The Nexus between Labor Diversity and Firm's Innovation

Pierpaolo Parrotta,[*] Dario Pozzoli[†] and Mariola Pytlikova[‡]

Abstract

In this paper we investigate the nexus between firm labor diversity and innovation using a linked employer-employee data from Denmark. Specifically, exploiting information retrieved from the comprehensive database and implementing a proper instrumental variable strategy, we are able to identify the contribution of workers diversity in cultural background, education and demographic characteristics to valuable firm's innovation activity. The latter is measured by: (1) the firm's propensity to apply for a patent, (2) the number of patent applications (intensive margin) and (3) the firm's ability to patent in different technological areas (extensive margin). We find that ethnic diversity plays an important role in propelling firm's innovation outcomes.

JEL Classification: C26, J20, L20.

Keywords: Labor diversity, patenting activity, extensive and intensive margins.

[*]Business and Social Sciences, Aarhus University; Department of Economics, University of Lausanne.

[†]Corresponding author. Business and Social Sciences, Aarhus University. Department of Economics, Hermodsvej 22, DK, 8230 Aabyhoj, Denmark. E-mail: dpozzoli@asb.dk.

[‡]Business and Social Sciences, Aarhus University.

1 Introduction

Many developed and developing countries have experienced several changes in the workforce composition which has lead to an increased heterogeneity of the labor force in terms of age, gender, skills and ethnicity. This is partly the result of policies adopted to counteract the problem of population aging, anti-discrimination measures, immigration and the worldwide globalization process (Pedersen et al. 2008). From the demand side, we observe increasing diversity across many workplaces and we often hear about the importance of further internationalization and demographic diversification. The promotion of diversity is often perceived as a chance to improve learning and knowledge management capabilities and then enhance firm productivity (Parrotta et al. 2010). Besides, workforce diversity is believed to be an important source of innovation. For instance, in a relatively recent survey conducted by the European Commission, a large number of respondents identified innovation as a key benefit of having diversity policies and practices (European Commission, 2005). If this is the case, firms could benefit from the growing diverse cultural backgrounds, demographic, and knowledge bases of the workforces. Moreover, since there is a widespread consensus that innovation is crucial for sustainable growth and economic development (as suggested in the new growth theories), knowing the link between workforce diversity and innovation seems to be essential for policy makers.

From a theoretical point of view, a paradox has been recognized: whereas labor diversity can be a source of creativity and therefore foster innovation activity, a high degree of heterogeneity among workers may induce misunderstanding, conflicts and uncooperative behaviors within workplaces and in this way hinder innovation (Basset-Jones, 2005). There is no general agreement on which effect may prevail. Specifically, differences in skills, education and more broadly in knowledge among employees seem to be beneficial rather than detrimental. According to Lazear (1999), positive effects may prevail as long as workers' information sets are not overlapping but relevant to one another. Ambiguity instead persists for diversity in ethnic and demographic characteristics of employees. On the one hand, people of different cultural backgrounds, age and gender may provide diverse perspectives, valuable ideas, problem-solving abilities, and in this way facilitate the achievement of optimal creative solutions and therefore stimulate innovations (Watson et al. 1993; Drach-Zahavy and Somech, 2001; Hong and Page, 2001). As people of different ethnic backgrounds also possess knowledge about global markets and customers tastes, they may stimulate firm to improve or develop products sold abroad accordingly (Osborne, 2000; Berliant and Fujita, 2008). Further, foreign employees may present different ways of searching for solutions to problems, i.e. heuristics, and therefore they are more likely to come up with innovative solutions than ethnically homogenous teams of workers (Hong and Page, 2004). On the other hand, such heterogeneities might create communication barriers, reduce the workforce cohesion and prevent cooperative participation in research activities, bringing high costs of "cross-cultural dealing" (Williams and O'Reilly, 1998; Zajac et al., 1991; Lazear, 1999). Similarly, diversity in age may facilitate innovation because there are complementarities between the human capital of younger and older workers: younger employees have knowledge of new technologies and IT and older employees have a better understanding and experience with the intra-firm structures and the operating process (Lazear, 1998). But, demographic heterogeneity among workers may create communication frictions if workers are prejudiced, and therefore bring some cost connected to the frictions (Becker, 1957). Thus, it is still unclear whether more heterogeneous worforces outperform the relatively more homogeneous ones with respect to innovation.

