of firms with respect to expected benefits arising from agglomeration externalities are mitigated by characteristics specific to science parks. There exists survey-based evidence for science parks in the UK that indicates that firms locate in science parks mostly for other reasons than agglomeration externalities, most notably to benefit from the reputation of a science park and direct forms of support. Assuming that enough firms are willing to locate in a park, the selection of tenant firms is made by the park's management. This means that the pattern of firms located in a park is the outcome of the park's management decision rather than the firms' deliberate decision to co-locate with existing firms in a park in order to benefit from agglomeration externalities. The science parks used in this analysis are not specialized in any specific business area, which means that the parks' managers are unlikely to select firms based on their business activity and rather based on other observable characteristics, such as size, age, and business group affiliation. In addition, given that in my sample, science parks have been in place for a number of years, park managers can only adjust the composition of on-site tenant firms at the margin, i.e., they can only admit new firms whenever an existing tenant leaves the park or additional space becomes available. The fact that firms can only locate in a park if they are admitted by the park's managers, together with the assumption that there is a large enough set of firms willing to locate on-site due to the reputation of the park and available support measures, mitigates endogeneity concerns due to self-sorting of firms. Apart from relying on these science park specific features, I also employ fixed effects to account for correlated effects associated with a specific science park.

The analysis of the effect of the composition of firms in a science park on their innovative performance has practical implications for the successful management of science parks and informs the broader debate in the economic geography literature on intra- vs. inter-industry agglomeration externalities. The results suggest that firms within the same 2-digit SIC industry benefit from mutual knowledge spillovers. However, the evidence found for 'global', science park wide positive knowledge spillovers is at best weak. Taken at face-value, these findings suggest that science parks might be served best if they concentrated on certain industries by promoting a relatively homogenous group of firms to locate on-site.

The remainder of this paper is organized as follows: Section 2 reviews the existing literature on agglomeration externalities. Section 3 explains the concept of a science park and discusses the existing evidence. Section 4 discusses the data used for my analysis. Section 5 outlines the empirical approach used to identify the effect of agglomeration externalities on firm-level innovative activity. Section 6 discusses issues related to estimation. Section 7 reviews the corresponding results and Section 8 concludes.

2 Agglomeration Externalities

There are two different forms of externalities: pecuniary and non-pecuniary externalities. Non-pecuniary externalities are non-market externalities while pecuniary externalities are mediated through markets. Pecuniary externalities arise, for example, from thick labor markets for specialized labor, forward-backward/input-output linkages, and economies of scale in larger markets (Krugman, 1991; Fujita et al., 1999). Hence, pecuniary externalities affect a firm's profitability but do not affect directly output. Non-pecuniary externalities affect productivity directly, i.e., they shift a firm's production function. Independently of whether pecuniary or non-pecuniary externalities give rise to agglomeration, the resulting spatial equilibrium is self-reinforcing due to the presence of multiplier effects arising from the reciprocal nature of agglomeration externalities.

An important issue in the empirical investigation of agglomeration externalities relates to the definition of spatial proximity which can be defined in various ways: the product market, technological, geographical, and temporal (Rosenthal and Strange, 2004). Most of the debate in the literature has focused on the industrial scope of agglomeration externalities within somehow defined geographical areas, i.e., whether externalities arise within or across industries within a geographical area. This debate originates in Marshall (1920), who argued that externalities are confined within industries. The best known example of Marshallian *within* externalities is the clustering of companies of the same industry such as in Silicon Valley or Wall Street. Porter (1990) suggests that local within-industry competition increases productivity, i.e., firms within the same cluster and industry see their productivity increase due to local competition. Jacobs (1969), in contrast, argues that externalities arise from interaction of firms across industries. The original ideas by Marshall, Jacobs, and Porter have spurred an extensive body of empirical work (for a review see Rosenthal and Strange, 2004). For example, Glaeser et al. (1992) provide an empirical investigation of inter- versus intra-industry externalities for US cities at the industry-level. Glaeser et al. (1992) find for their cross-section of US city-industries that industries exhibit lower growth in cities in which they are over-represented. Similarly, firms grow faster in cities in which firms are smaller than the nationwide industry average. Moreover, they find city-industries to grow faster if the city is less specialized. Glaeser et al. interpret their findings as evidence in favor of Jacobs spillovers and Porter's local competition idea. The work by Glaeser et al. has spurred a large number of similar studies using different units of observation, industry-, and geographical area definitions. For example, van der Panne (2004) uses data on new products announced by firms in trade journals to assess whether inter- or intra-industry externalities affect firms' innovativeness. He finds evidence in favor of Marshal-type externalities, i.e., the number of new-product-announcing firms increases within regions and industry the more spatially agglomerated the industry is. Devereux et al. (2007) analyze the determinants of an entrant's location choice, specifically distinguishing the effect government grants and agglomeration externalities have in attracting new plants into counties and regions in Great Britain. They find that firms in relatively more agglomerated industries prefer to locate new plants close to existing plants within the same industry.

Nevertheless, the literature has not settled on a conclusion in favor of either intra- or inter-industry externalities. An important question in this debate is how to define proximity and the empirical measure of agglomeration. Most of the corresponding literature on agglomeration externalities uses data aggregated at the postcode-, county-, region-, or country-level and employs concentration indices to measure agglomeration. Working with aggregate data poses potentially serious difficulties in empirical work as the definition of an area might be endogenous to the economic activity within the area and somewhat arbitrary anyway. More recent papers address this problem by using a continuous measure of distance between individual firms (Duranton and Overman, 2008).

The debate on Marshal vs. Jacobs non-pecuniary externalities also suffers from an important empirical problem: while it may be possible to assert the presence or absence of externalities by verifying whether a firm's productivity is affected by agglomeration, it is impossible to infer from this observation on its own what exactly causes the increase in productivity. The only available conclusion is that an increase in the scale of an activity within a somehow defined spatial area leads to an increase in productivity.[4] A range of channels for these agglomeration externalities have been proposed including natural geographic advantages (Krugman, 1993; Ellison and Glaeser, 1999; Roos, 2005), the presence of local amenities and consumption opportunities (Glaeser et al., 2001), thick labor markets (Krugman, 1991, Overman and Puga, 2009), the sharing of inputs, forward and backward linkages (Abdel-Rahman and Fujita, 1990; Knarvik and Steen 1999), and knowledge spillovers (Jaffe et al., 1993). Within science parks, the most likely channel for agglomeration externalities are knowledge spillovers. The basic idea is that the (unintended) sharing of knowledge leads to increased productivity for individual firms. Transmitting information requires physical proximity of firms due to the tacit component inherent in knowledge. Given this characteristic of knowledge, firms cluster spatially in order to capture and exploit knowledge spillovers. Jaffe et al. (1993) show that patent citations are highly geographically localized, with a patent being more likely to be cited by inventors within the same metropolitan statistical area than inventors outside the metropolitan statistical area. Defining US states as their spatial units of analysis, Audretsch and Feldman (1996) find that after controlling for geographical clustering of production, propensities to cluster differ across sectors with knowledge intensive industries tending to cluster relatively more. They show that the larger share of the relatively higher propensity to cluster is attributable to the desire of

[4]See for example Ciccone and Hall (1996) who look at the density of economic activity measured as employment at the county level in the US to explain variation in productivity at the state-level without specifying a specific channel for these externalities to occur.

firms to capture knowledge spillovers. A similar, but much simpler, approach is taken by Acs et al. (2007). They assume that spillovers can be captured by a density measure by dividing the number of establishments of sector i by the population living within the defined regional area. In addition, the authors use a variable which represents the ratio of all private business establishments divided by the population in the geographical area. This is a very crude spillover measure and is based entirely on the assumption that spillovers emerge from geographical proximity within industries. Mariotti et al. (2010) show that MNEs in Italy tend to co-locate with other MNEs within the same industry instead of domestic companies due to knowledge spillovers.

Apart from the problem of observational equivalence of agglomeration externalities arising through different channels, another important problem relates to the endogenous sorting of firms. If firms are aware of the importance of co-location, firms will sort endogenously into locations where agglomeration externalities are present. Alcácer and Chung (2007), for example, provide evidence on new entrants in the US showing that firms choose their location to maximize inward knowledge spillovers and minimize outflows. Less technologically advanced firms prefer to locate in regions with generally relatively more advanced levels of technological capacity whereas technologically advanced firms favor specifically locations in proximity to universities. In empirical work, the problem of endogenous sorting implies that unless the determinants of a firm's location choice are observed, estimates associated with any variable representing agglomeration will be biased and inconsistent. Greenstone et al. (2010) provide a convincing identification approach to address the problem by exploiting information from location rankings used by 47 large manufacturing plants in choosing their location in the US. Comparing the chosen location with the runner-up location listed in the ranking provides a credible counterfactual and thus allows to circumvent the problem of location choice. They find that before the opening of a plant, the winning and runner-up locations have similar TFP trends, whereas five years after the establishment of the plant, firms in the winning location register a considerable increase in their TFP levels. The authors interpret this as strong evidence for the presence of positive agglomeration spillovers. Regarding the specific channels of these externalties, Greenstone et al. suggest that labor flows and technological links, measured by a firm's SIC 2-digit industry, patent citations and R&D expenditure, work as channels for agglomeration externalities.

3 Science Parks

3.1 Overview

The UK Science Park association (UKSPA) defines a science park as *[...] a cluster of knowledge-based businesses, where support and advice are supplied to assist in the growth of the companies. In most instances, Science Parks are associated with a centre of technology such as a university or research institute.*[5] In fact, most science parks are directly owned or linked to a university or public research institution. Link and Scott (2007) noted that in the UK, all science parks are located on or near to a university campus; however, this is no longer true today as I have found only 85% to be directly linked to a university.[6] Link and Scott (2006) define university science parks as *a cluster of technology-based organizations that locate on or near a university campus in*

[5]http://www.ukspa.org.uk/about/faq

[6]Examples for parks not directly linked to universities or research institutions are the 'Granta' science park near Cambridge or 'The Bridge' in Dartford.

order to benefit from the university's knowledge base and ongoing research. The university not only transfers knowledge but expects to develop knowledge more effectively given the association with the tenants in the research park. Both definitions stress the relevance of the university in assisting tenant firms in their research and business operations. Phan et al. (2003) provide a definition that focuses more on the knowledge exchange function of science parks defining them as *property-based organizations with identifiable administrative centers focused on the mission of business acceleration through knowledge agglomeration and resource sharing.* (Phan et al., 2003: 166). Note that science parks are different from incubators as science parks are not restricted to start-up companies and thus may host large and well-established companies or their subsidiaries. Funding for science parks is mostly public and comes from diverse sources: funding agencies such as Regional Development Agencies (RDAs), European funds such as the European Regional Development Fund (ERDF), and university/college funds such as in the case of Cambridge Science Park and St John's Innovation Centre. University science parks are commonly (partly) owned and managed by the university itself (or its technology transfer office). One of the important tasks in managing a science park is to define the selection criteria for new tenants and to conduct their selection.[7]

The description shows that there are at least three different stakeholders involved in science parks: tenant firms, universities or other public research institutions, and policy makers. Science parks are meant to fulfill a number of roles for these different stakeholders:

Policy makers see science parks as a means to promote the creation of new, innovative firms and to support their growth. The focus is on encouraging entry of firms involved in drastic, high-risk/high-return innovations that will grow fast into large companies generating high-skilled employment. Examples for such companies that originate from the Cambridge region include Autonomy Corporation plc, an enterprize software company founded in 1996 with revenues of US$ 740 million and 1,684 employees in 2009 (according to its Annual Report), or CSR plc, a developer of wireless technology founded in 2001 with slightly over 1000 employees in 2009. These goals are attained mostly by providing a special environment, in which firms can draw on a range of support measures that allow them to pursue business ideas that might otherwise not be viable and to perform better than they otherwise would. Another important element of science parks is the inter-firm links that are promoted 'by construction' due to the geographical proximity of firms on a park site and common facilities and events fostering formal as well as informal exchange between tenants. By promoting entry and performance of highly innovative companies, science parks are supposed to generate positive externalities beyond the park and improve the performance of the local and national economy. These externalities are particularly welcome in areas subject to structural change, e.g., areas that undergo a transformation from extraction of coal or textile manufacturing to building new industrial foundations with a view to job creation. This explains why a large number of science parks in the UK has emerged in northern regions coping with the consequences of a recent fundamental structural industrial change. Quintas et al. (1992) argue that already following the 1979-81 recession in the UK, local government embraced the science park concept as part of a local development agenda to revive the local economy and create jobs.

For universities, science parks play a notable role in university-industry technology transfer

[7]In the case of some (less prestigious) science parks, the park's management might have to actively search for new tenants.

(for evidence on US science parks, see Link and Scott, 2003). This occurs through various channels: academic spin-offs, hiring of graduate students by tenant firms, licensing of university-owned patents by tenant firms, consultancy services provided by academics for tenants, joint research projects by academics and tenants, as well as specific courses provided by the university for tenants and their employees. Science parks may also help universities build and improve their reputation. Knowledge transfer and reputational benefits may also result in obtaining public funding otherwise unavailable to universities (or avoiding cuts in funding). Science parks may also directly generate income for universities. This would explain why the number of science parks has increased substantially after a considerable drop in public funding for universities at the beginning of the 1980s.

Incentives for firms to locate in a science park are also manifold. The existing literature has found that firms value above all the 'brand' of science parks which helps them build reputation (Westhead and Storey, 1994; Salvador, 2010). The essential element here is the selection process that firms have to undergo. Another benefit for tenants is access to university facilities, involving both physical facilities, such as libraries and labs, and human resources (Monck et al., 1988). Moreover, parks usually provide physical facilities that can be used by tenants, such as meeting rooms, conference facilities including guest houses, a cafeteria, and even nurseries. Science parks also provide business advice and services and may assist in the management of companies. Another important element is the flexibility in rental agreements, which can involve flexibility with regard to rental payments and the possibility to move within the science park from smaller to larger facilities and vice versa depending on the development of the business. However, there are also costs involved in locating in a science park. Rents in science parks may significantly exceed rents paid for comparable premises in the surrounding area (see Westhead and Storey (1994) who provide some survey-based evidence for firms perceiving the science park rent to be higher than for a comparable off-park location in the same area). Moreover, clustering of a large number of highly innovative firms may also mean increased competition among firms for access to critical resources, most notably management assistance provided by the park, advice by university academics and hiring of university graduates.

