SERC DISCUSSION PAPER 92

Ethnic Inventors, Diversity and Innovation in the UK: Evidence from Patents Microdata

Max Nathan (SERC, Department of Geography & Environment, London School of Economics)

October 2011

This work is part of the research programme of the independent UK Spatial Economics Research Centre funded by the Economic and Social Research Council (ESRC), Department for Business, Innovation and Skills (BIS), the Department for Communities and Local Government (CLG), and the Welsh Assembly Government. The support of the funders is acknowledged. The views expressed are those of the authors and do not represent the views of the funders.

Ethnic Inventors, Diversity and Innovation in the UK: Evidence from Patents Microdata

Max Nathan*

October 2011

* SERC, Department of Geography & Environment, London School of Economics

Acknowledgements

Thanks to Gianluca Tarasconi and Francesco Lissoni at Università Bocconi for KITESPATASTAT patent data, to Pablo Mateos at UCL for ONOMAP coding, and to the ONS Virtual Microdata Lab for LFS datasets. Thanks for advice and support to Henry Overman, Ian Gordon, Maria Carvalho, Riccardo Crescenzi, Giulia Faggio, Ben Gidley, Simona Iammarino, Carlo Menon, Rosa Sanchis-Guarner-Herrero, Carlos Vargas-Silva and Vivek Wadhwa. Participants at a SERC seminar (November 2010), an RSA Early Careers Conference (November 2010) and a COMPAS workshop (May 2011) provided helpful comments on previous versions of the paper. All remaining errors and omissions are my own.

I am grateful to the ESRC and the Department of Communities and Local Government for research support. The views in this paper are my own, and do not necessarily represent those of the Department or the ESRC. This work contains statistical data from ONS which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Abstract

Ethnic inventors play important roles in US innovation systems, especially in high-tech regions like Silicon Valley. Do 'ethnicity-innovation' channels exist elsewhere? This paper investigates, using a new panel of UK patents microdata. In theory, ethnicity might affect positively innovation via 'star' migrants, network externalities from co-ethnic groups, or production complementarities from diverse inventor communities. I use the novel ONOMAP name classification system to identify ethnic inventors. Controlling for individuals' human capital, I find small positive effects of South Asian and Southern European co-ethnic group membership on individual patenting. The overall diversity of inventor communities also helps raise individual inventors' productivity. I find no hard evidence that ethnic inventors crowd out patenting by majority groups.

JEL Classification: J15, J24, J61, M13, O3, R11, R23
Keywords: ethnic inventors, innovation, patents, cultural diversity, diasporas, cities

1. Introduction

At first glance, ethnicity, diversity and innovation do not seem closely linked. However, in recent years there has been growing policy and public interest in the role of 'ethnic inventors' in innovative activity, both in the UK and elsewhere (Kerr and Kerr 2011, Leadbeater 2008, Page 2007, Legrain 2004). These discussions and debates have largely drawn on recent experience in the United States. Since the 1980s minority communities, particularly those of South / East Asian origin, have played increasingly important roles in ideas generation in the science and technology sectors (Chellaraj et al 2005, Stephan and Levin 2001). US ethnic inventors – who are often migrants – are spatially concentrated at city-region level (Kerr 2008). High-tech US clusters like Silicon Valley have benefited from 'ethnic entrepreneurs' who both help connect South Bay firms to global markets, and are responsible for 52% of the Bay Area's startups (Wadhwa et al 2007). There are positive links between the presence of migrants and US regional patenting (Hunt and Gauthier-Loiselle 2008, Peri 2007). Diasporic communities appear to play important roles in the diffusion of knowledge both across US cities, and between US regions and 'home' countries (Kerr 2009, 2008).

By contrast, surprisingly little is known about the role of 'ethnic inventors' on innovation in the UK. Over the past two decades Britain has become substantially more ethnically diverse. The number of people from non-white ethnic groups grew by 53% between 1991 and 2001. For England and Wales between 2001 and 2009, non 'White British' groups have grown from 6.6m to 9.1m and now stands at one in six of the population (ONS 2011). Immigration has been a main driver, with a number of 'new migrant communities' forming since the early 1990s (Kyambi 2005). This paper asks: has UK innovation benefited from these population shifts as it has in the US?

Changing demography could affect innovation in at least four complementary ways. First, migrants or individuals from minority communities may be positively selected on the basis of skills or entrepreneurial behaviour, although this needs to be distinguished from other human capital endowments (Borjas 1987). Second, by lowering transaction costs, co-ethnic groups can accelerate within-group ideas generation and transmission, although discrimination may constrain knowledge spillovers (Docquier and Rapaport 2011, Kloosterman and Rath 2001). Third, cultural diversity may improve ideas generation across all groups, if the benefits of a larger set of ideas, perspectives outweigh trust or communication difficulties between those groups (Berliant and Fujita 2009, Page 2007, Alesina and La Ferrara 2004).

Finally, these channels may be more pronounced in urban areas because of the spatial clustering of minority communities, agglomeration economies, or both. In addition, cosmopolitan urban populations may raise demand for new goods and services, especially in non-tradable sectors (Gordon et al 2007, Mazzolari and Neumark 2009).

This paper looks at the role of ethnic inventors in innovation in the UK, using a new 12-year panel of patents microdata. Using the novel ONOMAP name classification system to build on pioneering US work by Kerr (2008) and Agrawal et al (2007) I am able to explore all four 'population-innovation' channels. I estimate a knowledge production function linking inventors' patenting activity to individual, group and area-level characteristics. Using techniques popularised by Blundell et al (1995), I exploit historic patent information to fit inventor-level fixed effects.

Once human capital is controlled for, I find that simply being an ethnic inventor has no significant effect on individual patenting rates. Conversely I find some positive effects for members of specific co-ethnic groups: Indian, South Asian and Southern European inventors. I also find small positive effects of inventor group diversity on individual patenting activity. Effects on majority inventors are less clear: increasing ethnic diversity has some negative links to majority groups' patenting activity individual level, but I find no effects of crowding out at area level. Urban location has relatively small effects on individual patenting after other individual and area-level factors are included. The results survive extensive robustness checks, although alternative measures of area-level human capital weaken diversity effects.

The paper adds to a small but growing empirical literature on immigration, ethnicity and innovation (Kerr and Kerr 2011). It also contributes to the emerging field of inventor microdata analysis (OECD 2009). It is one of very few studies exploring multiple ethnicity-innovation channels, at individual, group and area level. As far as I am aware, this is the first research of its kind in Europe.

The paper is structured as follows. Section 2 set out research questions and key terms. Section 3 reviews relevant theoretical frameworks and empirics. Sections 4 and 5 introduce the main data sources and provide descriptive statistics. Section 6 outlines the model and estimation strategy. Sections 7 – 9 give results, extensions and robustness checks. Section 10 concludes.

2. Research questions

My research questions are:

- Do ethnic inventors or co-ethnic groups influence patenting rates in the UK?
- Does the cultural diversity of inventor groups influence patenting rates?
- Do urban environments affect ethnicity- or diversity-innovation effects?

'Innovation', 'ethnicity' and 'diversity' are fuzzy concepts that need to be carefully defined. The innovation process is commonly divided into three phases: invention, adoption and diffusion (Fagerberg 2005): a standard UK definition of innovation is thus 'the successful exploitation of new ideas' (Department of Innovation Universities and Skills 2008). My chosen measure of innovation, patenting, is primarily an indicator of invention (OECD 2009). Specifically, I look at shifts in individual patenting rates, or 'inventor productivity'.

Patent data has several advantages: it has a positive relationship with other indicators of overall innovation 'performance' such as productivity and market share; it provides detailed information on geography and patent owners, both inventors and applicant firms; and is available for long time periods at relatively low cost. Not all inventions are patented, however, and patents have variable coverage across industries (with a well-known bias towards manufacturing) (OECD 2009). Patenting also responds to policy shocks – for example, US Supreme Court decisions in the 1980s and 1990s (particularly *Re Alappat* in 1994) led to spikes in software and information technology patenting (Li and Pai 2010).

I am able to deal with most of these challenges through careful identification strategies (see section 4). Unlike the majority of patent data studies, I am able to work at individual inventor level – using the KITES-PATSTAT patents dataset developed at Bocconi University (more of which below).

'Ethnicity' is as hard to pin down. Ethnic identity is a multifaceted concept with objective, subjective and dynamic elements (Aspinall 2009). Quantitative measures of identity tend to be partial: they focus on identity's visible, objective components, assuming away self-ascription and endogeneity issues (Ottaviano, Bellini et al. 2007). Given these limitations, quantitative researchers working with ethnic identity will always need to use a 'least-worst' proxy. I deploy two such measures, using the ONOMAP system to analyse inventor name information and read off likely ethnicity characteristics (see Section 4 for details). The first proxy is the ethnic group classifications prepared by the UK Office of National Statistics (ONS). The ONS measures attempt to combine different aspects of ethnic identity, but operate at a high level of generality and tend to focus on 'visible minorities' such as Black and Asian communities (Mateos, Webber et al. 2007).

I use 'geographical origin' as a second proxy measure. Geographical origin can offer very fine-grained information, but is one-dimensional as a measure of identity. In this case, because name data conflates migrants and their descendents, origin effectively operates as a measure of geographical 'roots'.[1] As such, it offers an alternative way of identifying likely ethnicity and co-ethnic group membership.

To measure the diversity of these ethnic groups, I use a Fractionalisation Index as commonly used in the development literature. See Section 4 for details.

3. Theoretical frameworks and evidence

Conventional theories of innovation have relatively little to say to about ethnicity or the composition of inventor communities. Schumpeter (1962) focuses on the 'entrepreneurial function' inside and outside firms, and the role of individuals in identifying and commercialising new ideas, in the face of social inertia or resistance. National 'innovation

[1] Although not national identity: the vast majority of those born in the UK think of themselves as British (Manning and Roy 2007). More broadly, ethnicity, nationality, sexuality and class are all elements in a broader sense of self (Fanshawe and Sriskandarajah 2010).

systems' approaches explore relationships between firms and public institutions such as government agencies and universities (Freeman 1987). More recently, spatial approaches focus on clustering of innovative activity due to agglomeration-related externalities, particularly local knowledge spillovers (Jaffe, Trajtenberg et al. 1993; Audretsch and Feldman 1996).

Endogenous growth theories provide the basis for a number of newer studies linking demography to innovation. Endogenous growth models suggest that shifts in the technology frontier help determine economic development. They also highlight the importance of human capital stocks and knowledge spillovers to levels of innovation (Romer 1990). In practice, access to knowledge is likely to be uneven across locations, sectors and social groups (Agrawal, Kapur et al. 2008).

Recent work suggests four ways in which demographic factors could positively influence ideas generation and transmission. Building on the material in the introduction, theoretical frameworks and empirics are discussed for each in turn.

3.1 Individual selection

Migrants are mobile carriers of ideas – so high-skilled migrants, in particular, may positively contribute to overall innovation rates (Kerr and Lincoln 2010). More broadly, from an economic perspective, migration decisions reflect expected returns: potential migrants balance out economic gains from migration and costs of moving abroad (Borjas 1987). The income maximisation approach implies that migrants are 'pre-selected' – and are more likely to be entrepreneurial, seeking out new ideas (Wadhwa, Saxenian et al. 2007).

Both these factors suggest migrant status may positively predict patenting rates, over and above other human capital attributes. Discrimination has ambiguous effects. It may lead to 'lock-out' from conventional labour market opportunities (Gordon 2001). Conversely, it may operate as a spur to innovation if excluded minorities are forced to develop new economic opportunities (Rath and Kloosterman 2000). The challenge is to distinguish ethnicity from wider human capital endowments and relevant industry / area characteristics.

US experience suggests some positive selection effects in science and high-tech sectors of the economy, particularly for migrant workers. US employers in these sectors report heavy dependence on skilled migrants (Wadhwa, Saxenian et al. 2007; Kerr and Lincoln 2010). Indo- and Chinese-American communities make disproportionate contributions to US science and engineering, in terms of workforce membership as well as Nobel Prize counts, elections to scientific academies and patent citations (Stephan and Levin 2001).

Anderson and Platzer (2007) report that immigrants have founded 40% of venture capital-backed technology companies currently trading in the US, including Google, eBay, Yahoo and Sun Microsystems. Wadhwa et al (2007) find the national immigrant contribution to patenting rose from 7.3% in 1998 to 24.6% in 2006. Using time series data, Chellaraj et al (2005) report that foreign graduate students and skilled immigrants have a significant positive impact on patent applications and grants. However, in a recent US study on immigrant patenting, Hunt and Gauthier-Loiselle (2008) suggest that once education and industry characteristics are controlled for, effects of individual migrant status disappear.

There is much less evidence from the UK. Nathan and Lee (2011) report some evidence that migrant entrepreneurs in London are more likely to innovate than average company founders. Basu (2002; 2004) suggests considerable variation in levels of entrepreneurship across minority communities, with class, education and family status important mediating influences.

3.2 Social networks and diaspora effects

A second set of theories suggests that cultural ‘sameness’ or ‘proximity’ helps knowledge spillovers (Agrawal, Kapur et al. 2008). Co-ethnic social networks – such as diasporas or transnational communities – provide network externalities that accelerate ideas transmission (Docquier and Rapoport 2011).

Social networks offer their members higher social capital and levels of trust, lowering transaction costs and risk. In turn, networks seem to positively affect innovative activity (Rodríguez-Pose and Storper 2006; Kaiser, Kongsted et al. 2011). Co-ethnic networks such as diasporas may be an important channel for knowledge spillovers and ideas flow – improving awareness of new technologies and passing on tacit knowledge, both within and across countries (Kerr 2008; Kerr 2009).