The empirical literature exploring the relationship between labor diversity and firm's innovation consists mainly of business case studies that often look at work-team compositions (Horwitz and Horwitz, 2007; and Harrison and Klein, 2007) or even focus on diversity in top management teams only (Bantel and Jackson, 1989; Knight et al. 1999; Pitcher and Smith, 2001).[1] That may be imputed to differences in research aims and approaches, but also to the lack of more comprehensive employer-employee data, which provide a notable amount of information on the labor force composition at the firm level. To the best of our knowledge, the evidence using more comprehensive data is virtually non-existent.

In this paper, we investigate the nexus between labor diversity and innovation using a rich register-based linked employer-employee dataset (LEED) from Denmark for the years 1995-2003. Regarding measures of innovation, we follow previous literature and make use of information on patents to proxy for innovation (Griliches, 1990; Bloom and Van Reenen, 2002). Specifically, we use the following three measures: (1) firm's propensity to apply for a patent, (2) the number of patents introduced each year and (3) firm's propensity to apply in more than one technological area. We investigate the effect of labor diversity on firm innovation by looking at three dimensions of employee diversity: cultural background, skills/education and demographics. Further, we deal with several problems that previous literature studying the impact of workforce diversity on innovation did not address properly. Most importantly, it might be that firms are aware of the importance of labor diversity and leverage it to improve their performances; then the relationship under investigation may be affected by simultaneity or endogeneity. To address these concerns, we implement an instrumental variable (IV) strategy à la Card (2001) based on predicted levels of workforce diversity for each commuting area, where a firm is located, as an instrument for the firm labor diversity. Further, as broadly documented by industrial and knowledge economics literature, firms are characterized by a different propensity to innovate. Thus, there exist unobserved and observed firm-specific heterogeneity that should be taken into account to evaluate the effect of any labor diversity dimension on firm's innovation outcome. Following Blundell et al. (1995), we account for past firms' success in innovation and use pre-sample information as an observable

[1] There exists also some literature on the effects of diversity - typically ethnic labor diversity - on innovation using aggregate regional or industry data, for instance Kelley and Helper (1999), Feldman and Audretsch (1999), Anderson et al. (2005), Niebuhr (2010); Kerr and Lincoln (2010).

proxy for unobservable permanent firm characteristics. Finally, we control for the potential role of the external knowledge in favoring firms' patenting activity and compute knowledge spillovers indicators based on geographical and technological distances between firms.

Implementing alternative estimation techniques, we find an evidence of the key role of the ethnic diversity in promoting firm's innovation as measured by the probability to innovate in more than one technological field, propensity to start patenting or number of patent applications. Specifically, we find that a 10 percentage change in ethnic diversity increases the number of firms' patent applications by approximately 1.5 percent. Whereas the contribution of ethnic diversity to start patenting is almost negligible (a standard deviation change in its value turns to raise such a probability by 0.05 percent), the effect of educational diversity on extensive margins is substantial: a standard deviation change in skill diversity raises the firms' probability to apply for a patent in different technological areas by 7.3 percent. Effects of diversity in education and demographics turn to be mostly insignificant when either the full set of controls is included or endogenity is taken care of.

These results support the hypothesis that ethnic diverse workers tend to have a wider pool of different experiences, knowledge bases and heuristics boosting their problem-solving capacities and creativity, which in turn facilitate innovations. In this regard, our findings are consistent with the theoretical frameworks proposed by Hong and Page (2001 and 2004) and Berliant and Fujita (2008). Hence, our results suggest firms to focus on recruitment strategies that explicitly account for heterogeneity in ethnicity. This article may also provide some suggestions to public authorities in terms of innovation policies. Given that innovation is considered as one of the most important components for the long-term economic growth, hence investigating the determinants of the innovation process may also lead to the identification of the sources of a sustainable growth. In this regard, public institutions and policy makers could invest resources to promote ethnic diversity within workplaces and in such a way increase the innovation, and ultimately the economic growth.

The structure of the paper is as follows: section 2 briefly describes the data, section 3 provides details on the empirical strategy, sections 4 and 5 explain all the results of our empirical analyses and section 6 offers some concluding remarks.