Firms located in science parks have been found to be heterogeneous with the most notable distinction that some firms are stand-alone companies conducting all their business operations in the science park while others are part of large multinationals. Some firms are engaged in applied research, focusing entirely on R&D or combine R&D with light industrial production, while others might only provide simple end-user oriented services. Hence, firm heterogeneity is mostly observable by looking at firm's ownership structure and sector of activity. Yet, this existing empirical evidence is based on the population of tenant firms *across* science parks in the UK, rather than on the population of firms *within* a single science park. Another important characteristic of tenants that Monck et al. (1988) and Westhead and Storey (1994) found for their sample of UK firms surveyed in 1986, is that firms often re-located into science parks, rather than choosing the science park location at the moment of establishment which distinguishes them from incubators.

3.2 Existing research on tenants

The economics literature on science parks and firms located therein is embryonic. Most of the existing literature is case-study based.[8] The empirical literature on science parks in the UK is among the most active within the field and started with a survey by Monck et al. (1988). Using their survey data on 183 on- and 101 off-park firms, Monck et al. (1988) showed that tenant firms had lower employment levels than comparable off-park firms. They provided evidence suggesting that the average lower employment levels stemmed from underperforming firms founded and run by (former) academics. Quintas et al. (1992) report evidence from a similar survey showing a low level of formal links between tenants and universities; however, their data show a higher level of informal links than between off-park firms and local universities. Westhead and Storey (1994) updated the database constructed by Monck et al. (1988) and collected additional information on 71 firms which were randomly chosen from the population of firms located in UK science parks between 1986 and 1992. As done by Monck et al. (1988), based on industry, ownership type, age, and location, Westhead and Batstone (1998) matched a comparable sample of 71 off-park firms to the sample of on-park firms and conducted face-to-face interviews with both on- and off-park firms in 1992 and 1993. Westhead and Batstone (1998) use their survey data to shed some light on firms' motivation to locate in a science park. They find that the decisive factor to locate in a science park is the prestige of the park; although it remains unclear what the criteria are which determine prestige.[9] The survey also reveals another important reason for firms' choosing a specific location, namely the fact that 'the key founder lived locally' (Westhead and Batstone, 1998: 2208). Access to markets appears to play only a minor role in location decisions. More importantly for my analysis, while access to universities is highly ranked among the location determinants, only 15% of science park firms reported 'proximity to firms in similar industrial sectors/using same technology' (Westhead and Batstone, 1998: 2208) to be a matter of importance when locating in a science park. Yet, science-park firms value a 'friendly atmosphere among tenants on site' slightly more (22%). Among off-park firms, 6% and 8% consider proximity and a friendly atmosphere to be important respectively. When asking science park firms directly about the aspects of their on-park location that they valued most, access to communal space ranked highly, suggesting the importance of *in*formal links to other entrepreneurs and employees. From this survey evidence, it appears that science park firms seek closer links with universities by locating in a park than to exploit externalities from agglomerating with similarly innovative firms.

Westhead (1997) uses the same data set, i.e., the follow-on sample to the Monck et al. (1988) survey and the new data collected by Westhead and Storey (1994), to investigate differences in R&D inputs and outputs among on- and off-park firms. While tenants appear to employ slightly higher R&D input and output, Westhead (1997) does not find any statistically significant differences between on- and off-park firms in terms of R&D input and output. However, the lack of statistical significance is most likely due to the small samples (less than 50 on- and off- park firms). Westhead (1997) suggests the fact that *[t]o maintain rental income some Park managers have relaxed their selection criteria for tenants. As a result, some firms have moved to Science*

[8]See for example Park (2002) for a case study on the Ideon Science Park in Lund, Sweden, and Wai Yip So (2006) for a case study on Hsinchu Science Park in Taiwan.

[9]Ferguson (1994) finds for a sample of tenant firms in two Swedish science parks that prestige generates benefits of two types for tenants: first, outside firms, including customers, perceive firms located in science parks more positively; second, the professional environment of a science park can give entrepreneurs and academics incentives to take the entrepreneurial venture seriously.

Parks simply because of the image and overall prestige of the site [...] is responsible for tenants not performing better in terms of innovative activity than off-park firms. This view suggests that science parks can confer benefits to their tenant firms if properly managed, i.e., the 'right' firms are admitted to locate in the park.

Siegel et al. (2003a) have data on a random sample of stand-alone firms drawn from the population of firms located at all science parks in the UK in 1986 and 1992. The data set also contains answers to questionnaires and in-person interviews conducted with the sampled firms and information on comparable firms not located in science parks for 1986. The sample used in the analysis has 89 science park and 88 non-science park firms with data for 1992. Siegel et al. (2003a) estimate a Griliches knowledge production function using the number of new products/services, the number of patents and copyrights as measures for knowledge output and R&D expenditures and the number of scientists and engineers as measures for knowledge input. The main object of interest is a dummy variable which indicates whether a firm is located in a science park that is included in the knowledge production function. To address endogeneity of the location dummy, they use a measure of 'radicalness' of a firm's innovation as an instrument in a two-step procedure. The assumption that having a breakthrough innovation determines a firm's location in a science park but does not directly affect knowledge output is clearly open to debate. They find that firms located in university science parks produce more new products/services and patents for the same input as firms not located in university science parks, although the differences between science park and non-science park firms are small.

For the US, Leyden et al. (2008) find that publicly traded, R&D performing firms in the US are more likely to be located in a science park if they have larger R&D expenditures and are more diversified in the product market as measured by the number of 4-digit industry classifications reported by a firm.

A third set of studies from Sweden provides further insights. Ferguson and Olofsson (2004) use data from a small-scale survey of 30 on- and 36 off-park firms in 1995. Ferguson and Olofsson complement the qualitative survey data with quantitative performance data on sales and employment for 1991-2000 for their sample of 66 firms. They find higher survival rates among firms located in science parks than for off-park firms for the period 1995-2002. This finding is partly explained by firms exiting after having left the park. Firm growth measured in terms of sales and employment was found not to differ in a statistically significant way between on- and off-park firms. At the same time, Ferguson and Olofsson find larger variation in growth rates across science park firms than across off-park firms. Lindelöf and Löfsten (2003) follow a similar strategy, comparing samples of surveyed on- and off-park firms in 1999, albeit for a larger sample containing 134 on- and 139 off-park firms. Contrary to Ferguson and Olofsson (2004), they find tenants to grow faster both in terms of employment and sales than off-park firms. However, they do not find statistically significant differences with regard to patents, licenses, and new products launched within three years prior to the survey. Interestingly, on-park firms do not rate proximity to other firms higher than off-park firms and off-park firms even regard cooperation with other firms as more important than tenants. Hence, like in the UK data, the Swedish evidence suggests that tenant firms are not primarily interested in locating in a science park in order to benefit from agglomeration externalities.

In summary, the sparse existing empirical literature on firms located in science parks has focused on analyzing differences in performance, measured as survival and growth, between on-and off-park firms. There has been no quantitative study investigating the sources for such differences, notably agglomeration externalities between tenant companies. Existing qualitative work, reporting firms' self-perception of the benefits associated with location in a science park sheds some light on the channels through which locating in a science park may affect performance, but is ultimately limited in establishing quantifiable and reliable results as self-reported measures are prone to perception bias. Nevertheless, the survey-based evidence strongly suggests that agglomeration externalities are not an important determinant of a firm's decision to locate in a science park. The evidence also suggests that informal links between firms within a science park act as a potential channel for spillovers.

4 Data

The discussion of the literature in Section 3.2 shows that the existing empirical evidence has been obtained from relatively small samples of cross-sectional data. In this section, I describe in detail the construction of the data set used in my analysis, which is an eight-year panel (2000-2007) of a total of 412 tenant firms that represent the complete universe of firms located in two adjacent science parks: the Cambridge Science Park and St John's Innovation Centre.

4.1 Cambridge Science Park and St John's Innovation Centre

Along with the science park associated with Heriot-Watt University in Scotland, the Cambridge Science Park (CSP) was the first science park in the UK — its construction began in 1973. The main motivation for the establishment of CSP was the UK Government's desire to foster university-industry linkages in the late 1960s. The location was chosen mainly because the land owned by Trinity College (Cambridge) was lying idle since the end of World War II. Interestingly, the costs associated with the establishment of the park were born by Trinity College as no government grants were received. The park is still fully owned and managed by Trinity College. For a more detailed description of the history of CSP see Bradfield (1981). Currently, the park offers 145,540 sq meter in lettable space in form of a wide range of differently sized office space, allowing the accommodation of tenant firms at different stages of their development.

St. John's Innovation Centre (SJC) is also among UK's oldest science parks. It is owned by Cambridge's St. John's College and was opened in 1987. At the moment, SJC offers a total of 4,900 sq meter of lettable space with single units varying in size from 100 to 3,500 sq ft which allows the accommodation of various types of companies.

The two parks also offer a number of facilities, such as a cafe and restaurant, a fitness center, a nursery, and meeting facilities. Note that neither of the two parks is officially specialized in a certain area of business activity.

The decision to use these two science parks for my analysis was driven by the following considerations: (a) both parks are similar in terms of types of tenant firms and geographically adjacent, only separated by a road which makes it highly likely that firms from both parks interact; (b) a considerable number of firms can be observed to move back and forth between the two parks which suggests that firms are indifferent between the two parks and locate in the park in which appropriate office space is available; (c) while either park is large relative to the average park size in the UK, combining data from both parks guarantees a sample size that is sufficiently large to conduct my analysis; (d) both parks are among the best-known and most highly respected science parks in the UK, which lends credibility to the argument that firms are eager to locate there primarily to gain a positive reputation with their customers and financiers. Note that firms at the Bioscience Innovation Centre (BIC), which is located between the two science parks, are also included in the analysis as the BIC drew its tenants from the existing parks CSP and SJC. Hence, while the BIC is strictly speaking not a science park, I add the firms which located in the BIC to the sample (see Table 1). Also note that there are a number of other science parks in Cambridge which are all much smaller in size and are furthermore geographically distant from CSP and SJC.[10] For this reason, I will focus on CSP and SJC (and BIC) in this paper.

4.2 Firm-level data

4.3 Core data and financial information

The data used for this analysis comes mainly from two sources: Financial Analysis Made Easy (FAME) and the ICC British Company Directory. Both sources are complementary commercial databases that draw on the same original source and cover the entire population of registered UK firms.[11] However, these two data sets differ somewhat in terms of actual data coverage, in particular with respect to small and medium-sized companies and micro-firms.

The FAME database is a commercial database provided by Bureau van Dijk.[12] To construct the data set, two versions of the FAME database have been used: October 2005 and March 2009. These two editions of FAME give universal coverage of UK registered firms between 2000 and 2007. The main motivation for using two different versions of FAME is that FAME keeps details of 'inactive' firms (see below) for a period of four years. If I used only the 2009 version of FAME, I would be unable to identify any firm that has exited the market before 2005, which would cause me to miss a substantial number of tenant firms.

FAME contains basic information on all firms, such as name, registered address, firm type and industry code. Availability of financial information varies substantially across firms. In the UK, the smallest firms are legally required to report only basic balance sheet information (shareholders' funds and total assets). The largest firms provide a much broader range of profit and loss information, as well as detailed balance sheet data. Availability of certain variables, most notably employment and R&D expenditure data, is severely limited in FAME which restricts the set of covariates.[13] The most commonly reported variable is total assets. Turnover, intermediate inputs, fixed assets, and number of employees have less coverage since their reporting is not mandatory. For example only around 3% of firms in FAME report employment data. Companies can appeal to a small company exemption and not report these variables.

[10] The other science parks are: the Cambridge Biomedical Campus which is located in Cambridge and concentrates on biomedical research; the CardioThoracic Bioincubator (CTBI), which is located outside of Cambridge in Papworth and linked to the Papworth Hospital; Granta Park located outside of Cambridge in Great Abington; IQ Cambridge located outside of Cambridge near Waterbeach; Start IQ located in Cambridge; and the University of Cambridge West Cambridge Site in Cambridge.

[11] FAME and ICC download data from Companies House records. In the remainder of this work I 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. Potential issues resulting from this definition for my empirical analysis are discussed in Section 4.5.

[12] http://www.bvdep.com/en/FAME.html

[13] The large range of missing data as a function of firm size in FAME makes it also difficult to investigate standard firm performance measures, such as productivity. For further discussion see Eberhardt and Helmers (2010).

In terms of numbers of firms, FAME October 2005 contains information on around 3.1 million firms (of which 0.9 million are inactive) and FAME March 2009 on 3.8 million firms (of which 1 million are inactive). Inactive firms are those that have exited the market and belong to one of the following categories: dissolved, liquidated, entered receivership or declared non-trading. Also, FAME gives exact dates for market entry in the form of a firm's incorporation date.

ICC is a very similar database. However, the information available in FAME and ICC differs with ICC providing slightly more data for smaller firms. Most importantly, in contrast to FAME, ICC provides previous registered addresses and the date when firms changed addresses, which enables me to determine the period during which a firm was located at CSP or SJC.

