Journal of Economic Geography **15** (2015) pp. 129–168 doi:10.1093/jeg/lbu006
Advance Access Published on 10 May 2014

Same difference? Minority ethnic inventors, diversity and innovation in the UK

Max Nathan[*,**,†]

*Spatial Economics Research Centre, London School of Economics, Houghton St, London WC2A 2AE, UK
**National Institute of Economic and Social Research, 2 Dean Trench St, London, SW1P 3HE
†Corresponding author: Max Nathan, Spatial Economics Research Centre, London School of Economics, Houghton St, London WC2A 2AE, UK. *email* <m.a.nathan@lse.ac.uk>

Abstract

Minority ethnic inventors play important roles in US innovation, especially in high-tech regions such as Silicon Valley. Do 'ethnicity–innovation' channels exist elsewhere? Ethnicity could influence innovation via production complementarities from diverse inventor communities, co-ethnic network externalities or individual 'stars'. I explore these issues using new UK patents microdata and a novel name-classification system. UK minority ethnic inventors are spatially concentrated, as in the USA, but have different characteristics reflecting UK-specific geography and history. I find that the diversity of inventor communities helps raise individual patenting, with suggestive influence of East Asian-origin stars. Majority inventors may benefit from multiplier effects.

Keywords: Innovation, cultural diversity, minority ethnic inventors, patents, cities
JEL classifications: J15, O31, R11
Date submitted: 18 February 2012 **Date accepted:** 4 February 2014

1. Introduction

At first glance, ethnicity, diversity and innovation do not seem closely linked. However, in recent years there has been growing research and policy interest in the role of minority ethnic inventors (Saxenian, 2006; Legrain, 2006; Leadbeater, 2008; Hanson, 2012; Wadhwa, 2012). This largely stems from recent experience in the USA, where the impact of these groups is striking. Since the 1980s minority communities, particularly those of South/East Asian origin, have played increasingly important roles in US science and technology sectors (Stephan and Levin, 2001; Chellaraj et al., 2008; Stuen et al., 2012). Stephan and Levin, for example, find that minority ethnic scientists are over-represented among the 250 most-cited authors, authors of highly cited patents and individuals elected to the US National Academies of Sciences or Engineering. Minority inventors are spatially concentrated at city-region level (Kerr, 2008b): in high-tech US clusters such as Silicon Valley, so-called 'ethnic entrepreneurs' help connect South Bay firms to global markets, and are responsible for 52% of the Bay Area's start-ups (Saxenian, 2006). Research also suggests positive links between diverse populations and US regional patenting (Peri, 2007; Hunt and Gauthier-Loiselle, 2010), and between diasporic communities and knowledge diffusion, both across American cities and internationally (Kerr, 2008a, 2009).

By contrast, very little is known about the role of minority ethnic inventors in European countries. This matters because innovation is an established driver of long-term economic growth, and European policymakers are actively seeking to upgrade national innovation systems (McCann and Ortega-Arguilés, 2013). It also matters because many European countries have become more ethnically diverse in recent years, and immigration/integration policy design is a major focus of debate (Putnam, 2007; Caldwell, 2009; Syrett and Sepulveda, 2011).

This article explores whether the UK innovation system has benefited from minority ethnic inventors, and the diversity they introduce. I ask: does the cultural diversity of inventor groups influence patenting rates? 'Diversity effects' are especially under-explored in the literature, and are the focus of the article. I also look at possible effects of minority ethnic status, co-ethnic group membership and the role of urban location.

The UK case is particularly interesting to explore. Census data show that the non-white population in England and Wales grew from 5.9% to 14% of the population between 1991 and 2011; between 2001 and 2011 the non-'White British' share rose from 12.7% to 19.5%. Immigration has been an important driver, with a number of new communities forming since the mid-1990s; the migrant population share rose from 9% to 13% during 2001–2011 (Office of National Statistics, 2012). These patterns are highly urbanized, with London now a 'majority minority' city for the first time in its history. Such deep shifts have proved politically controversial, especially the role of immigration: the current UK Government has introduced a cap on non-European Union (EU) migrants and set up tight entry criteria for skilled arrivals from these countries.[1]

As with migrants and minorities in the wider population, minority ethnic inventors have become an important feature of the UK's inventor population. Figure 1 shows the population shares for minority ethnic inventors against shares for migrants and minority ethnic groups in the wider working-age population. Minority ethnic inventors' population shares are higher, and rising faster, than either of the 'base' working-age groups: by 2004 they comprised 12.7% of the inventor population, against 9.3% for migrant workers and 6.8% for minority workers.

Changing demography might affect innovation in three ways. These effects are ambiguous in sign, and channels may operate as substitutes or complements. First, cultural diversity may improve ideas generation in groups of inventors, if the benefits of a larger set of ideas or perspectives outweigh trust or communication difficulties between those groups (Alesina and Ferrara, 2005; Page, 2007; Berliant and Fujita, 2008). Second, co-ethnic group membership can improve information flow and lower transaction costs, accelerating within-group ideas generation and transmission (Docquier and Rapoport, 2012). However, group size may constrain knowledge spillovers. Third, demographic shifts may introduce highly skilled 'stars' who make a substantial difference to knowledge generation, or who are more willing to introduce disruptive ideas (Borjas, 1987; Zucker and Darby, 2007; Duleep et al., 2012); here, minority ethnic status needs to be disentangled from other endowments and contextual factors. All three channels may also be more pronounced in urban areas, through the clustering of minority groups, agglomeration economies or both.

1 The UK's Points Based System is organized in five Tiers. For Tier 1, 'exceptional talent' places are limited to 1000 per year, of which 700 can be scientists; in most cases candidates for an 'entrepreneur' place need at least £200,000 of backing; 'investors' need to demonstrate they can invest at least £1m. For postgraduate researchers, post-study leave to stay in the UK has been cut from 3 years to 3 months. In 2011/2012, Tier 2 allows for 27,000 places, restricted to a tightly defined set of 'shortage occupations'.

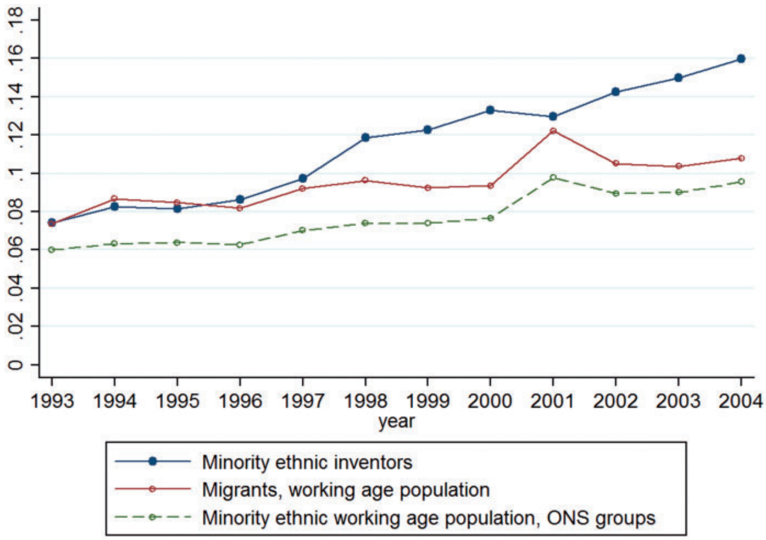

Figure 1. Growth in UK minority ethnic inventor population versus working-age migrant and minority ethnic populations, 1993–2004.

Source: KITES-PATSTAT/Office of National Statistics / Labour Force Survey.

Note: LFS data sample the working-age population, so will differ from Census estimates.

To explore, I construct a new 12-year panel of European Patent Office (EPO) patents microdata for the UK. I use the novel ONOMAP name-classification system to identify minority ethnic inventors, building on pioneering US work by Agrawal et al. (2008) and Kerr (2008b, 2010a). Descriptive analysis suggests that UK minority inventors have key differences from their American counterparts, reflecting the UK's distinctive geography, colonial and recent migration history. Although minority inventors are spatially clustered, as in the States, they are differently distributed from wider minority populations: many high-patenting areas do not have diverse inventor communities.

To explore effects on patenting I deploy a two-stage identification strategy, building on Oaxaca and Geisler (2003) and Combes et al. (2008). In the first stage, I estimate a knowledge production function linking counts of inventors' patenting activity to group diversity, controls and individual fixed effects. In the second stage, I decompose fixed effect estimates on minority ethnic status, co-ethnic group membership and other individual-level observables.

I find significant positive effects of inventor group diversity on individual patenting activity, worth about 0.025 patents per inventor. This result survives multiple robustness checks and tests for positive selection by mobile inventors. A back-of-the-envelope calculation suggests that increasing inventor diversity by around one standard deviation in a city such as Bristol could be worth around 40 extra patents in total. I also find suggestive evidence of positive contributions from minority ethnic high-patenting individuals, particularly East Asian-origin stars, once human capital is controlled for. Extensions imply some amplifying role of urban location and population density. Distributional tests indicate some multiplier 'effects' from minority to majority inventors, although these latter should be read as partial correlations, not causal links.

The article makes several contributions to the field. 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. It also adds to the growing empirical literature on immigration, ethnicity and innovation, and to the emerging field of inventor-level analysis (OECD, 2009).

The article is structured as follows. Section 2 sets out key concepts, theory and evidence. Section 3 introduces the data and identification strategy. Section 4 provides descriptive analysis. Section 5 outlines the identification and estimation strategy. Sections 6 and 7 give results, extensions and robustness checks. Section 8 concludes.

2. Definitions, framework, evidence

2.1. Key terms

'Innovation', 'ethnicity' and 'minority ethnic' all need careful definition. Innovation divides into invention, adoption and diffusion phases (Fagerberg, 2005). Patenting is primarily an indicator of invention (OECD, 2009). I look at shifts in individual patenting rates, hence 'inventor activity'.

Ethnic identity is a multifaceted notion with objective, subjective and dynamic elements (Aspinall, 2009). Robust quantitative measures of ethnicity therefore depend on stable, least-worst proxies, particularly as self-ascribed ethnicity information is not available from raw patents data (Ottaviano et al., 2007). I use inventor name information and the ONOMAP name-classification system developed by Mateos et al. (2007, 2011) to provide measures of inventor ethnicity, then use fractionalization indices to proxy inventor group diversity.

Ethnicity measures are based on (i) 12 geographical origin zones, where this origin is taken as a proxy for 'roots'; and (ii) nine 'macro-ethnic' categories similar to those used by the UK Office of National Statistics (ONS).[2] 'Minority ethnic' inventors are classified respectively as (i) those of likely non-UK roots and (ii) non-white inventors. Geographical origin data contain more detail and are less focused on visible appearance, so are my preferred measure (as Table 2 shows, under the ONS system 'other' is the second-largest ethnic category in the UK inventor population). In both cases, 'minority ethnic' combines UK and non-UK born groups, as my data cannot separately distinguish migrant inventors.

2.2. Literature review

Conventional theories of innovation have relatively little to say about ethnicity or diversity. For example, Schumpeter (1962) focuses on the individual 'entrepreneurial function' as a source of ideas; 'innovation systems' approaches highlight networks of firms and public institutions (Freeman, 1987); spatial approaches focus on the clustering of innovative activity due to agglomeration-related externalities, particularly local knowledge spillovers (Jaffe et al., 1993; Audretsch and Feldman, 1996). Endogenous growth theories help us to bridge demography and innovation. As Romer (1990) sets out, shifts in the technology frontier help determine economic

2 Geographic origin zones are 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. ONS groups are: White, Black Caribbean, Black African, Indian, Pakistani, Bangladeshi, Chinese and Other.

development, while human capital stocks and knowledge spillovers influence technological progress. However, access to knowledge is likely to be uneven across locations, sectors and social groups (Agrawal et al., 2008). Individual or group characteristics might then influence ideas generation and diffusion.

The existing literature identifies three potential ethnicity–innovation channels. First, the diversity of economic agents may influence innovative activity by acting as a production complementarity (Page, 2007; Berliant and Fujita, 2008, 2009). Specifically, individuals may benefit from group-level 'cognitive diversity' if this brings a richer mix of ideas and perspectives, which in turn helps members problem-solve and generate ideas. Ethnic or cultural mix may be a good proxy for cognitive diversity (Hong and Page, 2001, 2004). Such effects will be most likely observed in 'knowledge-intensive' environments (Fujita and Weber, 2003). Conversely, group-level cultural diversity may lead to lower trust and poor communication between individuals—for example, because of language barriers, misunderstandings or discriminatory attitudes. Co-operation (and thus spillovers) will be limited, leading to fewer, lower-quality solutions (Alesina and Ferrara, 2005).

Co-ethnicity may also offer advantages. Specifically, co-ethnic social networks—such as diasporas or transnational communities—may provide externalities (Agrawal et al., 2008; Docquier and Rapoport, 2012). Social networks offer their members higher social capital and trust, lowering transaction costs and risk, and helping ideas flow within the group (Rodríguez-Pose and Storper, 2006; Kaiser et al., 2011). In a closed setting, minority networks may be constrained by a small set of within-group possible matches (Zenou, 2011). In an open setting, such as under globalization, co-ethnic networks can be much larger and thus more influential. Again, in complex and/or research-intensive economic activities, diasporic communities may perform valuable roles both co-ordinating trans-national activity and facilitating information flows (Kapur and McHale, 2005; Saxenian and Sabel, 2008).