2 Data

2.1 Data sources

The data set we use for our analysis is obtained by merging three different data sources from Denmark. The first one is the 'Integrated Database for Labor Market Research' (IDA), which is a register-based LEED managed by Statistics Denmark, a Danish governmental institute responsible for creating statistics on the

Danish society and economy. IDA contains a broad set of information on individuals and firms for years 1980-2006. In particular, we are interested in gender, age, nationality, education, occupation and place of work, but also whether a firm is (partially or totally) foreign-owned and a multi-establishment. The second data source is a register of firms' business accounts (REGNSKAB) that provides information on a number of financial items, which we need in order to construct values of firms' capital stock, information on whether a firm is an exporter and the 3-digit industry, in which the firm operates. This database is also maintained by the Statistics Denmark and reports data for the period 1995-2006.[2] In REGNSKAB it is possible to identify partially and totally imputed values, which we exclude from our final data set in order to avoid any bias in the estimates. The last data source is a collection of patent applications sent to the European Patent Office (EPO) by Danish firms.[3] It covers a period of 26 years (1978-2003) and allows us to account for 2822 applicants and 2244 granted firms.[4] We disregard those industries[5] where there were no patenting firms during the period covered in our empirical analysis. We also exclude enterprises with less than 10 employees from our sample to allow all investigated firms potentially to reach the highest degree of (ethnic) diversity at least when an aggregated specification is used. Thus, our final data set contains information on approximately 14,000 firms per year over a period of 9 years (1995-2003).

2.2 Diversity measures

The workforce diversity (heterogeneity) measures used in this article are computed at the firm level and based on the Herfindahl index. The latter combines two important dimensions of diversity: the "richness", which refers to the number of defined categories within a firm, and the "evenness", which informs on how equally populated such categories are. Specifically, our diversity measures represent weighted averages of Herfindahl indexes computed at the workplace level:

$$Div\_h_{it} = \sum_{w=1}^{W} \frac{N_w}{N_i} \left(1 - \sum_{s=1}^{S} p_{wst}^2 \right) ,$$

where Div_h_{it} is the diversity index of firm i at time t for the dimension h, W is the total number of workplaces (w refers to a given workplace) constituting the firm, and therefore N_w and N_i denote the total number of workers at the workplace and firm levels, respectively. Thus, the ratio between the last two

[2] Part of the statistics in REGNSKAB refers to selected firms for direct surveying: all firms with more than 50 employees or profits higher than a given threshold. The rest is recorded in accordance with a stratified sample strategy. The surveyed firms can choose whether to submit their annual accounts and other specifications or to fill out a questionnaire. In order to facilitate responding, questions are formulated in the same way as required in the Danish annual accounts legislation.

[3] The access to these data has been made possible thanks to the Center for Economic and Business Research (CEBR), an independent research center affiliated with the Copenhagen Business School (CBS).

[4] More details concerning the construction and composition of the data set can be found in Kaiser, Kongsted and Rønde (2008).

[5] Agriculture, fishing and quarrying; electricity, gas and water supply; sale and repair of motor vehicles; hotels and restaurants; transports; and public services.

variables corresponds to the weighting function, while p_{wst} is the proportion of the workplace's employees falling into each category s at time t, with $s = 1, 2, ..., S$. The diversity index has a minimum value, which takes value on zero if there is only one category represented within the workplace, and a maximum value equal to $\left(1 - \frac{1}{S}\right)$ if all categories are equally represented. The index can be interpreted as the probability that two randomly drawn individuals in a workplace belong to different groups.

As we distinguish between cultural, educational (skill) and demographic diversity, a separate measure is computed along each of the three cited dimensions. Diversity in cultural background is associated with employees' country of origin[6] and is built by using the following eight categories: North America and Oceania, Central and South America, Africa, Western and Southern Europe, Formerly Communist Countries, East Asia, Other Asia, Muslim Countries.[7] Diversity in education is based on six categories. In particular, tertiary education (PhD, Master and Bachelor) is divided into the following four groups: engineering, humanities, natural sciences and social sciences. The other two categories are represented by secondary and compulsory education. Eight categories instead refer to the demographic diversity, which is computed by combining gender and four age dichotomous indicators associated with quartiles of the overall age distribution.