In the empirical analysis, due to item-nonresponse for all other financial data, the only financial variable used are total assets which I deflated using a gross output price index (the base year is 2000) at the sector-level provided by EUKLEMS (November 2009 release).[14]

4.4 Patent and trade-mark data

The firm-level data from FAME and ICC are combined with patent and trade-mark information. The patent data come from the European Patent Office's (EPO) PATSTAT database. Data on patent publications at the EPO, World Intellectual Property Organization (WIPO) and UK IP Office was downloaded from PATSTAT version April 2010. The three different types of patent publication are used since UK firms have a choice of how to approach patent protection. One method is to file an application at the UK IP office, which is relatively cheap and would, if ultimately successful, provide protection in the UK. Another option is to apply for the patent at the EPO. This is more expensive, but the advantage is that it provides a clear route to seek subsequent protection in member countries of the European Patent Convention (EPC). Another option is to ask WIPO to provide an initial examination and then publish the patent, after which there is a procedure to ask for full examination in the (current) 139 countries that are members of the Patent Cooperation Treaty (PCT). Patents published through WIPO are referred to as PCTs.[15] The trade-mark data come from Marquesa ltd. and cover UK trade-mark publications and Community (OHIM) marks registrations. For a more detailed description of the match of firm-level and patent data and the resulting integrated data set see Helmers et al. (2010). In the analysis, we bundle the different types of patents and trade-marks into single counts of patents and a single dummy variable of whether a firm has registered a trade-mark or not. This aggregation is necessary due to the relatively small numbers of firm-year observations reporting non-zero patent and trade-mark counts as will be discussed further in Section 7.1.

4.5 Tenant firms

In order to obtain the names of firms located at CSP and SJC during the period analyzed, I contacted CSP, SJC, and UKSPA to ask for a list of tenant firms and the associated entry and exit dates in the parks. However, none of these three organizations provided me with any information.

Therefore, I identified tenant firms located at CSP and SJC in the following way: First, I used the lists of tenant firms's names available on the parks websites to identify current tenants.

[14] http://www.euklems.net/

[15] It is also possible that some UK firms may apply to the USPTO or other countries' IP offices. However, it is likely that the UK firm would also apply for an equivalent at the EPO or UKIPO, in which case the invention is covered by the data used in this analysis.

The names of tenant firms were used to find their registered numbers in FAME and ICC using both an automated matching algorithm and manual checking of all matching results. Out of 183 firms listed on the parks websites, I was able to find 163 in FAME and ICC (96 out of 104 for CSP and 67 out of 79 for SJC). I also checked all firms' websites to ensure that these firms are/were located at either of the parks and dropped those for which I could not unambiguously ascertain their physical location.

Second, since the lists available on the parks' websites contain mostly current tenants, i.e., miss firms that were located in the park between 2000-2007 but have left by 2010, I trawled the 2005 and 2009 versions FAME separately for firms located at the addresses of CSP and SJC. Since I am using two versions of FAME, this search method also retrieves firms that were located at CSP and SJC at an earlier date, but had left by 2010. As will be shown below, there is an enormous amount of entry into and exit from the parks every year, which means that this second step in the collection of tenant firms is crucial to obtain reliable data. In order to avoid false matches, I manually checked all matches and verified on the companies websites that these firms were indeed located at any of the two parks during the time period of interest. I performed a similar exercise using ICC. ICC keeps in its records firms that have exited the market and provides information on changes in firms' registered addresses. I used this information to search for tenant firms that report registered addresses at one of the science parks that I had missed in steps one and two.

Finally, I use the information on changes in firms' registered addresses provided by ICC for all firms retrieved in the different ways described above to determine when a firm moved onto a science park site and when it left. For single-location firms, this provides fairly accurate information on the time period a firm spends in a science park.[16] However, this search algorithm may still miss firms that are neither listed among tenant firms on the parks' websites nor found in FAME/ICC to have their registered addresses at one of the science parks. Given the information available, I am unable to account for such multilocation firms, although I do not expect there to be a large number of such cases nor do I expect these firms to be systematically different from the firms included in the sample.

4.6 Data cleaning

I dropped 16 firms that do not report SIC codes.[17] Since it is compulsory for firms to report SIC codes, these firms are companies that either exited even before they could report SIC codes or never actually entered the market. Hence, by dropping these firms, I do not expect to bias the sample but to exclude firms that never undertook any economic activity. Table 1 contains a breakdown of the tenants that I identified through the search described above and after dropping firms with missing SIC codes. Note that in the empirical analysis, I also drop firms that report only a single year of data — this eliminates another 91 firms from the sample. The remaining difference between the 412 firms reported in Table 1 and the 275 firms employed in the regression analysis is accounted for by item-non response by firms.[18]

[16] There may still be cases where firms have a registered address that does not correspond to it's actual physical address, although this is rare for single-location firms. I also have firms' trading addresses and am thus able to account for such cases.

[17] Note that I also dropped six entities located in the parks that are not trading businesses, such as the Royal Society of Chemistry.

[18] And by dropping 'island firms', i.e., firms that do not share a SIC 2-digit code with any other in the parks — for more discussion see Footnote 25.

5 Empirical Strategy

5.1 Identification

As mentioned in the introduction, there are three main empirical challenges in the identification of the impact of agglomeration externalities on innovative performance of firms located in CSP and SJC.

The first issue regards the definition of the object of analysis. The literature review in Section 2 shows that the existing empirical analysis has been carried out on substantially different data sets where one might suspect that the motivation to use more or less aggregate units of analysis is mainly driven by data availability. In contrast, using firms within science parks as the unit of analysis provides an intuitive solution to this problem. All firms located in the two science parks are included in the analysis as I am interested in the effect of agglomeration externalities only within the setting of science parks.[19]

Second, the problem of identification of agglomeration externalities is similar to the well-known difficulty of identification of peer effects as laid out in Manski (1993).[20] Manski shows that within a linear framework for a specific network structure consisting of mutually exclusive groups in which all individuals are interrelated, i.e., linked through a regular connected network,[21] it is impossible to infer from the observed mean distribution of a sample whether average behavior within a group affects the individual behavior of members of that group. In other words, the expected mean outcome of a peer group and its mean characteristics are perfectly collinear due to the simultaneity induced by social interaction. This fundamental identification problem was termed *reflection problem* by Manski. As will be shown below, there exists substantial heterogeneity among firms in terms of their connections to other firms in the network. In this case, endogenous peer effects are not simply equal to the group average of the dependent variable. Hence, as shown by Bramoullé et al. (2009) and Calvó-Armengol et al. (2009), peer effects are identified as they vary across firms as a function of their relative position in the network. Under the assumption that a firm's location within the network is exogenous with regard to peer effects, the reflection problem can thus be overcome. This implies that agglomeration effects are mediated through a firm's relative position in the network spanned by all firms within the science park.

In order to define a firm's location in the network, a definition of firms' similarity is needed. Since all firms within CSP and SJC are geographically located very close to each other and I am lacking information for a large number of tenants on their specific physical location within the science park across time, I focus on similarity measured in the product market, i.e., as defined by their SIC industry classifications. While SIC classes are a product-based classification, they also contain information on the technological position of firms. As will be shown below, this is reflected in considerable variation in patenting propensities across sectors. Importantly, slightly over 20% of all firms in the parks report more than a single SIC code (see Table 3), which enables me to create a network connecting all firms in the science parks. This measure

[19] While this seems to be intuitive given the question of interest, i.e., to identify a 'treatment effect on the treated', it nevertheless requires the assumption that firms located outside the park do not generate any agglomeration externalities affecting individual on-park firms differently. In future research, I will explore potential implications of this assumption.

[20] See also Moffitt (2001) as well as Blume and Durlauf (2006).

[21] A network is connected if every two nodes in the network are connected by some path in the network. That is network (N, g) is connected if for each $i \in N$ and $j \in N$ there exists a path in (N, g) between i and j.

of similarity allows me to directly explore the question of whether externalities occur within or across industries as discussed in Section 2. It is important to note that this involves an assumption: externalities do not affect all firms identically, but the effect differs according to firms' relative network location where the network is spanned by overlaps in firms' multiple industry classifications. The underlying intuition is that externalities occur through face-to-face contact of entrepreneurs and employees on the science park site. Since all firms are located in the park, they are all similarly close in geographical terms. However, entrepreneurs and employees differ considerably according to the activity carried out by their companies, i.e., a software engineer might find it easier and more insightful to exchange knowledge with another software engineer from a different company than to exchange ideas with a biochemist. While the discussion of the existing literature in Section 3.2 has shown that firms are not primarily attracted to science parks because of agglomeration externalities, the discussion pointed out that firms value informal links between firms which are made easier by the availability of shared facilities, such as a cafeteria or even a gym which are available on CSP and SJC. I therefore conjecture that (unintended) knowledge spillovers across firms occur through informal contacts between entrepreneurs and employees and that such spillovers are more likely to occur between people working in similar business and technology areas.

In the existing literature, starting with Jaffe (1986), technology spillovers have been captured by proximity measures of the patent portfolios held by firms (defined by their IPCs)[22] within and across markets. Bloom et al. (2007) combine information on firms' patents (using IPCs) and the industry in which they operate (using SICs). The interpretation is that IPCs measure technological proximity whereas SICs measure proximity in the product market, i.e., competition. The main difficulty of applying Jaffe's technological proximity measure in practice is the availability of data. In order to allow for sufficient similarity among firms' patent portfolios in his sample, Jaffe (1986), for example, aggregates 328 patent classes into 49 categories. As pointed out by Fox-Kean and Thompson (2005), in order to be meaningful, similarity should be measured at least at the technological subclass level.[23] This poses a enormous challenge to the available data. Bloom et al. (2007), therefore, restrict their sample to large patenting firms with sizeable patent portfolios and existing patent stocks. Moreover, within the context of a science park, it is unlikely that SIC industries measure actual competition between firms in the product market. Considering the unrealistic data requirements to apply an IPC-based proximity measure for the given sample of science park firms and the fact that product market competition is likely to take place outside of the park, using SICs to measure technological proximity between firms appears to be a defendable choice.

The third empirical challenge relates to the process that generates the observed equilibrium

[22] The International Patent Classification (IPC) is a hierarchical classification allocating patents into technology areas.

[23] The discussion by Fox-Kean and Thompson (2005) considers the classification at the USPTO. The main issue, however, carries over to the IPC classification. Consider for example, IPC $C12C11/00$ ('Fermentation processes for beer'). The first letter C is the 'section symbol': 'Chemistry and Metallurgy'. There are 20 classes within Section C. Classes are represented by a two-digit number following the section symbol, i.e., $C12$ ('Biochemistry [...]'). This is the level at which for example Jaffe et al. (1993) use patent classifications - Jaffe's (1986) original classification is even broader. Fox-Kean and Thompson (2005) note that there still exists enormous technological heterogeneity within classes. For example $C12$ contains 11 subclasses. Subclasses are represented by a letter following the two-digit class number, i.e., $C12C$ represents 'Brewing of Beer' whereas $C12N$ represents 'Micro-organisms or enzymes' which contains a number of diverse technologies such as antigens and antibodies. Hence, similarity measures based on IPCs depend crucially on the level of disaggregation of IPCs used. In order to obtain sufficient overlap of firms' patent portfolios at the subclass-level, firms need to have large patent portfolios, which would limit the analysis to a tiny subset of large companies.

allocation of firms in the science parks. As discussed above, if firms choose to locate in a science park because they anticipate agglomeration externalities, proximity between firms is endogenous and hence the corresponding coefficients will be biased and inconsistent. Leyden et al. (2008) discuss the decision process of admitting new firms to locate in an existing science park. In their model, firms in the park decide whether to admit new members based on the expected change in their profits following entry of the new firm into the park. While in the model by Leyden et al. (2008) existing tenants decide on admission of new tenants, in reality, the manager of a science park picks new tenants whenever space becomes available assuming that there is a sufficiently large pool of potential tenants willing to locate in the park. This implies that the composition of firms is mostly exogenous with regard to existing tenant firms' decisions. Indeed, the survey evidence discussed above shows that firms want to locate in science parks mostly because of reputation gains conferred by the park and the associated university. Location shows up only at the very bottom of firms' ranking of benefits from locating in a science park. Westhead (1998) for example reports from his survey of firms located in UK science parks that firms rank "proximity to firms in similar industries/sectors using same technology" only 17th out of 20 factors influencing location decisions (only 15% of firms stated this as an important factor). In contrast, "prestige and overall image of site" is ranked first with nearly 83% of firms regarding this as a decisive location determinant. This suggests that firms do not factor in externalities in their decision to apply for admission to the park. This is corroborated by a broader analysis asking any type of firm located in the Cambridge area for the principal reasons to locate in the area. These surveys have found that apart from the university's reputation acting as a major attraction, more than three-quarters of the interviewed start-ups located in Cambridge because the founder was already living there (Segal et al., 1986; CBR, 1996).

Moreover, the park's manager faces a constraint imposed by the availability of space on-site, which means that the maximum number of firms is capped by the size of the park (at least in the short- to medium-term). Since I analyze science parks a substantial number of years after their establishment (27 years in case of CSP and 14 years in case of SJC), it means that in practice, the manager can only adjust the composition of firms at the margin, i.e., only when existing tenants exit or new space becomes available, the manager is able to choose a new tenant and places the tenant into the space that has become available. In summary, the available evidence suggests that the specific characteristics of science parks attract firms for reasons other than expected gains from co-location with other firms. Such gains could be most likely achieved more easily in other environments without having to undergo a potentially onerous selection process. Moreover, neither CSP nor SJC have an official policy to select only certain types of firms active in specific business areas. This mitigates concerns that estimates of agglomeration effects might be confounded with contextual effects.[24] Apart from relying on these arguments, I also include a CSP dummy variable, which is equal to one if a firm is located at CSP in a specific year. The dummy variable, therefore, accounts for potential correlated effects specific to the science parks. Unobserved common shocks over time are captured through a time trend.