A third view is that individual characteristics matter, especially if minority ethnic inventors are migrants. From an economic perspective, migration decisions reflect expected returns: potential migrants balance out gains from migration and costs of moving abroad (Borjas, 1987). This implies that some migrants are 'pre-selected' on the basis of skill and entrepreneurialism (Wadhwa et al., 2007). Minority ethnic inventors who are migrants may also be more willing to invest in host country-relevant human capital, as they face lower opportunity costs than natives (Duleep et al., 2012). Migrant/minority status may thus positively predict patenting, over and above other human capital attributes, and regardless of diasporic ties or group composition. Here, the challenge is to distinguish ethnicity from other human capital endowments.

In theory, each of these channels has an ambiguous effect on innovation, and channels may operate as substitutes or complements (for example, group-level diversity effects may co-exist with individual 'stars'). The empirical literature is still sparse, but available evidence largely suggests net positive effects. Diversity channels remain the least-thoroughly explored, beyond a management literature testing small-sample correlations between team mix and business performance (see Page (2007) for a review). A few robust studies link ethnic diversity and innovation at group or workforce level. Some find correlations or causal links between team composition and product or process innovation (Ostergaard et al., 2011; Ozgen et al., 2011; Parrotta et al., 2013; Nathan and Lee, 2013). Others find no such connections (Maré et al., 2011). A couple of area-level studies also identifies links between skilled migrant diversity and innovation, for example Ozgen et al. (2012) for EU regions and Gagliardi (2011) for the UK.[3]

Co-ethnicity channels are better covered (see Docquier and Rapoport (2012) for a recent review of this literature). Several qualitative case studies trace links between specific US-based diasporas and 'home' countries such as India, China, Taiwan, Ireland and Israel (Kapur and McHale, 2005; Saxenian, 2006; Saxenian and Sabel, 2008). A range of quantitative studies identify links between co-ethnic communities and industrial performance in home countries (Kerr, 2008a), trade and FDI flows (Rauch and Trindade, 2002; Rauch and Casella, 2003; Kugler and Rapoport, 2007; Javorcik et al., 2011) and US multinational activity (Foley and Kerr, 2013). By contrast, Agrawal et al. (2008) find that physical location is up to four times more important for knowledge diffusion than co-ethnic connections.

A few recent studies test for individual-level 'star' effects. In the US Stephan and Levin (2001), Chellaraj et al. (2008) and Wadhwa et al. (2008) highlight the contributions of Indo and Chinese-American scientists to US science, particularly foreign graduate students. Kerr and Lincoln (2010) identify positive effects of US skilled migrant visas to patenting by ethnic Indian and Chinese inventors. Stuen et al. (2012) identify causal links between foreign PHD presence and subsequent highly cited publications. However, Hunt (2011) and Hunt and Gauthier-Loiselle (2010) find that individual 'migrant effects' are largely or wholly explained by education and industry hiring patterns.

This brief review highlights three empirical gaps. First, as mentioned, diversity–innovation channels are under-explored. Second, the vast bulk of the literature is focused on the USA, with only a handful of European studies exploring ethnicity–innovation connections: I am only aware of two area-level studies on diversity and patenting outcomes, Ozgen et al. (2012) and Niebuhr (2010), and no analysis at the individual or group level, where channels are most likely sited. Third, the interaction between individual, group and area factors is poorly covered. Innovative activity and minority communities tend to be concentrated in urban locations. Urban areas may amplify ethnicity–innovation channels, for example via localized knowledge spillovers; alternately, minority inventor communities may be physically isolated, limiting the opportunity for interaction (Jacobs, 1969; Zenou, 2009). I am aware of only two relevant empirical studies: Hunt and Gauthier-Loiselle (2010) find suggestive evidence of positive amplifying effects for US metros; Kerr (2010b) tracks breakthrough inventions across US cities, with co-ethnic networks aiding diffusion.

3. Data

I have three main data sources. Patents information comes from the European Patent Office (EPO). Raw patent data cannot typically be used at inventor level, because of common/misspelled names or changes of address: I use the KITES-PATSTAT cleaned dataset, which allows robust identification of individual UK-resident inventors (see Appendix A for details of the cleaning process). The raw data cover the period 1978–2007, dated by priority year, and contain geocoded information on 141,267 unique British-resident inventors and 123,030 patents with at least one British-resident

3 Other firm-level studies test links between workforce diversity and productivity: these include Maré and Fabling (2011), Hoogendoorn et al. (2013), Malchow-Møller et al. (2011) and Trax et al. (2012).

inventor.[4] Ethnicity information is then derived from inventor names using the ONOMAP name-classification system (see below and Appendix B). Finally, I combine this individual-level information with data on area-level characteristics, assembled from the UK Labour Force Survey (Office of National Statistics, 2013).

3.1. Working with patents data

I make several changes to the raw data. First, following Hall et al. (2001), I truncate the dataset by 3 years to end in 2004.[5] Second, I group patent observations in 4-year 'yeargroups'. Invention is a process, not an event, and inventors typically work on an invention for some time before filing a patent. Following Menon (2009), I use the mean citation lag of EPO patents to proxy the invention process.[6] Third, the main regressions use unweighted patent counts; area-level analysis uses weighted patents to avoid double-counting (OECD, 2009). Fourth, patents also have variable coverage across industries (with a well-known bias towards manufacturing) and are sensitive to policy shocks (OECD, 2009; Li and Pai, 2010).[7] I use technology field dummies and area-level industry shares to control for structural biases in patenting activity. Finally, I restrict the sample to 1993–2004. This allows me to fit precise area-level controls from the LFS, and to use pre-1993 inventor data to construct individual-level controls based on 'historic' activity (see Section 7).

3.2. Identifying ethnic inventors

I use the ONOMAP name-classification system (Mateos et al., 2007, 2011) to generate ethnicity information for individual inventors, building on similar approaches in US studies by Kerr (2008b, 2010a) and Agrawal et al. (2008). ONOMAP is developed from a very large names database extracted from Electoral Registers and telephone directories, covering 500,000 forenames and a million surnames across 28 countries. It classifies individuals according to most likely 'cultural–ethnic–linguistic' (CEL) characteristics, identified from forenames, surnames and forename–surname combinations. Essentially, ONOMAP exploits structural similarities and differences between name families, which reflect underlying cultural, ethnic and linguistic features—for example, 'John Smith' is more likely to be ethnically British than French. It also exploits the fact that 'distinctive naming practices in cultural and ethnic groups are persistent even long after immigration to different social contexts' (Mateos et al., 2011, p. e22943). Full details of ONOMAP are in Appendix B.

ONOMAP has the advantage of providing objective information at several levels of detail and across several dimensions of identity. It is also able to deal with Anglicisation of names, and names with multiple origins. Individual-level validation exercises suggest that

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). The full dataset has 160,929 unique UK-resident inventors: 19,492 observations lack postcode information.

5 There is typically a lag between applying for a patent and its being granted. This means that in a panel of patents, missing values appear in final periods.

6 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 4 years (OECD, 2009), so I group patents into 4-year periods.

7 Patents data also have some inherent limitations: not all inventions are patented, and patents may not record everyone involved in an invention.

ONOMAP matches almost all names and gives <5% measurement error (Lakha et al., 2011). For the KITES-PATSTAT data, ONOMAP matches over 99% of inventor names, and provides classification at various levels: after discussions with the ONOMAP team the inventor data were classified into 68 CEL 'subgroups', as well as two simpler typologies based on 12 geographical origin zones and nine 'macro-ethnic' groups based on the Office of National Statistics (ONS) 1991 Census classification. The descriptive analysis uses all three classifications (see Section 4). However, as many CEL subgroups are small, the regression analysis uses the less detailed groupings to minimize measurement error from small cells, and to allow easy matching with information from area-level controls.

4. Descriptive analysis

Tables 1–5 provide some initial descriptive analysis. Table 1 breaks down inventors by CEL subgroup, showing the 30 largest groups. We can see that although English,

Table 1. Inventors by 30 biggest CEL subgroups, 1993–2004

CEL subgroup	Frequency	%	Cumulative %
English	48,101	68.71	68.71
Celtic	5799	8.28	76.99
Scottish	3641	5.2	82.19
Irish	2034	2.91	85.1
Welsh	1452	2.07	87.17
Indian Hindi	751	1.07	88.25
German	731	1.04	89.29
Italian	600	0.86	90.15
French	572	0.82	90.96
Chinese	560	0.8	91.76
Polish	529	0.76	92.52
Muslim	483	0.69	93.21
European	387	0.55	93.76
Greek	340	0.49	94.25
Hong Kongese	335	0.48	94.73
Pakistani	326	0.47	95.19
Sikh	299	0.43	95.62
Spanish	244	0.35	95.97
Vietnamese	244	0.35	96.32
Jewish	205	0.29	96.61
Japanese	205	0.29	96.9
Portuguese	197	0.28	97.18
East Asian and Pacific	159	0.23	97.41
Danish	138	0.2	97.61
Sri Lankan	133	0.19	97.8
Dutch	115	0.16	97.96
South Asian	114	0.16	98.12
Swedish	109	0.16	98.28
Turkish	108	0.15	98.43
Pakistani Kashmir	78	0.11	98.55
Russian	78	0.11	98.66
Total	70,007	N/A	100

Source: KITES-PATSTAT/ONOMAP.

Table 2. Inventors by geographical origin and ONS ethnic groups, 1993–2004

	Frequency	%	Cumulative
Probable geographic area of origin			
British Isles	61,025	87.17	87.17
South Asia	1841	2.63	89.8
Central Europe	1804	2.58	92.38
East Asia	1539	2.2	94.57
Southern Europe	1442	2.06	96.63
Eastern Europe	801	1.14	97.78
Middle East	638	0.91	98.69
Northern Europe	374	0.53	99.22
Rest of World	337	0.48	99.7
Africa	177	0.25	99.88
Central Asia	.	.	.
Americas	.	.	100
Total	70,077	100	
Probable ethnic group, 1991 Census categories			
White	65,744	93.91	93.91
Any other ethnic group	1323	1.89	95.8
Indian	1262	1.8	97.6
Chinese	1046	1.49	99.1
Pakistani	404	0.58	99.67
Black-African	163	0.23	99.91
Bangladeshi	.	.	.
Black-Caribbean	.	.	100
Total	70,077	100	

Source: KITES-PATSTAT/ONOMAP.
Notes: Ethnic groups typology taken from 1991 Census to allow comparability with pre- and post-2001 area conditions. Some frequencies are suppressed to avoid disclosure and are marked by '.'.

Welsh, Scottish and Celtic[8] 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), and some largely recent migrant communities (e.g. Polish, Vietnamese).

Table 2 recuts the sample by geographical origin zones and by ONS macro-ethnic groups. Geographical origin zones (top panel) 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 earlier, ONS ethnic groups (bottom panel) are much less flexible, with 'other' the next largest inventor group after 'white'.

Table 3 sets out some differences in patenting activity between minority ethnic and majority inventor groups. Minority ethnic inventors, on average, patent slightly less than majority inventors (0.51 patents per yeargroup versus 0.54). As a whole, minority inventors are also less likely to be 'multiple' and 'star' inventors (who patent 2–4 times

8 'Celtic' denotes names common to Scottish, Welsh and Irish CEL types.

Table 3. Comparing patenting activity by majority and minority ethnic inventors, 1993–2004

	Observations (%)	% multiple inventors	% star inventors
All inventors	70,007 (100)	9.10	2.59
Of which			
Majority inventors	61,025 (87.2)	9.25	2.67
Minority inventors	8982 (12.8)	8.10	2.02
Different?	N/A	***	***
	Patent counts	**Patents by multiples**	**Patents by stars**
All inventors	0.536	1.917	4.384
Of which			
Majority inventors	0.539	1.909	4.358
Minority inventors	0.510	1.975	4.616
Different?	***	***	**

Source: KITES-PATSTAT/ONOMAP.
Notes: Multiple inventors patent 2–4 times in at least one 4-year period. Star inventors patent at least five times in at least one 4-year period. 'Patenting' is unweighted patenting activity per inventor per 4-year period. Differences between populations from *t*-tests and rank-sum tests.
Significant at *10%, **5% and ***1%.

per period and at least five times per period, respectively). However, minority multiple and star inventors patent significantly more than their majority counterparts (for stars, 4.616 versus 4.358 patents, respectively). All of these differences are statistically significant, as measured by *t*-tests and rank-sum tests. I return to this in Section 7 with more formal decomposition of individual characteristics.

Minority and majority ethnic inventors also differ in the type of patenting they are most likely to do. Table 4 decomposes minority and majority patenting by the groups' most common Observatoire des Sciences and des Techniques (OST30) technology fields (so that, for example, 0.12% of minority inventors most often patent in biotechnology (OST field 15), against 0.072% of majority inventors). Chi-square tests confirm that the two distributions are independent. The two groups are fairly close together across most technology fields, but minority inventors are more concentrated in information technology, semi-conductors, pharmaceutical and cosmetics, and agriculture and food products.

Next, I use postcode information to locate inventors in UK Travel to Work Areas (TTWAs), which are designed to cover self-contained labour markets: TTWAs are a good approximation of a local functional economy, and superior to administrative units such as local authority districts (Robson et al., 2006).[9] I then fit a simple urban/rural typology of TTWAs developed in Gibbons et al. (2011), allowing me to explore the

9 Formally, 75% of those living in a given TTWA also work in the TTWA, and vice versa. Matching is done by postcode sector, which minimizes observations lost through incomplete or mistyped postcode information (matching on full postcodes drops around 12% of observations; matching on postcode sector drops 5.77%). I exclude inventors resident in Northern Ireland.