However, given that the overall categorization might be somehow arbitrary, we decide to use a more disaggregated one, too. The alternative cultural background diversity is based on linguistic classification.[8] Specifically, we group foreign employees together by family of languages, to which the language spoken in their home country belongs. Using the third linguistic tree level language classification drawn from Ethnologue, we end up having 40 linguistic groups.[9] Further, our disaggregated diversity indexes in education and demographics are based on eight and ten categories, respectively. Differently from the former classification, the secondary education is split into 3 sub-groups: high school, business high school and vocational education. Demographic diversity is computed by combining gender and five age dichotomous indicators associated with quintiles of the overall age distribution.

2.3 Descriptive statistics

Table 1 reports some descriptive statistics (median, mean and standard deviation) of the variables used in our empirical analysis. The firm population is divided into two groups based on whether a firm applied for at least one patent (patenting firm) or did not. Patenting firms are characterized by notably higher values of capital and labor inputs: the average capital stock is almost 9.5 times the value of the non-patenting firms.

[6] Native Danes are excluded.

[7] See Appendix1 for more details about the countries belonging to each ethnic category.

[8] Previous literature argues that linguistic distance serves as a good proxy for cultural distance (Guiso et al., 2009; Adsera and Pytlikova, 2010).

[9] The linguistic classification is more detailed than the grouping by nationality. Specifically, we group countries (their major official language spoken by the majority in a particular country) by the third linguistic tree level, e.g. Germanic West vs. Germanic North vs. Romance languages. The information on languages is drawn from the encyclopedia of languages "Ethnologue: Languages of the World", see the Appendix section for more details about the list of countries and the linguistic groups included.

The latter are more likely to be single-establishment companies and markedly less export-oriented: on average the share of exporters halves among those firms that have never applied for a patent. Small differences are shown instead for the foreign ownership status: the foreign capital penetration is quite low among Danish firms. For the purposes of our analysis it appears relevant to take into account the role of external sources of knowledge since they may facilitate firms' innovation activity. Although we already control (using the export dummy) whether firms compete in the international arena and then have access to foreign knowledge, more precise indexes of knowledge spillovers can be defined at the national level. Specifically, we construct two measures of knowledge spillovers, one based on the geographical distance and the other on the technological proximity, see Appendix 2 for a detailed description of the external knowledge indexes. Looking at these measures of knowledge spillovers, see Table 1, we find no evidence of diffused clustering behavior or huge differences in technological distance between the two groups of firms.

As evident from the Table 1, there are remarkable differences between patenting and non-patenting firms with respect to firms' workforce composition. Not surprisingly, patenting firms are characterized by larger shares of highly educated employees, white-collar workers and managers, whereas the opposite holds true for middle managers. Interestingly, patenting firms also record a higher share of female and foreign employees. Workers in these knowledge-based firms are slightly older on average terms: presumably the share of young employees is lower because patenting firms hire a wider proportion of well trained and experienced people. As a matter of fact long tenure profiles are more common within patenting firms' environment. Diversity indexes register higher values for patenting firms. Particularly evident is the differential in the ethnic heterogeneity that is 3.5 times larger on average with respect to non-patenting firms. These indexes also report substantial lower education diversity, which is 16% poorer in mean values. Thus, the presented descriptives raise a reasonable interest in evaluating the "nexus" between firms' patenting behavior and diversity in ethnicity, education and demographics.

3 Econometric methods

3.1 Propensity to innovate

To investigate the effect of labor diversity on firm's propensity to innovate, we employ a standard binomial regression technique in our analyses. Specifically, we estimate the following probit model:

$$\begin{cases} z_{it} = 1 \quad if \quad z_{it}^* > 0 \\ z_{it} = 0 \quad otherwise \end{cases}$$

$$with \quad z_{it}^* = \gamma_c Div\_c_{it} + \gamma_s Div\_s_{it} + \gamma_d Div\_d_{it} + x_{it}^{'}\beta + \eta_i + v_{it}$$

where z_{it}^* denotes the unobservable variable inducing firm i to apply at least once for a patent at time t; z_{it} indicates whether firm i concretely has applied at time t; the first three terms at the right-hand side are diversity in cultural background, skills and demographics respectively. The vector $x_{it}^{'}$ includes an extensive set of observable characteristics, like, among others, the external knowledge indexes and the firm-specific characteristics described in section 2.3; η_i denotes the firm-specific unobservable effect and v_{it} is the error term. Similar to Blundell, Griffith and Van Reenen (2002), we proxy for the unobserved heterogeneity η_i by arguing that the main source of unobserved permanent differences in firms' capabilities to innovate can be captured by the pre-sample history of innovative successes. In line with that, we assume that the firms' average number of patent applications provides a good approximation of the above unobservable heterogeneity component η_i. However, an overall increase in the number of patent applications is recorded during the pre-sample period. Thus, as in Kaiser et al. (2008) we deal with that by normalizing a firm's number of patents in a pre- sample year by the total number of patents applied for during that year:

$$\eta_i = \frac{1}{T} \sum_{t=\tau}^{T+\tau} \left(\frac{y_{it}}{\sum_{i=1}^{I} y_{it}} \right)$$