[24]In the dedicated peer effects literature, the problem of correlated effects is often addressed by exploiting random assignment of units into peer groups. However, to the best of my knowledge, this identification strategy has only been employed to identify peer effects among students (Sacerdote, 2001; Zimmerman, 2003; de Giorgi et al., 2010). Random assignment of firms into a location is unlikely to exist in practice. The argument, therefore, is that science parks come close to an environment in which firms co-locate 'randomly', that is they do not make their location decision based on peer characteristics, which mitigates potential endogeneity issues arising from selection into peer groups.

5.2 Empirical specification

Agglomeration externalities arise from the co-location of firms. I assume that externalities do not affect all firms in the parks equally but depend on a firm's relative position within the inter-firm network spanned by all firms in both science parks. To capture this position, I focus on summary statistics of the network topology mapped out by the structure of firms' overlapping SIC codes. Define the network (N, g) as a finite set of nodes $N = \{1, ..., n\}$ which are the firms located in the science parks and a real value $n \times n$ adjacency matrix g. The adjacency matrix describes the links between nodes which are denoted as $g_{..}$, where $g_{ij} = g_{ji} = 1$ indicates that node i is linked with node j and vice versa, i.e., it is an undirected and symmetric network. The entries of g are zero if firms are not direct neighbors or one if firms are direct neighbors. I set $g_{ii} = 0$ because self-loops have no economic meaning in this context. If firm i shares a SIC code at the 2-digit level with firm j, firms i and j are considered to be direct neighbors. Consequently, if firms i and j do not share a 2-digit SIC code, they are assumed to have no direct connection (but they can still be related through firm k if firm k is a neighbor to both). In the resulting network all tenant firms will be related to each other either directly or indirectly because about 20% of tenant firms report more than one SIC code (see Table 3).[25] Note also that the adjacency matrix can only assume values zero or one ($g_{ij} \in \{0, 1\}$).

Figure 1 shows a simple stylized example of how the network is constructed. The network is formed by six firms; one firm reports two SIC codes and thus shares an SIC code with the component consisting of the two black triangles as well as with the component consisting of the three yellow pentagons. The corresponding undirected symmetric adjacency matrix g looks as follows:

$$g = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \tag{1}$$

There are a number of network summary measures which can be used to summarize a firm's network location, i.e., its similarity vis-à-vis the entire network (Goyal, 2007; Jackson, 2008). The most simple measure is the degree of a node, which is the number of direct links that involve that node. Nodes that share a direct link are referred to as neighbors, i.e., $N_i(g) = \{j \in N | g_{ij} = 1\}$. Node i's degree in network g is defined as[26]

$$d_i(g) = \#\{j : g_{ji} = 1\} = \#N_i(g) \tag{2}$$

The density of a network, also referred to as *degree centrality*, is then defined as the average degree divided by the number of possible links: $density = d_i(g)/(n-1)$ where $density \in \{0, 1\}$. The degree distribution of a network is a way of measuring heterogeneity of nodes within a given network as it measures the relative frequencies of nodes that have different degrees. If there is no heterogeneity, all nodes have the same degree, which corresponds to a connected network. For

[25] Note that I keep only nodes that have at least one link in the network, i.e., define $N(g) = \{i | \exists j \text{ s.t. } ij \in g$, or $ji \in g\}$. This leads me to drop eight such 'island' companies from the sample.

[26] Since the network is undirected, in- and out-degree coincide.

example, the nodes in the network depicted in Figure 1 and the corresponding adjacency matrix (1), differ in terms of their degree. The black triangles have $d_i(g) = 2$, the yellow pentagons have $d_i(g) = 3$ and the grey square has $d_i(g) = 5$. The degree measures how well connected a node is in terms of its direct connections, but does not provide any information on a node's indirect links and the relative importance of a node within the network.

When considering the presence and effect of peer effects, a firm's network location should depend on its neighboring firms' locations. There is a range of centrality measures that define a node's centrality based on how central a node's neighbors are which captures the basic self-referential nature of peer effects because node i's own centrality depends on its neighbors' centrality, which in turn is a function node i's centrality.

Katz (1953) proposed a centrality measure that accounts for the possibility that the influence exerted by nodes further away declines with distance. The Katz centrality measure gives lower weight to longer walks that emanate from an individual node.[27] A walk of length 1 is worth a, a walk of length 2 is worth a^2 etc. for some discount factor $0 < a < 1$. In vector notation, $g\mathbf{1}$ denotes the vector of degrees of nodes, which shows how many walks of length 1 emanate from each node (where $\mathbf{1}$ is the $n \times 1$ vector of 1s). Then, $g^k\mathbf{1}$ denotes the vector whose ith entry is the total number of walks of length k that emanate from each node (g^k denotes the kth power of adjacency matrix g). Hence, the vector of Katz centrality is given by $P^{K2}(g,a) = ag\mathbf{1} + a^2g^2\mathbf{1} + a^3g^3\mathbf{1} + ... = (1 + ag + a^2g^2...)ag\mathbf{1}$. Under certain conditions,[28] this sum is finite and can thus be written as $P^{K2}(g,a) = (I - ag)^{-1}ag\mathbf{1}$. Bonacich (1987) extended the Katz centrality by adding a weight parameter b such that

$$Katz - Bonacich = \phi(g,a,b) = (\mathbf{I} - bg)^{-1}ag\mathbf{1} \tag{3}$$

where $\mathbf{I}$ is the identity matrix and $a > 0$, $b > 0$. b is a factor that captures how the value of being connected to another node decays with distance while a captures the base value on each node. Thus, the Katz-Bonacich measure simply counts for each node the total number of direct and indirect paths of any length in a given network originating from that node where paths are discounted using a geometrically decaying factor b. However, the values of a and b are generally unknown and in practice chosen somewhat arbitrarily.

Ballester et al. (2006) provide a useful result showing that the Katz-Bonacich network centrality measure described above is proportional for each node in a network to its equilibrium strategy in a peer effects game where players choose simultaneously their actions. This means that the collection of Katz-Bonacich measures within a network describes a unique Nash equilibrium of firms' strategies within the network that contains the structure of interaction between players. This result provides an economic justification to use the Katz-Bonacich centrality to capture a node's network location.

Calvó-Armengol et al. (2009) extend the peer effects game proposed by Ballester et al. (2006) to allow payoff functions to shift with individual characteristics. Individual payoffs are interdependent in this game as defined by the adjacency matrix. Importantly, Calvó-Armengol et al. assume that payoffs can be separated into an idiosyncratic and a peer component. Denote each firm's own idiosyncratic innovative activity by y_i with $y_i \in \mathbb{R}_+$ which is additively separable

[27] A walk is a sequence of nodes in which two nodes have a link between them (Goyal, 2007).

[28] A sufficient condition is that a be smaller than 1 divided by the norm of the largest eigenvalue of g. This condition holds whenever a is smaller than 1 divided by the maximum degree of any node (Jackson, 2008).

from innovative effort induced by the presence of peers z_i. It is important to stress that in the context of firms in a science park, I regard firms' patenting activity as an indicator of innovative activity omitting issues related to strategic interaction among tenant firms in terms of patenting decisions. Anecdotal evidence suggests that they are of little relevance within this specific setting in practice. Each firm i chooses its own and peer-induced innovative effort levels $y_i \geq 0$ and $z_i \geq 0$ to maximize its linear-quadratic payoff function (omitting time subscripts):

$$\pi_i(y_i, z_i, g) = \delta_i y_i - \frac{1}{2}y_i^2 + \mu \sum_{j=1}^{n} g_{ij} z_i - \frac{1}{2}z_i^2 + b \sum_{j=1}^{n} g_{ij} z_i z_j \tag{4}$$

where $b > 0$, $\mu > 0$. $\delta_i = \sum_{m=1}^{M} \beta^m x_i^m$ captures the effect of M exogenous observable firm and sector characteristics. δ_i can also include contextual effects, i.e., neighbors' exogenous characteristics. Equation (4) implies that a firm's patenting activity is separably additive in its own innovative efforts y_i and the effect other firms have on firm i, i.e., z_i. The effect of other firms is captured through two terms. The term $\sum_{j=1}^{n} g_{ij} z_i$ captures the effect of the number of direct links i has within the network, i.e., how well it is directly connected. The second component, captures also the impact of weighted indirect links, i.e., the effect of a neighbor's neighbors and so on. This is necessary because in equilibrium, not only the effect of direct neighbors, but also that of indirect neighbors determines strategies of node i, but as proposed by Bonacich (1987), the effect of neighbors is weighted with geometrically increasing weights with distance. The cross-derivative with respect to z_i and z_j where i and j are directly connected, therefore, captures such peer effects that diffuse through the entire network:

$$\frac{\partial^2 \pi_i(y_i, z_i, g)}{\partial z_i \partial z_j} = b g_{ij} \geq 0 \tag{5}$$

where the cross-derivative is zero if $g_{ij} = 0$. The cross-derivative assumes that peer effects are complementary, i.e., if i increases its innovative efforts, j benefits if j also increases its efforts. To maximize expected payoffs, firms choose their optimal innovative activity $y_i \geq 0$ and $z_i \geq 0$ simultaneously, which yields unique best response functions in pure strategies for each $i = 1, ..., n$. The first-order conditions corresponding to Equation (8) are:

$$y_i^* = \sum_{m=1}^{M} \beta^m x_i^m \tag{6}$$

and

$$z_i^* = \mu \sum_{j=1}^{n} g_{ij} + b \sum_{j=1}^{n} g_{ij} z_j \tag{7}$$

Using these first-order conditions (6) and (7), the optimal level of innovative activity is composed of a firm's own and peer efforts:

$$y_i^* = \sum_{m=1}^{M} \beta^m x_i^m + \mu \sum_{j=1}^{n} g_{ij} + b \sum_{j=1}^{n} g_{ij} z_j \tag{8}$$

Calvó-Armengol et al. (2009) show that the individual equilibrium outcome described in Equation (8) is uniquely defined by

$$y_i^* = \sum_{m=1}^{M} \beta^m x_i^m + \frac{\mu}{b}\phi_i(g,b) \tag{9}$$

if the largest eigenvalue of the adjacency matrix $\omega(g)$ satisfies $b\omega(g) < 1$.[29] Calvó-Armengol et al. (2009) demonstrate that expression (9) is directly related to the Katz-Bonacich measure in Equation (3) under the assumption that $a = b$. The effect of a node's centrality as captured by the Katz-Bonacich centrality is therefore in equilibrium equal to the ratio $\frac{\mu}{b}$.

6 Identification and Estimation

The model in Equation (8) corresponds to a spatial autoregressive error model (Anselin, 1988). Gibbons and Overman (2010) note that the different Cliff and Ord type spatial model specifications, notably the spatial autoregressive autoregressive model (SARAR) and the spatial autoregressive error model (SEM) are nested. This implies that the identification problem in a SARAR model that contains both endogenous (spatial lag in the dependent variable) and contextual (spatial lag in the independent variables) effects apply also to the SEM specification. To see this more clearly, rewrite Equation (8) in matrix notation (omitting time and firm subscripts) and add a stochastic error term u:

$$Y = X\beta + e \tag{10}$$

where e captures linear and additively separable peer effects:

$$e = \mu g\mathbf{1} + bge + u \tag{11}$$

where $u \sim N(0, \sigma^2 I)$. Now apply the spatial counterpart to the Cochrane-Orcutt transformation to the model in Equation (10) using expression (11):

$$\begin{aligned} Y - X\beta &= \mu g\mathbf{1} + bge + u \\ \iff Y - X\beta &= \mu g\mathbf{1} + bg(Y - X\beta) + u \\ \iff Y &= bgY + X\beta - bgX\beta + \mu g\mathbf{1} + u \end{aligned} \tag{12}$$

which is equivalent to the SARAR model, i.e., a model specification that contains a spatial lag in the dependent and independent variables. Therefore, without theoretical foundations motivating the spatial error model specification, it would be difficult to attribute an economic interpretation to the estimates for b. Moreover, the fact that the different spatial autoregressive models are nested, poses a challenge to the identification akin to the Manski (1993) reflection problem. To achieve identification, parameters β, μ and b must uniquely define the reduced form of Equation (12), which is

$$Y = (I - bg)^{-1}X\beta + (I - bg)^{-1}b\beta gX + (I - bg)^{-1}\mu g\mathbf{1} + (I - bg)^{-1}u \tag{13}$$

Bramoullé et al. (2009) show that identification of (13) is possible under fairly mild condi-

[29] This condition rules out explosive feedback processes between indirect neighbors within the network. Note that this condition only restricts the relative magnitude of peer effects and not their absolute value.

tions: for identification, a small degree of intransitivity of the spatial weights matrix in the sense that node i is a neighbor to nodes j and k while nodes j and k are not directly connected with each other suffices.[30] This means that the model in Equation (13) is identified if matrices I, g and g^2 are linearly independent. As pointed out above, the adjacency matrix produced by using firms' overlapping SIC codes is undirected and thus symmetric. Yet, due to heterogeneity across firms in terms of their degree, the matrix is irregular and therefore satisfies the condition for identification given by Bramoullé et al. (2009). Calvó-Armengol et al. (2009) provide a variant of this condition, namely $\frac{g_i^2}{g_i} \neq \frac{g_j^2}{g_j}$ (assuming $\mu \neq 0$), that is straightforward to test by checking whether the $2 \times n$ matrix formed by the two column vectors $g\mathbf{1}$ (own degree vector) and $g^2\mathbf{1}$ (vector containing total of neighbors' degrees) is of rank two. For example, for the network depicted in Figure 1 and the corresponding adjacency matrix (1), the rank of the matrix composed of $g^2\mathbf{1}$ and $g\mathbf{1}$ is indeed equal to two. This condition requires to have at least two nodes in the network that differ in terms of the average connectivity of their direct neighbors. Therefore, identification comes from the heterogenous structure of interaction between firms contained in the network g.

The main threat to identification arises from the possible misspecification of the adjacency matrix g. Given the lack of actual data on interaction between firms, I assume that firms' interact based on the overlap in SIC codes between firms. This is, if firms operate in the same business area, I assume interaction between employees occurs whereas such interaction is absent between firms in unrelated business areas. I discuss the issue of misspecification and potential implications in more detail in Section 7.3. The section also reports results from a robustness exercise that addresses the issue of misspecification.