Of course, other social networks – such as family or kinship networks, or professional associations – might be equally or more important. And co-ethnic effects on individual patenting are ambiguous. Matching and learning economies may be present within the group (‘enclave’ activity) and between different groups (‘middleman minority’ activities) (Bonacich 1973). But externalities will be constrained by group size, majority attitudes and links between groups. First, within a minority group, individual members are less likely to match ideas than those in the majority group since there will be a smaller set of possible matches. Second, if members of majority group(s) discriminate against minority groups, or if minority groups lack social connections to majority actors, this will limit matches across groups and ‘middleman minority’ activity (Zenou 2011).

In a closed economy, effects of co-ethnic groups are determined by group size and the level of interaction between groups. Under globalisation, co-ethnic communities may be more influential. Increasing numbers of businesses in high-cost countries are looking to relocate research and development (R&D) activity into lower-cost countries (Mowery 2001; Archibugi and Iammarino 2002; Cantwell 2005; Yeung 2009). Diasporic communities with members present in high-cost ‘host’ countries may help firms move into lower-cost ‘home’ countries, identifying collaborators or accelerating joint ventures (Kapur and McHale 2005; Saxenian and Sabel 2008). This raises both the size of the innovating co-ethnic community and the rate of information flow between its members, in both ‘home’ and ‘host’ locations.

A number of case studies suggest that diasporas are important influences on knowledge flows (Bresnahan and Gambardella 2004; Saxenian 2006; Docquier and Rapoport 2011). In a 2002 survey, Saxenian finds that 82% of Chinese and Indian immigrant scientists and engineers exchange technological information with colleagues in ‘home’ countries. Jaffe and Tratjenberg (1999) find that countries with a common language have larger R&D spillovers and international patent citation rates. Kerr (2008) , studying co-ethnic inventors, finds that own-ethnicity citations are 50% higher than citations to other ethnicities, controlling for industry: co-ethnic communities in ‘host’ countries positively influence industrial performance in ‘home’ countries.

Patenting growth in US cities is also faster for technologies that depend heavily on communities of immigrant inventors (Kerr 2009). By contrast, Agrawal et al (2008; 2011) compare co-ethnic and co-location effects on patent citations, finding that physical location is up to four times more important.

US ethnic inventor communities are relatively recent phenomena largely shaped by migration flows since the 1960s (Saxenian 2006). The UK's immigration story is very different: migrant and minority communities are the result of both colonial history (Australasia, some African and South-East Asian groups) and geographical proximity (many European countries). British-based diasporas may not, therefore, share the characteristics of US-style transnational communities.

The existing UK evidence base is mixed. I am unaware of any European studies that explicitly link co-ethnicity to patenting. Fairlie et al (2009) find some support for co-ethnicity effects on British-Indian business performance, although innovation is not considered. Qualitative work by Nakhaie et al (2009) confirms that co-ethnicity effects both vary significantly across groups, and are shaped by wider socio-economic contexts.

3.3 Diversity effects

'Cultural distance' between economic agents may also influence innovation rates. Specifically, individual inventors in a group may benefit from group-level diversity if this brings a richer mix of ideas and perspectives. Berliant and Fujita (2009) model a system of firm-level knowledge creation, showing that worker heterogeneity can accelerate ideas generation through individual-level production complementarities. Hong and Page (2001; 2004) similarly model scenarios in which 'cognitively diverse' teams exploit a larger pool of ideas and skills, suggesting that cultural mix is a good proxy for cognitive diversity.

On the other hand, group-level cultural diversity may have a negative effect if it leads to lower trust and poor communication between individuals – for example, because of language barriers, misunderstandings, discriminatory attitudes or both. Spillovers (and co-operation) will be limited, leading to fewer, lower-quality solutions (Alesina and La Ferrara 2004). Fujita and Weber (2003) argue that positive diversity effects will be most likely observed in research-based or 'knowledge-intensive' activities – such as those leading to patenting. Parrotta et al (2011) suggest that while diversity of knowledge is likely to be positive for innovation, especially in research-intensive tasks, cultural diversity's effects are much harder to predict.

The overall empirical evidence here is positive, though not uniformly so. At organisation level, several recent studies link workforce diversity and innovation in knowledge-intensive environments. Parrotta et al (2011) find positive effects of workforce cognitive and cultural diversity on Danish firms' patenting rates. Studying London firms, Nathan and Lee (2011) find that both management and workforce diversity help raise product and process innovation. However Ozgen et al (2011) find weaker links between cultural diversity and product/process innovation in 'white collar' Dutch firms. Maré et al (2011) find no systematic links between workforce characteristics and innovation among businesses in New Zealand.

More broadly, reviews of organisational and management literature find a small but significant workplace 'diversity advantage' on measures of business performance. Negative communication and trust effects are present in the short term but progressively decline (Landry and Wood 2008).

3.4 Urban effects

We might observe bigger co-ethnicity and diversity effects on innovative activity in cities because of population mix, agglomeration economies or both. Innovative activity, migrant and minority communities tend to be spatially clustered in urban areas. Kerr (2008) finds that US ethnic inventors are spatially concentrated, largely in the biggest urban agglomerations.

Urban areas may also have positive or negative 'amplifying' effects. For example, if cultural diversity contributes to economic diversity, it may help foster knowledge spillovers across sectors at urban level (Jacobs 1969). Jacobs also argues that cities accelerate innovation by fostering the recycling and recombination of existing products and ideas into new forms. The more cosmopolitan the urban population, the greater the potential for hybridisation (Hall 1998; Gordon, Whitehead et al. 2007). Conversely, members of minority communities may be physically isolated in particular urban neighbourhoods. Spatial segregation may limit the opportunity for knowledge spillovers and interaction with other groups (Zenou 2011).

A number of US and European studies suggest a link between area level diversity and innovative activity, although none look at the UK case. Peri (2007) finds that US states' share of foreign-born PhDs is positively associated with levels of patenting. Hunt and Gauthier-Loiselle (2008) find that immigrant population shares raise state-level patenting, and that these effects are greater than individual-level effects – suggesting urban, group and individual-level dynamics are all in play. Kerr and Lincoln (2010) use shifts in US visa quotas to identify effects of immigrant scientists on patenting in US cities, suggesting positive effects of skilled migrants on both 'ethnic' and overall innovative activity at urban level. Ozgen et al (2010), studying EU NUTS2 regions, find positive connections between migration, immigrant diversity and regional patenting. Niebuhr (2006) finds a positive link between the diversity of German regions and regional innovation, especially for highly skilled employees.

4. Data and identification strategy

I have three main data sources for the analysis. Patents information comes from the European Patent Office (EPO), which is made available through the OECD PATSTAT database.[2] Raw patent data cannot typically be used at inventor level, because of common/misspelled names, or changes of address: I use a cleaned form of the data provided by the KITES team at Bocconi University, which allows robust identification of individual UK-resident inventors (Lissoni, Tarasconi et al. 2006).[3] Ethnicity information is then derived from inventor names using the ONOMAP name classification system (see below). Finally, I combine this individual-level information with area-level controls, assembled from UK Labour Force Survey held in the Office of National Statistics Virtual Microdata Lab. My data assembly strategy builds on pioneering US studies of inventor activity by Kerr (2008; 2008; 2009; 2010), but makes important adaptations to the UK case. This is because of a number of methodological challenges linked to both the patents and diversity data, which are dealt with briefly below.

[2] In full: EPO Worldwide Patent Statistical Database.

[3] Microdata from the PATSTAT-KITES database (http://db.kites.unibocconi.it/). For details of the algorithmic cleaning of the raw data, see Lissoni et al (2006).

4.1 Working with patents data

The raw patents data covers the period 1977-2010, dated by priority year.[4] The dataset contains geocoded information on 141,267 unique British-resident inventors and 131,738 patents with at least one British-resident inventor.[5] During this time the UK experienced generally low levels of immigration (from the late 1970s to the mid-90s), followed by an upshift from the late 1990s onwards (Wadsworth 2010).

I make a number of changes to the patents data to make it fit for purpose. First, there is typically a lag between applying for a patent and its being granted. This means that in a panel of patents, missing values typically appear in final periods. Following Hall et al (2001), I truncate the dataset by three years to end in 2004.

Second, innovation and invention are processes, not events. Inventors typically work on an invention for some time before filing a patent. This means that year-on-year variations in patenting will not be driven simply by year-on-year variation in the things that drive innovation. In principle, the simplest way of dealing with this issue is to lag independent and control variables. However, it is not obvious *a priori* which length of lag should be fitted and there is also the problem that current drivers may still *partly* explain current patenting levels, even if other factors act with a lag.

I therefore follow the alternative approach of Menon (2009) and group patent observations together, using mean citation lags to specify the appropriate interval. If patent B cites patent A, the 'citation lag' between the two is the time period between the filing of A and the filing of B: the lag offers a rough way to capture the relevant external conditions affecting patenting. The mean citation lag for EPO patents is four years (Harhoff et al 2006, in OECD, 2009), so I group patents into four-year periods or 'yeargroups'. I organise independent variables and controls along the same lines (except for areas' historic patent stocks, where lags are straightforward to apply).

[4] 'Priority dates' represent the first date the patent application was filed anywhere in the world. The OECD recommends using priority years as the closest to the actual time of invention (OECD 2009).
[5] The full dataset has 160,929 unique UK-resident inventors: 19,492 observations lack postcode information. In total 201,016 inventors are attached to these patents, indicating significant co-patenting.

Third, the main analysis uses unweighted patent counts to measure 'inventor productivity', that is, the number of times an inventor engages in patenting activity in a given time period. Some of the extensions and robustness checks are done at area level. In this case I use weighted patent counts to avoid double-counting innovative activity: raw counts are divided by the number of inventors involved (OECD 2009). For clarity, henceforth all patent counts are unweighted unless stated otherwise.

Finally, I use a combination of technology field dummies and area-level industrial structure controls to control for structural biases in patenting activity across different industrial sectors. These are described further in section 6.

4.2 Identifying ethnic inventors

I use the ONOMAP name classification system to generate ethnicity information for individual inventors. ONOMAP was originally designed for mining NHS patient data and classifies individuals according to most likely cultural, ethnic and linguistic (CEL) characteristics identified from forenames, surnames and forename-surname combinations.[6]

ONOMAP is built from a very large names database drawn from UK Electoral Registers plus a number of other contemporary and historical sources, covering 500,000 forenames and a million surnames across 28 countries (Mateos et al 2007). These are then algorithmically grouped together, combining information on geographical area, religion, language and language family. Separate classifications of surnames, forenames and surname-forename combinations are produced. This gives 185 basic CEL categories, which can be aggregated at different levels of detail, broken down into constituent parts (such as likely religion and language) and crosswalked onto other classifications (such as ONS ethnic groups).[7]

[6] For a brief summary see http://www.onomap.org/FAQ.aspx.

[7] Names information is drawn from 1998 and 2004 GB Electoral Registers, Northern Ireland Electoral Register 2003, Irish Electoral Register 2003, plus electoral data from Australia (2002), NZ (2002), United States (1997) and Canada (1996). Experian MOSAIC geo-demographic data and the Experian Consumer Dynamics datafile are used to boost the sample. This produces 25360 surnames and 299797 first names. These are classified using a combination of triage, spatio-temporal analysis, geo-demographic analysis, text mining, 'name-to-ethnicity' techniques from population registers and researching international name frequencies. 'British names' are taken as those originating in the British Isles (including Ireland) or arriving there before 1700. For full details see Mateos et al (2007).

ONOMAP exploits similarities and differences between name families – so that 'John Smith' is more likely to be ethnically British than French:

> *Each name … [is] assigned an Onomap type (the lowest level in the classification) together with a probability score that summarises the likelihood of a particular name belonging to such a type. Such probability score is derived from the share of the population with that (fore/sur)name that also has a (sur/fore)name belonging to the same Onomap type. When classifying a list of names, the Onomap software assesses both components of a person's name (forename and surname). In cases of conflict between … forename and surname it assigns the Onomap type with the highest probability score.* Lahka et al (2011), p3

Because ONOMAP uses surname *and* forename information, it is able to deal with many names with multiple cultural origins; the historically fuzzy boundaries of many states (e.g. Germany and the Netherlands), and the alteration and/or adoption of names traditional to the UK.[8] Like Kerr's similar work on US patents data (Kerr 2008), ONOMAP has the drawback of only observing objective characteristics of identity – the most conservative interpretation is that it provides information on *most likely* ethnicity. However, unlike the MELISSA commercial database used by Kerr, which only identifies high-level ethnicities, the ONOMAP system allows me to examine inventor characteristics from several angles and at several levels of detail. ONOMAP also matches 99% of inventor names (compared with Kerr's 92-98% success rates).

For the descriptive analysis I exploit the full range of CEL information, as well as ONS ethnic groups and geographical origin. For the regressions, I use ONS ethnic groups and geographical origin only. This is because it is not possible to use the CEL typology in the controls, which would leave me unable to explore the influence of area-level demographic characteristics on inventor characteristics.

---

[8] The author's name is one of the more challenging to classify. According to Mateos (by email), 'Nathan is unclassified at the moment in Onomap, perhaps because there are conflicting frequencies in India, New Zealand and the UK. "Max" is classified as "Jewish", probably because it is common in this community in the UK compared to the national average. Therefore you would be classified as 'Jewish'." This is a good proxy for my actual British/English/secular Jewish sense of self.

ONS ethnic group information is based on the nine categories developed for the 1991 Census. These are relatively dated and lose some important detail – for example, the second largest inventor group after 'white' is 'other' – so are likely to be subject to some measurement error.[9]

Geographical origin information provides finer-grained information on twelve zones across Europe, Africa, Asia and the Americas.[10] Because name information does not distinguish migrants from their descendents, I use likely geographical origin as a measure of geographical 'roots' – an important, albeit partial, aspect of ethnicity. I use this as my preferred measure of ethnicity, as geographical origin is objective and provides a greater level of detail.