As firms can leverage labor diversity to improve their innovation performances, we also instrument our variables of interest in order to obtain a causal effect of workforce diversity on firm innovation activities. More specifically, we implement an instrumental variable (IV) strategy based on the predicted levels of workforce diversity in cultural background, skills and demographic characteristics at the commuting area where the firm is located. In this approach, the prediction of the current composition of a commuting areas' labour supply is computed using its early 90s composition and the current stocks (Card, 2001).[10] The so-called functional economic regions or commuting areas are identified using a specific algorithm based on the following two criteria. Firstly, a group of municipalities constitute a commuting area if the interaction within the group of municipalities is high compared to the interaction with other areas. Furthermore, at least one municipality in the area must be a center, i.e. a certain share of the employees living in the municipality must work in the municipality, too (Andersen, 2000).[11]

We believe that diversity at the commuting area level presents a suitable supply driven instrument for workplace level diversity because commuting areas in Denmark (except for the area around Copenhagen) are typically relatively small and therefore firms very likely recruit workers from a given local supply of labor, which is characterized by a certain degree of heterogeneity. This argument is further reinforced by the role

[10]Since firm diversity is computed as weighted average of the workplace diversity measures, the instruments here are a weighted average of diversity measures related to the commuting area, where workplaces are located.

[11] In total 104 commuting areas are identified.

of networks in the employment process (Montgomery, 1991, Munshi, 2003). Thus firms placed in areas with a high labor diversity are also more likely to employ more a diverse workforce. Moreover, the rather low residential mobility in Denmark (Deding et al. 2009) seems to support the properness of our IV strategy.[12]We use the same IV strategy for the analyses of intensive and extensive margins.

3.2 Intensive margins

As the number of patents is by definition restricted to non-negative integers, the econometric strategy used to analyze the relationship between intensive margins of patenting activity and labor diversity is grounded on the family of count models. As a starting point we assume that the data generating process follows a Poisson distribution. If the random variable Y_{it}, in our case number of patent applications filed by firm i at time t, is Poisson distributed, then the probability that exactly y applications are observed is as follows

$$P\left(Y_{it} = y \mid \lambda_{it}\right) = \frac{e^{-\lambda_{it}}\lambda^{y}}{y!} \ .$$

Covariates can be introduced by specifying the individual (firm) mean as

$$\lambda_{it} = exp\left(\beta_c Div\_c_{it} + \beta_s Div\_s_{it} + \beta_d Div\_d_{it} + w_{it}^{'}\beta_w + \eta_i\right) \ , \tag{1}$$

where η_i stands for the unobserved time-invariant firm-specific heterogeneity term and w_{it} is a vector of patent production determinants, as specified in subsection 3.1. Following Blundell et al. (1995), we also include, among the covariates w_{it}, the discounted patent stock of firm i at period $t-1$ in order to account for potential state dependence in patenting activity. This is calculated as

$$disc\_stock_{it-1} = y_{it-1} + (1-\delta)disc\_stock_{it-2} \ ,$$

where y_{it-1} is the lagged number of patent applications and δ is the depreciation rate set equal to 30 per cent as in Blundell et al. (1995).

We also add a dummy variable taking value on zero if the firm had never innovated prior to 1995, to capture persistent differences between patenting and non-patenting firms (Blundell et al., 1995; Blundell et

[12]Furthermore, the validity of our instrument is even strengthened by the spatial dispersion policy implemented for immigrants between 1986 and 1998 by the Danish authorities. The dispersal policy implied that new refugees were randomly distributed across locations in Denmark, see e.g. Damm (2009).

al., 1999). In addition, this dummy variable represents a remedy for the so-called "zero-inflation problem" given that in our data many firms never applied for a single patent. The pre-sample information technique is feasible in a study like ours because we have a long series for the dependent variable (1977-1994) prior to the starting period (1995) of the final sample in use.