Under the assumption of spatially correlated errors, the error variance-covariance (VC) matrix with non-zero off-diagonal elements is given by (Anselin, 1988):

$$E[\omega\omega'] = E\{[(I - bg)^{-1}u][(I - bg)^{-1}u]'\} = \sigma_u^2\Omega_\omega \tag{14}$$

The SEM model of Equation (10) is estimated using a ML approach where the log likelihood is given by[31]

$$\ln L = \frac{1}{2\sigma_u^2}(Y^* - X^*\beta - \mu g\mathbf{1})'(Y^* - X^*\beta - \mu g\mathbf{1}) - \frac{N}{2}\ln(\sigma_u^2) + \ln(\mid I - bg \mid_+) - N\ln(\sqrt{2\pi}) \tag{15}$$

where β and μ are estimated parameters and Y^* and X^* are the spatial Cochrane-Orcutt transformed dependent and independent variables, i.e., $Y^* = Y - bgY$ and $X^* = X - bgX$. In practice, the term $\ln(\mid I - bg \mid_+)$ is estimated as $\ln(\mid I - bg \mid_+) = \sum_{i=1}^N \ln(1 - b\lambda_i)$ where λ are the roots of g (Ord, 1975).[32] The coefficient b associated with spatial autocorrelation of the error term is the decay parameter in the Katz-Bonacich measure and interpreted as the 'global' peer effect on a firm's own innovative effort measured by its patenting activity.

[30]This produces exclusion restrictions, i.e., gx is a valid instrument for gy because x_j affects y_i (since they are direct neighbors) but x_j affects y_k only through its effect on y_i (since j and k are not direct neighbors). However, depending on the degree of intransitivity prevalent in the network, identification may be weak as g and g^2 will be collinear.

[31]ML is preferred over a spatial 2SLS approach as Lee (2009) shows that in the case of misspecification of the adjacency matrix g, the resulting bias is lower for the ML estimator relative to the 2SLS approach.

[32]For estimation, the adjacency matrix is row-standardized, i.e., each element of row i is divided by the degree of i, which implies that $(I - bg)^{-1}$ exists for all $|b| < 1$.

7 Results

7.1 Descriptives

Table 1 shows the total number of tenants identified and displays the number of firms entering and exiting the park as indicated by changes in firms' registered addresses provided by the ICC database. The table shows a substantial increase in tenant firms within the eight-year period analyzed. The total of tenant firms more than doubled from 127 in 2000 to 266 by 2007. Moreover, there appears to be a lot of churning, i.e., entry and exit of firms. Entry and exit in most cases do not coincide with market entry and exit. Out of the 169 firms exiting, 84 (50%) left the park as a consequence of market exit. This is a remarkably high number of firm failures as this implies that approximately 20% of firms that were located at some point in time at CSP or SJC during the period 2000-2007 went out of business. This is a relatively high number because in contrast to incubators, science parks are not a start-up firm environment. The mean and median age of firms in CSP and SJC that went out of business is 8.7 and 5 years respectively whereas the available empirical evidence for the UK suggests that most firms go out of business within the first four to five years of their existence (Disney et al., 2003; Helmers and Rogers, 2010). Table 2 complements the information provided in Table 1 by showing separately the number of firms located in CSP and SJC (and the Bioscience Innovation Centre - BIC). There are less firms located in CSP than SJC. However, firms located in CSP are considerably larger than firms in SJC. The median size measured by a firm's total assets (deflated to 2000) in CSP is around £2 million whereas it is only £27,000 for firms at SJC, which explains the presence of a larger number of firms despite the smaller available space in SJC. Table 2 also shows that the increase in the absolute number of firms observed in the aggregate data shown in Table 1 is due to an increase in tenant firms located at SJC. While CSP expanded only moderately by approximately 20% growing from 73 to 91 firms, SJC grew from 48 to 175 firms.

Table 3 shows SIC codes by year and indicates the share of firms reporting more than a single SIC code. The table shows that the set of firms located in CSP and SJC covers a broad range of activities. Nevertheless, nearly 70% of firms are concentrated in sectors 72 (computer), 73 (R&D), and 74 (other business activities), which are generally considered to be more knowledge-intensive sectors. Firms in high- and medium-to-high-tech manufacturing industries (SIC 24, 29-35) account for 11% of tenants. This table shows how inter-firm networks are constructed as around 20% of firms report more than a single SIC code, which creates firm-level heterogeneity in terms of the number of direct links of firm i with firm j (where $j \neq i$).

Figure 2 visualizes the network structure of tenant firms contained in the sample used in the regression analysis for 2000 and 2007. The circles are the nodes of the network which represent tenant firms. The edges, i.e., the lines drawn between nodes, show whether firms are close to each other as defined by the overlap in SIC codes. Figure 2 displays a core-periphery structure with vertices in the cores being connected. There are also a number of nodes that are linked only through a single node with the other components of the network. While there appears to be little change in this network structure over time, the number of nodes in the periphery increases slightly between 2000 and 2007. The yellow squares in the network represent firms that apply for a patent in a given year. Most yellow squares are hardly visible because they are buried in the clouds of red circles, which implies that most patentees are part of one of the big components composed of firms in the same SIC 2-digit sector.

Figure 3 plots the degree distributions for 2000, 2004, and 2007. A degree is defined as the number of edges per node and the density plot shows the distribution of relative frequencies across tenant firms. The plot shows that the networks for the different years are irregular and points to the presence of substantial heterogeneity in firms' degrees. The distribution for 2000 is bimodal, whereas this bimodal distribution appears to smooth out over time. Goyal (2007) lists two characteristics of the degree distribution of networks that have been observed in empirical work: first, there is substantial variation in the number of degrees across firms; second, the average number of degrees is considerably lower than the number of nodes of the network. Figure 3 suggests that the data on tenant firms also bears out these features. Table 4 complements the figure and provides summary statistics for firms' degrees over the entire period 2000-2007. The table shows that the average degree varies between 33 (in 2000) and 63 (in 2006) which is considerably lower than the number of nodes (101 in 2000 and 198 in 2006). The network density, which is equivalent to the average degree centrality remains relatively constant throughout the period, suggesting that the increase in the average degree over time is due to the increase in the number of nodes.

Finally, Table 5 shows some descriptive statistics for the sample of firms used in the regression analysis. It shows that firms are on average 9.5 years old, although there is substantial variation as firms are aged between one and 85 years. This suggests that conditioning for age in the regression analysis is important to account for heterogeneity across firms. The table also shows that the firms included in the sample are highly innovative with patent application counts of up to 64 in a single year. Overall, the mean is relatively high with .7 patent applications per firm-year observation. The trade-mark dummy indicates that slightly more than 8% of firm-year observations report the registration of a trade-mark. In the sample used in the regression analysis, nearly 19% of firms are found to have registered a trade-mark between 2000 and 2007. Another control variable is the patenting intensity at the 3-digit SIC-level, which is computed as the ratio between the total number of patent applications and turnover for *all* firms contained in FAME. The ratio is normalized to lie between zero and one. The variation of the patenting intensity variable across firms suggested by the descriptive statistics confirms the picture drawn by Table 3, i.e., considerable diversity of industries in which science park firms are active. However, the mean of .22 is relatively high pointing to the fact that most firms in CSP and SJC are in patenting-intensive sectors, such as SIC 73.[33] The subsidiary dummy shows that a large fraction of tenant firms (42.5%) belongs to a holding company. This characteristic distinguishes science parks from incubators as science parks host a large share of companies that are part of well-established large business groups.

7.2 Agglomeration externalities

The results of estimating the model in Equation (10) are easily summarized. Table 6 shows the results when estimating the model using the pooled sample for the period 2000-2007 assuming that externalities occur between firms *within* the same 2-digit SIC sector and diffuse 'globally' throughout the entire park.

Table 6 reports results in Columns (1)-(3) when ignoring network-wide inter-firm externalities, i.e., the effect that node i's neighbors' neighbors have on node i and contextual effects. The specifications in Columns (1)-(3) only allow for spillovers from direct neighbors captured by the

[33]For more evidence on the use of intellectual property across sectors in the UK see Helmers et al. (2010).

parameter μ. Columns (4)-(6) in addition account for science park wide peer effects by allowing for spatial autocorrelation in the error term. The specification in Columns (4)-(6) also includes a firm's degree as a measure for direct effects arising from technological proximity to other companies. The coefficient associated with the effect of direct neighbors reported in Column (1) is .045, when adding sector-level fixed effects, the point estimate increases to .105. Since I employ the natural logarithm of degree and of patent count as conditioning and dependent variables respectively, the coefficient in Column (2) can be interpreted as a 10% increase in a firm's degree leading to an increase in its patenting activity of slightly over 10%, which is a sizeable economic effect. In Column (3), I add the CSP dummy variable to the specification in order to account for potential correlated effects. The dummy variable turns out to be statistically highly significant, although the estimate of μ remains nearly unchanged. This provides some reassurance that the estimate of the effect of direct links is not sensitive to potentially omitted common unobservables.

Columns (4)-(6) report the results when allowing for 'global' externalities in the form of spatial autocorrelation. Column (3) reports a statistically significant (at 1% level) b of .172 pointing to the presence of substantial positive 'global' inter-firm spillovers. However, once also sector-level fixed effects are employed in addition to the time dummies in the specifications reported in Columns (5) and (6), the peer effect estimate falls considerably in magnitude and is no longer statistically significant although it remains positive. The estimate associated with a firm's degree is positive and statistically significant for all model specifications. The coefficients reported in Columns (4)-(6) are remarkably similar to those reported in Columns (1)-(3). Hence, the results in Columns (1)-(6) provide evidence for the presence of knowledge spillovers between firms that are directly linked within the same 2-digit industry, i.e., MAR externalities, and the results shown in Column (4) provide some, albeit much less robust, evidence for network-wide externalities propagating across nodes linked through indirect paths within the network.

Among the firm-level covariates included in the specification, only a firm's trade-mark dummy and size measured as total assets are statistically significant. Trade-marking firms are substantially more likely to patent, which confirms the hypothesis that this variable may capture a firm's IP management and potentially also its more general greater scope for innovation. The estimates associated with a firm's total assets suggest that larger companies are more likely to patent. The sector-level patenting intensity variable is only statistically different from zero when no sector-level fixed effects are included in the specification. This implies that there exists substantial heterogeneity of firms across sectors, which is either captured by the sector-level patenting intensity variable or sector-level fixed effects. Somewhat surprisingly, a firm's age and business group affiliation do not appear to have any effect on a firm's patenting behavior.

Table 7 reports results when allowing also for contextual effects, i.e., firm i's neighbor's exogenous characteristics ($g \times X_j$ where $i \neq j$), to affect the outcome of firm i. Qualitatively the results remain largely unchanged, although the peer effects coefficient b is not statistically significantly different from zero for any of the three specifications reported in Table 7 Columns (4)-(6). The coefficients of the spatial lag of the independent variables suggest that neighbors' trade-marks, business group affiliation, age and the size (measured as total assets) are statistically significantly associated with a firm's patent count. Trade-marking and older neighbors are positively associated with a firm's own number of patent applications. The negative effect of being surrounded by larger companies is economically negligible with a 10% increase in neighbors' size translating only into a .1% drop in the number of patent applications holding everything else constant.

The estimate of $\hat{b}$ reported in Table 6 Column (4) is used to compute the Katz-Bonacich centrality measure defined in Equation (3). Figure 4 shows the empirical cumulative distribution functions of the Katz-Bonacich centrality measures for 2000, 2004, and 2007. The figure shows the absence of first-order stochastic dominance of any distribution over the other which implies that mean centrality does not differ across the distributions over time. Figure 5 shows the correlation between the ln number of patent applications and the Katz-Bonacich centrality measure. The lines fitted through the point cloud show the positive relationship between the number of patents and the centrality measure. Interestingly, when excluding firm-year observations reporting a zero patent count, the slope of the fitted line remains unchanged and it simply shifts upwards. This results from the wide distribution of Katz-Bonacich centrality measures for firm-year observations with zero patents, which mitigates concerns about the effect of the large number of zeros in the sample.

To sum up, the results provide some evidence that firms within the same SIC 2-digit industries affect each other's innovative activity measured as patent application counts positively. However, most of the spillovers appear to occur between firms that are directly linked, rather than through 'global' spillovers that propagate throughout the entire network.

7.3 Robustness

.

The definition of interactions between firms is based on firms' overlapping 2-digit SIC codes. This is an assumption required due to the absence of actual information on interaction patterns between firms in the science parks. While such assumptions are common in the literature on spillovers (Audretsch and Feldman, 1996), there is ample scope for misspecification. As is well known, if a model is misspecified, the ML estimator yields biased and inconsistent estimates. Despite the importance of this issue for identification, the spatial econometrics literature provides very little guidance on the magnitude and direction of the bias and the circumstances under which it is most likely to affect the estimates. Páez et al. (2008) use Monte Carlos exercises to analyze bias in a SARAR model from under-specification of the adjacency matrix, i.e., assuming that a given node's degree is smaller than in the true model, and over-specification, i.e., assuming that a given node's degree is larger than in the true model. For the SARAR specification, Páez et al. find that bias from under-specification is particularly severe when average degree and/or clustering in a network is low, i.e., the adjacency matrix is sparse, and true underlying spatial autocorrelation autocorrelation high. Over-specification results in pronounced bias when average degree is low, but clustering high, which means that in networks where there are connected components but with little connections between components, adding false links results in a particularly severe bias. Lee (2009) derives theoretically the bias arising from misspecification of the adjacency matrix in a model with spatial lags in the independent variables and provides Monte Carlos results for a SARAR model suggesting that the bias from under-specification points downward whereas in the case of over-specification estimates are upward biased. Generally, Lee finds bias from over-specification to be lower than bias from under-specification.