Table 4. Comparing patenting for minority ethnic and majority inventors, 1993–2004

Modal OST30 field	% share of patenting by		
	Majority	Minority ethnic	All
Biotechnologies	7.39	12.03	7.99
Telecommunications	7.04	10.09	7.43
Information technology	6.05	9.18	6.46
Organic chemistry	10	8.94	9.86
Pharmaceuticals/cosmetics	7.06	8.83	7.29
Control/measure/analysis tools	9.12	8.4	9.03
Medical engineering	4.91	4.4	4.84
Optics	2.8	4.21	2.98
Basic chemistry	4.2	3.61	4.12
Audiovisual technology	2.94	3.37	2.99
Semi-conductors	1.13	3.05	1.38
Electrical engineering	3.68	2.84	3.57
Handling/printing	4.13	2.23	3.88
Consumer goods	3.88	2.16	3.66
Macromolecular chemistry	1.88	2.01	1.9
Mechanical engineering	2.86	2	2.75
Civil engineering	3.18	1.72	2.99
Materials processing	2.16	1.53	2.08
Engines/pumps/turbines	2.02	1.39	1.94
Materials/metallurgy	1.47	1.35	1.45
Transport technology	3.12	1.31	2.88
Mechanical elements	2.33	1.2	2.19
Agricultural and food products	1.41	1.11	1.37
Surface technology	1.14	0.99	1.12
Machine tools	1.21	0.57	1.13
Agricultural and food apparatuses	0.88	0.43	0.82
Thermal processes	0.63	0.34	0.59
Environmental technology	0.58	0.33	0.55
Nuclear technology	0.49	0.32	0.47
Space technology/weapons	0.32	0.08	0.28
Total	100	100	100

Source: KITES-PATSTAT.
Notes: OST30 reclassification of IPC technology fields.

potential effects of urban environments: ‘primary urban’ TTWAs are defined as those containing an urban core of at least 125,000 people.

Table 5 presents location quotients (LQs) for the 35 TTWAs with the largest shares of minority ethnic inventors by geographical origin, plus comparator LQs for the wider minority ethnic population (the latter defined by ONS ethnic groups).[10] The table confirms that minority ethnic inventors are spatially clustered, with a long tail of TTWAs with LQs under 1. High-ranking TTWAs for minority ethnic inventors are predominantly

10 Location quotients compare the local area share of a group *i* with the national share. Formally, $LQ_{ia} = (p_{ia}/p_a)/(p_i/p)$, where p_{ia}/p_a is the local population share of *i* in area *a*, and p_i / p is *i*'s national population share. An LQ of above 1 indicates concentration; scores below 1 indicate dispersion.

Table 5. Minority ethnic inventor LQs, 1993–2004. Top 35 TTWAs

LQ (minority population)	LQ (minority inventors)	TTWA name	TTWA type
1.332	4.009	Crawley	Primary urban
1.137	3.552	Southampton	Primary urban
8.663	3.219	London	Primary urban
0.267	2.779	Bangor, Caernarfon and Llangefni	Welsh rural
1.482	2.599	Oxford	Primary urban
0.621	2.499	Dundee	Primary urban
1.006	2.417	Swindon	Primary urban
1.163	2.374	Cambridge	Primary urban
0.197	2.254	St Andrews and Cupar	N Scotland rural
0.829	2.130	Colchester	Primary urban
0.155	2.124	Inverness and Dingwall	N Scotland rural
0.183	2.111	Carlisle	N England rural
1.380	2.050	Guildford and Aldershot	Primary urban
0.698	2.033	Edinburgh	Primary urban
1.276	2.009	Glasgow	Primary urban
6.453	1.931	Birmingham	Primary urban
3.055	1.850	Bedford	Primary urban
1.114	1.821	Lancaster and Morecambe	N England rural
0.427	1.817	Livingston and Bathgate	N Scotland rural
7.268	1.793	Bradford	Primary urban
1.676	1.773	Cardiff	Primary urban
0.990	1.765	Canterbury	Rest England rural
0.483	1.743	Aberdeen	Primary urban
0.349	1.741	Norwich	Primary urban
0.400	1.730	Wirral and Ellesmere Port	Primary urban
0.386	1.726	Lanarkshire	Primary urban
4.056	1.708	Wycombe and Slough	Primary urban
5.239	1.678	Leicester	Primary urban
0.986	1.678	Liverpool	Primary urban
0.719	1.671	Eastbourne	Rest England rural
0.825	1.662	Newbury	SW England rural
0.205	1.659	St Austell	SW England rural
3.117	1.635	Leeds	Primary urban
1.209	1.626	Brighton	Primary urban
2.068	1.619	Reading and Bracknell	Primary urban

Source: KITES-PATSTAT/ONOMAP/ONS.
Notes: TTWAs use 2001 boundaries. 'Primary urban' TTWAs contain an urban core with at least 125,000 people, 'rural' TTWAs may contain smaller urban settlements. Cells with fewer than 10 inventors suppressed. Population LQs from ONS minority ethnic groups in working-age population, not CEL data.

'primary urban', although a number of less dense and rural areas also feature, predominantly university towns (St Andrews, Lancaster, Inverness, Carlisle, Bangor) or areas adjoining TTWAs with universities (Honiton and Axminster, adjoining Exeter).[11]

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

Table 6. Weighted patent counts by TTWA, 1993–2004. Top 35 areas

Weighted patent count	TTWA name	TTWA type
1613.33	London	Primary urban
1021.22	Cambridge	Primary urban
617.47	Oxford	Primary urban
533.29	Harlow and Bishop's Stortford	Rest England rural
507.08	Manchester	Primary urban
496.12	Guildford and Aldershot	Primary urban
456.90	Bristol	Primary urban
424.77	Southampton	Primary urban
414.35	Crawley	Primary urban
370.59	Reading and Bracknell	Primary urban
366.80	Ipswich	Primary urban
344.94	Wycombe and Slough	Primary urban
344.17	Swindon	Primary urban
303.09	Birmingham	Primary urban
265.75	Newcastle and Durham	Primary urban
254.54	Stevenage	Primary urban
254.23	Nottingham	Primary urban
252.37	Leicester	Primary urban
235.58	Wirral and Ellesmere Port	Primary urban
210.11	Worcester and Malvern	Primary urban
206.02	Edinburgh	Primary urban
203.80	Leeds	Primary urban
167.67	Coventry	Primary urban
167.36	Luton and Watford	Primary urban
166.46	Warwick and Stratford-upon-Avon	Rest England rural
151.64	Aberdeen	Primary urban
151.24	Portsmouth	Primary urban
149.98	Bedford	Primary urban
147.75	Margate, Ramsgate and Sandwich	Rest England rural
144.87	Derby	Primary urban
143.20	Warrington and Wigan	Primary urban
142.31	Glasgow	Primary urban
139.42	Cardiff	Primary urban
138.46	Maidstone & North Kent	Primary urban
135.11	Hull	Primary urban

Source: KITES-PATSTAT/ONOMAP/ONS.
Notes: TTWAs use 2001 boundaries. Primary urban TTWAs defined as Table 4. Weighted patents stocks averaged 1993–2004. Weighting by inventors/patent and based on inventor address, not applicant address.

Overall, minority ethnic inventors follow the same urbanized spatial distribution as wider minority populations, but they are less concentrated in the largest and most diverse cities (such as London, Birmingham and Manchester) and more concentrated in second-tier cities and university towns (such as Oxford, Cambridge, Southampton and Guildford): the corresponding pairwise correlation of minority inventors to minority population LQs is 0.348. Note that wider populations are not identified using CEL data, so these comparisons should be used with care.

Table 6 gives weighted counts for the 35 TTWAs with the highest patenting activity: to minimize 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 minority ethnic inventor shares and diverse inventor groups (for example, London, Southampton, Crawley, Oxford and Cambridge). However, another group of high-patenting TTWAs have rather more homogenous inventor and general populations (for example, Bristol, Manchester and Reading). The pairwise correlation between minority inventor LQ and weighted patent stocks is 0.560.

Four broad lessons emerge from the descriptives. First, the UK's population of minority ethnic inventors appears substantially different from that of the USA, where minority ethnic inventor communities are dominated by South and East Asian groups (Kerr, 2008b, 2010a). By contrast, the UK has a number of European groups, South Asian and East Asian inventors drawn in large part from former colonies, plus recent migrant communities. Second, minority inventors are under-represented in the upper tail of multiple and star inventors; but those who are present patent significantly more than their 'majority ethnic' counterparts. There are also some differences in patenting fields, with minority inventors more likely to focus on semi-conductors and IT (as in the USA) as well as chemistry and food/agriculture fields (distinctive). Third, as in the USA, minority ethnic inventors are spatially concentrated, but the link to wider population diversity is relatively weak. Fourth, although minority ethnic inventor presence is positively correlated with high patent stocks, not all high-patenting locations have large minority inventor shares or diverse inventor communities.

5. Econometric analysis

For the regression analysis, I build a panel of UK-resident inventors' patenting activity between 1993 and 2004 inclusive. The sample includes all and only those inventors who patent at least once during this period. Each inventor-yeargroup-area cell records how many times an inventor patents in each 4-year phase. The basic panel covers 70,007 inventors across three 'yeargroups', giving 210,021 observations in the raw sample. Cell counts vary from 0 to 36, with a mean of 0.53 (see Table 6). Note that inventors are only observed when patenting. Blanking all cells where the inventor is not active—the most conservative response—would radically reduce sample size, as most inventors patent only once (and would miss instances where inventors were constrained from patenting for some reason). I thus zero all cells when no inventor activity is recorded, and test 'blanking' in robustness checks.

5.1. Identification strategy

This panel setting allows me to explore how changes in inventor group ethnic diversity might affect individual patenting activity, and to look at possible roles of minority ethnic status and co-ethnic group membership. To reliably identify group-level 'diversity effects', I need to control for individual ethnicity and unobserved individual characteristics as well as wider influencing factors (such as area-level demographic and economic conditions, technology field and time trends). Individual fixed effects are the most robust way to control for individual-level unobservables. However, as minority ethnic status and ethnic group membership are time-invariant, they drop out of any subsequent fixed effects regression. I therefore develop a two-stage identification strategy, drawing on Oaxaca and Geisler (2003) and Combes et al. (2008).

The first stage focuses on diversity. The estimating model is a modified knowledge production function, regressing counts of individual patenting activity on inventor group diversity, plus area-level controls, technology field-time effects and individual fixed effects. Group diversity effects on individual patenting activity should then reflect a combination of (i) externalities of ethnic diversity, (ii) changes in TTWA composition or (iii) inventors moving between TTWAs. The first of these is my variable of interest, and the second is captured in the area-level controls vector. Movers are a potential omitted variable if between-TTWA movement is a strong feature of the data, particularly if inventors select into high-innovation clusters. To deal with this, I identify the set of moving inventors in the panel (see Appendix A). In the main regressions, movers are constrained to one location: I then run a series of separate checks, exploring overall patterns of movement and testing the extent to which changes in area patent counts are explained by in-movers versus other factors (see Section 6).

For the second stage of the analysis I retrieve estimates of the individual fixed effect, then regress this on individuals' observable characteristics.[12] Here the variable of interest is minority ethnic status or co-ethnic group membership, and controls cover individual patenting intensity and scope as well as historical patenting activity (see Section 7).

5.2. Empirical strategy

The first stage model is set out below. For inventor i in area j and yeargroup t, I estimate:

$$\mathrm{PCOUNT}_{ijt} = a + b\mathrm{DIV}_{jt} + \mathbf{VCTRLS}_{jt}c + \mathbf{ICTRLS}_{j}d + \mathrm{I}_i + \mathrm{TF}^{*}\mathrm{YG}_{pt} + \mathrm{e}_i, \qquad (5.1)$$

where PCOUNT_{ijt} is a count of the number of times an inventor engages in patenting during a given 4-year period (patenting activity), the variable of interest is DIV_{jt}, the diversity of active inventors in a given TTWA and time period, and I_i is the individual fixed effect. As movers are constrained to a single location, all area-invariant information is absorbed in the individual fixed effect.[13] The model thus effectively fits inventor-area fixed effects:

$$\mathrm{PCOUNT}_{ijt} = a + b\mathrm{DIV}_{jt} + \mathbf{VCTRLS}_{jt}c + \mathbf{ICTRLS}_{j}d + \mathrm{I}_{ia} + \mathrm{TF}^{*}\mathrm{YG}_{pt} + \mathrm{e}_i \qquad (5.2)$$

For group a in area j in year t, DIV_{jt} is given by:

$$\mathrm{DIV}_{jt} = 1 - \sum\nolimits_a \left[\mathrm{SHARE}_{ajt}\right]^2, \qquad (5.3)$$

where SHARE_{ajt} 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 geographical origin or ethnic groups. Similar measures are used widely in the development literature, as well as some area-level studies (Easterley and Levine, 1997; Alesina and Ferrara, 2005; Ottaviano and Peri, 2005, 2006).

12 My preferred estimator is a negative binomial fixed effects estimator, which should permit me to fit time-invariant individual-level regressors in the stage 1 model: in practice, identification is very unstable and so the two-stage process is preferred.

13 In a linear estimator with both sets of fixed effects, area dummies drop out. The conditional fixed effects negative binomial estimator does allow time-invariant regressors, but adding in a large number of right-hand side dummies to a model with only three time periods is likely to create an 'incidental parameters problem' (Heckman 1981), which in turn leads to inconsistent estimates.