As we have mentioned before, one may argue that the relationship between firm-patenting activity and diversity could be affected by endogeneity. The latter issue might arise because there could be unobserved firm-specific factors influencing both the number of patent applications and the degree of labor diversity. To address these concerns, we apply a two-stage IV procedure to the Poisson model as suggested by Vuong (1984). In this case, equation (1) is specified as follows:

$$\lambda_{it} = exp\left(\beta_c Div\_c_{it} + \beta_s Div\_s_{it} + \beta_d Div\_d_{it} + w_{it}^{'}\beta_w + \eta_i + u_{it}\right) \tag{2}$$

where the term u_{it} can be interpreted as unobserved heterogeneity correlated with the diversity indexes but uncorrelated with the vector of patent production determinants w_{it}.[13] To model the correlation between the endogenous variables and u_{it}, we specify a system of linear reduced-form equations, one for each diversity index. This is

$$\begin{cases} Div\_c_{it} = w_{it}^{'}\gamma_w + z_{it}^{'}\gamma_z + \varepsilon_{cit} \\ Div\_s_{it} = w_{it}\gamma_w + z_{it}^{'}\gamma_z + \varepsilon_{sit} \\ Div\_d_{it} = w_{it}^{'}\gamma_w + z_{it}^{'}\gamma_z + \varepsilon_{dit} \end{cases}$$

where z_{it} is the vector of exogenous variables that affects firm level diversity, but does not directly affect the number of patent applications. As in section 3.1, the excluded variables are the diversity indexes computed at the commuting area where the firm is located and the model is just-identified. The error terms ε are assumed to have zero mean and to be correlated across equations for a given firm i, but uncorrelated across observations. Furthermore, we assume that the errors u and ε are related via

$$u_{it} = \rho_c\varepsilon_{cit} + \rho_s\varepsilon_{sit} + \rho_d\varepsilon_{dit} + \zeta_{it} \tag{3}$$

where $\zeta_{it} \sim \left[0, \sigma_\zeta^2\right]$ is independent of ε_{cit}, ε_{sit} and ε_{cit}.[14] Substituting equation (3) in equation (2) for u_{it} and taking the expectation with respect to ζ yields

$$E_\zeta(\lambda) = exp(\beta_c Div\_c + \beta_s Div\_s + \beta_d Div\_d + w^{'}\beta + \eta + lnE(e^\zeta) + \rho_c\varepsilon_c + \rho_s\varepsilon_s + \rho_d\varepsilon_d)\,.$$

[13]The error term u_{it} is added to allow for endogeneity. It also induces overdispersion, so that the Poisson model and the Negative binomial model are empirically equivalent.

[14]This assumption means that ε is a common latent factor that affects both diversity and patent applications and is the only source of dependence between them, after controlling for the influence of the observed variables.

The constant term $lnE(e^{\zeta})$ can be absorbed in the coefficient of the intercept as an element of w. It follows that

$$\lambda_{it} = exp\left(\beta_c Div\_c_{it} + \beta_s Div\_s_{it} + \beta_d Div\_d_{it} + w_{it}^{'}\beta_w + \eta_i + \rho_c \varepsilon_{cit} + \rho_s \varepsilon_{sit} + \rho_d \varepsilon_{dit}\right) ,$$

where ε_{cit}, ε_{sit} and ε_{cit} are the new additional variables. Given that the former variables are unobservable, we follow a two-step estimation procedure where we first estimate and generate them and second we estimate parameters of the Poisson model after replacing ε_{cit}, ε_{sit} and ε_{cit} with $\hat{\varepsilon}_{cit}$, $\hat{\varepsilon}_{sit}$ and $\hat{\varepsilon}_{cit}$. Obviously, the variance and covariance matrix of the two-step estimator needs to be adjusted for the above replacement by bootstrapping the sequential two-step estimator.

3.3 Extensive margins

The estimation approach used to evaluate the extensive margins of firms' patenting behavior is similar to the one adopted for the firms' propensity to patent. Although the count data models would be more suitable for the analyses of relationship between workforce diversity and the number of different technological areas of patent application, our data and concretely the lack of minimum observations required to run count data models do not allow us to use them. Instead, we evaluate whether more labor diversity increases the probability of a firm to (apply for a) patent in more than one technological area.