I conduct two robustness checks. I specify the adjacency matrix g based on the overlap in SIC 3-digit codes as well as 1-digit codes. This allows me to under- and over-specify the adjacency matrix relative to the matrix defined at the SIC 2-digit level. When defining interactions at the SIC 3-digit level, the number of each firm's direct neighbors drops drastically (there are also slightly less firms in the sample as the number of 'island' firms increases, which are dropped from the sample). While the average degree varied between 32.85 (in 2000) and 63.34 (in 2000) for the definition at the 2-digit level, employing the 3-digit level yields an average degree between 15.17 (in 2000) and 31.52 (in 2006), i.e., less than half the number of direct neighbors. This means that the social interaction matrix is under-specified relative to the matrix at the 2-digit level. The corresponding results are reported in Table 8. When I use a definition at the SIC 1-digit level, the number of observations increases as there are no 'island' firms which drop from the sample (the sample size increases from 1,299 to 1,352 observations). Also the average degree increases to a range between 65.29 (in 2000) and 140.38 (in 2006). The corresponding results are reported in Table 9 and demonstrate the robustness of the main findings to changes in the definition of the adjacency matrix.

8 Conclusion

This paper analyzes the effect of the composition of tenant firms in two adjacent science parks in the UK, CSP and SJC, on inter-firm knowledge spillovers using an eight-year panel of firms. The objective is to evaluate whether agglomeration externalities among firms in science parks arise from concentrating innovative firms pursuing similar or diverse activities. Identification of agglomeration effects is obtained from heterogeneity across firms with regard to their relative position within the network spanned by firms' SIC codes. Heterogeneity in firms' network position implies that the effect of agglomeration externalities varies across firms according to their level of exposure to other firms' influences. This offers a way of linking the network topology, more precisely the position of each firm within the network as represented by its Katz-Bonacich centrality, to the outcome variable of interest: firms' patent applications as a measure for their innovative activity. The main results show evidence for substantial spillovers between firms within the same industries at the 2-digit SIC level, which also obtain at the 1- and 3-digit SIC levels. Spillovers, which I interpret as knowledge spillovers within the given context, appear to occur mostly among directly linked firms, whereas 'global' science-park wide spillovers do not seem to have a robust effect on patenting behavior.

This evidence suggests that science parks might benefit from selecting tenant firms that pursue similar activities rather than from locating a very heterogeneous set of firms in a park. Regarding the debate in the economic geography literature on localization vs. urbanization externalities, this paper adds empirical evidence in favor of localization externalities benefitting firms within science parks.

References

[1] Abdel-Rahman H. and M. Fujita (1990): 'Product Variety, Marshallian Externalities, and City Sizes,' *Journal of Regional Science*, Vol. 30, No. 2, pp. 165-183.

[2] Acs Z., C. Armington and T. Zhang (2007): 'The determinants of new-firm survival across regional economies: The role of human capital stock and knowledge spillover', *Papers in Regional Science*, Vol. 86, No 3.

[3] Alcácer J. and W. Chung (2007): 'Location Strategies and Knowledge Spillovers,' *Management Science*, Vol. 53, No. 5, pp. 760-776.

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

[5] Arrow K. (1962): 'Economic Welfare and the Allocation of Resources for Inventions', in *The Rate and Direction of Inventive Activity*, R. Nelson ed., Princeton University Press.

[6] Audretsch D. and M. Feldman (1996): 'R&D Spillovers and the Geography of Innovation and Production', *American Economic Review*, Vol. 86, Nr. 3, pp. 630-640.

[7] Ballester C., Calvó-Armengol A., and Y. Zenou (2006): 'Who's Who in Networks. Wanted: The Key Player,' *Econometrica*, Vol. 74, pp. 1403-1417.

[8] Baumgardner J. (1988): 'Physicians' Services and the Division of Labor Across Local Markets,' *Journal of Political Economy*, Vol. 96, No. 5, pp. 948-982.

[9] Bloom N., M. Schankerman, and J. Van Reenen (2007): 'Identifying Technology Spillovers and Product Market Rivalry', NBER Working Paper No 13060.

[10] Blume L. and S. Durlauf (2006): 'Identifying Social Interactions: A Review', Michael Oakes and Jay Kaufman (eds.) Methods in Social Epidemiology, Wiley.

[11] Bonacich P. (1987): 'Power and Centrality: A Family of Measures,' *American Journal of Sociology*, Vol. 92, pp. 1170-1182.

[12] Bradfield J.R.G. (1981): 'The Cambridge Science Park,' *TrAC Trends in Analytical Chemistry*, Vol. 2, Issue 2, pp. 6-9.

[13] Bramoullé Yann, Djebbari Habiba, and Fortin Bernard (2009): 'Identification of peer effects through social networks', *Journal of Econometrics*, Vol. 150, pp. 41-55.

[14] Butts Carter T. (2010): 'sna: Tools for Social Network Analysis,' R package version 2.1.

[15] Cabral R. and S. Sapolnik Dahab (1998): 'Science parks in developing countries: the case of BIORIO in Brazil,' *International Journal of Technology Management*, Vol. 16, No.8, pp. 726-739.

[16] Calvó-Armengol A., E. Patacchini, and Y. Zenou (2009): 'Peer Effects and Social Networks in Education,' Review of Economics Studies, Vol. 76, pp. 1239-1267.

[17] Ciccone A. and R. Hall (1996): 'Productivity and the Density of Economic Activity,' *American Economic Review*, Vol. 86, No. 1, pp. 54-70.

[18] Coles M. G. and E. Smith (1998): 'Marketplaces and Matching,' *International Economic Review*, Vol. 39, No. 1, pp. 239-254.

[19] De Giorgi G., M. Pellizzari, and S. Redaelli (2010): 'BIdentification of Social Interactions through Partially Overlapping Peer Groups,' *American Economic Journal: Applied Economics*, Vol. 2, No. 2, pp. 241-275(35).

[20] Devereux M., Griffith R., and H. Simpson (2004): 'The geographic distribution of production activity in the UK,' *Regional Science & Urban Economics*, Vol. 34, pp. 533-564.

[21] Devereux M., Griffith R., and H. Simpson (2007): 'Firm location decisions, regional grants and agglomeration externalities,' *Journal of Public Economics*, Vol. 91, pp. 413435.

[22] Disney R., J. Haskel and Y. Heden (2003): 'Entry, exit and establishment survival in UK manufacturing,' *Journal of Industrial Economics*, Vol. 51(1), pp. 91-112.

[23] Döring T. and J. and Schnellenbach (2006): 'What Do We Know about Geographical Knowledge Spillovers and Regional Growth?: A Survey of the Literature', *Regional Studies*, Vol. 40, Nr. 3, pp. 375-395.

[24] Dorfman N. (1983): 'Route 128: The development of a regional high technology economy,' *Research Policy*, Vol. 12, pp. 299-316.

[25] Duranton G. and H. Overman (2008): 'Exploring the Detailed Location Patterns of UK Manufacturing Industries using Microgeographic Data,' Journal of Regional Science, Vol. 48, No. 1, pp. 213243.

[26] Eberhardt M. and C. Helmers (2010): 'Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators,' Oxford University, Department of Economics Discussion Paper No. 513.

[27] Ellison G. and E. Glaeser (1997): 'Geographic concentration in U.S. manufacturing industries: A dartboard approach,' *Journal of Political Economy*, Vol. 105, pp. 889-927.

[28] Ellison G. and E. Glaeser (1999): 'The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?,' *American Economic Review*, Vol. 89, No. 2, Papers and Proceedings of the One Hundred Eleventh Annual Meeting of the American Economic Association, pp. 311-316.

[29] EU Commission (2008): 'Towards world-class clusters in the European Union: Implementing the broad-based innovation strategy,' COM(2008) 652.

[30] Ferguson R. and C. Olofsson (2004): 'Science Parks and the Development of NTBFs Location, Survival and Growth,' *Journal of Technology Transfer*, Vol. 29, pp. 517.

[31] Fox-Kean M. and P. Thompson (2005): 'Patent Citations and the Geography of Knowledge Spillovers: A Reassessment,' *American Economic Review*, Vol. 95, No. 1, pp. 450-460.

[32] Fujita M. and J.-F. Thisse (1996): 'Economics of Agglomeration,' *Journal of the Japanese and International Economies*, Vol. 10, pp. 339-378.

[33] Fujita M., P. Krugman, and A. Venables (1999): 'The Spatial Economy: Cities, Regions, and International Trade', *MIT Press.*

[34] Gibbons and Overman (2010): 'Mostly Pointless Spatial Econometrics?,' mimeo Spatial Economics Research Centre LSE.

[35] Glaeser Edward, Kallal Hedi, Scheinkman José, and Shleifer Andrei (1992): 'Growth in Cities', *Journal of Political Economy*, Vol. 100, No. 6, pp. 1126-1152.

[36] Glaeser E., J. Kolko, and A. Saiz (2001): 'Consumer City,' *Journal of Economic Geography*, Vol. 1, 27-50.

[37] Goyal S. (2007): 'Connections,' Princeton University Press.

[38] Greenstone M., R. Hornbeck, and E. Moretti (2010): 'Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,' *Journal of Political Economy*, forthcoming.

[39] Hanson G. (2005): 'Market potential, increasing returns and geographic concentration,' *Journal of International Economics*, Vol. 67, pp. 1 24.

[40] Helmers, C., M. Rogers and P. Schautschick (2010): 'Firm-Level Analysis of Intellectual Property in the UK: The Oxford Firm-level Intellectual Property Database,' mimeo University of Oxford.

[41] Helmers C. and M. Rogers (2010): 'Innovation and the Survival of New Firms in the UK,' *Review of Industrial Organization*, Vol. 36, No. 3, pp. 227-248.

[42] Henderson J.V. (1986): 'Efficiency of resource usage and city size,' *Journal of Urban Economics*, Vol. 19, pp. 47-70.

[43] Henderson J.V. (1997): 'Externalities and industrial development,' *Journal of Urban Economics*, Vol. 42, pp. 449-470.

[44] Holmes T. (1999): 'Localization of industry and vertical disintegration,' *Review of Economics and Statistics*, Vol. 81(2), pp. 314-325.

[45] Jacobs J. (1969): 'The Economy of Cities,' Vintage, New York.

[46] Jackson M. O. (2007): 'Social and Economic Networks,' Princeton University Press.

[47] Jaffe A. (1986): 'Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value', *American Economic Review*, Vol. 76, No. 5, pp. 984-1001.

[48] Jaffe A., M. Trajtenberg, R. and Henderson (1993): 'Geographic localization of knowledge spillovers as evidenced by patent citations,' *Quarterly Journal of Economics*, Vol. 108, pp. 577-598.

[49] Katz L. (1953): 'A New Status Index Derived from Sociometric Analysis,' *Psychometrica*, Vol. 18, pp. 39-43.

[50] Kelejian H. and Prucha I. (1999): 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review*, Vol. 40, pp 509533.

[51] Krugman P. (1980): 'Scale Economies, Product Differentiation, and the Pattern of Trade,' *American Economic Review*, Vol. 70, No. 5, pp. 950-959.

[52] Krugman P. (1991): 'Geography and Trade', *Gaston Eyskens Lecture Series*, MIT Press.

[53] Krugman P. (1993): 'First nature, second nature, and metropolitan location,' *Journal of Regional Science*, Vol. 33, pp. 129144.

[54] Leake J. and P. Treloar (2010): 'Driving Business Success through the Power of the network,' UKSPA Good Practice Note: Networking, available under http://www.ukspa.org.uk/UKSPA%20Case%20Study%20Networking.pdf

[55] Lee S.-Y. (2009): 'Bias from Misspecified Spatial Weight Matrices in SAR Models: Theory and Simulation Studies,' mimeo Department of Economics,San Francisco State University.

[56] Leyden D., Link A., and D. Siegel (2008): 'A Theoretical and Empirical Analysis of the Decision to Locate on a University Research Park,' *IEEE Transactions on Engineering Management*, Vol. 55, No. 1, 23-28.

[57] Lindelöf P. and Löfsten H. (2003): 'Science Park Location and New Technology-Based Firms in Sweden Implications for Strategy and Performance,' *Small Business Economics*, Vol. 20, pp. 245258.

[58] Link A. and J. Scott (2003): 'U .S. science parks: the diffusion of an innovation and its effects on the academic missions of universities,' *International Journal of Industrial Organization*, Vol. 21, pp. 13231356.

[59] Link A. and J. Scott (2006): 'US University Research Parks,' *Journal of Productivity Analysis*, Vol. 25(1), pp. 4355.

[60] Link A. and Scott J. (2007): 'The economics of university research parks, *Oxford Review of Economic Policy*, Vol. 23, No. 4, pp. 661-674.

[61] Manski C. (1993): 'Identification of Endogenous Social Effects: The Reflection Problem', *Review of Economic Studies*, Vol. 60, No. 3, pp. 531-542.

[62] Mariotti S., L. Piscitello, and S. Elia (2010): 'Spatial agglomeration of multinational enteprises: the role of information externalities and knowledge spillovers,' *Journal of Economic Geography*, doi:10.1093/jeg/lbq011

[63] Marschak J. and W.H. Andrews (1944): 'Random Simultaneous Equations and the Theory of Production,' *Econometrica*, Vol. 12, No. 3-4, pp. 143-205.

[64] Marshall A. (1920): 'Principles of Economics,' Macmillan, London.

[65] Midelfart Knarvik K. H. and F. Steen (1999): 'Self-reinforcing Agglomerations? An Empirical Industry Study,' *Scandinavian Journal of Economics*, Vol. 101(4), pp. 515-532.

[66] Moffitt (2001): 'Policy interventions, low-level equilibria, and social interactions', in Social Dynamics, eds. Steven Durlauf and Peyton Young, MIT Press.