Combining geography and name information in this way is not problem-free. ONOMAP does not distinguish geography if countries share a common language, so that North American and Australasian-origin inventors are largely identified as British-origin inventors (or unclassified). This may understate the true extent of inventor diversity. In practice, resulting measurement error is likely to be small. First, although the largest concentrations of these groups are in London, their spatial distribution is not very different from minority communities as a whole. Second, they represent a relatively small share of the UK's minority population. I use the LFS to explore the prevalence of American, Canadian, Australian and New Zealand migrants. In 1994 these groups comprised just 8.84% of migrants, falling to 7.98% in 2004.

To measure diversity of ethnic groups, I use a Fractionalisation Index. For identity group *a* in area *j* in year *t*, the Index is given by:

$$FRAC_{jt} = 1 - \sum_a [SHARE_{ajt}]^2 \quad (1)$$

Where SHARE is *a*'s share of the relevant population (here, all active inventors in a given area). The Index measures the probability that two individuals in an area come from different

[9] The full set of ONS 1991 groups is White, Black Caribbean, Black African, Indian, Pakistani, Bangladeshi, Chinese and Other.
[10] The full set of twelve geographical origin zones is Africa, Americas, British Isles, Central Asia, Central Europe, East Asia, Eastern Europe, Middle East, Northern Europe, South Asia, Southern Europe and Rest of the World.

geographical origin or ethnic groups. Similar measures are used widely in the development literature, as well as some city and state-level studies (Easterley and Levine 1997, Alesina and La Ferrara 2004, Ottaviano and Peri 2005, 2006).

5. Data assembly and descriptive analysis

I assemble a panel of UK-resident inventors' patenting activity between 1993 and 2004 inclusive, dividing the time period into three four-year 'yeargroups' as explained in the previous section. Each inventor-yeargroup cell records how many times an inventor patents in that time period. After cleaning, the basic panel covers 125,502 inventors across three four-year yeargroups, giving 376,506 observations. Cell counts vary from zero to 36, with a mean of 0.318.[11]

I use postcode information to locate inventors in UK Travel to Work Areas (TTWAs), which are good approximations of local economies (and superior to administrative units such as local authority districts).[12] Matching is done by postcode sector, which minimises the number of observations lost through incomplete or mistyped postcode information.[13] I then fit an urban / rural typology of TTWAs developed in Gibbons et al (2011), allowing me to explore the potential effects of urban environments (see Appendix C for details and maps).

Working with inventors (rather than patents or applicants) presents three linked areas where measurement error may arise. The first issue is robustly identifying individuals. I minimise this risk by using appropriately cleaned data. The second issue is about inventor activity. Inventors are only visible when patenting, and we do not know for certain what they are doing the rest of the time. The most conservative solution is to blank all cells where the inventor is not active. However, as most inventors – in the UK and elsewhere – patent only once, this would radically reduce sample size (and would miss instances where inventors were constrained from patenting for some reason). For the main analysis I thus zero all cells

[11] Just over 39% of inventors invent pre-1993, but do not invent during 1993-2004.

[12] TTWAs are designed to cover largely self-contained labour markets: 75% of those living in a given TTWA also work in the TTWA, and vice versa. TTWAs are thus a good approximation for local spatial economies and for city regions (Robson et al 2006).

[13] Matching on full postcodes drops around 12% of observations. Matching on postcode sector drops 5.77% of observations. I exclude information on inventors resident in Northern Ireland. A small number of postcode sectors cross TTWA boundaries, so matching is not perfect.

when no inventor activity is recorded. Using a sub-sample of inventors, I run robustness checks comparing both ‘zeroed’ and ‘blanked’ approaches. I find sample construction has no effect on the results (see Section 8).

The third issue is about inventor location. We cannot be sure where inventors are when they are not actively patenting; and we need to identify those inventors who have moved location. I explore this issue by identifying likely movers. Following Agrawal et al (2006), I define movers as inventors with the same forename and surname, who patent in the same technology fields, in different TTWAs, at different points in time. As Agrawal and colleagues point out, this strategy minimises the risk of false positive errors – identifying inventors who are movers who are not – but does not deal with false negatives (identifying movers as non-movers). Measurement error from the latter is random, so will reduce the precision of, but not bias, my main results. The conservative estimates that result suggest around 14% of the sample are likely movers. This suggests firstly that the vast majority of inventors do not move during the sample period; and therefore it is reasonable to count non-movers as present in the same TTWA in which they first patent.

5.1 Descriptive statistics

Some basic descriptives are set out in Tables 1-8, along with some wider population data from the Labour Force Survey.

Table 1 breaks down inventors by CEL subgroup, showing the 30 largest groups. Because CEL classifications are not available in the LFS, I do not present comparison data for the wider population here (although see my first paper for some simple area-level analysis). We can see that while English, Welsh, Scottish and Celtic[14] inventors make up the bulk of the sample, other inventor groups divide fairly evenly into geographically proximate communities (e.g. Irish, plus a series of European groups), groups reflecting the UK’s colonial history in South and East Asia (e.g. Indian Hindi, Sikh, Pakistani, Hong Kong Chinese) plus some largely recent migrant communities (e.g. Polish, Vietnamese).

[14] ‘Celtic’ denotes names common to Scottish, Welsh and Irish CEL types.

Tables 2 and 3 recut the sample by probable geographical origin zones and by 1991 ONS ethnic groups. Geographical origin zones (Table 2) allow me to preserve some of the detail from the full CEL classification, including several areas of Europe as well as South and East Asia. As highlighted in the previous section, ONS ethnic groups (Table 3) are much less flexible, focusing on visible majorities and minorities, relegating the rest of the inventors to 'other'.

Tables 4 – 6 cut the sample geographically. Table 4 presents the 40 Travel to Work Areas with the largest shares of ethnic inventors by geographical origin, and for comparison provides migrant shares in the wider TTWA working-age population. High-ranking TTWAs are predominantly urban, although a number of rural areas also feature, predominantly university towns (St Andrews, Lancaster, Canterbury) or areas adjoining TTWAs with universities (Bude and Holsworthy) and/or manufacturing clusters (Holyhead, Pembroke and Tenby, Louth and Horncastle).[15] Comparing ethnic inventors with migrants in the overall population, we can see that areas in the top half of the table mostly have bigger shares of ethnic inventors than in the wider working-age population – London is one notable example. Table 5 presents the same data as location quotients, confirming that ethnic inventors are more spatially clustered than the wider migrant population.

Table 6 compares Fractionalisation Index scores for active inventors and wider working age populations. The cultural diversity of inventors is greater than that of the wider population in most TTWAs (London, Bradford, Birmingham, Brighton, Leicester and Reading are the six exceptions in the top 40). Again, there are a number of rural areas in the table. As some rural areas have fairly few inventors, a small sample may lead to high values of the Fractionalisation Index.

Finally, Table 7 gives weighted counts for the 40 TTWAs with the highest patenting activity: to minimise double counting, I weight each patent by the number of inventors involved. The results follow the familiar geography of UK innovative activity. A number of these high-patenting areas also have large ethnic inventor shares and diverse inventor groups (for example London, Southampton, Crawley, Oxford and Cambridge). However, another

[15] Many inventors will work in professional / technical occupations, which are characterised by longer-than-average commuting distances. Building commuting zones on the basis of these workers' commuting patterns substantially reduces the total number of zones (Robson et al 2006), suggesting that commuting across conventional TTWAs is not uncommon.

group of high-patenting TTWAs have rather more homogenous inventor and general populations (for example, Bristol, Manchester, Reading and Ipswich).

A number of broad lessons emerge from the descriptives. First, the UK's population of ethnic inventors appears substantially different from that of the US. American ethnic inventor communities are dominated by South and East Asian groups (Kerr 2008). By contrast, the UK has a number of European groups, with South Asian and East Asian inventors drawn in large part from former colonies. Second, as in the US ethnic inventors are spatially concentrated, and more clustered than minority populations in general. Third, not all high-patenting locations have large ethnic inventor shares or diverse inventor communities.

6. Regression analysis: estimation strategy

I now explore whether these inventor, group and area-level characteristics influence individual inventor productivity. The descriptives highlight the distinctive composition of UK ethnic inventors, as well as their spatial concentration. I therefore use the data to estimate a modified knowledge production function, linking counts of patenting activity to individual, group and area characteristics. I use aggregated LFS client file microdata to construct a range controls. As LFS microdata is only provided with local administrative district-level identifiers, I aggregate to TTWA level using a postcode weighting system.[16] Summary statistics for the 12-year panel are given in Table 8.

For inventor *i* in area *j* and yeargroup *t*, I estimate:

$$\mathrm{PCOUNT}_{ijt} = a\mathrm{INV}_i + b\mathrm{DIV}_{jt} + \mathbf{CONTROLS}_{\mathbf{jt}}\boldsymbol{c} + \mathrm{P}_i + \mathrm{U}_j + \mathrm{YG}_t + \mathrm{e}_i \qquad (2)$$

[16] I aggregate individual-level data to local authority-level averages, and then aggregate these to TTWA-level using postcode shares. Local Authority District (LAD) boundaries are not congruent with TTWA boundaries, so straightforward aggregation is not possible. Using the November 2008 National Postcode Sector Database (NSPD), I calculate the number of postcodes in each 2001 TTWA and in each of its constituent LADs. For each TTWA, I then calculate constituent LADs' 'postcode shares'. Shares sum to one, and are used as weights to construct TTWA-level averages. *Example:* suppose a TTWA consists of parts of three LADs. The TTWA has 100 postcodes, 60 of which are in LAD_a, 30 in LAD_b and 10 in LAD_c. The relevant LAD weights are 0.6, 0.3 and 0.1 respectively. The TTWA-level average of variable *x* is given by (*x*)TTWA = 0.6*(*x*)_a + 0.3*(*x*)_b + 0.1*(*x*)_c.

Where PCOUNT is a simple count of the number of times an inventor engages in patenting during a given four-year period. My first variable of interest is INV, a dummy variable taking the value one if the inventor is a likely ethnic inventor. (I later extend the model replacing INV with a set of dummies for various co-ethnic groups.) My second key variable is DIV, the diversity of active inventors in a given TTWA and time period. DIV is given by the Fractionalisation Index in Section 4.

CONTROLS represents a vector of largely TTWA-level controls covering key spatial, economic, and demographic characteristics affecting relationships between INV and innovation, DIV and innovation or both. Unless otherwise stated, all controls are for the same 1993 – 2004 period as the patent data.

For example, innovative activity and patenting are both spatially concentrated, reflecting benefits from agglomeration that may persist over time (Simmie, Carpenter et al. 2008). Co-ethnicity or diversity effects on patenting might then simply reflect agglomeration and path-dependence. I fit a dummy for primary urban areas, U, and fit log of population density to explore agglomeration effects more broadly. I also fit the model with measures of 1981-84 area weighted patent stocks to control for historic asset effects, and experiment using different lags of the historic patent stocks control.

Inventor demographic characteristics may be entirely explained by area demographic characteristics: for example, places with more diverse populations may produce more diverse inventor groups. Failing to account for this leads to bias on DIV. I control for this by using area-level fractionalisation indices (and cross-check using migrant population shares).

Human capital stocks are closely correlated with innovative activity (Romer 1990) and as discussed in Section 3, may account for apparent ethnicity effects on patenting. Given the role of 'ethnic scientists' in the US and elsewhere, area-level human capital controls include the share of degree-holders with Science, Technology, Engineering and Mathematics (STEM) qualifications in the local working-age population. (The share of degree-holders with PhDs in any subject is used as an alternative control, as it is less precise in terms of subject.) Patent data provides very little individual-level information on human capital, but I am also able to fit P, an individual-level fixed effect explained below.

I fit various further controls for precision. Patenting is known to be higher in 'knowledge-intensive' high-tech and manufacturing sectors, so I include measures of the share of workers employed in 'knowledge-intensive' manufacturing, following The Work Foundation's definition of 'knowledge-intensive' firms (Brinkley 2008).[17] Patenting activity is also vulnerable to sector-specific shocks, and the spike in software patenting since the mid-1990s is well-covered in the literature (Li and Pai 2010). To account for this I fit a dummy for the IPC technology field 'electrical engineering and electronics'.[18] Patenting is likely to be lower in areas with a lot of entry-level jobs or areas of joblessness, so I include the share of workers in entry-level occupations and the share of long term unemployed as further controls.

6.1 Inventor fixed effects

Area-level controls for human capital may not fully account for differences in human capital between inventors. The panel data structure should allow this to be controlled through individual fixed effects (Hausman, Hall et al. 1984). However, fixed effects panel estimators for nonlinear models require observations to have a non-zero value in at least one time period (Cameron and Trivedi 2009). As I am as interested in whether or not inventors patent as the number of times they patent, such an approach is not ideal.[19]

Blundell et al (1995) develop a now widely-used[20] alternative, exploiting historic information to control for permanent unobserved differences between agents. They argue that firms' capacity to innovate is largely explained by the build-up of knowledge in the firm at the point in which it enters the sample. With long enough time series data, pre-sample activity approximates an individual fixed effect.

---

[17] This follows standard OECD definitions but adjusts for the UK context. The final list of 3-digit SIC sectors includes medium and high-tech manufacturing (pharmaceuticals, aerospace, computers and office machinery, electronic communications, software, other chemicals, non-electrical machinery, motors and transport equipment).

[18] I also experiment with a more precise information technology dummy (OST30_4), with similar results.

[19] Random effects estimators are a potential alternative strategy, but Hausman tests (chi-squared = 19979.75, pr = 0.000) suggest these are not justifiable.

[20] A Google Scholar search turns up 351 citations. Highly cited examples include Baptista and Swann (1998), Katila and Ahuja (2002), Beaudry and Breschi (2003), Dushinitsky and Knox (2005), O'Shea et al (2005) and Aghion and Howitt (2006).

For individual inventors, historic patenting activity is likely to work in a similar way. The patent data provides information on inventor activity from 1977, 16 years before the start of the regressions panel in 1993: around 23% of inventors in the sample period also invent before 1993, covering 40% of cells. I replicate this 'entry stock' estimator, using the pre-sample mean of inventors' patent counts as an approximate individual fixed effect.