To deal with sectoral and industry patenting shocks, the model includes technology field-by-yeargroup fixed effects (TF^*YG_{pt}), where p indexes shares of patenting in one of the 30 OST-defined technology fields. $\mathbf{VCTRLS}_{jt}$ and $\mathbf{ICTRLS}_{j}$ are vectors of, respectively, time-varying and time-invariant TTWA-level controls covering key spatial, economic and demographic characteristics affecting relationships between DIV and innovation: all controls are for the same 1993–2004 period as the patent data. I use aggregated ONS population and LFS client file microdata to build these.[14]

Patenting and population diversity are spatially concentrated, reflecting benefits from agglomeration that may persist over time (Simmie et al., 2008). Diversity effects on patenting might then simply reflect agglomeration and path-dependence. $\mathbf{ICTRLS}_{j}$ includes a dummy for urban TTWAs, and 1981–1984 area weighted patents to control for historic 'knowledge stocks' (robustness checks explore different lags). $\mathbf{VCTRLS}_{jt}$ includes the log of population density to explore wider agglomeration effects, plus a series of other variables. Inventor demographic characteristics may be entirely explained by area demographic characteristics: for example, places with more diverse populations may produce more diverse inventor groups. I control for this by using area-level fractionalization indices of ONS macro-ethnic groups (and cross-check using migrant population shares). Third, human capital stocks are closely correlated with innovative activity (Romer, 1990) and may account for apparent ethnicity effects on patenting. To deal with this, I fit areas' share of science, technology, engineering and maths (STEM) degree-holders in the local working-age population.

I fit 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 Brinkley (2008).[15] Patenting may also be lower in areas with a lot of entry-level jobs, so I include the share of workers in entry-level occupations as a control. Summary statistics are given in Table 7.

My panel exhibits excess zeroes (63.2%) and slight over-dispersion (the variance of PCOUNT, 1.129, is over twice the mean, 0.529). As the assumptions of the standard Poisson model are not met, I fit the model as a conditional fixed effects negative binomial (Hausman et al., 1984).[16]

14 I aggregate individual-level data to local authority-level averages, and then aggregate these to TTWA-level means using postcode shares. Local Authority Districts (LADs) 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: relevant LAD weights are 0.6, 0.3 and 0.1, respectively. The TTWA-level average of X is given by $X_{TTWA} = 0.6^*(X)_a + 0.3^*(X)_b + 0.1^*(X)_c$.

15 This adjusts OECD definitions for the UK context. The final list of three-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).

16 Hausman tests strongly suggest that the conditional fixed effects estimator is preferred to random effects ($chi^2 = 734.21$, $P = 0.000$). Given the large sample size, a conditional fixed effects estimator is preferred to an unconditional estimator with individual-level dummies.

Table 7. Summary statistics

Variable	*N*	Mean	SD	Min	Max
Inventor patent count/4-year period	210,010	0.536	1.074	0	36
Inventor patents 2–4 times/YG	210,010	0.091	0.288	0	1
Inventor patents at least 5 times/YG	210,010	0.026	0.159	0	1
Inventor patents pre-1993	210,010	0.05	0.218	0	1
Inventor mean patent count pre-1993	210,010	0.028	0.174	0	9.429
Inventor is TTWA mover, same YG	210,010	0.013	0.115	0	1
Inventor moves across TTWAs	210,010	0.025	0.157	0	1
Inventor patents across OST30 fields	210,010	0.096	0.294	0	1
Minority ethnic inventor (geography)	210,010	0.128	0.334	0	1
Minority ethnic inventor (ONS ethnic)	210,010	0.061	0.239	0	1
Inventor UK origin	210,010	0.872	0.334	0	1
Inventor Central Europe origin	210,010	0.026	0.158	0	1
Inventor East Asian origin	210,010	0.022	0.147	0	1
Inventor Eastern Europe origin	210,010	0.011	0.106	0	1
Inventor South Asian origin	210,010	0.026	0.16	0	1
Inventor Southern Europe origin	210,010	0.021	0.142	0	1
Inventor Rest of world origin	210,010	0.022	0.147	0	1
Frac Index, geographic origin groups	210,010	0.215	0.112	0	0.571
Inventor White ethnicity	210,010	0.939	0.239	0	1
Inventor Black Caribbean ethnicity	210,010	0.000	0.007	0	1
Inventor Black African ethnicity	210,010	0.002	0.048	0	1
Inventor Indian ethnicity	210,010	0.018	0.133	0	1
Inventor Pakistani ethnicity	210,010	0.006	0.076	0	1
Inventor Bangladeshi ethnicity	210,010	0.001	0.03	0	1
Inventor Chinese ethnicity	210,010	0.015	0.121	0	1
Inventor Other ethnic group	210,010	0.019	0.136	0	1
Frac Index, ONS ethnic groups	210,010	0.108	0.062	0	0.56
TTWA Frac Index, geo. groups	210,010	0.159	0.117	0.017	0.526
% Graduates	210,010	0.237	0.051	0.09	0.358
% Graduates with STEM degrees	210,010	0.121	0.031	0.035	0.186
% Graduates with PhDs	210,010	0.008	0.007	0	0.031
% Employed high-tech manufacturing	210,010	0.029	0.014	0	0.189
% Employed medium-tech manuf.	210,010	0.045	0.022	0.006	0.154
% In entry-level occupations	210,010	0.34	0.048	0.251	0.521
% Unemployed at least 12 months	210,010	0.015	0.011	0	0.052
Log(population density)	210,010	6.469	0.976	2.06	8.359
Log(TTWA w/patents, 1981–1984)	210,010	4.028	1.439	−1.386	6.543

Source: KITES-PATSTAT/ONS.
Note: Statistics for estimation sample. For reasons of space, country of origin dummies are shown for UK-origin and the six largest minority ethnic groups.

6. Main results

The main results for the first stage model are given in Table 8. The dependent variable is the count of patenting activity, or unweighted patent counts (results for weighted patents are almost identical). The left hand panel shows results for DIV measured with geographic origin zones, my preferred specification; the right hand

Table 8. First stage regression: individual patent counts and inventor group diversity

Inventor patent counts	Geo origin zones			ONS groups		
	(1)	(2)	(3)	(1)	(2)	(3)
Frac Index of inventors	0.075	0.221***	0.248***	0.111	0.312***	0.337***
	(0.100)	(0.020)	(0.023)	(0.165)	(0.011)	(0.014)
Frac Index of TTWA pop			0.028			0.061
			(0.058)			(0.054)
% STEM degrees, TTWA			0.323***			0.308***
			(0.106)			(0.106)
Log of TTWA population density			0.015**			0.010
			(0.007)			(0.007)
% Employed in hi-tech mf (OECD)			0.237			0.107
			(0.164)			(0.149)
% Employed in medium-tech mf (OECD)			−0.106			−0.075
			(0.110)			(0.115)
% Workers in entry-level occupations			−0.053			−0.090**
			(0.036)			(0.042)
Log of area weighted patent stocks (1981–1984)			−0.024***			−0.023***
			(0.006)			(0.007)
Urban TTWA			−0.051***			−0.047***
			(0.015)			(0.015)
ln(alpha)	−1.016***			−1.010***		
	(0.048)			(0.046)		
Individual fixed effect	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
Observations	210,008	210,008	210,008	210,008	210,008	210,008
Log-likelihood	−20,6721.358	−91,887.733	−91,829.454	−206,723.863	−91,913.822	−91,861.933
Chi-squared	167.855	21,597.972	.	169.380	10,830.210	.

Source: KITES-PATSTAT/ONS.
Notes: Constant not shown. Model (1) uses yeargroup dummies. Models (2) and (3) use OST30 technology field*yeargroup dummies. Bootstrapped standards errors are clustered on TTWAs. Results are marginal effects at the mean.
Significant at *10%, **5% and ***1%.

panel repeats the regression using the simpler Index built with ONS macro-ethnic groups. In each case, column 1 shows a bivariate regression for the main variables of interest only, column 2 adds individual fixed effects and column 3 adds controls. Coefficients are presented as marginal effects at the mean. Column 1 indicates a significant log alpha term, confirming over-dispersion. Controls are generally of the expected size and sign. Bootstrapped, cluster-robust standard errors are fitted in all cases.

For geographic origin zones, estimates of DIV in the bivariate regression are small and close to zero (column 1). Including individual fixed effects increases the effect of DIV, which is now significant at 1% (column 2). As expected, model fit is also substantially better. Once controls are added, model fit improves further: the marginal effect of DIV is 0.248, significant at 1%. A 10-point increase in the Fractionalization Index—increasing inventor diversity in Bristol to that in Oxford, for example—would then raise each Bristol inventor's patenting activity by just under 0.025 patents in a given 4-year period. A back-of-the-envelope calculation of the aggregate effect across the area's 1628 inventors is then 40.4 unweighted patents.[17] For DIV measured by ONS groups, the pattern of results is similar but marginal effects of DIV are rather bigger, at 0.337 (also significant at 1%). Interestingly, coefficients of wider population diversity are small and close to zero in the preferred specification, small and positive significant in the ONS models. The urban area dummy is negative, but population density has a positive link to patenting activity. I explore these urban and density connections further in the next section.

To put the main result into perspective, note that effects of DIV are rather smaller than for human capital and technology field-time dummies. For example, the marginal effect of area-level science, engineering, technology and maths degree-holders is 0.323, significant at 1%. That implies that a 10% rise in STEM graduates in Bristol is linked to 0.032 extra patents per inventor (or over 65 unweighted patents at the area level, almost a third larger than the diversity result). This chimes with the existing empirical literature, which suggests that 'diversity effects' are relatively small where they exist.

As a basic crosscheck, I compare the negative binomial estimates with linear fixed effects regressions. Angrist and Pischke (2009) argue that once raw coefficients are converted into marginal effects, non-linear modelling offers little over standard linear regression. OLS regressions give results with a similar sign and significance, but with marginal effects around twice as large. Results are given in Appendix C, Table C1.

6.1. Robustness checks

I conduct a number of robustness checks. Results are summarized in Table 9. I first fit some basic specification checks against the main result (column 1). Some of the inventor geographical origin groups are small, so the Fractionalization Index may be affected by measurement error. Column 2 refits the Index as seven categories, aggregating the six smallest groups into a single 'other' category. Marginal effects of DIV are identical though the model fit changes slightly. I also run a falsification test on ONOMAP. I randomly assign ethnicity, with 'fake' categories following the same underlying structure as the ONOMAP classification, and build a fake Fractionalization Index: if this gives the same results as the ONOMAP Index, it suggests that ONOMAP is no better than random assignment. Results are shown in column 3: fake DIV is −0.050 rather than 0.248, significant at 5% rather than 1%, and with reduced model fit. Inventor diversity effects might also collapse to simple size effects, not least because Fractionalization Indices tend to be highly correlated with group population shares (the pairwise correlation here is 0.779). Column 4 fits the share of minority ethnic inventors; column 5 fits the Fractionalization Index and share together. In both cases, marginal effects of minority ethnic inventor shares are negative, whereas those of DIV stay positive.

Next, I check for omitted variables. Column 6 refits the Equation (5.1) with area-by-technology field-by-yeargroup dummies, which capture localized industry/sector trends. Effects of DIV shrink to 0.231, but remain positive significant. Column 7 fits the model without inventors from London—a city with high levels of cultural diversity; column 8

17 The average weighted patent count per inventor is 0.235, versus 0.535 for unweighted patents. Again, a back of the envelope calculation suggests approximate aggregate *weighted patent* effect of (0.235/0.535)*40.4 = 17.7 weighted patents.

Table 9. Individual patent counts and inventor group diversity, robustness checks

Individual patent counts	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Frac Index of inventors (geo origin groups)	0.248***				0.293***	0.231***	0.268***	0.250***	0.366***	0.020	0.812***	0.248***
	(0.023)				(0.025)	(0.023)	(0.014)	(0.022)	(0.025)	(0.033)	(0.098)	(0.022)
Frac Index of inventors (x7 geo origin groups)		0.248***										
		(0.023)										
Fake Frac Index of inventors (x12 randomized groups)			−0.050**									
			(0.025)									
% Minority ethnic inventors				−0.654***	−1.018***							
				(0.066)	(0.081)							
Urban TTWA dummy	−0.055***	−0.055***	−0.046***	−0.029*	−0.033*	0.001	−0.083***	−0.077***	−0.003	−0.115***	−0.063***	−0.058***
	(0.018)	(0.018)	(0.018)	(0.017)	(0.017)	(0.019)	(0.013)	(0.019)	(0.014)	(0.026)	(0.018)	(0.009)
Frac Index of inventors * urban TTWA										0.285***		
										(0.023)		
% STEM degrees, TTWA	0.323***	0.321***	0.306***	0.349***	0.411***	−1.429***	−0.052		−1.318***	0.313***	0.187*	0.306**
	(0.106)	(0.106)	(0.106)	(0.107)	(0.103)	(0.055)	(0.092)		(0.059)	(0.106)	(0.106)	(0.137)
% PHDs, TTWA								2.872***				
								(0.210)				
Log of TTWA population density	0.015**	0.015**	0.011	0.007	0.009	−0.009	0.020***	0.032***	−0.006	0.019***	0.029***	0.016*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.009)
Frac Index of inventors * log of TTWA pop density											−0.259***	
											(0.067)	
Log of area weighted stock of patents (1989–1992)												−0.025***
												(0.004)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	210,008	210,008	210,008	210,008	210,008	210,008	188,786	210,008	210,008	210,008	210,008	210,008
Log-likelihood	−91,829.454	−91,830.107	−91,940.012	−91,911.029	−91,758.801	−91,853.839	−82,718.388	−91,796.241	−92,174.193	−91,799.246	−91,812.706	−93,888.356

Source: KITES-PATSTAT/ONS.
Notes: Controls as in Table 7. Bootstrapped standard errors in parentheses, clustered on TTWAs. Results are marginal effects at the mean. Significant at *10%, **5% and ***1%.

fits the area share of PHD-holders as an alternative area-level human capital control. Removing London raises the effect of DIV to 0.268; switching to PHDs also raises estimates of DIV, to 0.250. Both are significant at 1%. Column 9 adds the share of 'stars' in the TTWA inventor population, where stars are defined as inventors patenting at least five times during a given period. This raises the marginal effect of DIV from 0.248 to 0.366 and is still significant at 1%.[18]

I then test for urban amplifying effects. Minority ethnic inventors are spatially concentrated in urban locations: as discussed in Section 2, agglomeration economies might generate some of the diversity result. Columns 10 and 11 test for amplifying effects of urban and high-density areas, respectively, fitting interactions of the Fractionalization Index with the urban TTWA dummy and with logged population density. In the first case, the effect of DIV alone falls to zero, but the joint effect of urban DIV is 0.285, significant at 1%. Effects of urban status remain negative, as before. In the second case, estimates of DIV grow substantially to 0.812, whereas the joint effect of DIV and population density is negative at −0.259. Population density marginal effects are 0.029, larger than in the main regressions. All are significant at 1%. Together, this suggests an amplifying effect of urban areas, which disappears in the biggest and most dense cities. This may partly reflect the spatial distribution of minority ethnic inventors, who are most densely clustered in second tier cities and university towns, rather than the largest urban cores. Note also that removing London-based inventors raises marginal effects of inventor diversity, which is compatible with these results.