[67] Openshaw S. (1984): 'The Modifiable Areal Unit Problem,' Norwich Geo Books.

[68] Ord K. (1975): 'Estimation Methods for Models of Spatial Interaction,' *Journal of the American Statistical Association*, Vol. 70, No. 349, pp. 120-126.

[69] Overman H. and D. Puga (2009): 'Labour Pooling as a Source of Agglomeration: An Empirical Investigation,' CEPR Discussion Paper No. DP7174.

[70] Páez A., D. Scott and . Volz (2008): 'Weight matrices for social influence analysis: An investigation of measurement errors and their effect on model identification and estimation quality,' *Social Networks*, Vol. 30, pp. 309317.

[71] Park S.-C. (2002): 'Science Parks in Sweden as Regional Development Strategies: A Case Study on Ideon Science Park,' *AI & Soc*, Vol. 16, pp. 288-298.

[72] Phan P., Siegel D., and Wright M. (2003): 'Science parks and incubators: observations, synthesis and future research,' *Journal of Business Venturing*, Vol. 20, pp. 165-182.

[73] Porter M. (1990): 'The Competitive Advantage of Nations', *Free Press*, New york.

[74] Puga D. (2010): 'The magnitude and causes of agglomeration economies,' *Journal of Regional Science*, Vol. 50, No. 1, pp. 203-219.

[75] Quintas P., Wield D., and D. Massey (1992): 'Academic-industry links and innovation: questioning the science park model,' *Technovation*, Vol. 12(3), pp. 161-175.

[76] Rice P., A. J. Venables, and E. Patacchini (2006): 'Spatial determinants of productivity: Analysis for the regions of Great Britain,' *Regional Science & Urban Economics*, Vol. 36, pp. 727752.

[77] Romer Paul (1990): 'Endogenous Technological Change', *Journal of Political Economy*, Vol. 98, No. 5, pp. 71-101.

[78] Roos M. W. M. (2005): 'How important is geography for agglomeration?,' *Journal of Economic Geography*, Vol. 5, pp. 605-620.

[79] Rosenthal S. and W. Strange (2004): 'Evidence on the nature and sources of agglomeration economies,' in Henderson J.V. and J.F. Thisse (eds.) Handbook of Regional and Urban Economics, Elsevier.

[80] Sacerdote B. (2001): 'Peer Effects with Random Assignment: Results for Dartmouth Roommates', *Quarterly Journal of Economics*, Vol. 116, No. 2, pp. 681-704.

[81] Salvador E. (2010): 'Are science parks and incubators good "brand names" for spin-offs? The case study of Turin,' *Journal of Technology Transfer*, DOI 10.1007/s10961-010-9152-0

[82] Schumpeter J. (1942): 'Capitalism, Socialism, and Democracy', *Harper*, New York.

[83] Siegel D., P. Westhead, and M. Wright (2003a): 'A ssessing the impact of university science parks on research productivity: exploratory firm-level evidence from the United Kingdom,' *International Journal of Industrial Organization*, Vol. 21, pp. 13571369.

[84] Siegel D., Westhead P., and M. Wright (2003b): 'Science Parks and the Performance of New Technology-Based Firms: A Review of Recent U.K. Evidence and an Agenda for Future Research,' *Small Business Economics*, Vol. 20, pp. 177184.

[85] Tan J. (2006): 'Growth of industry clusters and innovation: Lessons from Beijing Zhongguancun Science Park,' *Journal of Business Venturing*, Vol. 21, pp. 827-850.

[86] Thompson P. and M. Fox-Kean (2005): 'Patent Citations and the Geography of Knowledge Spillovers: A Reassessment', *American Economic Review*, Vol. 95, No. 1, pp. 450-460.

[87] van der Panne G. (2004): 'Agglomeration Externalities: Marshall versus Jacobs,' *Journal of Evolutionary Economics*, Vol. 14, pp. 593-604.

[88] Wai Yip So B. (2006): 'Reassessment of the State Role in the Development of High-Tech Industry: A Case Study of Taiwan's Hsinchu Science Park,' *East Asia*, Vol. 23, No. 2, pp. 61-86.

[89] Westhead P. (1997): 'R&D 'inputs' and 'outputs' of technologybased firms located on and off Science Parks,' *R&D Management*, Vol. 27, No. 1., pp. 45-62.

[90] Westhead P. and Batstone S. (1998): 'Independent Technology-based Firms: The Perceived Benefits of a Science Park Location,' *Urban Studies*, Vol. 35, No. 12, pp. 2197-2219.

[91] Zimmerman D. (2003): 'Peer Effects in Higher Education: Evidence from a Natural Experiment, Williams Project on the Economics of Higher Education,' *Review of Economics and Statistics*, Vol. 85(1), pp. 9-23.

Figure 1: Example network

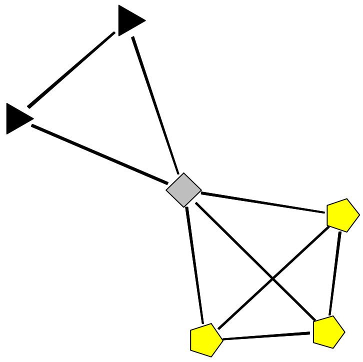

Figure 2: Network of sample - 2000 and 2007

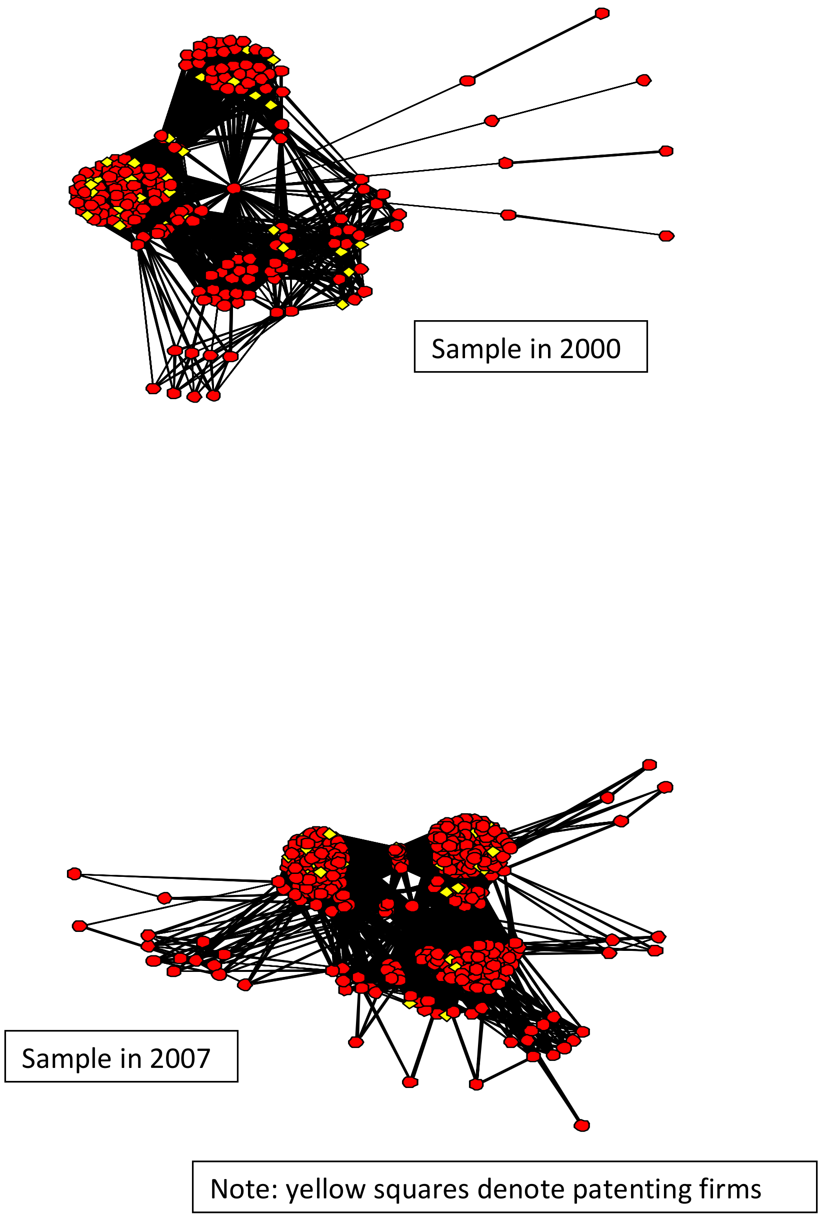

Figure 3: Degree distribution for sample

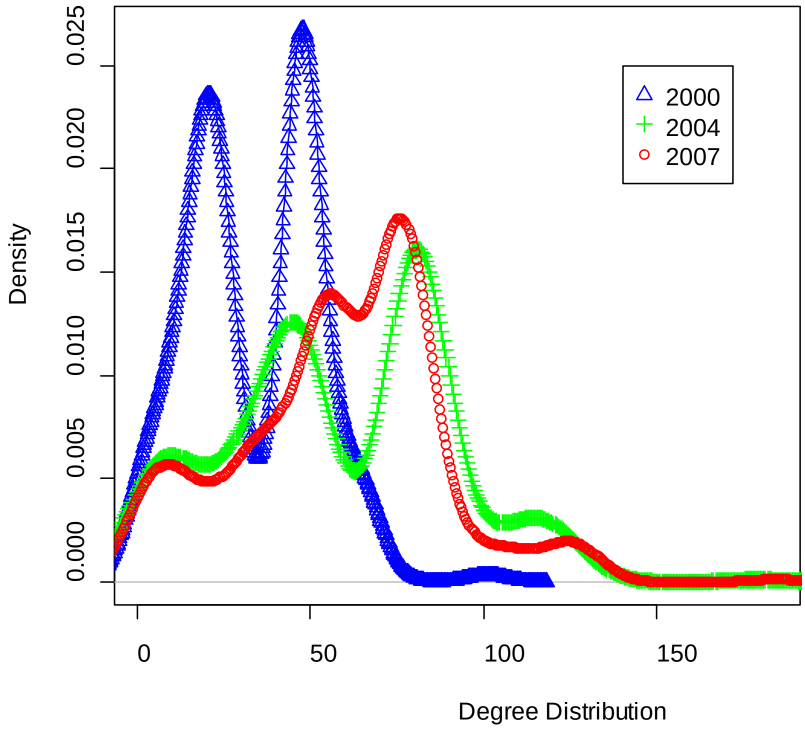

Figure 4: Empirical CDF of Katz-Bonacich Centrality

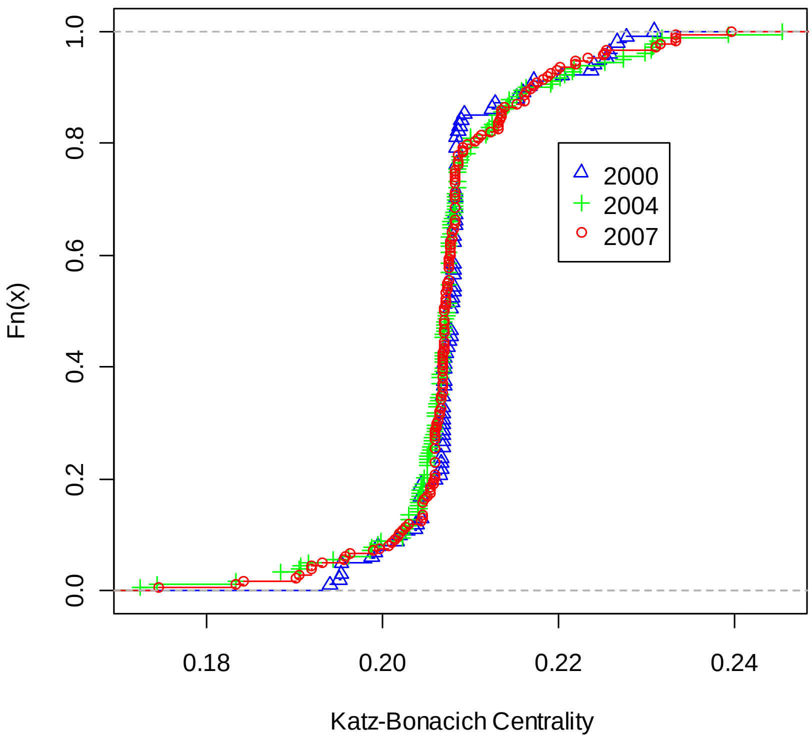

Figure 5: Scatter plot Katz-Bonacich Centrality vs. ln No. Patents

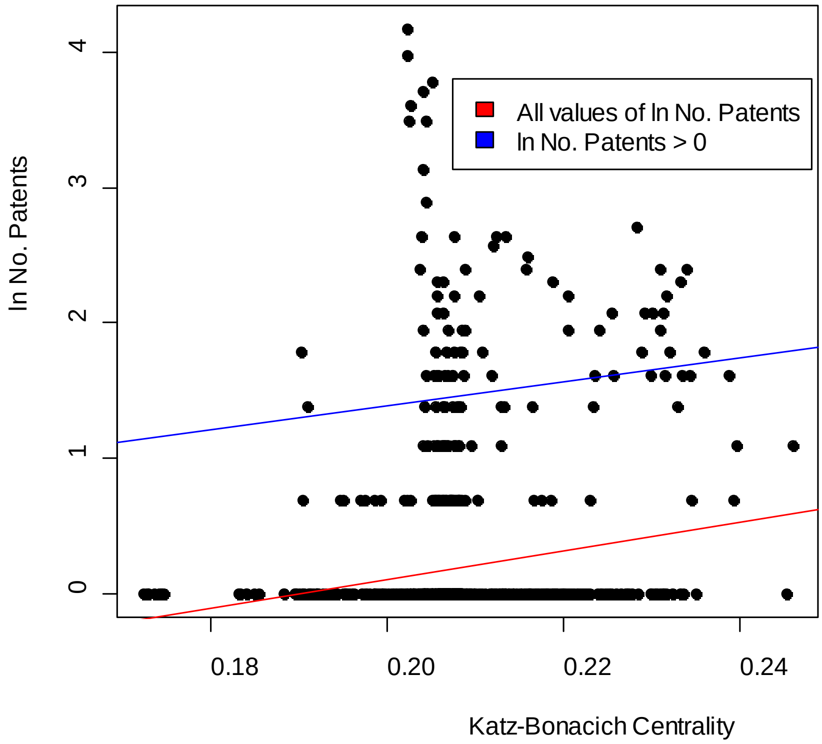

Table 1: Tenant Firms - Entry & Exit from/to CSP, SJC, and BIC

Year	# Tenants	# Entry	# Exit
(1)	(2)	(3)	(4)
2000	127	24	3
2001	170	46	27
2002	182	39	15
2003	199	32	23
2004	221	45	31
2005	230	40	28
2006	239	37	19
2007	266	46	23
# Firms	412	309	169