I exclude inventors with no pre-sampling history – they may have been inactive or not in the labour force – and run the model on a reduced sample of 89,309 observations. The new sample removes younger inventors and recent migrants. As such it may understate true inventor diversity (or indirectly affect results if younger people are more open to diverse environments). Critically, however, the restriction allows me to distinguish ethnicity, diversity and human capital effects. I experiment with the full sample to check robustness, finding key variables and overall model fit are poor.[21]

6.2 Model specification

Count data is usually modelled using Poisson or negative binomial estimators. My panel exhibits excess zeroes (78%) and over-dispersion (the variance of PCOUNT is over 2.5 times the mean). This means the basic assumptions of the Poisson model are not met, leading to likely inefficient estimates (Greene 1994). As such, a negative binomial or zero-inflated model may be preferred.

Diagnostic tests suggest the negative binomial is the better fit, and has the added benefit of running a Poisson model as a base case.[22] Against this, Angrist and Pischke (2009) argue that once raw coefficients are converted into marginal effects, non-linear modelling offers little over standard linear regression. I therefore fit the model with both negative binomial and OLS estimators.

[21] Fundamentally, I argue the reduced sample preferable to running a bigger sample of inventors for whom historic patenting information is ambiguous. Firm-level studies, in contrast, typically have information on exactly when agents enter/exit the market.
[22] Log-likelihood tests and AIC scores. I also experiment with zero-inflated models (ZIP and ZINB). Both perform well on diagnostic tests, although interpretation is extremely complex. Results from Poisson regressions are available on request.

7. Regression analysis: results

Results from the main regressions are given in Tables 9 (negative binomial) and 10 (OLS). In each table, column 1 shows a bivariate regression for the main variables of interest only, column 2 adds controls and column 3 adds the fixed effect. For ease of interpretation and comparison with OLS models, negative binomial results are presented as marginal effects at the mean. Negative binomial models show a significant log alpha term, confirming over-dispersion. Controls are generally of the expected size and sign.

7.1 All inventors

Ethnic status and inventor group composition have no significant effect on individual inventor productivity (column 1). The coefficient of INV is close to zero and DIV is negative insignificant. When controls are added (column 2), both INV and DIV become positive. Coefficients get bigger, and in the OLS results DIV is now significant at 5%.

As explained above, the aim of the individual-level fixed effects is to control for individual inventors' human capital endowments, allowing identification of the various ethnicity channels. As expected, once the fixed effects are included (column 3) overall model fit improves and the results change substantially. INV remains insignificant but its coefficient more than doubles, for both sets of models. For negative binomial models, the marginal effect of DIV is now 0.087, significant at 5%.

Specifically, a 10-point increase in the inventor Fractionalisation Index – increasing active inventor diversity in Bristol to that in Oxford, for example – is linked to an average marginal effect of 10*(0.087) = 0.87 extra patents per inventor. For OLS models, diversity effects are slightly larger. DIV is 0.099, significant at 10%: a 10-point rise in inventor group diversity is associated with a 0.99 unit increase in expected patenting, or an extra patent per inventor. Interestingly, coefficients of *area population* diversity are negative (significant at 10% for negative binomials, not for OLS).

To put this into perspective, effects of diversity on patent counts are smaller and/or weaker than human capital, whether the latter is measured at the area level or at individual level. This fits with the existing empirical evidence that diversity effects on innovation are generally fairly small, where they exist (see Section 3). For negative binomial models, for example, the marginal effect of STEM degrees is 0.304, significant at 5%. This suggests that a 10-point increase in the area's share of science graduates is linked to 3 extra patents per inventor. This is as expected given that patenting is concentrated in science and technology sectors. The marginal effect of the individual fixed effect 0.101, significant at 1%: past patenting activity is strongly linked to current patenting rates.

Results for ONS ethnic groups function as a basic cross-check (Table 11). These broadly confirm the main findings. For negative binomial models, INV remains close to zero throughout; with controls and fixed effects the marginal effect of ethnic DIV is 0.125, significant at 5%. For OLS models, coefficient sizes and magnitudes are similar but none of the results is significant.

Table 12 shows results from three initial robustness checks. First, I fit the TTWA share of degree holders with PhDs in any subject as an alternative area-level human capital control (column 2). PhDs are a prerequisite in many research positions, and as specialists, PhD-holders may be more likely to patent. I find that an area's share of PHDs strongly positively associated with inventor productivity, and dominates DIV in both model specifications. One interpretation of this result is that places that are attractive to PHDs also attract a diverse group of inventors, due to some other factor – such as a 'tolerant' milieu as suggested by Florida (2002).

An alternative explanation is that high-patenting PHDs are themselves ethnic inventors, as suggested by US studies on star scientists (Stephan and Levin 2001; Chellaraj, Maskus et al. 2005). In this case, diversity is the fundamental driver and the PhD variable is a so-called 'bad control' (Angrist and Pischke 2009). As discussed in section 3, one then needs to disentangle the ethnic and human capital components of stars' performance. I am unable to observe whether or not inventors have PHDs, so am unable to make these checks. Further research is needed here, perhaps with a subset of inventors in academic institutions where PHDs are more or less essential. I continue to focus on diversity because this is my main interest. But the results when including the PHD variable urge caution in interpreting these results as purely causal (of course, this is not the only identification challenge, as discussed further below).

Second, I fit the model with a lagged dependent variable to control for effects of past patenting within the sample (column 3). Diversity effects persist: coefficients are now rather smaller but also more precise, with DIV significant at 1% (negative binomial) and 5% (OLS). Third, I fit the model without London – a city with high levels of cultural diversity and relatively low levels of patenting per head of population (Wilson 2007).[23] Results, in column 4, show that diversity effects persist in the negative binomial specification (significant at 5%), but are insignificant in OLS.

Overall, the main results suggest no significant effect of ethnic inventor status on inventor productivity relative to other inventors, once individuals' human capital and area conditions are accounted for. However, the composition of the inventor group matters: more diverse inventor communities have a small positive effect on individual inventor productivity. The rest of this section examines other channels –urban location and co-ethnicity – in more detail.

7.2 Urban areas and urban inventors

The evidence review (Section 3) suggests that urban areas may 'amplify' ethnicity-innovation processes via population composition effects, agglomeration effects or a combination of the two. However, the main results (Tables 9 and 10) find a weakly negative relationship between urban TTWAs and inventor productivity. In the negative binomial, for example, the marginal effect the urban TTWA dummy is -0.021, significant at 10%; in the OLS results the coefficient is not significant and is close to zero. By contrast the agglomeration control, log population density, is positive at 0.0005 in the negative binomial specification, 0.008 in OLS, although neither is significant.

In order to identify the separate effects of urban location and urban density, I fit the two separately and then interact them. The pairwise correlation between the urban TTWA dummy and log population density is 0.565, suggesting some differences in urban characteristics. Results are given in Table 13. Column 2 includes urban TTWA dummies only, column 3 log population density only, column 4 an interaction effect. We can see that fitted separately, each is negative on inventor productivity (although marginally significant at

[23] Although London has relatively *high* patenting per inventor – see Table 7.

best). Fitted together, each is positive – with a negative interaction effect, suggesting some diseconomies of agglomeration on inventor productivity in the largest conurbations.

Columns 5-7 explore specific effects of diverse urban areas. Column 5 interacts the Fractionalisation Index with the urban TTWA dummy. The coefficient of DIV is now higher (0.136, significant at 5%) but the interaction term is negative insignificant at -0.066. Column 6 repeats the exercise with population density. DIV is now much larger (0.284), but is insignificant with large standard errors: the interaction term is also negative insignificant. Finally, column 7 includes both urban variables and interacts the Fractionalisation Index with population density. DIV is now very large and significant, but noisy: the interaction term is negative and marginally significant.

Taken together, these results suggest that agglomeration is helpful for inventor productivity, although has some diseconomies in bigger urban areas. Diverse urban areas do not seem to amplify inventor productivity, however. Overall, I find a weak effect of urban areas on inventor productivity, which is perhaps surprising given the emphasis on geographical proximity in the innovation literature. The UK context helps explain the discrepancy. Raw patent counts are highest in relatively small cities, notably Oxford and Cambridge. Conurbations, particularly London, are dominated by service sector activities where patenting is less likely to occur. The next chapter explores the London experience in more detail, using survey data which captures a broader range of innovative activity.

7.3 Co-ethnicity / diaspora effects

The data also allows me to explore co-ethnic / diasporic group effects. Specifically, rather than estimating INV as a single ‘ethnic inventor’ dummy, I now include a series of dummies taking the value one if the inventor is a member of each geographical origin zone. I run the model for all minority co-ethnic groups, taking UK-origin as the reference category. Results for negative binomial models are given in Table 14: for simplicity I restrict my analysis to the five biggest geographical origin zones (South Asia, Central Europe, East Asia, Southern Europe and Eastern Europe). Results are interpreted as the marginal effect of being in one of these co-ethnic groups, relative to membership of the majority group.

I find significant positive effects of South Asian- and Southern European–origin inventors on expected patenting rates, and negative significant effects of East Asian-origin inventors, relative to UK-origin inventors. Specifically, marginal effects are 0.025 for South Asian inventors, significant at 10%, -0.037 (1%) for East Asian inventors; and 0.053(10%) for Southern European inventors. The South Asian result is intuitively plausible given the strong historic connections between the UK and South Asian countries (India, Pakistan, and Bangladesh) and the presence of large migrant and established minority communities here. It also accords with US research showing significant diaspora effects of Indo-American communities. The Southern European result is likely to reflect the relatively large shares of inventors in the UK with Spanish, Italian or Portuguese backgrounds (Table 1).

The East Asian result is in stark contrast to US research showing strong diaspora effects for Chinese and Taiwanese communities (Saxenian 2006; Dahlman 2010). This may reflect the lack of strong diasporas in the UK outside Hong Kong-origin Chinese, and the different circumstances behind recent community formation in the US (economic migration of skilled workers) and the UK (handover of Hong Kong to China between 1984 and 1997).

Results may also be driven by the large geographical origin zones I am using to proxy diasporic communities. I experiment with ONS ethnicity measures of Indian and Chinese inventors to conduct a partial cross-check using more tightly-defined groups, confirming my main result.[24] Overall, then, these results suggest that co-ethnic group membership, as well as the diversity of the local inventor community, both have small positive effects on individual patenting rates.

8. Further robustness checks

I conduct checks on a series of potential endogeneity problems. These fall into two broad categories: robustly identifying diaspora and diversity channels, and dealing with path-dependence. Results are shown in Tables 15 and 16.

[24] Indian inventors make up just over three quarters of South Asian inventors (see Table 9), so I also break down the South Asian result in more detail. I find a positive non-significant link between Pakistani inventors and inventor productivity, but a very strong negative link with Bangladeshi inventors. Given their small representation in the sample, this may be largely explained by measurement error.

8.1 Identifying human capital, diversity and diaspora effects

I face two immediate identification challenges. First, the combination of area-level controls and individual fixed effects may not be fully capturing inventors' human capital. Assuming that human capital has a positive effect on patenting, the resulting omitted variable bias will overstate effects of DIV, pushing coefficients of DIV upwards.

To explore, I include an alternative fixed effect in the main model, again exploiting pre-sample information. Alongside overall output, intellectual range is another plausible indicator of overall human capital. My original fixed effect measures knowledge accumulation by summing pre-sample patents. In addition, I identify 'generalists' as inventors patenting across at least two technology fields (for example, filing patents in both electronics and biotechnology). The fixed effect is a dummy with value one if an inventor patents across technology fields in the pre-sample period.[25]

Results are given in Table 15. Columns 1-3 compare the original fixed effect, the 'generalist' fixed effect and both together. INV remains insignificant throughout; marginal effects of DIV fall from 0.087 to 0.05, 10% significance with the generalist fixed effect (column 2). Fitting both fixed effects together (column 3) slightly increases the size and strength of the DIV marginal effect (to 0.055, 5% significance) and improves model fit. Columns 4-5 rerun this model for co-ethnic groups: with both fixed effects in play, the main co-ethnic group effects remain significant albeit smaller.

Second, inventor diversity effects might collapse to simple size effects. Fractionalisation Indices tend to be highly correlated with group population shares (in this case, the pairwise correlation between DIV and the share of non-UK origin inventors in the TTWA is 0.8039). To test this, I replace the Fractionalisation Index of inventors with the share of ethnic inventors in the local inventor population. Results, in Table 16, show that the coefficient on ethnic inventor share is similar to group diversity, but is not significant on individuals' expected patent rates either when fitted individually (column 2) or with DIV (column 3). Interacting the two raises the marginal effect of DIV, which stays significant at 5%, but with a large negative value for the interaction term (column 4). This suggests that the

[25] The dummy will also be capturing the minority of inventors who patent more than once.

overall diversity of inventors, rather than an aggregation of ethnic inventors, drives the main results. Column 5 repeats the analysis for diasporic groups, with similar outcomes.

8.2 Historic patent stocks / path-dependence

As explained in section 6, innovative activity is spatially concentrated, and these concentrations tend to persist over time as inventors and firms select into innovative locations, as areas progressively build innovative 'capacity'. If the historic patent stocks term in the main model is mis-specified, agglomeration and path-dependence will not be adequately controlled for. To test for this I plug a range of pre-sample historic patent counts into the main model.

Negative binomial results are given in Table 17. I find as that as the historic lag decreases, the coefficient and significance of historic patenting activity rises (from -0.000 for 1981-84 to 0.001 for 1993-96, significant at 5%). The marginal effect of inventor diversity get smaller and weaker as the historic lag shortens – from 0.087, significant at 5%, for 1981-84 stocks to 0.067 (10%) for 1989-92 stocks. This suggests that historic area-level characteristics help explain some of the diversity effect – but do not eliminate it.