Finally, I check for appropriate historical settings. If the historic patent stocks term in the main model is mis-specified, path-dependence will not be adequately controlled for. Column 12 shows results for the most conservative specification (when the lag is dropped to the 4-year period before the sample). Effects of DIV barely change, and results for other lags also show no change.

I also conduct three further structural tests. First, my results might be particular to the choice of time period, in which the UK experienced substantial rises in net migration and minority ethnic populations (Graph 1). To test this I run a reduced-form model on the full set of inventors active between 1981 and 2004, and on the sub-group active between 1981 and 1992. Results (Appendix C, Table C2) show positive significant effects of DIV in the long sample: in the earlier period, DIV is non-significant and close to zero. National demographic changes, then, help explain my results.

Next, I reconstruct my sample by blanking all inventor-yeargroup cells when an inventor is not patenting. This is a more conservative way of treating inactive inventors, and will deal with any measurement error introduced by zeroing. My choice of estimator means that blanking out non-activity has the effect of restricting the sample to inventors who patent more than once. I compare estimates for multiple inventors across two different samples, one with zeroed and one with missing observations for non-activity. Reduced-form results show that estimates for the two sub-samples are identical (Appendix C, Table C3). This strongly suggests that sample construction has no effect on my main findings.

18 I exclude inventors who are themselves stars, so as to capture any effect of the presence of stars around that inventor. I also run tests for the sum of stars, the sum of multiple inventors (inventing more than once) and the share of multiple inventors, none of which change my main result.

Finally, I transform the model into a wholly area-level specification: this loses individual fixed effects but allows for an alternative estimation of aggregate effects. I collapse the panel to area level and estimate:

$$\mathrm{Y}_{jt} = \mathrm{a} + b\mathrm{DIV}_{jt} + \mathbf{VCTRLS}c_{jt} + \mathrm{A_j} + \mathrm{YG} * \mathrm{TF}_{pt} + \mathrm{e}_{jt}, \tag{6.4}$$

where Y is the total count of unweighted patents for area j in yeargroup t, A is the area-level fixed effect and all other terms are defined as in Equation (5.2). The two models are not identical, and we should expect estimates of b to differ: Equation (6.4) substitutes area fixed effects for individual-area fixed effects, and this loses important variation, as the main results suggest that individual characteristics help drive patenting. Sample construction is also different; in the individual panel DIV is effectively 'weighted' across inventor populations in each area, whereas the area-level panel cleans this out (means of DIV differ quite a lot, at 0.213 for the individual panel and 0.109 for the area panel).

I estimate Equation (6.4) in OLS, with Poisson results included for comparison (as shares of zeroes are low and mean–variance assumptions are met). Results are shown in Appendix C, Table C4. In the OLS model the beta of DIV is 335.48, which implies that a 0.1 shift in area DIV is linked to 33.5 extra patents in that area. This compares to a (rough) aggregate effect of 40.4 patents from the individual-level model. This suggests that (i) my main result holds in an area-level specification, (ii) this specification misses out salient individual-level factors and (iii) sample construction issues may also be in play. Area-level results should also be treated as associations: unobserved area-level factors might affect aggregate patenting (but not individual inventors). For all these reasons, my main, individual-level results are preferred.

6.2. Moving inventors

If inventors select into high-innovation clusters that help them become more productive, this might create upwards bias on coefficients of DIV or, in extremis, explain the result entirely. To explore this issue, I use information from the KITES-PATSTAT cleaning process to identify inventors who move between TTWAs (see Appendix A). The group of movers comprises 1781 individuals (around 2.5% of the sample), of who 963 (1.33%) move within the same yeargroup. I then run a series of checks on the influence of movers. First, I re-assign movers from their first to their second locations and re-run model (Equation 5.2), with almost no change to coefficients of DIV (see Appendix C, Table C5). Next, I manually examine mover origin and destination points. Specifically, I look for whether moves are between contiguous TTWAs or across greater distances. Contiguous moves, especially from an urban to a rural TTWA might suggest lifecycle-related relocation, for example a new family moving from a city to a less dense area. Moves across greater distances might suggest job-related motives. I find that over 90% of moves are between contiguous TTWAs (for example, Cambridge–Huntingdon, Reading–Newbury, Middlesborough and Stockton–Hartlepool–Bishop Auckland).

Finally, I construct an area-level panel and regress the change in area-level weighted patent counts on the change in movers to a given TTWA. For TTWA j, I estimate:

$$\Delta\mathrm{WPATENTS}_j = a + b\Delta\mathrm{MOVERS}_j + \Delta\mathbf{VCTRLS}c_j + \mathrm{e}_j, \tag{6.5}$$

Table 10. Testing for the role of moving inventors in the first stage model

% Change in total weighted patents, 1993–2004	(1)	(2)	(3)	(4)
% Change in moving inventors	0.056**	0.050*	0.082**	0.082**
	(0.028)	(0.026)	(0.037)	(0.038)
% Change, TTWA Fractionalization Index		−0.521	−0.355	−0.361
		(0.335)	(0.255)	(0.256)
% Change, TTWA STEM degrees		0.893	1.202	1.192
		(0.726)	(0.754)	(0.756)
% Change, TTWA high-tech manufacturing		0.848	0.564	0.552
		(0.793)	(0.894)	(0.891)
% Change, TTWA medium-tech manufacturing		0.169	0.573	0.574
		(0.505)	(0.366)	(0.370)
% Change, TTWA population density		10.445	12.189	
		(16.729)	(15.488)	
% Change, TTWA entry-level occupations		−1.130	−0.454	−0.713
		(1.088)	(1.180)	(1.201)
OST30 technology field effects	N	N	Y	Y
Observations	206	202	198	198
F-statistic	3.989	1.707	2.824	2.753
R^2	0.003	0.096	0.318	0.317

Source: KITES-PATSTAT/ONS.
Notes: Standard errors are in parentheses, are heteroskedasticity and autocorrelation-robust and clustered on TTWAs.
Significant at *10%, **5% and ***1%.

where

$$\Delta \text{WPATENTS}_j = (\text{WPATENTS}_{j2004} - \text{WPATENTS}_{j1993})/\text{WPATENTS}_{j1993}. \quad (6.6)$$

And WMOVERS_j is assembled similarly. **VCTRLS** contains the same set of area-level variables from model (5.2), with time-varying variables expressed as percentage changes. This horse-race setting allows me to test the relative contribution of movers to overall patenting. A large and significant value of b compared with c would suggest that positive selection is an issue at the area level (although these are associations, not causal effects). Results are given in Table 10. I find small, positive significant coefficients of movers on changes in area patenting (0.082, significant at 5%) but these are dwarfed by changes in other area-level characteristics (such as STEM degrees and high-tech manufacturing) that are fitted as controls in the main model. For instance, a 10% rise in moving inventors is linked to a 0.8% rise in total patenting; a similar increase in STEM degrees is associated with an 11.9% rise. This also suggests that impacts of movers at the area level on *individual* inventors' outcomes are likely to be minimal.

7. Extensions

7.1. Minority ethnic status and co-ethnic group membership

The second stage analysis explores roles of minority ethnic status and co-ethnic group membership in individuals' patenting activity in more detail. To do this, I retrieve estimates of the individual fixed effects from Equation (5.2) and regress these on observable individual characteristics. The fixed effects are capturing all time-invariant individual factors, which may include ethnicity elements (see Section 2.2). I therefore aim to separate coefficients of minority ethnic status, group membership and other salient individual-level factors (such as human capital and previous experience). These results are associations, not causal links. Note that because I do not observe how individual fixed effects are scaled, I am unable to interpret point estimates in relation to the dependent variable.[19] However, I am able to discuss the sign and significance of the independent variables, as well as their sizes relative to each other.

Specifically, I estimate the following cross-sectional model for inventor i:

$$\mathrm{IHAT}_i = a + \mathbf{ETH}b_i + \mathrm{cMULTIPLE}_i + d\mathrm{STAR}_i + \mathrm{PRE}_i + e\mathrm{PRECOUNT}_i + u_i, \quad (7.7)$$

where IHAT_i is the estimated fixed effect and $\mathbf{ETH}_{\boldsymbol{i}}$ is either a dummy for minority ethnic status, or a vector of co-ethnic group dummies. In the latter case I take UK origin as the reference category and estimate coefficients of the five largest minority ethnic groups, aggregating the six smaller groups into a 'rest of the world' category. Control variables are dummies for inventors who patent between two and four times in a given yeargroup ($\mathrm{MULTIPLE}_i$), over five times (STAR_i), plus two controls which use historic patenting activity to approximate human capital characteristics. (Note that as IHAT is derived from a patent counts regression, results using MULTIPLE and STAR have to be interpreted with caution.) Historic patenting controls draw on a widely used approach developed by Blundell et al. (1995), who argue that agents' capacity to innovate is largely explained by their cumulatively generated knowledge at the point in which they enter a sample. With long enough time-series data, pre-sample activity thus approximates agent-level human capital. Following this logic, I fit a dummy for whether inventors patented in the pre-1993 period (PRE_i), and for those that did, $\mathrm{PRECOUNT}_i$, is the mean of historic patenting activity. As before, summary statistics are given in Table 7 (top panel).

I estimate the model in OLS, using bootstrapped standard errors to deal with heteroskedasticity arising from first stage sampling error.[20] Results are set out in Table 11: Feasible Generalised Least Squares (FGLS) regressions give almost identical coefficients (see Appendix C, Table C6). Coefficients of minority ethnic status are negative and significant at 1% in all specifications; by contrast, pre-sample patenting activity has a positive link, also significant at 1% (with a significant 'penalty' for those not patenting pre-sample). Multiple and 'star' inventors also show positive coefficients, significant at 1%. Estimates of minority status are substantially smaller than these latter two variables.

Columns 2 through 4 fit interactions of minority ethnic status with multiple and star inventor status. The latter finds positive joint coefficients, which are net positive and 10% significant (columns 3 and 4). This is in line with the earlier descriptive analysis, and suggests that individual-level links between minority ethnic status and patenting exist, at least for higher-patenting inventors, even after human capital is controlled for.

Table 12 explores further for the five largest co-ethnic groups, plus a rest of the world group. Coefficients should be interpreted as associations and as relative to UK origin, the reference category. Co-ethnic group membership coefficients are negative significant

19 Results are also robust to using fixed effects derived from the OLS regressions.
20 A Breusch–Pagan test on the basic OLS regression gives a Chi^2 statistic of 63.98 ($P = 0.000$), suggesting that heteroskedasticity is present.

Table 11. Second stage regressions: decomposing fixed effect estimates from first stage

Inventor fixed effects (estimated)	(1)	(2)	(3)	(4)
Minority ethnic inventor (geo groups)	−0.199***	−0.201***	−0.206***	−0.209***
	(0.010)	(0.011)	(0.010)	(0.011)
Inventor patents 2–4 times (multiple)	1.097***	1.095***	1.097***	1.093***
	(0.019)	(0.019)	(0.019)	(0.019)
Minority ethnic * multiple inventor		0.022		0.040
		(0.064)		(0.062)
Inventor patents at least 5 times (star)	3.695***	3.695***	3.664***	3.663***
	(0.059)	(0.059)	(0.061)	(0.061)
Minority ethnic * star inventor			0.320*	0.325*
			(0.192)	(0.191)
Average patenting, pre-1993	0.199***	0.199***	0.202***	0.202***
	(0.076)	(0.076)	(0.076)	(0.076)
Dummy, inventor patents pre-1993	−0.113**	−0.113**	−0.113**	−0.113**
	(0.044)	(0.044)	(0.044)	(0.044)
Constant	−0.170***	−0.169***	−0.169***	−0.168***
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	70,007	70,007	70,007	70,007
R^2	0.253	0.253	0.253	0.253

Source: KITES-PATSTAT/ONS.
Notes: Robust standard errors in parentheses, bootstrapped, 50 repetitions.
Significant at *10%, **5% and ***1%.

as before; joint effects of most co-ethnic group stars are positive, and are 10% significant for East Asian-origin stars. Cross-checking using ONS ethnic groups finds a stronger result for Chinese star inventors (1.639, significant at 5%). There is some variation in coefficient size between co-ethnic groups, suggestive of differing diaspora resources and capacity.