Table 2: Tenant Firms - CSP, SJC, BIC

Year	# Tenants		
	SJC	CSP	BIC
(1)	(2)	(3)	(4)
2000	48	73	6
2001	74	86	10
2002	98	78	6
2003	106	87	6
2004	122	93	6
2005	133	91	6
2006	146	89	4
2007	175	91	
# Firms	267	144	10

Table 3: Distribution of SIC codes for full sample (2000-2008)

SIC 2-digit	2000	2001	2002	2003	2004	2005	2006	2007	Total	%
Agriculture (01)								1	1	0.2
Mining (13)			1	1	1				1	0.2
Publishing etc. (22)	1	2	2	2	3	3	3	3	4	0.8
Chemicals (24)	13	14	15	16	17	18	18	16	21	4.0
Rubber and plastic (25)					1	1	1		1	0.2
Machinery nes (29)	1	1	2	2	2	3	3	2	3	0.6
Office machinery & computers (30)	2	2	2	3	3	1	1	2	4	0.8
Electrical Machinery (31)	4	4	4	4	5	6	5	8	9	1.7
Radio, & television (32)	1	3	4	4	5	5	4	3	7	1.4
Medical & optical instruments (33)	6	7	8	8	9	9	10	10	12	2.3
Other transport equipment (35)					1				1	0.2
Other manufacture (36)	2	2	3	4	5	4	6	6	7	1.4
Recycling (37)						1	1	1	1	0.2
Construction (45)	1	1	2	2			2	2	4	0.8
Repair of Motor vehicles (50)					1	1	1	1	1	0.2
Wholesale trade (51)	12	15	14	15	17	18	18	19	24	4.6
Retail trade (52)			1	1	1	1	1	1	1	0.2
Hotels & restaurants (55)				1	1	1	1	1	1	0.2
Transport activities (63)	1	1	1	1	1	1	1	1	1	0.2
Telecommunications (64)	1	1	2	4	7	8	6	6	11	2.1
Financial intermediation (65)	1	1	1	1	1	1	1	1	1	0.2
Insurance (66)	1	1	1	1	1	1	1	1	1	0.2
Auxiliary to financial intermediation (67)	1		2	3	1	1	1	1	4	0.8
Real estate (70)	1	4	5	2	2	2	2	2	9	1.7
Renting of machinery (71)			2	2	1	1	1	1	2	0.4
Computer (72)	28	40	45	52	58	63	68	80	111	21.5
R&D (73)	24	30	30	34	38	44	45	44	69	13.3
Other business activities (74)	56	75	81	84	94	96	100	114	180	34.8
Education (80)	1	3	4	4	4	6	5	5	7	1.4
Health (85)	3	4	3	4	5	5	5	5	6	1.2
Activities nes (91)	1	1	1	1	2	3	1	1	4	0.8
Motion picture production (92)	1	1	1	1	1		1	1	2	0.4
Other services (93)	2	4	3	3	3	3	3	3	6	1.2
# Tenants	127	170	182	199	221	230	239	266	412	
# Tenants > 1 SIC code	32	41	47	50	55	60	60	60	84	
% Tenants > 1 SIC code	25.2	24.1	25.8	25.1	24.88	26.1	25.8	22.6	20.5	

Table 4: Degree - Summary Statistics

Year	Network Density[♭]	Degree Mean	Std.Dev.	Med.	Min	Max	Firms
2000	0.166	32.85	17.88	27	1	67	101
2001	0.167	43.94	23.59	38	1	87	134
2002	0.166	48.58	24.27	40	1	100	148
2003	0.168	54.61	28.19	45	1	112	165
2004	0.164	57.91	31.64	47	1	130	179
2005	0.158	59.20	32.44	55	1	135	190
2006	0.162	63.34	32.99	57	1	145	198
2007	0.159	57.73	28.70	55	2	125	184

Notes:
[♭] Network Density is defined as the average degree divided by the number of possible links $n-1$.

Table 5: Summary Statistics

	Mean	Std.Dev.	Med.	Min	Max	Obs
No. Patent Applications	0.700	3.661	0	0	64	1,299
ln No. Patent Applications	0.185	0.563	0	0	4.174	1,299
Trade-mark	0.081	0.273	0	0	1	1,299
Subsidiary	0.444	0.497	0	0	1	1,299
CSP	0.452	0.498	0	0	1	1,299
Age	9.570	12.118	6	1	85	1,299
ln Age	2.006	0.772	1.945	0.693	4.454	1,299
Total Assets$^\flat$ £1000	7,012	24,127	114.603	0	235,716	1,299
ln Total Assets$^\flat$ £1000	4.887	3.402	4.750	0	12.370	1,299
SIC 3-digit Patenting Intensity$^\natural$	0.222	0.332	0.070	0.003	1	1,299

Notes:
The sample contains 275 firms.
$^\flat$ Deflated using sector-level asset deflator (base year=2000).
$^\natural$ Ratio of no. of patents and turnover in SIC 3-digit sector (normalized such that $\in \{0, 1\}$).

Table 6: Estimates excluding contextual effects

	OLS			SEM		
Dependent variable: ln No. Patent Applications						
	(1)	(2)	(3)	(4)	(5)	(6)
ln **Degree** (μ)	0.045**	0.105*	0.100**	0.041**	0.105***	0.101***
	(0.022)	(0.059)	(0.058)	(0.016)	(0.027)	(0.027)
Peer effects (b)				0.172**	0.079	0.085
				(0.086)	(0.096)	(0.098)
CSP			0.137**			0.137***
			(0.068)			(0.036)
Trade-mark	0.383***	0.362***	0.353***	0.383***	0.362***	0.353***
	(0.108)	(0.097)	(0.099)	(0.051)	(0.050)	(0.050)
Subsidiary	0.028	0.026	-0.023	0.034	0.026	-0.021
	(0.059)	(0.065)	(0.069)	(0.029)	(0.030)	(0.033)
ln Age	-0.003	0.032	0.022	-0.002	0.032	0.023
	(0.035)	(0.044)	(0.043)	(0.022)	(0.021)	(0.022)
ln Total Assets$^\flat$	0.021**	0.023**	0.016*	0.021***	0.023***	0.017***
	(0.009)	(0.010)	(0.010)	(0.004)	(0.005)	(0.005)
Patenting Intensity$^\natural$	0.655***	0.115	0.104	0.657***	0.110	0.097
	(0.184)	(0.158)	(0.159)	(0.043)	(0.285)	(0.283)
Constant	-0.211	-0.281	-0.245	-0.212	-0.525	-0.602
	(0.124)	(0.151)	(0.144)	(0.085)	(0.350)	(0.349)
Year dummies	Included	Included	Included	Included	Included	Included
Sector fixed effects		Included	Included		Included	Included
R^2	0.258	0.302	0.309			

Notes:
1,299 observations for all specifications.
Clustered standard errors in Columns (1)-(3).
$^\flat$ in £1000 deflated using sector-level asset deflator (base year=2000).
$^\natural$ Ratio of no. of patents and turnover in SIC 3-digit sector (normalized such that $\in \{0, 1\}$).

Table 7: Estimates (including contextual effects)

	OLS			SEM		
Dependent variable: ln No. Patent Applications						
	(1)	(2)	(3)	(4)	(5)	(6)
ln **Degree** (μ)	0.039*	0.106*	0.101*	0.038**	0.106***	0.101***
	(0.022)	(0.058)	(0.058)	(0.016)	(0.027)	(0.027)
Peer effects (b)				0.061	-0.021	-0.015
				(0.095)	(0.104)	(0.105)
CSP			0.133**			0.132***
			(0.067)			(0.036)
Trade-mark	0.386***	0.361***	0.353***	0.387***	0.360***	0.353***
	(0.106)	(0.097)	(0.099)	(0.051)	(0.050)	(0.050)
Subsidiary	0.036	0.027	-0.019	0.036	0.027	-0.019
	(0.059)	(0.066)	(0.069)	(0.029)	(0.031)	(0.033)
ln Age	-0.018	0.013	0.005	-0.017	0.013	0.004
	(0.035)	(0.039)	(0.039)	(0.024)	(0.024)	(0.024)
ln Total Assets$^\flat$	0.020**	0.023**	0.017*	0.020***	0.023***	0.017***
	(0.008)	(0.009)	(0.009)	(0.005)	(0.005)	(0.005)
Patenting Intensity$^\natural$	0.656***	0.122	0.111	0.656***	0.123	0.112
	(0.187)	(0.163)	(0.163)	(0.043)	(0.284)	(0.283)
$g\times$Trade-mark	0.008	0.012	0.011	0.008	0.012**	0.011*
	(0.009)	(0.010)	(0.010)	(0.006)	(0.006)	(0.283)
$g\times$Subsidiary	-0.011*	-0.007	-0.007	-0.011***	-0.007**	-0.008**
	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)
$g\times$ln Age	0.004	0.003	0.003	0.004***	0.003**	0.003**
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
$g\times$ln Total Assets$^\flat$	-0.001*	-0.001	-0.001	-0.001*	-0.001*	-0.001*
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
$g\times$Patenting Intensity$^\natural$	0.0001	-0.001	-0.001	0.000	-0.001	-0.001
	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)
Constant	-0.192	-0.289	-0.248	-0.193	-0.567	-0.635
	(0.117)	(0.157)	(0.152)	(0.089)	(0.350)	(0.349)
Year dummies	Included	Included	Included	Included	Included	Included
Sector fixed effects		Included	Included		Included	Included
R^2	0.266	0.308	0.315			

Notes:
1,299 observations for all specifications.
Clustered standard errors in Columns (1)-(3).
$^\flat$ in £1000 deflated using sector-level asset deflator (base year=2000).
$^\natural$ Ratio of no. of patents and turnover in SIC 3-digit sector (normalized such that $\in \{0, 1\}$).

Table 8: Robustness: Estimates excluding contextual effects — SIC **3**-digit level

	OLS		SEM	
Dependent variable: ln No. Patent Applications				
	(1)	(2)	(3)	(4)
ln **Degree** (μ)	0.048**	0.123**	0.044**	0.113***
	(0.023)	(0.051)	(0.017)	(0.028)
Peer effects (b)			0.181***	0.106
			(0.063)	(0.070)
Trade-mark	0.393***	0.351***	0.395***	0.354***
	(0.110)	(0.102)	(0.051)	(0.051)
Subsidiary	0.029	0.004	0.032	0.007
	(0.059)	(0.077)	(0.029)	(0.034)
ln Age	0.003	0.035	0.005	0.038
	(0.037)	(0.048)	(0.023)	(0.023)
ln Total Assets$^\flat$	0.019***	0.026**	0.019***	0.026***
	(0.009)	(0.011)	(0.005)	(0.005)
Patenting Intensity$^\natural$	0.630***	0.097	0.634***	0.066
	(0.181)	(0.164)	(0.044)	(0.286)
Constant	-0.194	-0.156	-0.202	-0.279
	(0.121)	(0.106)	(0.081)	(0.159)
Year dummies	Included	Included	Included	Included
Sector fixed effects		Included		Included
R^2	0.260	0.316		

Notes:
1,284 observations for all specifications.
Clustered standard errors in Columns (1) and (2).
$^\flat$ in £1000 deflated using sector-level asset deflator (base year=2000).
$^\natural$ Ratio of no. of patents and turnover in SIC 3-digit sector (normalized such that $\in \{0, 1\}$).

Table 9: Robustness: Estimates excluding contextual effects — SIC **1**-digit level

	OLS		SEM	
Dependent variable: ln No. Patent Applications				
	(1)	(2)	(3)	(4)
ln **Degree** (μ)	0.045**	0.127**	0.040***	0.124***
	(0.018)	(0.049)	(0.014)	(0.026)
Peer effects (b)			0.223**	0.185*
			(0.103)	(0.108)
Trade-mark	0.393***	0.380***	0.393***	0.379***
	(0.109)	(0.103)	(0.050)	(0.049)
Subsidiary	0.031	0.032	0.035	0.034
	(0.057)	(0.059)	(0.028)	(0.028)
ln Age	0.004	0.023	0.006	0.025
	(0.035)	(0.038)	(0.022)	(0.022)
ln Total Assets$^{\flat}$	0.019**	0.021**	0.019***	0.020***
	(0.008)	(0.008)	(0.004)	(0.005)
Patenting Intensity$^{\natural}$	0.624***	0.631***	0.635***	0.641
	(0.179)	(0.182)	(0.042)	(0.043)
Constant	-0.234	-0.678	-0.234	-0.347
	(0.126)	(0.225)	(0.087)	(0.129)
Year dummies	Included	Included	Included	Included
Sector fixed effects		Included		Included
R^2	0.263	0.278		

Notes:
1,352 observations for all specifications.
Clustered standard errors in Columns (1) and (2).
$^{\flat}$ in £1000 deflated using sector-level asset deflator (base year=2000).
$^{\natural}$ Ratio of no. of patents and turnover in SIC 3-digit sector (normalized such that $\in \{0, 1\}$).