8.3 Sample construction

I construct my sample by zeroing all inventor-yeargroup cells when an inventor is not patenting. As discussed in Section 5, this is not the most conservative way of treating inventors when they are not active, and there is some risk it may introduce measurement error into the results. To check for this I compare results from two samples – one with zeroed observations and one with non-active periods set as missing observations.

My identification strategy depends on using inventors' historic patenting activity, so blanking out non-activity has the effect of restricting the sample to inventors who patent more than once. I thus compare estimates for the set of multiple inventors across two different samples, one with zeroed and one with missing observations for non-activity. Results are given in Table 18. We can see that estimates for the two sub-samples are identical, suggesting that sample construction has no effect on my main results.

Overall, the results from these cross-checks suggest that my main results are robust to the main endogeneity challenges: omitted variables, path-dependence and sample construction issues. However, further research is required to identify the relative contribution of majority and ethnic PHDs to patenting.

9. Impacts on majority groups

The analysis has established some positive connections between inventor group composition, the presence of diasporic groups and individual inventor productivity. However, this has ignored distributional effects – that is, specific impacts of ethnic inventors on majority inventors. Given that immigration is a major driver of cultural diversity, it is important to look at these distributional impacts.

A number of studies in the immigration literature look at 'native outflows', in which UK-born physically leave an area after migrants arrive (Borjas 1994). 'Geographical crowd-out' of this kind is hard to assess here – as explained in section 5, although the number of mobile inventors seems low, movers cannot be definitively identified. I conduct exploratory logit regressions to identify individual and area-level factors which might influence mover status. Results suggest individual human capital (measured by the fixed effect) has a substantial, significant positive link to mover status. By contrast, coefficients for areas' share of migrant inventors are much smaller and statistically insignificant.

'Resource crowd-out' is a potentially more serious issue. There are two ways in which this might happen. First, the presence of ethnic inventors might affect majority patenting rates at the individual level. A given majority inventor may benefit from ethnic inventors via the production complementarities outlined in section 3, or may 'lose' from disbenefits such as lower trust or communications difficulties. The balance of these two effects on the average majority inventor needs to be identified.

Second, even if there are human capital externalities at the group level, majority individuals may lose out from the presence of minority inventors (Borjas 2011). In this case, ethnic inventors might crowd out majority inventors from relevant jobs and resources, such as space in R&D labs; or diaspora benefits might only be accessible to group members. This will affect the composition of overall patenting at area level. At the extreme, increases in area-level patent counts might be partly or wholly explained by a rising share of 'ethnic' patents – majority patenting shares could be static or even falling. Conversely, there might be multiplier effects from ethnic to majority group inventors, raising everyone's patent counts.

I test for both forms of resource crowd-out. At the individual level, I first re-run model (1) for majority inventors only. Results are given in the first panel of Table 18. The marginal effect of DIV on majority inventor productivity is 0.072, significant at 10%. This implies a positive multiplier effect of inventor diversity on majority groups – but it is smaller and weaker than on all inventors.

Next, I run model (1) for the whole sample but fit INV as a majority inventor dummy. Results are given in the second panel of Table 18. As with minority status, majority status has no significant effect on inventor productivity when other factors are controlled for (columns 1 and 2). However, interacting majority status with inventor diversity produces a positive significant effect of majority status, a larger and stronger effect of diversity – but a significant negative effect on majority inventors in diverse areas (column 3). Unlike the previous test, this suggests that while inventor diversity brings benefits, majority inventors in diverse inventor communities lose out.

To explore area-level effects, I draw on recent work by Card (2005), Kerr and Lincoln (2010) and Faggio and Overman (2011). I assemble a panel of TTWA-level weighted patent counts for 1993-2004. I define 'ethnic' patents as patents with at least one ethnic inventor; all other patents are 'majority' patents. Following Faggio and Overman (2011), I then regress the percentage change in total weighted patents during the period on the percentage change in ethnic patents. For TTWA *j* I estimate:

$$\Delta \mathrm{TPATENTS}_j = a + b\Delta \mathrm{EPATENTS}_j + \mathbf{CONTROLS}\boldsymbol{c}_{\mathrm{jtbase}} + e_j \qquad (3)$$

Where:

$$\Delta \mathrm{TPATENTS}_j = \mathrm{TPATENTS}_{j2004} - \mathrm{TPATENTS}_{j1993} / \mathrm{TPATENTS}_{j1993} \qquad (4)$$

And $\Delta EPATENTS_j$ is assembled similarly. **CONTROLS** is a vector of area-level controls for the base period 1993.[26] The coefficient of interest is *b*. As explained by Card (2005), if estimates of *b* are less than one, increases in ethnic patenting lead to a smaller increase in overall patenting, implying some crowd-out of majority patenting by ethnic inventors. Estimates of *b* larger than one imply multiplier effects; if *b* is equal to one, there are no distributional impacts either way.

OLS results are given in Table 19. The simplest specifications of (4) suggest some crowd-out, with *b* estimated at 0.199 and 0.259, significant at 1%. However, *b* becomes insignificant once controls and standard errors clustered on TTWAs are introduced (column 4). An alternative specification using shifts in TTWAs' technology field shares delivers very similar results (column 5). This suggests there is little evidence of crowd-out.

Model (4) does not fully control for simultaneity or reverse causality. I experiment with lags of ethnic patents as an instrument, but none pass the required first-stage tests. Results should therefore be interpreted with caution.

10. Conclusions

In recent years there has been growing academic and policy interest in links between immigration, ethnic diversity and innovation. This paper looks at the role of ethnic inventors on innovative activity in the UK, using a new 12-year panel of patents microdata. I have been able to explore a number of potential 'ethnicity-innovation' channels – individual positive selection, externalities from diasporic groups and from the cultural diversity of inventor communities, as well as 'amplifying' effects of urban environments. The research is one of very few studies to explore these links, and as far as I am aware is the first outside the US.

The results suggest that individual minority status has no significant effect on inventor patenting rates once other factors are controlled for. Conversely some diasporic groups, and group cultural composition, have small positive effects on inventor productivity. Effects on

[26] Log of population density, % STEM degree, % employed in knowledge-intensive manufacturing, % migrant working-age population, % entry-level occupations, % long term unemployed, urban dummy. Alternative specifications control for TTWA change in OST7 technology field shares 1993-2004.

‘majority’ inventors are unclear: there are some indications of individual-level crowd-out, but not at area level. Although patenting activity is very spatially clustered in the UK, in contrast to the wider literature, I find little evidence that urban environments improve individuals’ patenting activity once other individual and area-level controls are taken into account.

Overall, ethnic inventors are a net positive for patenting in the UK, although the British experience is significantly different from the US. This partly reflects distinctive patterns of US migrant settlement: most notably, the recent emergence of ethnic inventor communities from Cold War science research, which have attracted very large numbers of skilled workers into a small number of locations (Saxenian 2006). By contrast, recent ‘calls’ for migrant workers in the UK since the mid-20th century have been largely focused on less skilled occupations, although policy is now becoming more skill-biased. Results may also reflect culturally distinctive US attitudes to entrepreneurship, as evidenced by sociological studies of Jewish and Afro-Caribbean migrant communities in New York and London (Gordon, Whitehead et al. 2007), and by the complex interplay between class, skills, resources and attitudes that influence real-world entrepreneurial behaviour (Basu 2002).

There are three important caveats to these results. First, diversity and diaspora effects are relatively small – human capital and patent field / industry effects are more important determinants of inventors’ productivity. This is intuitive, and echoes much of the existing literature (see above). Second, working with inventor data presents a number of potential measurement error challenges. Most seriously, my data only allows a fuzzy identification of ethnic inventors and diasporic groups. Using geographical origin as a proxy for co-ethnicity also presents conceptual challenges, although cross-checks support my results. Third, although the results survive a number of robustness checks, alternative measures of area-level human capital weaken effects of DIV. Further work is needed on the relative contribution of majority and ethnic PHDs to patenting. Conversely, data restrictions mean that my sample understates the true numbers of ethnic inventors. The real benefits of ethnic inventors may thus be larger.

The results may have implications for the current Coalition government’s migration policies. Net immigration is one of the main factors behind the growth of ethnic inventor communities in the UK: a phenomenon which appears to raise rates of innovation through a combination of diversity and diaspora effects, with no hard evidence of negative distributional effects on native inventors. A migration cap that places restrictions on skilled immigration from outside Europe is likely to put some constraints on innovative activity, leading to welfare losses both to the UK and to UK-born workers. Similar welfare losses may arise from proposed restrictions on post-study routes to work for non-EU students.

The paper leaves a number of questions for future research. Further work could explore social networks, co-ethnicity and geographical location in more detail – via analysis of patent citations and international co-invention / co-patenting. Within the UK, data offering better identification of ethnic and migrant inventors, in particular recent immigrants, would provide a clearer picture of current developments. Alternatively, qualitative methods could shine further light on migrant and diaspora dynamics. Further work could also examine sectoral and area differences, as well as distributional impacts in more detail.

Table 1. UK-resident inventors: 30 biggest CEL subgroups, 1993-2004.

CEL subgroup	**Freq.**	**%**	**Cumulative %**
ENGLISH	86,118	69.17	69.17
CELTIC	10,653	8.56	77.73
SCOTTISH	6,557	5.27	82.99
IRISH	3,583	2.88	85.87
WELSH	2,523	2.03	87.9
INDIAN HINDI	1,255	1.01	88.91
GERMAN	1,205	0.97	89.87
ITALIAN	975	0.78	90.66
FRENCH	958	0.77	91.43
CHINESE	920	0.74	92.16
POLISH	886	0.71	92.88
OTHER MUSLIM	793	0.64	93.51
OTHER EUROPEAN	665	0.53	94.05
HONG KONGESE	588	0.47	94.52
GREEK	574	0.46	94.98
PAKISTANI	551	0.44	95.42
SIKH	500	0.4	95.82
SPANISH	438	0.35	96.18
VIETNAMESE	427	0.34	96.52
JEWISH	351	0.28	96.8
PORTUGUESE	326	0.26	97.06
JAPANESE	293	0.24	97.3
EAST ASIAN & PACIFIC	263	0.21	97.51
DANISH	216	0.17	97.68
OTHER SOUTH ASIAN	209	0.17	97.85
SRI LANKAN	209	0.17	98.02
DUTCH	207	0.17	98.19
TURKISH	198	0.16	98.34
SWEDISH	191	0.15	98.5
RUSSIAN	138	0.11	98.61

Source: ONOMAP/KITES-PATSTAT.

Notes:

1) ‘OTHER MUSLIM’ subgroup includes CEL name types ‘BALKAN MUSLIM’, ‘MALAYSIAN MUSLIM’, ‘MUSLIM INDIAN’, ‘SUDANESE’, ‘WEST AFRICAN MUSLIM’, ‘OTHER MUSLIM’ (SMALLER MIDDLE EASTERN COUNTRIES, N/AFRICAN COUNTRIES, CENTRAL ASIAN REPS)
2) 'JEWISH' includes CEL name types ‘JEWISH / ASHKENAZI’, ‘SEPHARDIC JEWISH’
3) ‘EAST ASIAN AND PACIFIC’ includes CEL name types ‘BURMESE’, ‘CAMBODIAN’, ‘FIJIAN’, ‘HAWAIIAN’, ‘LAOTIAN’,‘MAORI’, ‘MAURITIAN’, ‘POLYNESIAN’, ‘SAMOAN’, ‘SINGAPOREAN’, ‘SOLOMON ISLANDER’, ‘SOUTH EAST ASIAN’ , ‘THAI’, ‘TIBETIAN’, ‘TONGAN’, ‘TUVALUAN’, ‘EAST ASIAN & PACIFIC OTHER’
4) ‘OTHER SOUTH ASIAN’ includes CEL name types ‘ASIAN CARIBBEAN’, ‘BENGALI’, ‘BHUTANESE’, ‘GUYANESE ASIAN’, ‘KENYAN ASIAN’, ‘NEPALESE’, ‘PARSI’, ‘SEYCHELLOIS’, ‘SOUTH ASIAN’, ‘TAMIL’

Table 2. UK-resident inventors: geographical origin groups, 1993-2004.

Probable geog area of origin, CEL	Freq.	%	Cumulative %
BRITISH ISLES	109,429	87.89	87.89
SOUTH ASIA	3,074	2.47	90.36
CENTRAL EUROPE	3,035	2.44	92.8
EAST ASIA	2,557	2.05	94.85
SOUTHERN EUROPE	2,394	1.92	96.78
EASTERN EUROPE	1,395	1.12	97.9
MIDDLE EAST	1,060	0.85	98.75
NORTHERN EUROPE	606	0.49	99.24
REST OF WORLD	568	0.46	99.70
AFRICA	324	0.26	99.96
CENTRAL ASIA	31	0.02	99.98
AMERICAS	29	0.02	100.00

Source: ONOMAP/KITES-PATSTAT.

Table 3. UK-resident inventors: biggest ONS ethnic groups, 1993-2004.

Probable ethnic group in 1991 Census categories, CEL	%	Cumulative %
WHITE	94.28	94.28
ANY OTHER ETHNIC GROUP	1.76	96.04
INDIAN	1.69	97.73
CHINESE	1.41	99.14
PAKISTANI	0.54	99.68
BLACK - AFRICAN	0.24	99.92
BANGLADESHI	0.08	100
BLACK - CARIBBEAN	0	100

Source: ONOMAP/KITES-PATSTAT.
Notes: Ethnic groups typology taken from 1991 Census to allow comparability pre and post-2001. Frequencies have been supressed to avoid disclosure.

Table 4. Shares of migrants and ethnic inventors in TTWA working-age populations, 1993-2004. Top 40 areas.