I then run a series of robustness tests. I first check for omitted variables, fitting dummies for moving inventors and for inventors patenting across at least two OST30 fields (a measure of 'generalists' that captures intellectual range). I also fit the count of within-sample patenting alongside historic patent counts. Results show minimal change compared with my main findings. Next, I run a falsification test on main results with fake ethnic group dummies generated by random assignment. Coefficients of 'fake' minority ethnic and co-ethnic group variables are generally non-significant, and model fit is substantially worse. Results are shown in Appendix C, Tables C6 and C7.

7.2. Distributional analysis

Finally, I briefly explore potential impacts of minority ethnic inventors on majority groups. This might involve physical outflows, in which UK-origin inventors leave an area after minority groups arrive (Borjas, 1994), or 'resource crowd-out', in which minority ethnic inventors displace majority inventors from jobs or (say) lab space (Borjas and Doran, 2012). Analysis of moving inventors suggests that they have minimal impact on the main results. However, resource crowd-out could co-exist with

Table 12. Second stage regressions, co-ethnic groups

Inventor fixed effects (estimated)	(1)	(2)
Inventor South Asian origin	−0.314***	−0.310***
	(0.021)	(0.020)
Star * South Asian		−0.219
		(0.277)
Inventor Central Europe origin	−0.112***	−0.117***
	(0.019)	(0.021)
Star * Central European		0.256
		(0.485)
Inventor East Asian origin	−0.142***	−0.157***
	(0.027)	(0.025)
Star * East Asian		1.053*
		(0.576)
Inventor Southern Europe origin	−0.175***	−0.183***
	(0.030)	(0.030)
Star * Southern European		0.359
		(0.408)
Inventor Eastern Europe origin	−0.112***	−0.127***
	(0.029)	(0.029)
Star * Eastern European		0.559
		(0.575)
Inventor rest of world origin	−0.289***	−0.298***
	(0.027)	(0.025)
Star * Rest of world		0.380
		(0.546)
Inventor patents at least 5 times (star)	3.695***	3.663***
	(0.060)	(0.061)
Controls	Y	Y
Observations	70,007	70,007
R^2	0.254	0.254

Source: KITES-PATSTAT/ONS.
Notes: Constant not shown. Controls as in Table 10, plus multiple inventor dummy. All models use bootstrapped standard errors, 50 repetitions.
Significant at *10%, **5% and ***1%.

externalities at the inventor group level. At the extreme, 'minority ethnic' patents might wholly explain increases in area-level patent counts. Conversely, there might be multiplier effects from minority ethnic inventors to majority group inventors, leading to crowd-in (Hunt and Gauthier-Loiselle, 2010).

To explore, I draw on work by Card (2010) and Faggio and Overman (2014). I define 'minority' patents as those with at least one minority ethnic inventor; all other patents are 'majority' patents. I assemble a panel of TTWA-level weighted patent counts for 1993–2004 and regress the percentage change in majority patents on that in minority patents, with both expressed as a share of *all* patenting in the base year. Specifically, for TTWA j I estimate:

$$((\mathrm{MP}_{j04} - \mathrm{MP}_{j93})/\mathrm{TP}_{j93}) = a + b((\mathrm{EP}_{j04} - \mathrm{EP}_{j93})/\mathrm{TP}_{j93}) + \mathbf{CTRLS}c_{jtbase} + \mathrm{e}_j, \quad (7.8)$$

where MP refers to majority patents, EP refers to minority ethnic patents, and **CTRLS** is a vector of area-level controls for the base period 1993, including the previous stock of weighted patents. The coefficient b expresses the relationship between majority and minority patenting. If b is 0, a 1-unit change in minority patenting has no consequences for majority patenting, simply adding 1 to total patenting. Estimates above 0 indicate multiplier effects of size b, resulting in a more-than-proportionate increases in total patenting. Conversely, estimates below 0 indicate crowding-out.

Results are given in Appendix C, Table C8. It is important to emphasize that these should be interpreted as partial correlations, not as causal links. Unobserved factors such as area-level shocks may influence both sides of the equation—and running the regression in reverse also indicates some connections from majority to minority patents. In fully specified form, results from Equation (7.8) give b at around 1.9, significant at 1%. This suggests that each additional minority patent is linked to just almost two additional majority patents, implying a multiplier 'effect'. However, the confidential interval is between 0.92 and 2.22, so the connection is not observed with much precision, and omitted variables are also likely to be in play. Coefficients should thus be interpreted with caution.[21]

8. Conclusions

In recent years, there has been growing interest in the links between minority ethnic communities, diversity and innovation. The contribution of minority ethnic inventors and 'ethnic entrepreneurs' to US innovation is substantial, suggesting that European countries' innovation systems could also benefit from these groups' presence and activity.

This article looks at the role of minority ethnic inventors on innovative activity in the UK, using a new 12-year panel of patents microdata and a powerful name-classification system. I uncover some distinctive features of the UK inventor community, and explore different potential 'ethnicity–innovation' channels—individual selection, externalities from diasporic groups and from the cultural diversity of inventor communities, as well as 'amplifying' roles 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 of its kind outside the USA.

The descriptive analysis suggests that the UK's minority ethnic inventor community has a few important commonalities with the USA—with large South and East Asian-origin groups, plus groups of multiple and star inventors who patent significantly more than majority counterparts. Minority inventors patent most often in semi-conductors, IT, pharmaceutical and agriculture/food fields: these modal shares are somewhat higher than majority inventors'. I also find differences: UK inventor demographics reflect proximity to Continental Europe, colonial history and recent immigration trends. Minority ethnic inventors are spatially clustered, as in the USA, but seem to follow a different distribution from wider minority populations. Not all high-patenting regions have diverse inventor communities.

Regressions find a small, positive effect of inventor group diversity on individual patenting activity, which is not driven by inventor mobility or the crowding-out of majority inventors (rather, I find suggestive evidence of crowding-in from minority to majority patenting). This suggests that learning externalities exist for diverse inventor

21 I experiment with lags of minority ethnic patents as an instrument, but none pass first-stage tests.

groups, over and above simple size/co-location effects. Tests also suggest an amplifying role of urban location, but this dies away in the densest environments where minority inventors are less clustered than the wider population.

Do inventor characteristics such as human capital or co-ethnic group membership help explain the diversity result? Some tentative positive associations emerge for minority ethnic and East Asian-origin stars, especially those of Chinese ethnicity (the latter both relatively large groups in the UK inventor community). This suggests the existence of network externalities within (some) diasporic groups, which may operate as a complement to the across-group effect. I speculate that stars might also generate substantive knowledge spillovers, as well as having a motivating effect on those around them: minority stars patent significantly more than their majority counterparts. Certainly, larger shares of star inventors in an area increase the diversity effect, suggesting that these channels operate as complements.

Overall, the results suggest that minority ethnic inventors are a net positive for patenting in the UK, and imply that policymakers should aim to increase both the skills and the mix of the country's research communities. They also highlight some distinctive features of the UK innovation system. In the USA, minority ethnic inventor communities have been historically shaped by Cold War science, which attracted very large numbers of skilled workers into a small number of high-tech locations (Saxenian, 2006). By contrast, until recently 'calls' for migrant workers in the UK have focused on less skilled occupations and on Commonwealth countries, especially in Africa and South/East Asia (Somerville, 2007). 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 et al., 2007), and by the complex interplay between class, skills, resources and attitudes that influence real-world entrepreneurial behaviour (Clark and Drinkwater, 2010). The rigidities of some European labour markets could also explain UK inventor demographics, as young researchers seek new opportunities in more open environments.[22]

There are two important caveats to the results. First, diversity and diaspora effects are relatively small—human capital and technology effects are more important determinants of inventors' productivity. This is intuitive, and echoes much of the existing literature. Second, working with inventor data presents a number of measurement challenges: most seriously, my data only allow a fuzzy identification of ethnic inventors and diasporic groups.

This leaves a number of areas for future research. We need to better understand what is driving these results—not least, the scale(s) at which the diversity effect is operating (teams, departments, communities of interest). Understanding the quality and influence of minority patenting (for example, through citations data) is also a priority. Better individual-level data would allow the identification of migrants, as well as revealing other salient characteristics (such as age, gender, qualifications, experience): linking inventor information to academic or professional curricula vitae (CVs) would be one way to achieve this. Research could also explore the detailed roles of minority inventors in the technology fields where they are most active, and in specific locations where they are clustered. Finally, the analysis should be extended to other European countries.

22 Thanks to a referee for this last point.

Acknowledgements

Thanks to Gianluca Tarasconi and Francesco Lissoni at Università Bocconi/APE-INV for KITES-PATSTAT patent data, to Pablo Mateos at UCL for ONOMAP coding, and to the ONS Secure Data Service for LFS datasets. Thanks for advice and support to Alex Bryson, Maria Carvalho, Riccardo Crescenzi, Giulia Faggio, Ben Gidley, Ian Gordon, Simona Iammarino, Bill Kerr, Carlo Menon, Henry Overman, Rebecca Riley, Rosa Sanchis-Guarner, Carlos Vargas-Silva and Vivek Wadhwa. Seminar participants at IZA, EUROLIO, COMPAS, SERC, the Urban Economics Association and the Regional Studies Association all provided helpful comments on previous versions, as did three anonymous referees. This work includes analysis based on data from the Labour Force Survey, produced by the Office for National Statistics (ONS) and supplied by the Secure Data Service at the UK Data Archive. The data are 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 or the Secure Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research datasets that may not exactly reproduce National Statistics aggregates. All the outputs have been granted final clearance by the staff of the SDS-UKDA. 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. Any remaining errors and omissions are my own.

Funding

This paper draws on PHD research funded by the UK Economic and Social Research Council (ESRC) and the Department of Communities and Local Government (DCLG). I am grateful to the ESRC and the DCLG for this support. The views in this paper are my own, and do not necessarily represent those of the Department or the ESRC.

References

Abramovsky, L., Griffith, R., Macartney, G., Miller, H. (2008) *The Location of Innovative Activity in Europe*. IFS Working Papers WP08/10. London: Institute for Fiscal Studies.

Agrawal, A., Cockburn, I., McHale, J. (2006) Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, 6: 571–591.

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

Alesina, A., Ferrara, E. L. (2005) Ethnic diversity and economic performance. *Journal of Economic Literature*, 43: 762–800.

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

Aspinall, P. (2009) The future of ethnicity classifications. *Journal of Ethnic and Migration Studies*, 35: 1417–1435.

Audretsch, D., Feldman, M. (1996) R&D spillovers and the geography of innovation and production. *American Economic Review*, 86: 630–640.

Berliant, M., Fujita, M. (2008) Knowledge creation as a square dance on the Hilbert cube. *International Economic Review*, 49: 1251–1295.

Berliant, M., Fujita, M. (2009) Dynamics of knowledge creation and transfer: the two person case. *International Journal of Economic Theory*, 5: 155–179.

Blundell, R., Griffith, R., Van Reenen, J. (1995) Dynamic count data models of technological innovation. *The Economic Journal*, 105: 333–344.

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: 1667–1717.

Borjas, G., Doran, K. (2012) The collapse of the Soviet Union and the productivity of American mathematicians. *Quarterly Journal of Economics*, 127: 1143–1203.