% ethnic inventors	% migrants /population	TTWA name	TTWA type
0.287	0.158	Crawley	Primary Urban
0.241	0.148	Southampton	Primary Urban
0.206	0.359	London	Primary Urban
0.171	0.173	Oxford	Primary Urban
0.169	0.169	Cambridge	Primary Urban
0.166	0.113	Dundee	Primary Urban
0.158	0.101	Oban	N Scotland rural
0.153	0.174	Guildford & Aldershot	Primary Urban
0.152	0.147	Swindon	Primary Urban
0.147	0.113	St Andrews & Cupar	N Scotland rural
0.147	0.143	Edinburgh	Primary Urban
0.143	0.141	Colchester	Primary Urban
0.143	0.092	Pembroke & Tenby	Welsh rural
0.141	0.104	Carlisle	N England rural
0.138	0.114	Bude & Holsworthy	SW England rural
0.136	0.127	Aberdeen	Primary Urban
0.133	0.106	Holyhead	Welsh rural
0.129	0.174	Brighton	Primary Urban
0.126	0.122	Lancaster & Morecambe	N England rural
0.124	0.170	Bedford	Primary Urban
0.122	0.107	Livingston & Bathgate	N Scotland rural
0.121	0.136	Cardiff	Primary Urban
0.120	0.128	Glasgow	Primary Urban
0.120	0.098	Inverness & Dingwall	N Scotland rural
0.119	0.101	Lanarkshire	Primary Urban
0.119	0.114	Newcastle & Durham	Primary Urban
0.116	0.210	Birmingham	Primary Urban
0.115	0.092	Haverfordwest & Fishguard	Welsh rural
0.114	0.119	York	Primary Urban
0.114	0.200	Leicester	Primary Urban
0.114	0.184	Reading & Bracknell	Primary Urban
0.113	0.215	Wycombe & Slough	Primary Urban
0.111	0.109	Wirral & Ellesmere Port	Primary Urban
0.109	0.157	Leeds	Primary Urban
0.109	0.143	Newbury	SW England rural
0.108	0.111	Louth & Horncastle	Rest England rural
0.107	0.108	Liverpool	Primary Urban
0.106	0.139	Canterbury	Rest England rural
0.106	0.129	Margate, Ramsgate & Sandwich	Rest England rural
0.106	0.144	Harlow & Bishop's Stortford	Rest England rural

Source: ONOMAP/KITES-PATSTAT/ONS.
Note: TTWAs use 2001 boundaries. ‘Primary urban’ TTWAs contain an urban core with at least 125,000 people. TTWAs with fewer than 10 inventors suppressed.

Table 5. Ethnic inventor Location Quotients, 1993-2004. Top 40 areas.

LQ	TTWA name	TTWA type
2.372	Crawley	Primary Urban
1.989	Southampton	Primary Urban
1.703	London	Primary Urban
1.414	Oxford	Primary Urban
1.394	Cambridge	Primary Urban
1.375	Dundee	Primary Urban
1.304	Oban	N Scotland rural
1.266	Guildford & Aldershot	Primary Urban
1.252	Swindon	Primary Urban
1.216	St Andrews & Cupar	N Scotland rural
1.213	Edinburgh	Primary Urban
1.180	Pembroke & Tenby	Welsh rural
1.180	Colchester	Primary Urban
1.162	Carlisle	N England rural
1.139	Bude & Holsworthy	SW England rural
1.122	Aberdeen	Primary Urban
1.101	Holyhead	Welsh rural
1.062	Brighton	Primary Urban
1.044	Lancaster & Morecambe	N England rural
1.024	Bedford	Primary Urban
1.005	Livingston & Bathgate	N Scotland rural
1.000	Cardiff	Primary Urban
0.995	Glasgow	Primary Urban
0.988	Inverness & Dingwall	N Scotland rural
0.981	Lanarkshire	Primary Urban
0.980	Newcastle & Durham	Primary Urban
0.955	Birmingham	Primary Urban
0.953	Haverfordwest & Fishguard	Welsh rural
0.941	York	Primary Urban
0.940	Leicester	Primary Urban
0.938	Reading & Bracknell	Primary Urban
0.932	Wycombe & Slough	Primary Urban
0.917	Wirral & Ellesmere Port	Primary Urban
0.898	Leeds	Primary Urban
0.897	Newbury	SW England rural
0.893	Louth & Horncastle	Rest England rural
0.886	Liverpool	Primary Urban
0.876	Canterbury	Rest England rural
0.875	Margate, Ramsgate & Sandwich	Rest England rural
0.872	Harlow & Bishop's Stortford	Rest England rural

Source: ONOMAP/KITES-PATSTAT/ONS.
Note: TTWAs use 2001 boundaries. ‘Primary urban’ TTWAs contain an urban core with at least 125,000 people. TTWAs with fewer than 10 inventors suppressed.

Table 6. Fractionalisation Index scores for inventors and TTWA working-age populations, 1993-2004. Top 40 areas.

Inventor FRAC	Population FRAC	TTWA name	TTWA type
0.384	0.498	London	Primary Urban
0.354	0.188	Southampton	Primary Urban
0.310	0.206	Crawley	Primary Urban
0.308	0.225	Oxford	Primary Urban
0.305	0.133	Dundee	Primary Urban
0.293	0.139	Honiton & Axminster	SW England rural
0.288	0.122	Lancaster & Morecambe	N England rural
0.283	0.226	Cambridge	Primary Urban
0.282	0.184	Swindon	Primary Urban
0.279	0.099	Bangor, Caernarfon & Llangefni	Welsh rural
0.273	0.168	Colchester	Primary Urban
0.256	0.106	Carlisle	N England rural
0.255	0.126	St Andrews & Cupar	N Scotland rural
0.255	0.122	Bude & Holsworthy	SW England rural
0.250	0.234	Guildford & Aldershot	Primary Urban
0.244	0.179	Edinburgh	Primary Urban
0.241	0.275	Bradford	Primary Urban
0.239	0.143	Glasgow	Primary Urban
0.237	0.263	Birmingham	Primary Urban
0.234	0.148	Aberdeen	Primary Urban
0.226	0.104	Wirral & Ellesmere Port	Primary Urban
0.225	0.164	Cardiff	Primary Urban
0.224	0.104	Livingston & Bathgate	N Scotland rural
0.222	0.206	Bedford	Primary Urban
0.218	0.135	Lincoln	Rest England rural
0.217	0.121	Liverpool	Primary Urban
0.215	0.225	Brighton	Primary Urban
0.213	0.289	Wycombe & Slough	Primary Urban
0.210	0.126	Newcastle & Durham	Primary Urban
0.208	0.172	Bristol	Primary Urban
0.208	0.269	Leicester	Primary Urban
0.207	0.184	Eastbourne	Rest England rural
0.203	0.134	Monmouth & Cinderford	Rest England rural
0.202	0.190	Leeds	Primary Urban
0.201	0.244	Luton & Watford	Primary Urban
0.199	0.142	Norwich	Primary Urban
0.194	0.158	Rugby	Rest England rural
0.194	0.239	Reading & Bracknell	Primary Urban
0.193	0.169	Harlow & Bishop's Stortford	Rest England rural
0.192	0.114	Stafford	Rest England rural

Source: ONOMAP/KITES-PATSTAT/ONS.
Note: TTWAs use 2001 boundaries. ‘Primary urban’ TTWAs contain an urban core with at least 125,000 people. TTWAs with fewer than 10 inventors suppressed.

Table 7. TTWAs' weighted patent stocks, 1993-2004. Top 40 areas.

Weighted patent count	TTWA name	TTWA type
1697.14	London	Primary Urban
1155.59	Cambridge	Primary Urban
719.36	Oxford	Primary Urban
705.62	Harlow & Bishop's Stortford	Rest England rural
531.69	Manchester	Primary Urban
489.87	Guildford & Aldershot	Primary Urban
483.41	Southampton	Primary Urban
440.96	Bristol	Primary Urban
428.15	Reading & Bracknell	Primary Urban
416.01	Crawley	Primary Urban
379.21	Ipswich	Primary Urban
365.63	Swindon	Primary Urban
342.90	Wycombe & Slough	Primary Urban
341.67	Stevenage	Primary Urban
312.93	Newcastle & Durham	Primary Urban
309.40	Wirral & Ellesmere Port	Primary Urban
301.75	Leicester	Primary Urban
289.82	Birmingham	Primary Urban
260.66	Nottingham	Primary Urban
223.87	Leeds	Primary Urban
218.49	Edinburgh	Primary Urban
213.60	Worcester & Malvern	Primary Urban
183.83	Margate, Ramsgate & Sandwich	Rest England rural
181.10	Coventry	Primary Urban
169.36	Bedford	Primary Urban
167.98	Luton & Watford	Primary Urban
165.09	Cardiff	Primary Urban
163.87	Glasgow	Primary Urban
161.37	Warwick & Stratford-upon-Avon	Rest England rural
161.20	Warrington & Wigan	Primary Urban
152.70	Hull	Primary Urban
148.04	Derby	Primary Urban
147.14	Aberdeen	Primary Urban
138.16	Portsmouth	Primary Urban
136.70	Milton Keynes & Aylesbury	Primary Urban
130.99	Middlesbrough & Stockton	Primary Urban
121.67	Chelmsford & Braintree	Primary Urban
121.35	Chester & Flint	Welsh rural
118.13	Northampton & Wellingborough	Primary Urban
113.95	Maidstone & North Kent	Primary Urban

Source: KITES-PATSTAT/ONS.
Note: TTWAs use 2001 boundaries. 'Primary urban' TTWAs contain an urban core with at least 125,000 people. Patents are weighted by number of inventors, not area population.

Table 8. Summary statistics.

Variable	N	Mean	SD	Min	Max
Inventor patent count / 4-year period	89312	0.114	0.694	0	25
Inventors' ave patent count, pre-1993	89312	0.405	0.351	0.286	11.143
Inventor likely techfield mover	89312	0.256	0.437	0	1
Inventor likely TTWA mover	89312	0.143	0.35	0	1
Inventor is UK geog. origin	89312	0.937	0.243	0	1
Inventor is foreign geog. origin	89312	0.063	0.243	0	1
Inventor African origin	89312	0.002	0.041	0	1
Inventor Americas origin	89312	0.000	0.013	0	1
Inventor Central Asia origin	89312	0.000	0.018	0	1
Inventor Central Europe origin	89312	0.012	0.107	0	1
Inventor rest of world origin	89312	0.003	0.058	0	1
Inventor East Asian origin	89312	0.007	0.084	0	1
Inventor East Europe origin	89312	0.007	0.086	0	1
Inventor Middle East origin	89312	0.006	0.075	0	1
Inventor Northern Europe origin	89312	0.003	0.052	0	1
Inventor South Asian origin	89312	0.015	0.123	0	1
Inventor South European origin	89312	0.007	0.086	0	1
Frac. Index, geog. origin groups	89312	0.209	0.118	0	0.612
Inventor is white ethnicity	89312	0.97	0.172	0	1
Inventor is minority ethnic	89312	0.03	0.172	0	1
Inventor Black Caribbean	89312	0	0.01	0	1
Inventor Black African	89312	0.002	0.04	0	1
Inventor Indian	89312	0.012	0.107	0	1
Inventor Pakistani	89312	0.003	0.052	0	1
Inventor Chinese	89312	0.004	0.064	0	1
Inventor other ethnic group	89312	0.01	0.099	0	1
Frac. Index, ethnic groups	89312	0.108	0.066	0	0.449
TTWA Frac Index, geog. groups	89309	0.225	0.142	0	0.528
TTWA Frac Index, ethnic groups	89309	0.169	0.141	0	0.459
% graduates	89309	0.238	0.051	0.106	0.362
% graduates with STEM degrees	89309	0.121	0.032	0.041	0.196
% graduates with PhDs	89309	0.007	0.005	0	0.029
% employed hi-tech manufacturing	89309	0.027	0.014	0	0.194
% employed medium-tech m'facturing	89309	0.046	0.023	0	0.135
% in entry level occupations	89309	0.338	0.049	0.25	0.667
% unemployed >=12 months	89309	0.016	0.012	0	0.08
log(population density)	89309	6.605	1.053	2.06	8.359
Electronics patent	89312	0.009	0.093	0	1
TTWA weighted patent count	89312	493.094	578.301	0	1888.03
TTWA weighted patents, 1981-84	88726	144.814	201.789	0.25	613.859

Source: KITES-PATSTAT/ONS/LFS
Note: Area-level controls not available for all TTWAs.

Table 9. Patent counts, geographical origin zones, negative binomial results.

Individual patent counts	**(1)**	**(2)**	**(3)**
Ethnic inventor, geog.	-0.000	0.004	0.009
	(0.011)	(0.008)	(0.007)
Frac Index of inventors, geog. origin groups	-0.061	0.079	0.087**
	(0.101)	(0.050)	(0.042)
Frac Index, TTWA country of birth		-0.203*	-0.140*
		(0.110)	(0.085)
% STEM degrees, TTWA		0.372**	0.304**
		(0.176)	(0.147)
Log of TTWA population density		0.005	0.005
		(0.008)	(0.007)
Area weighted patents, 1981-84		-0.000	-0.000
		(0.000)	(0.000)
% hi-tech mf empl, OECD defn.		-0.159	-0.111
		(0.281)	(0.226)
% medium-tech mf, OECD defn.		0.048	0.051
		(0.172)	(0.134)
% entry-level occupations		0.042	0.113
		(0.123)	(0.106)
% unemployed >=12 months		-0.313	-0.000
		(0.441)	(0.354)
Electronics / OST7 type 1 patents		2.074***	1.697***
		(0.132)	(0.176)
Urban TTWA		-0.018*	-0.021*
		(0.015)	(0.015)
Fixed effect			0.101***
			(0.007)
ln(alpha)			
Constant	2.991***	2.683***	2.491***
	(0.052)	(0.063)	(0.069)
Observations	89312	88726	88726
Log-likelihood	-25328.463	-24379.554	-23859.107
Chi^2 fit statistic (Wald)	376.947	3520.345	2693.200

Source: KITES-PATSTAT/ONS/LFS

Notes: Notes: All models use time dummies. Heteroskedasticity and autocorrelation-robust standard errors clustered on TTWA. Except for ln(alpha) term, coefficients are marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%

Table 10. Patent counts, geographical origin zones, OLS results.

Individual patent counts	**(1)**	**(2)**	**(3)**
Ethnic inventor, geog.	-0.002	0.004	0.011
	(0.011)	(0.011)	(0.010)
Frac Index of inventors, geog. origin groups	-0.055	0.119**	0.099*
	(0.088)	(0.058)	(0.055)
Frac Index, TTWA country of birth		-0.137	-0.079
		(0.127)	(0.115)
% STEM degrees, TTWA		0.302	0.334
		(0.292)	(0.278)
Log of TTWA population density		0.006	0.008
		(0.010)	(0.009)
Area weighted patents, 1981-84		-0.000	-0.000
		(0.000)	(0.000)
% hi-tech mf empl, OECD defn.		-0.166	-0.245
		(0.385)	(0.367)
% medium-tech mf, OECD defn.		0.120	0.093
		(0.240)	(0.216)
% entry-level occupations		0.084	0.149
		(0.166)	(0.154)
% unemployed >=12 months		-1.211	-0.934
		(0.747)	(0.719)
Electronics / OST7 type 1 patents		2.356***	2.305***
		(0.139)	(0.135)
Urban TTWA		-0.024	-0.028
		(0.019)	(0.017)
Fixed effect			0.266***
			(0.036)
Constant	0.196***	0.122	-0.034
	(0.010)	(0.107)	(0.105)
Observations	89312	88726	88726
F-statistic	76.283	52.523	50.226
R^2	0.007	0.107	0.125

Source: KITES-PATSTAT/ONS/LFS
Notes: All models use time dummies. Heteroskedasticity and autocorrelation-robust standard errors clustered on TTWA. * = significant at 10%, ** 5%, *** 1%.

Table 11. Patent counts, all inventors, ONS ethnic groups.

Negative binomial

Individual patent counts	**(1)**	**(2)**	**(3)**
Ethnic inventor, ONS minority ethnic group	-0.006	-0.000	0.000
	(0.016)	(0.014)	(0.012)
Frac Index of inventors, ONS ethnic groups	-0.165	0.101	0.125**
	(0.145)	(0.067)	(0.056)
Controls	N	Y	Y
Fixed effects	N	N	Y
Observations	89312	88726	88726
Log-likelihood	-25319.277	-24386.644	-23864.136
Chi^2 goodness of fit statistic (Wald)	414.921	2706.003	2426.458

OLS

Individual patent counts	**(1)**	**(2)**	**(3)**
Ethnic inventor, ONS minority ethnic group	-0.010	-0.002	0.003
	(0.015)	(0.014)	(0.013)
Frac Index of inventors, ONS ethnic groups	-0.155	0.123	0.097
	(0.131)	(0.082)	(0.077)
Controls	N	Y	Y
Fixed effects	N	N	Y
Observations	89312	88726	88726
F-statistic	75.337	54.477	58.197
R^2	0.007	0.107	0.125

Source: KITES-PATSTAT/ONS/LFS.
Notes: All models use time dummies. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of area birth country groups, % entry-level occupations, % long term unemployed, urban TTWA dummy. Heteroskedasticity and autocorrelation-robust standard errors clustered on TTWA. Negative binomial models show marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 12. Robustness checks. Negative binomial and OLS results.

Negative Binomial

Individual patent counts	**(1)**	**(2)**	**(3)**	**(4)**
Ethnic inventor, geographic origin	0.009 (0.007)	0.007 (0.007)	-0.000 (0.001)	-0.002 (0.001)
Frac Index of inventors, geog. origin groups	0.087** (0.042)	0.046 (0.039)	0.016*** (0.006)	0.016*** (0.006)
% with PhDs in TTWA		2.649*** (0.504)		
#times inventor patents in previous YG within sample			0.053*** (0.002)	0.057*** (0.002)
Controls	Y	Y	Y	Y
Fixed effects	Y	Y	Y	Y
Include London?	Y	Y	Y	N
Observations	88726	88726	88726	75571
Log-likelihood	-23859.107	-23821.523	-16507.273	-21524.746
Chi^2 fit statistic (Wald)	2693.200	2181.073	4008.364	2095.403

Source: KITES-PATSTAT/ONS/LFS
Notes: All models use time dummies. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of area birth country groups, % entry-level occupations, % long term unemployed, urban TTWA dummy. Heteroskedasticity and autocorrelation-robust standard errors clustered on TTWA. Negative binomial models show marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 13. Urban areas. Negative binomial results.

Individual patent counts	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Ethnic inventor, geographic origin	0.009	0.009	0.008	0.008	0.009	0.008	0.009
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Frac Index of inventors, geog. origin groups	0.087**	0.085**	0.066	0.080*	0.136**	0.284	0.494**
	(0.042)	(0.043)	(0.041)	(0.041)	(0.067)	(0.201)	(0.231)
urban TTWA	-0.021	-0.016		0.054	-0.007		-0.028*
	(0.015)	(0.010)		(0.043)	(0.010)		(0.015)
log of TTWA population density	0.005		-0.002	0.016		0.004	0.016**
	(0.007)		(0.005)	(0.012)		(0.005)	(0.008)
urban TTWA * ln(pop density)				-0.016			
				(0.014)			
Frac Index * urban TTWA					-0.066		
					(0.076)		
Frac Index * ln(pop density)						-0.037	-0.067*
						(0.033)	(0.037)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	88726	88726	88726	88726	88726	88726	88726
Log-likelihood	-23859.107	-23861.196	-23871.085	-23853.923	-23859.802	-23868.311	-23850.578
Chi^2 fit statistic (Wald)	2693.200	2594.921	3234.725	2754.837	2720.994	4245.201	3717.697

Source: KITES-PATSTAT/ONS/LFS
Notes: all models use time dummies and individual fixed effects. Robust standard errors clustered on TTWA. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, frac. index of birth country groups, % entry-level occupations, % long term unemployed. Coefficients are marginal effects at the mean.
* = significant at 10%, ** 5%, *** 1%.

Table 14. Inventor groups, negative binomial results.

Inventor patent count	Marginal effect
Africa origin	-0.037* (0.022)
Americas origin	0.176 (0.166)
Central Asia origin	0.045 (0.055)
Central Europe origin	-0.003 (0.014)
Diasporic origin	-0.019 (0.014)
East Asia origin	-0.037*** (0.007)
Eastern Europe origin	0.032 (0.034)
Middle East origin	-0.008 (0.025)
Northern Europe origin	0.001 (0.045)
South Asia origin	0.025* (0.015)
Southern Europe origin	0.053* (0.040)
Frac Index of inventors, geog. origin groups	0.087** (0.042)
Controls	Y
Observations	88726
Log-likelihood	-23843.642
Chi-squared	4438.933

Source: KITES-PATSTAT/ONS/LFS
Notes: all models use time dummies. Robust standard errors clustered on TTWA. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of ONS ethnic groups, % entry-level occupations, % long term unemployed, urban TTWA dummy. Coefficients are marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 15. Alternative fixed effects, negative binomial results.

Individual patent counts	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Ethnic inventor, geog. origin	0.009 (0.007)	0.003 (0.004)	0.005 (0.004)		
Central Europe origin				-0.003 (0.014)	-0.001 (0.009)
East Asia origin				-0.037*** (0.007)	-0.016*** (0.006)
Eastern Europe origin				0.032 (0.034)	0.013 (0.022)
South Asia origin				0.025* (0.015)	0.012* (0.009)
Southern Europe origin				0.053* (0.040)	0.024 (0.017)
Frac Index of inventors, geog. origin groups	0.087** (0.042)	0.050* (0.028)	0.055** (0.027)	0.087** (0.042)	0.055** (0.026)
Fixed effect, average patents pre-sample	0.101*** (0.007)		0.028*** (0.004)	0.100*** (0.007)	0.028*** (0.004)
Fixed effect, patents in >1 IPC7 field		0.217*** (0.010)	0.184*** (0.009)		0.183*** (0.009)
Controls	Y	Y	Y	Y	Y
Observations	88726	88726	88726	88726	88726
Log-likelihood	-23859.107	-22138.191	-21926.052	-23843.642	-21917.627
Chi-squared	2693.200	3670.001	5323.670	4438.933	6041.785

Source: KITES-PATSTAT/ONS/LFS

Notes: all models use time dummies. Robust standard errors clustered on TTWA. In models (4) and (5) I fit dummies for all minority co-ethnic groups with UK-origin the reference category. To save space results for the five largest minority groups only are shown here. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of birth country / ONS ethnic groups, % entry-level occupations, % long term unemployed, urban dummy. Coefficients are marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 16. Diversity effects versus size effects, negative binomial results

Individual patent counts	(1)	(2)	(3)	(4)	(5)
Ethnic inventor, geog. origin	0.009 (0.007)	0.009 (0.007)	0.009 (0.007)	0.008 (0.007)	
Central Europe origin					-0.003 (0.014)
East Asia origin					-0.037*** (0.008)
Eastern Europe origin					0.032 (0.034)
South Asia origin					0.024* (0.015)
Southern Europe origin					0.054* (0.041)
Frac Index of inventors, geog. origin groups	0.087** (0.042)		0.108*** (0.041)	0.191** (0.080)	0.189** (0.079)
% ethnic inventors, geog. origin as share of all inventors		0.068 (0.145)	-0.058 (0.138)	0.060 (0.121)	0.057 (0.121)
Frac index * % ethnic inventors				-0.676* (0.345)	-0.662** (0.336)
Controls	Y	Y	Y	Y	Y
Observations	88726	88726	88726	88726	88726
Log-likelihood	-23859.107	-23868.208	-23858.221	-23851.433	-23836.126
Chi-squared	2693.200	3064.329	2830.487	3853.584	5748.078

Source: KITES-PATSTAT/ONS/LFS

Notes: all models use time dummies. Robust standard errors clustered on TTWA. In model (5) I fit dummies for all minority co-ethnic groups with UK-origin the reference category. To save space results for the five largest minority groups only are shown here. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of birth country / ONS ethnic groups, % entry-level occupations, % long term unemployed, urban dummy. Coefficients are marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 17. Alternative historic patent stocks: influence on inventor productivity.

Individual patent counts	**(1)**	**(2)**	**(3)**
Ethnic inventor, geog.	0.009	0.008	0.008
	(0.007)	(0.007)	(0.007)
Frac Index of inventors, geog. origin groups	0.087**	0.083**	0.067*
	(0.042)	(0.041)	(0.040)
Area historic weighted stock of patents, 1981-1984	-0.000		
	(0.000)		
Area historic weighted stock of patents, 1985-1988		-0.000	
		(0.000)	
Area historic weighted stock of patents, 1989-1992			0.000
			(0.000)
Controls	Y	Y	Y
Fixed effects	Y	Y	Y
Observations	88726	89196	89268
Log-likelihood	-23859.107	-23994.163	-24030.991
Chi^2 fit statistic (Wald)	2693.200	2720.995	2865.519

Source: KITES-PATSTAT/ONS/LFS
Notes: All models use time dummies. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of area birth country groups, % entry-level occupations, % long term unemployed, urban TTWA dummy. Heteroskedasticity and autocorrelation-robust standard errors clustered on TTWA. Coefficients are marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 18. Sample construction test for multiple inventor sub-sample.

Individual patent counts	All, zeroed	Multiple, zeroed		Multiple, blanked	
	(1)	(2)	(3)	(4)	(5)
Ethnic inventor, geographic origin	0.009	-0.095	-0.093	-0.095	-0.093
	(0.007)	(0.110)	(0.110)	(0.110)	(0.110)
Frac Index of inventors, geog. origin groups	0.087**	0.856	4.103**	0.856	4.103**
	(0.042)	(0.575)	(2.057)	(0.575)	(2.057)
Urban TTWA	-0.021	-0.170		-0.170	
	(0.015)	(0.134)		(0.134)	
Log of TTWA population density	0.005	0.025	0.056	0.025	0.056
	(0.007)	(0.058)	(0.072)	(0.058)	(0.072)
Frac Index * log population density			-0.579		-0.579
			(0.370)		(0.370)
Controls	Y	Y	Y	Y	Y
Observations	88726	4842	4842	4842	4842
Log-likelihood	-23859.107	-8526.503	-8527.051	-8526.503	-8527.051
Chi-squared	2693.200	173.503	185.897	173.503	185.897

Source: KITES-PATSTAT/ONS/LFS
Notes: All models use time dummies. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of area birth country groups, % entry-level occupations, % long term unemployed, urban TTWA dummy. Heteroskedasticity and autocorrelation-robust standard errors clustered on TTWA. Coefficients are marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 19. Distributional effects: individual level

Native patent counts	**(1)**	**(2)**	**(3)**
Frac Index of inventors, geog. origin groups	-0.069 (0.097)	-0.057 (0.077)	0.072* (0.041)
Controls	N	N	Y
Individual fixed effects	N	Y	Y
Observations	83672	83672	83098
Log-likelihood	-23726.567	-23236.532	-22334.827
Chi^2 fit statistic (Wald)	343.508	628.231	2536.289

Individual patent counts	**(1)**	**(2)**	**(3)**
UK inventor	-0.010 (0.008)	-0.009 (0.007)	0.027*** (0.008)
Frac Index of inventors, geog. origin groups		0.087** (0.042)	0.253*** (0.077)
UK * Frac Index			-0.172*** (0.056)
Controls	Y	Y	Y
Observations	88726	88726	88726
Log-likelihood	-23870.231	-23859.107	-23852.425
Chi^2 fit statistic (Wald)	3421.238	2693.200	2866.909

Source: KITES-PATSTAT/ONS/LFS
Notes: All models use time dummies. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, fractionalisation index of area birth country groups, % entry-level occupations, % long term unemployed, urban TTWA dummy. Heteroskedasticity and autocorrelation-robust standard errors clustered on TTWA. Coefficients are marginal effects at the mean. * = significant at 10%, ** 5%, *** 1%.