Brinkley, I. (2008) *The Knowledge Economy: How Knowledge Is Reshaping the Economic Life of Nations*. London: The Work Foundation.
Caldwell, C. (2009) *Reflections on the Revolution in Europe: Immigration, Islam and the West*. London: Allen Lane.
Card, D. (2010) How immigration affects US cities. In R. P. Inman (ed.) *Making Cities Work: Prospects and Policies for Urban America*, pp. 158–200. Princeton: Princeton University Press.
Chellaraj, G., Maskus, K., Mattoo, A. (2008) The contribution of international graduate students to U.S. innovation. *Review of International Economics*, 16: 444–462.
Clark, K., Drinkwater, S. (2010) Recent trends in minority ethnic entrepreneurship in Britain. *International Small Business Journal*, 28: 136–146.
Combes, P.-P., Duranton, G., Gobillon, L. (2008) Spatial wage disparities: sorting matters! *Journal of Urban Economics*, 63: 723–742.
Docquier, F., Rapoport, H. (2012) Globalization, brain drain, and development. *Journal of Economic Literature*, 50: 681–730.
Duleep, H. O., Jaeger, D., Regets, M. (2012) *How Immigration May Affect U.S. Native Entrepreneurship: Theoretical Building Blocks and Preliminary Results*. IZA Discussion Paper 6677. Bonn: IZA.
Easterley, W., Levine, R. (1997) Africa's growth tragedy: policies and ethnic divisions. *Quarterly Journal of Economics*, 112: 1203–1250.
Fagerberg, J. (2005) Innovation: a guide to the literature. In J. Fagerberg, D. Mowery, R. Nelson (eds) *The Oxford Handbook of Innovation*, pp. 1–27. Oxford: OUP.
Faggio, G., Overman, H. (2014) The effect of public sector employment on local labour markets. *Journal of Urban Economics*, 79: 91–107.
Foley, C. F., Kerr, W. R. (2013) Ethnic innovation and U.S. multinational firm activity. *Management Science*, 59: 1529–1544.
Freeman, C. (1987) *Technology Policy and Economic Policy: Lessons from Japan*. London: Pinter.
Fujita, M., Weber, S. (2003) *Strategic Immigration Policies and Welfare in Heterogenous Countries*. Institute of Economic Research Working Papers. Kyoto: Kyoto University.
Gagliardi, L. (2011) *Does Skilled Migration Foster Innovative Performance? Evidence from British Local Areas*. SERC Discussion Paper DP0097. London: SERC.
Gibbons, S., Overman, H. G., Resende, G. (2011) *Real Earnings Disparities in Britain*. SERC Discussion Paper DP0065. London: SERC.
Gordon, I., Whitehead, C., Travers, T. (2007) *The Impact of Recent Immigration on the London Economy*. London: City of London Corporation.
Hall, B., Jaffe, A., Trajtenberg, M. (2001) The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. Cambridge, MA: NBER.
Hanks, P., Tucker, D. K. (2000) A diagnostic database of American personal names. *Names: A Journal of Onomastics*, 48: 59–69.
Hanson, G. H. (2012) Immigration, productivity and competitiveness in American industry. *Paper Presented at American Enterprise Institute Conference Competing for Talent: The United States and High-Skilled Immigration*. Washington, DC, 31 January.
Hausman, J. A., Hall, B. H., Griliches, Z. (1984) Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52: 909–938.
Heckman, J. (1981) The incidental parameters problem and the problem of initial condition in estimating a discrete time-discrete data stochastic process. In C. F. Manski, D. L. McFadden (eds) *Structural Analysis of Discrete Data and Econometric Applications*, pp. 179–197. Cambridge, MA: MIT Press.
Hong, L., Page, S. (2001) Problem solving by heterogeneous agents. *Journal of Economic Theory*, 97: 123–163.
Hong, L., Page, S. (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: 16385–16389.
Hoogendoorn, S., Oosterbeek, H., van Praag, M. (2013) The impact of gender diversity on the performance of business teams: evidence from a field experiment. *Management Science*, 59: 1514–1528.
Hunt, J. (2011) Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. *Journal of Labor Economics*, 29: 417–457.

Hunt, J., Gauthier-Loiselle, M. (2010) How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, 2: 31–56.
Jacobs, J. (1969) *The Economy of Cities*. London: Vintage.
Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108: 577–598.
Javorcik, B. S., Özden, Ç., Spatareanu, M., Neagu, C. (2011) Migrant networks and foreign direct investment. *Journal of Development Economics*, 94: 231–241.
Kaiser, U., Kongsted, H. C., Rønde, T. (2011) *Labor Mobility, Social Network Effects, and Innovative Activity*. IZA Discussion Paper 5654. Bonn: IZA.
Kapur, D., McHale, J. (2005) Sojourns and software: internationally mobile human capital and high tech industry development in India, Ireland and Israel. In A. Arora, A. Gambardella (eds) *From Underdogs to Tigers: The Rise and Growth of the Software Industry in Brazil, China, India, Ireland and Israel*, pp. 236–274. Oxford: OUP.
Kerr, W. (2008a) Ethnic scientific communities and international technology diffusion. *Review of Economics and Statistics*, 90: 518–537.
Kerr, W. (2008b) *The Agglomeration of US Ethnic Inventors*. HBS Working Paper 09-003. Boston, MA: Harvard Business School.
Kerr, W. (2009) *Breakthrough Innovations and Migrating Clusters of Innovation*. NBER Working Paper 15443. Cambridge, MA: NBER.
Kerr, W. (2010a) The agglomeration of US ethnic inventors. In E. Glaeser (ed.) *Agglomeration Economics*, pp. 237–276. Chicago: University of Chicago Press.
Kerr, W., Lincoln, W. (2010) *The Supply Side of Innovation: H-1b Visa Reforms and US Ethnic Invention*. NBER Working Paper 15768. Cambridge, MA: NBER.
Kerr, W. R. (2010b) Breakthrough inventions and migrating clusters of innovation. *Journal of Urban Economics*, 67: 46–60.
Kugler, M., Rapoport, H. (2007) International labor and capital flows: complements or substitutes? *Economics Letters*, 94: 155–162.
Lakha, F., Gorman, D., Mateos, P. (2011) Name analysis to classify populations by ethnicity in public health: validation of Onomap in Scotland. *Public Health*, 125: 688–696.
Leadbeater, C. (2008) *The Difference Dividend: Why Immigration is Vital to Innovation*. NESTA Provocation. London: NESTA.
Legrain, P. (2006) *Immigrants: Your Country Needs Them*. London: Little, Brown.
Li, X., Pai, Y. (2010) The changing geography of innovation activities: what do patent indicators imply? In X. Fu, L. Soete (eds) *The Rise of Technological Power in the South*, pp. 69–88. Basingstoke: Palgrave MacMillan.
Lissoni, F., Tarasconi, G., Sanditov, B. (2006) *The KEINS Database on Academic Inventors: Methodology and Contents*. CESPRI Working Paper 181. Milan: Universita' Bocconi.
Malchow-Møller, N., Munch, J. R., Skaksen, J. R. (2011) *Do Foreign Experts Increase the Productivity of Domestic Firms?* IZA Discussion Paper 6001. Bonn: IZA.
Maré, D. C., Fabling, R. (2011) *Productivity and Local Workforce Composition*. Motu Working Paper 11-10. Wellington, NZ: Motu Economic and Public Policy Research.
Maré, D. C., Fabling, R., Stillman, S. (2011) *Immigration and Innovation*. IZA Discussion Paper 5686. Bonn: IZA.
Mateos, P. (2007) A review of name-based ethnicity classification methods and their potential in population studies. *Population, Space and Place*, 13: 243–263.
Mateos, P., Longley, P., O'Sullivan, D. (2011) Ethnicity and population structure in personal naming networks. *PLoS One*, 6: e22943.
Mateos, P., Webber, R., Longley, P. (2007) *The Cultural, Ethnic and Linguistic Classification of Populations and Neighbourhoods Using Personal Names*. CASA Working Paper. London: UCL.
McCann, P., Ortega-Arguilés, R. (2013) Modern regional innovation policy. *Cambridge Journal of Regions, Economy and Society*, 6: 187–216.
Menon, C. (2009) *Stars and Comets: An Exploration of the Patent Universe*. SERC DP0037. London: LSE.
Nathan, M., Lee, N. (2013) Cultural diversity, innovation and entrepreneurship: firm-level evidence from London. *Economic Geography*, 89: 367–394.

Niebuhr, A. (2010) Migration and innovation: does cultural diversity matter for regional R&D activity? *Papers in Regional Science*, 89: 563–585.
Oaxaca, R. L., Geisler, I. (2003) Fixed effects models with time invariant variables: a theoretical note. *Economics Letters*, 80: 373–377.
OECD, (2009) *OECD Patent Statistics Manual*. Paris: OECD.
Office of National Statistics, (2012) *Ethnicity and National Identity in England and Wales 2011*. Newport: ONS.
Office of National Statistics, (2013) *Labour Force Survey 1993–2011 SN: 6727. Secure Data Service Access [computer file]*. Colchester: UK Data Archive.
Ostergaard, C. R., Timmermans, B., Kristinsson, K. (2011) Does a different view create something new? The effect of employee diversity on innovation. *Research Policy*, 40: 500–509.
Ottaviano, G., Bellini, E., Maglietta, A. (2007) *Diversity and the Creative Capacity of Cities and Regions*. SUSDIV Paper 2.2007. Bologna: FEEM.
Ottaviano, G., Peri, G. (2005) Cities and cultures. *Journal of Urban Economics*, 58: 304–337.
Ottaviano, G., Peri, G. (2006) The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography*, 6: 9–44.
Ozgen, C., Nijkamp, P., Poot, J. (2011) *The Impact of Cultural Diversity on Innovation: Evidence from Dutch Firm-Level Data*. IZA Discussion Paper 6000. Bonn: IZA.
Ozgen, C., Nijkamp, P., Poot, J. (2012) Immigration and innovation in European regions. In P. Nijkamp, J. Poot, M. Sahin (eds) *Migration Impact Assessment: New Horizons*. Cheltenham: Edward Elgar.
Page, S. (2007) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies*. Princeton: Princeton University Press.
Parrotta, P., Pozzoli, D., Pytlikova, M. (2013) The nexus between labor diversity and firm's innovation. *Journal of Population Economics*, 27: 303–364.
Peri, G. (2007) Higher education, innovation and growth. In G. Brunello, P. Garibaldi, E. Wasmer (eds) *Education and Training in Europe*, pp. 56–70. Oxford: Oxford University Press.
Petersen, J., Longley, P., Gibin, M., Mateos, P., Atkinson, P. (2011) Names-based classification of accident and emergency department users. *Health & Place*, 17: 1162–1169.
Putnam, R. (2007) E pluribus unum: diversity and community in the twenty-first century. *Scandinavian Political Studies*, 30: 137–174.
Rauch, J. E., Casella, A. (2003) Overcoming informational barriers to international resource allocation: prices and ties. *The Economic Journal*, 113: 21–42.
Rauch, J. E., Trindade, V. (2002) Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84: 116–130.
Robson, B., Barr, R., Lymperopoulou, K., Rees, J., Coombes, M. (2006) *A Framework for City-Regions*. Working Paper 1: Mapping City-Regions. London: Office of the Deputy Prime Minister.
Rodríguez-Pose, A., Storper, M. (2006) Better rules or stronger communities? On the social foundations of institutional change and its economic effects. *Economic Geography*, 82: 1–25.
Romer, P. (1990) Endogenous technological change. *Journal of Political Economy*, 98: 71–102.
Saxenian, A. (2006) *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge, MA: Harvard University Press.
Saxenian, A., Sabel, C. (2008) Venture capital in the 'periphery': the new argonauts, global search and local institution-building. *Economic Geography*, 84: 379–394.
Schumpeter, J. (1962) *The Theory of Economic Development*. Berlin: Springer.
Simmie, J., Carpenter, J., Chadwick, A., Martin, R. (2008) *History Matters: Path Dependence and Innovation in British City-Regions*. London: NESTA.
Somerville, W. (2007) *Immigration under New Labour*. Bristol: Policy Press.
Stephan, P., Levin, S. (2001) Exceptional contributions to US science by the foreign-born and foreign-educated. *Population Research and Policy Review*, 20: 59–79.
Stuen, E. T., Mobarak, A. M., Maskus, K. E. (2012) Skilled immigration and innovation: evidence from enrolment fluctuations in US doctoral programmes. *The Economic Journal*, 122: 1143–1176.
Syrett, S., Sepulveda, L. (2011) Realising the diversity dividend: population diversity and urban economic development. *Environment and Planning A*, 43: 487–504.
Trajtenberg, M., Schiff, G., Melamed, R. (2006) *The 'Names Game': Harnessing Inventors' Patent Data for Economic Research*. NBER Working Papers 12479. Cambridge, MA: NBER.

Trax, M., Brunow, S., Suedekum, J. (2012) *Cultural Diversity and Plant-Level Productivity*. IZA Discussion Paper 6845. Bonn: IZA.
Wadhwa, V. (2012) *The Immigrant Exodus: Why America Is Losing the Global Race to Capture Entrepreneurial Talent*. Philadelphia: Wharton.
Wadhwa, V., Saxenian, A., Rissing, B., Gereffi, G. (2007) *America's New Immigrant Entrepreneurs*. Durham, NC: Duke University/iSchool, UC Berkeley.
Wadhwa, V., Saxenian, A., Rissing, B. A., Gereffi, G. (2008) Skilled immigration and economic growth. *Applied Research in Economic Development*, 5: 6–14.
Zenou, Y. (2009) *Urban Labor Economics*. Cambridge: Cambridge University Press.
Zenou, Y. (2011) *Spatial versus Social Mismatch: The Strength of Weak Ties*. IZA Discussion Papers 5507. Bonn: Institute for the Study of Labor (IZA).
Zucker, L. G., Darby, M. R. (2007) Star Scientists, Innovation and Regional and National Immigration. National Bureau of Economic Research Working Paper 13547. Cambridge, MA: NBER.

Appendix

A. The KITES-PATSTAT database

Raw patent data cannot typically be used at inventor level because of common/misspelled names, changes of address or duplication (Trajtenberg et al., 2006). The dataset of inventors used in this article is taken from the KITES-PATSTAT database, developed by the KITES centre at Bocconi University under the APE-INV initiative.[23] The original patents data come from the EPO PATSTAT system, and are cleaned by the KITES team to allow robust identification of individual inventors.

The KITES cleaning procedure has three stages (Lissoni et al., 2006). First, inventor name and address fields are cleaned and standardized, and a unique CODINV code is applied to all inventors with the same names, surnames and address. Second, 'similarity scores' are assigned for pairs of inventors with the same name and surname but different addresses. Scores are higher for pairs whose dyads are located in the same city/province/region, who patent in the same technological fields (this is measured separately at 4-, 6-, 10- and 12-digit level, with corresponding weights), who share the same applicant organization, who share co-inventors or who are in 'small world' networks with third parties, and who cite each other. Scores are lower for pairs whose dyads patent 20 or more years apart, or who share common surnames (assessed by name frequency analysis for each country). Third, a threshold for similarity scores is generated for each country, over which inventor pairs are considered the same person and given the same identifier.

This cleaning procedure deals with the 'who is who' problem, and indirectly, allows me to identify moving inventors—the 'who is where' problem—much more precisely than some previous studies such as Agrawal et al. (2006). Comparing the original and cleaned CODINV codes allows me to see cases where inventors with different addresses have been subsequently coded as the same individual but at different addresses. I define these as cases of moving inventors. This group turns out to comprise 1781 individuals (2.45% of my sample). Within this group, I also identify a set of smaller inventors who appear in two different TTWAs in same yeargroup. This is a group of 963 individuals (1.33% of my sample).

23 See http://db.kites.unibocconi.it/

B. ONOMAP and minority ethnic inventors

'Ethnicity' is not straightforward to frame or measure (see main text). The 'gold standard' scenario is when there is a rich and flexible typology and where individuals can self-ascribe, for example in a Census (Aspinall, 2009). In many cases, such as health or patents data, this is not available, and name-based approaches have emerged as a powerful alternative (see Mateos (2007) for a review of this literature). The intuition behind name-based approaches is that naming relates to cultural, ethnic, linguistic features of individuals, families and communities. Mateos et al. (2011) point out that 'naming practices are far from random, instead reflecting social norms and cultural customs. They exist in all human groups, and follow distinct geographical and ethno-cultural patterns ... distinctive naming practices in cultural and ethnic groups are persistent even long after immigration to different social contexts' (p. e22943). Working with a database of 18 million names from over 17 countries, the authors show the persistence of naming networks in migrant and minority communities in 'host' and new 'home' environments (Mateos et al., 2011). Note that this last feature of 'naming networks' makes name-based systems suitable for identifying minority ethnic inventors, in particular.

B.1 The ONOMAP system

One of the limitations in early name-based classification systems was a restricted number of names (Mateos, 2007). The ONOMAP system, built at University College London, has designed to deal with this problem.[24] ONOMAP uses a reference population of 500,000 forenames and a million surnames, derived from electoral register or telephone directory name frequency data for 28 countries. Names are then classified into groups, exploiting name-network clustering between surname and forename pairings. Techniques used include forename–surname triage, spatio-temporal analysis, geo-demographic analysis, text mining, 'name-to-ethnicity' analysis from population registers, international name frequency and genealogy resources, and individual name research for hard cases (see Mateos et al. (2007) for details). The final classification comprises 185 'cultural–ethnic–linguistic' (CEL) groups, building on frameworks developed by Hanks and Tucker (2000). At its finest level, this gives 185 CEL 'types': given the frequency distribution of these types in the inventor data, inventors were eventually classified at a higher level based on 68 CEL 'subgroups'. ONOMAP also provides detail on CEL component criteria, including 12 geographical origin groups and nine 'macro-ethnic' groups that derive from the UK Office of National Statistics 1991 Census classification.

ONOMAP is used to classify inventor names via an algorithm that uses surname, forename and surname–forename combinations. In most cases, both elements of a person's name share the same CEL type; in other cases there will be multiple possibilities (such as the author's own name), in which case the system assigns the most likely type based on the underlying name networks in the reference population. In a few cases, names are unclassified (in the case of the KITES-PATSTAT dataset, this is under 1% of inventors). ONOMAP has also been extensively tested with individual-level datasets where ethnicity is known. Petersen et al. (2011) analyse over 107,000 patients for a London hospital; ONOMAP matches over 95% of names. Lakha et al. (2011) test birth registration, pupil census and health data for 260,748 individuals: ONOMAP matches

24 See http://www.onomap.org/

over 99% of names and gives <5% measurement error. For this article, I also subject ONOMAP to a falsification test where ethnicity is assigned at random. In both cases, ONOMAP performs better than random assignment (see Section 7).

B.2 Potential limitations of ONOMAP

There are two potential limitations of ONOMAP relevant to research on inventor demographics. First, the system is unable to distinguish migrants from second-plus generation communities. This article thus focuses on the larger group of minority ethnic inventors. A second limitation stems from international languages, such as Spanish and English, where similar names may be found across several communities and countries. This could be a source of measurement error. In practice, ONOMAP explicitly models Spanish, Mexican, Filipino, Latin American and other Spanish-language names; it also distinguishes English mainland, Cornwall and Channel Islands; Scottish and Welsh; Black Caribbean, American and British; British South African; American Indian and 'American Other' groups. Australasian names are not separately classified, but in the 2011 Census, Australasians make up just 2.4% of the wider migrant population in England and Wales (versus 3.3% from North/South America and 36.3% from Continental Europe). This suggests any remaining misclassification is residual noise rather than a structural problem in the data.

C. Additional results

Table C1. First stage estimator tests: individual patent counts and inventor group diversity

	Geo origin zones			ONS ethnic groups		
	(1)	(2)	(3)	(1)	(2)	(3)
Negative binomial						
Frac Index of inventors	0.075	0.221***	0.248***	0.111	0.312***	0.337***
	(0.100)	(0.020)	(0.023)	(0.165)	(0.011)	(0.014)
Individual fixed effect	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
Log-likelihood	−20,6721.358	−91,887.733	−91,829.454	−20,6723.863	−91,913.822	−91,861.933
OLS						
Frac Index of inventors	0.089	0.644**	0.623**	0.122	0.814*	0.758*
	(0.115)	(0.272)	(0.282)	(0.181)	(0.424)	(0.423)
Individual fixed effects	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
F-statistic	68.238	89.492	49.994	69.024	46.575	46.575
R^2	0.012	0.018	0.018	0.012	0.018	0.018

Source: KITES-PATSTAT/ONS.

Notes: 210,008 observations. Negative binomial coefficients are marginal effects on the mean. In each panel, column (1) uses yeargroup dummies, columns (2) and (3) use technology field*yeargroup dummies and individual fixed effects. Column (3) controls include Fractional Index of TTWA population, % STEM degree holders in TTWA, log of TTWA population density, % high-tech manufacturing in TTWA, % medium-tech manufacturing in TTWA, % workers in entry-level occupations, log of area weighted patent stock 1981–1984, urban TTWA dummy. Bootstrapped standard errors are in parentheses and are clustered on TTWAs.
Significant at *10%, **5% and ***1%.

Table C2. First stage regressions: choice of time period test, reduced form model

Individual patent counts	(1)	(2)	(3)	(4)
Frac Index of inventors by geographical origin	0.623**	0.644***	0.237***	−0.022
	(0.282)	(0.048)	(0.019)	(0.022)
Controls	Y	Y	Y	Y
Observations	210,008	210,008	587,805	293,266
R^2	0.018	0.018	0.038	0.016

Source: KITES-PATSTAT/ONS.
Notes: Model is estimated in OLS. Column (1) fits the full regression. Columns (2)–(4) fit reduced form model with individual*area fixed effects and technology-field*year fixed effects, as detailed controls are not available pre-1993. Column (2) fits inventors active 1993–2004, column (3) fits inventors active 1981–2004, column (4) fits inventors active 1981–1992 only. Standard errors are in parentheses, are heteroskedasticity and autocorrelation-robust and clustered on TTWAs.
Significant at *10%, **5% and ***1%.

Table C3. First stage regressions: sample construction test, reduced form model

Individual patent counts	(1) All	(2) Multiple	(3) Blanks
Frac Index of inventors by geographical origin	0.623**	0.210	0.210
	(0.282)	(0.185)	(0.185)
Controls	Y	Y	Y
Observations	210,008	19,118	19,118
R^2	0.018	0.004	0.004

Source: KITES-PATSTAT/ONS.
Notes: Model is estimated in OLS. Columns (1) and (2) use a sample where inventor*area*time cells are marked as zero when inventors are not patenting; column (2) restricts this sample to inventors who patent more than once. Column (3) uses a sample of multiple inventors where non-active cells are marked as missing rather than zero. All models use technology field*yeargroup dummies and individual fixed effects. Controls as per Table C1. Standard errors are in parentheses, are heteroskedasticity and autocorrelation-robust and clustered on TTWAs.
Significant at *10%, **5% and ***1%.

Table C4. Area-level alternative specification for the first stage model

Aggregate patent counts	OLS		Poisson	
	Unweighted	Weighted	Unweighted	Weighted
Frac Index of inventors (geo origin)	335.481**	124.173*	88.630**	38.920*
	(158.083)	(63.563)	(39.646)	(20.364)
Controls	Y	Y	Y	Y
Observations	532	532	532	532
Log-likelihood	−3269.429	−2712.868	−3485.019	−2173.729
R^2	0.936	0.952		

Source: KITES-PATSTAT/ONS.
Notes: Dependent variables are various unweighted and weighted area-level patent counts. Poisson coefficients are marginal effects at the mean. All models use technology field*yeargroup dummies and area (TTWA) fixed effects. Controls as per Table C1. Standard errors are in parentheses, are heteroskedasticity and autocorrelation-robust and clustered on TTWAs.
Significant at *10%, **5% and ***1%.

Table C5. Moving inventors test: reassigning primary location for moving inventors

Individual patent counts	Location 1	Location 2
Frac Index of inventors by geographical origin	0.248***	0.262***
	(0.023)	(0.015)
Controls	Y	Y
Observations	210,008	210,008
Log-likelihood	−91,829.454	−91,772.246

Source: KITES-PATSTAT/ONS.
Notes: All models use technology field*yeargroup dummies and individual fixed effects. Controls as per Table C1. Bootstrapped standard errors in parentheses. Coefficients are marginal effects at the mean.
Significant at *10%, **5% and ***1%.

Table C6. Second stage regressions: robustness tests on fixed effects decomposition

Estimated individual fixed effect	(1)	(2)	(3)	(4)	(5)	(6)
Minority ethnic inventor	−0.199***	−0.194***	−0.196***	−0.200***	−0.198***	
	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	
Moving inventor, same yeargroup			0.512***			
			(0.036)			
Moving inventor				0.044*		
				(0.025)		
Inventor patents in ≥1 technology field					0.213***	
					(0.015)	
Fake minority ethnic						0.016
						(0.010)
Controls	Y	Y	Y	Y	Y	Y
Observations	70,007	70,007	70,007	70,007	70,007	70,007
R^2	0.253	0.343	0.256	0.253	0.256	0.249

Source: KITES-PATSTAT/ONS.
Notes: Column (1) fits the main regression as per Table 10 in the article. Column (2) uses an FGLS estimator instead of bootstrapped OLS. Columns (3) and (4) introduce additional controls for moving inventors. Column (5) includes a control for 'generalists' (patenting across at least one technology fields). Column (6) uses 'fake' (randomly assigned) minority ethnic status. For all models, controls include multiple inventor dummy, star dummy, inventor average patent count, pre-1993 and dummy for inventor activity, pre-1993. All models use robust standard errors, bootstrapped, 50 repetitions. Constant not shown.
Significant at *10%, **5% and ***1%.

Table C7. Second stage regressions: falsification test

Estimated individual fixed effect	(1)	(2)
Inventor Central European origin	−0.112***	
	(0.019)	
Inventor East Asian origin	−0.142***	
	(0.027)	
Inventor East European origin	−0.112***	
	(0.029)	
Inventor rest of world origin	−0.289***	
	(0.027)	
Inventor South Asian origin	−0.314***	
	(0.021)	
Inventor South European origin	−0.175***	
	(0.030)	
Fake origin group 2 dummy		0.047**
		(0.020)
Fake origin group 3 dummy		0.022
		(0.022)
Fake origin group 4 dummy		−0.017
		(0.023)
Fake origin group 5 dummy		−0.021
		(0.022)
Fake origin group 6 dummy		0.022
		(0.030)
Fake origin group 7 dummy		0.016
		(0.026)
Controls	Y	Y
Observations	70,007	70,007
R^2	0.254	0.249

Source: KITES-PATSTAT/ONS.
Notes: Column (1) fits the main regression, column (2) fits randomly assigned categories. Controls as in Table C6. All models use robust standard errors, bootstrapped 50 repetitions.
Significant at *10%, **5% and ***1%.

Table C8. Distributional analysis. Resource crowd-out/-in

Change in majority weighted patents 1993–2004	(1)	(2)	(3)	(4)	(5)
% Change in minority ethnic weighted patents 1993–2004	1.645***	1.576***	1.907***	1.988***	1.908***
	(0.341)	(0.330)	(0.104)	(0.073)	(0.088)
TTWA population Frac Index, 1993		0.943	1.046	1.431	−1.085
		(1.594)	(1.761)	(1.621)	(1.396)
TTWA share of STEM graduates, 1993		−4.492	−2.398	4.295	−2.057
		(3.951)	(3.021)	(3.090)	(2.993)
TTWA high-tech manufacturing, 1993		−4.203	−7.638	−5.771	0.037
		(4.202)	(4.735)	(4.660)	(3.842)
TTWA medium-tech manufacturing, 1993		−4.475	−3.114	−3.927	1.041
		(4.009)	(4.301)	(3.991)	(3.422)
Log(TTWA population density, 1993)		−0.204	−0.041	0.128	0.112
		(0.170)	(0.130)	(0.108)	(0.099)
Urban TTWA		−0.070	−0.466**	−0.163	−0.494**
		(0.226)	(0.211)	(0.228)	(0.194)
Log(area patent stocks, 1989–1992)				−0.327***	
				(0.104)	
Log(area patent stocks, 1981–1984)					0.026
					(0.077)
OST30 technology field dummies	N	N	Y	Y	Y
Observations	203	203	201	196	176
R^2	0.391	0.427	0.712	0.768	0.798

Source: KITES-PATSTAT/ONS.

Notes: Column (1) fits the variable of interest, column (2) adds controls, column (3) adds technology field dummies. Constant not shown. Heteroskedasticity and autocorrelation-robust standards errors are clustered on TTWAs.

Significant at *10%, **5%, *** 1%.