Table 20. Distributional effects: area level

% change in total weighted patents, 1993-2004	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
% change in weighted ethnic patents, 1993-2004	0.199*** (0.065)	0.259*** (0.066)	0.248*** (0.068)	0.248 (0.177)	0.259 (0.178)
Controls	N	Y	Y	Y	Y
OST7 technology field dummies	N	N	Y	Y	N
HAC standard errors	N	N	N	Y	Y
Observations	220	220	210	210	206
F-statistic	9.299	1.467	3.646	1.144	0.966
R^2	0.041	0.041	0.141	0.141	0.151

Source: KITES-PATSTAT/ONS/LFS

Notes: All models use time dummies. Controls fitted: log of population density, % STEM degrees, % employed in knowledge-intensive manufacturing, % migrant working-age population, % entry-level occupations, % long term unemployed, urban dummy. Technology field dummies cover OST7 fields 1 -6: electrical engineering and electronics; instruments; chemicals and materials; pharmaceuticals and biotechnology; industrial processes; mechanical engineering, machines and transport. Consumer goods and civil engineering patents are used as the reference category. * = significant at 10%, ** 5%, *** 1%.

REFERENCES

Agrawal, A., I. Cockburn, et al. (2006). "Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships." Journal of Economic Geography **6**(5): 571-591.

Agrawal, A., D. Kapur, et al. (2008). "How do spatial and social proximity influence knowledge flows? Evidence from patent data." Journal of Urban Economics **64**(2): 258-269.

Agrawal, A., D. Kapur, et al. (2011). "Brain Drain or Brain Bank? The impact of skilled emigration on poor country innovation." Journal of Urban Economics **69**(1): 43-55.

Alesina, A. and E. La Ferrara (2004). Ethnic Diversity and Economic Performance. NBER Working Paper 10313. Cambridge, MA, NBER.

Anderson, S. and M. Platzer (2007). American Made: The Impact of Immigrant Entrepreneurs and Professionals Arlington, National Venture Capital Association.

Angrist, J. and J.-S. Pischke (2009). Mostly Harmless Econometrics. Princeton, Princeton University Press.

Archibugi, D. and S. Iammarino (2002). "The globalization of technological innovation: definition and evidence." Review of International Political Economy **9**(1): 98–122.

Aspinall, P. (2009). "The Future of Ethnicity Classifications." Journal of Ethnic and Migration Studies **35**(9): 1417-1435.

Audretsch, D. and M. Feldman (1996). " R&D Spillovers and the Geography of Innovation and Production." American Economic Review **86**: 630-640.

Basu, A. (2002). "The Interaction between Culture and Entrepreneurship in London's Immigrant Businesses." International Small Business Journal **20**(4): 371-393

Basu, A. (2004). "Entrepreneurial aspirations among family business owners: An analysis of ethnic business owners in the UK." International Journal of Entrepreneurial Behaviour & Research **10**(1/2): 12-33.

Berliant, M. and M. Fujita (2009). The Dynamics of Knowledge Diversity and Economic Growth. 56th Annual North American Meeting, Regional Science Association International. San Francisco.

Blundell, R., R. Griffith, et al. (1995). "Dynamic Count Data Models of Technological Innovation." The Economic Journal **105**(429): 333-344.

Bonacich, E. (1973). "A Theory of Middleman Minorities." American Sociological Review **38**(5): 583-594.

Borjas, G. (1987). "Self-Selection and the Earnings of Immigrants." American Economic Review **77**: 531-553.

Borjas, G. (1994). "The Economics of Immigration." Journal of Economic Literature **32**(4): 1667-1717.

Borjas, G. (2011). The Benefits and Costs of High Skill Immigration. Migration: Economic Change, Social Challenge London UCL.

Bresnahan, T. and A. Gambardella (2004). Old-Economy Inputs for new-Economy Outcomes: What have we learned? Building High-Tech Clusters. T. Bresnahan and A. Gambardella. Cambridge CUP 331-358.

Brinkley, I. (2008). The Knowledge Economy: How knowledge is reshaping the economic life of nations. London, The Work Foundation.

Cameron, A. C. and P. Trivedi (2009). Microeconometrics using Stata. College Station, Stata Press.

Cantwell, J. (2005). MNCs, local clustering and science-technology relationships. Technological Change and Economic Catch-Up: The Role of Science and Multinationals. G. Santangelo. Cheltenham, Edward Elgar.

Card, D. (2005). "Is the New Immigration Really So Bad?" The Economic Journal **115**(November): 300-323.

Chellaraj, G., K. Maskus, et al. (2005). The Contribution of Skilled Immigration and International Graduate Students to U.S. Innovation World Bank Policy Research Working Paper 3588. Washington DC, World Bank.

Dahlman, C. (2010). Innovation Strategies of Three of the BRICS: Brazil, India and China- What Can We Learn From Three Different Approaches. The Rise of Technological Power in the South. X. Fu and L. Soete. Basingtoke, Palgrave MacMillan.

Department of Innovation Universities and Skills (2008). Innovation Nation. London, Department of Innovation, Universities and Skills.

Docquier, F. and H. Rapoport (2011). Globalisation, Brain Drain and Development. IZA DP 5590. Bonn, IZA.

Fagerberg, J. (2005). Innovation: A guide to the literature. The Oxford Handbook of Innovation. J. Fagerberg, D. Mowery and R. Nelson. Oxford, OUP.

Faggio, G. and H. Overman (2011). The effect of public sector employment on local labour markets. Spatial Economics Research Centre Annual Conference. London.

Fairlie, R., H. Krashinsky, et al. (2009). Indian Entrepreneurial Success in the United States, Canada and the United Kingdom. RAND Corporation Working Paper 727. Santa Monica, CA, RAND Corporation.

Florida, R. (2002). The Rise of the Creative Class. New York, Basic Books.

Freeman, C. (1987). Technology Policy and Economic Policy: Lessons from Japan London, Pinter.

Fujita, M. and S. Weber (2003). Strategic Immigration Policies and Welfare in Heterogenous Countries. Institute of Economic Research Working Papers. Kyoto, Kyoto University.

Gibbons, S., H. Overman, et al. (2011). Real Earnings Disparities in Britain. SERC Discussion Paper SERCDP0065. London, LSE.

Gordon, I. (2001). Unemployment and spatial labour markets: strong adjustment and persistent concentration. Geographies of Labour Market Inequality. R. Martin and P. Morrison. London, Routledge.

Gordon, I., C. Whitehead, et al. (2007). The Impact of Recent Immigration on the London Economy. London, City of London Corporation.

Greene, W. H. (1994). Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. NYU Working Paper No. EC-94-10. New York NYU.

Hall, B., A. Jaffe, et al. (2001). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. Cambridge, Mass., NBER.

Hall, P. (1998). Cities in Civilisation: Culture, Innovation and Urban Order. London, Weidenfeld and Nicholson.

Hausman, J. A., B. H. Hall, et al. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. NBER Technical Working Paper No. 17. Cambridge, MA, NBER.

Hong, L. and S. Page (2001). "Problem Solving by Heterogeneous Agents." Journal of Economic Theory **97**(1): 123-163.

Hong, L. and S. Page (2004). "Groups of diverse problem solvers can outperform groups of high-ability problem solvers." Proceedings of the National Academy of Sciences of the United States of America **101**(46): 16385-16389.

Hunt, J. and M. Gauthier-Loiselle (2008). How much Does Immigration Boost Innovation? NBER Working Paper 14312. Cambridge, Mass., NBER.

Jacobs, J. (1969). The Economy of Cities. London, Vintage.

Jaffe, A. and M. Trajtenberg (1999). "International Knowledge Flows: Evidence from patent citations." International Journal of Knowledge Organisation **8**: 105-136.

Jaffe, A. B., M. Trajtenberg, et al. (1993). "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." The Quarterly Journal of Economics **108**(3): 577-598.

Kaiser, U., H. C. Kongsted, et al. (2011). Labor Mobility, Social Network Effects, and Innovative Activity. I. D. N. 5654. Bonn, IZA.

Kapur, D. and J. McHale (2005). Sojourns and Software: Internationally mobile human capital and high tech industry development in India, Ireland and Israel. From Underdogs to Tigers: The Rise and Growth of the Software Industry in Brazil, China, India, Ireland and Israel. A. Arora and A. Gambardella. Oxford, OUP.

Kerr, W. (2008). The Agglomeration of US Ethnic Inventors. HBS Working Paper 09-003. Boston, MA, Harvard Business School.

Kerr, W. (2008). "Ethnic Scientific Communities and International Technology Diffusion." Review of Economics and Statistics **90**(3): 518-537.

Kerr, W. (2009). Breakthrough Innovations and Migrating Clusters of Innovation. NBER Working Paper 15443. Cambridge, MA, NBER.

Kerr, W. and W. Lincoln (2010). The Supply Side of Innovation: H-1b Visa Reforms and US Ethnic Invention NBER Working Paper 15768. Cambridge, Mass., NBER

Landry, C. and P. Wood (2008). The Intercultural City: Planning for Diversity Advantage. London, Earthscan.

Li, X. and Y. Pai (2010). The Changing Geography of Innovation Activities: What do Patent Indicators Imply? The Rise of Technological Power in the South. X. Fu and L. Soete. Basingstoke, Palgrave MacMillan.

Lissoni, F., G. Tarasconi, et al. (2006). The KEINS Database on Academic Inventors: Methodology and Contents. CESPRI Working Paper 181. Milan, Universita' Bocconi.

Maré, D. C., R. Fabling, et al. (2011). Immigration and Innovation. IZA Discussion Paper 5686. Bonn, IZA.

Mateos, P., R. Webber, et al. (2007). The Cultural, Ethnic and Linguistic Classification of Populations and Neighbourhoods using Personal Names. CASA Working Paper. London, UCL.

Menon, C. (2009). Stars and Comets: An Exploration of the Patent Universe. SERC DP0037. London, LSE.

Mowery, D. C. (2001). "Technological Innovation in a Multipolar System: Analysis and Implications for U.S. Policy." Technological Forecasting and Social Change **67**(2-3): 143-157.

Nakhaie, R., X. Lin, et al. (2009). "Social Capital and the Myth of Minority Self-Employment: Evidence from Canada." Journal of Ethnic and Migration Studies **35**(4): 625-644.

Nathan, M. and N. Lee (2011). Does Cultural Diversity Help Innovation in Cities? Evidence from London Firms. Spatial Economics Research Centre Discussion Paper SERCDP0069. London, LSE.

Niebuhr, A. (2006). Migration and Innovation: Does Cultural Diversity Matter for Regional R&D Activity? IAB Discussion Paper 14/2006. Nuernberg, Institut fuer Arbeitsmarkt- und Berufsforschung.

OECD (2009). OECD Patent Statistics Manual. Paris, OECD.

Ottaviano, G., E. Bellini, et al. (2007). Diversity and the Creative Capacity of Cities and Regions. SUSDIV Paper 2.2007. Bologna, FEEM.

Ozgen, C., P. Nijkamp, et al. (2010). Immigration and Innovation in European Regions. Department of Spatial Economics working paper. Amsterdam, VU University.

Ozgen, C., P. Nijkamp, et al. (2011). The Impact of Foreign Workers on Innovation and Knowledge Spillovers: Analysis of Survey Data from the Netherlands. Migration: Economic Change, Social Challenge. London, UCL.

Parrotta, P., D. Pozzoli, et al. (2011). The Nexus Between Labor Diversity and Firm's Innovation [sic]. NORFACE Discussion Paper 2011-5.

Peri, G. (2007). Higher Education, Innovation and Growth. Education and Training in Europe. G. Brunello, P. Garibaldi and E. Wasmer. Oxford, Oxford University Press.

Rath, J. and R. Kloosterman (2000). "Outsiders' Business: A Critical Review of Research on Immigrant Entrepreneurship." International Migration Review **34**(3): 657-681.

Rodríguez-Pose, A. and M. Storper (2006). "Better Rules or Stronger Communities? On the social foundations of institutional change and its economic effects." Economic Geography **82**(1): 1-25.

Romer, P. (1990). "Endogenous Technological Change." Journal of Political Economy **98**(5): 71-102.

Saxenian, A.-L. (2006). The New Argonauts: Regional Advantage in a Global Economy. Cambridge, MA, Harvard University Press.

Saxenian, A.-L. and C. Sabel (2008). "Venture Capital in the 'Periphery': The New Argonauts, Global Search and Local Institution-Building." Economic Geography **84**(4): 379-394.

Schumpeter, J. (1962). The Theory of Economic Development. Berlin, Springer.

Simmie, J., J. Carpenter, et al. (2008). History Matters: Path dependence and innovation in British city-regions. London, NESTA.

Stephan, P. and S. Levin (2001). "Exceptional Contributions to US Science by the foreign-born and foreign-educated." Population Research and Policy Review **20**: 59-79.

Wadhwa, V., A.-L. Saxenian, et al. (2007). America's New Immigrant Entrepreneurs. Durham, NC, Duke University / iSchool, UC Berkeley.

Wadsworth, J. (2010). Immigration and the UK Labour Market: The Evidence from Economic Research. CEP Election Analysis. London, LSE.

Wilson, R. (2007). Innovation in London. Working Paper 19. London, GLA Economics.

Yeung, H. (2009). "Regional Development and the Competitive Dynamics of Global Production Networks: An East Asian Perspective." Regional Studies **43**(3): 325-351.

Zenou, Y. (2011). Spatial versus Social Mismatch: The Strength of Weak Ties. IZA Discussion Papers 5507. Bonn, Institute for the Study of Labor (IZA).

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Spatial Economics Research Centre (SERC)
London School of Economics
Houghton Street
London WC2A 2AE

Tel: 020 7852 3565
Fax: 020 7955 6848
Web: www.spatialeconomics.ac.uk

SERC is an independent research centre funded by the Economic and Social Research Council (ESRC), Department for Business Innovation and Skills (BIS), the Department for Communities and Local Government (CLG) and the Welsh Assembly Government.