4 Results

This section reports findings for each of the outcome dimensions we look at: propensity to innovate, intensive and extensive margins. Several specifications among the different econometric models here employed help in understanding the strength of our results. Further, in the sensitivity analyses subsection we examine whether the results differ between white- and blue-collar occupations, and across alternative diversity and innovation measures.

4.1 Results on labor diversity and propensity to innovate

Table 2 reports estimates concerning the propensity to patent. As explained in the previous section, we implement probit models having as dependent variable the dummy indicating whether a firm has applied for a patent in a given year. In column 1 we show a model with the three workforce diversity indexes as the only regressors. The workforce diversity can explain about 14% of the overall variation in the dependent variable and is associated with sizable and significantly positive effects. Columns 2 and 3 show results from probit models with all other covariates; while the former treats the diversity indexes as exogenous variables, the latter shows the IV specification with predicted workforce diversity levels at commuting areas as instruments for the firm workforce diversity. Columns 4 to 6 report models with single diversity dimensions: we check whether a diversity index captures characteristics associated with other indexes, i.e. ethnic diversity may pick up some of the skill diversity effects as individuals with same education but coming from different countries may present degrees of educational heterogeneity as well. Results between the IV and full specification models are rather similar and imply that a standard deviation change in the ethnic diversity increases the probability to apply for patent by 0.05 percentage points. On the contrary, the effects related to education and demographic diversity vanish.

Turning to the other control variables, the inclusion of pre-sample fixed effects turns out to be important to deal with time invariant unobserved heterogeneity among firms. The variable attaches statistically significant positive coefficient and it also corrects the estimates on labor diversity. Further, firms with higher shares of highly skilled and vocational workers, and exporting firms have higher propensity to patent. Instead, the knowledge spillovers based on technological and geographical distances, and the average firm tenure do not explain much of such a propensity.

As explained above, we run additionally the models using diversity indexes based on more detailed category specification; the results are shown in the Table 2, columns 7 to 12. Now the effect of a standard deviation change in the ethnic diversity produces an increase in the probability to apply for a patent by 0.07 percentage points, whereas the effects of education and demographic diversity appear negligible.

4.2 Results on labor diversity and intensive margins

In the next step, we analyze how firm workforce diversity contribute to the number of patent application. Tables 3 reports the results of the intensive margins analyses, here the estimated coefficients represent elasticities. The first column in Table 3 shows the output of a Poisson regression having only the diversity measures as regressors: the coefficients to all diversity indexes are large, positive and significant. Once more, after including all the other control variables (column 2) their dimension and statistical significance decreases. Nonetheless, except for the ethnic heterogeneity, the diversity indexes don't retain their statistically significant positive coefficients. Taking the IV Poisson specifications as the most reliable, we find that ten percent increase in the ethnic diversity leads to 1.5 percentage points increase in the number of patent applications. This effect is quite sizable given that the elasticity associated to a production input like human capital (proxied by the share of highly skilled workers) is less than twice larger. Important effects are also related to the shares of technicians, capital and labor stock, while knowledge spillovers variables do not show significant contributions to the overall number of patent applications. As in the case of patenting propensity, exporters benefit from the knowledge gained in the international markets.

Columns 7 to 12 report findings for the disaggregated classification of labor diversity dimensions, which are very similar to the results using aggregate diversity specifications. Specifically, in the IV Poisson (column 3) a ten percent increase in ethnic diversity implies a 2.07 percent increase in the number of patent applications.[15]

4.3 Results on labor diversity and extensive margins

Table 4 reports the effects of labor diversity on the probability of applying for a patent in different technological areas in a given year. The structure of this table is similar to Table 2. The low number of annual patent applications in each technological area does not allow us to use potentially more suited count models. Regarding the variables of interest, we find that the diversity indexes alone explain 7 percent of the overall variation in the dependent variable, coefficients to ethnic and educational diversity indexes are positive and statistically significant. However, the significance of the diversity in education vanishes when endogeneity is taken care of. Overall, we find that ethnic diversity is important for patenting in different technological areas. Taking the lowest estimate between the full IV specifications, it turns out that a standard deviation increase in skill diversity is associated to a raise of about 7.3 percent in the probability to patent in different technological fields. Thus it seems as the ethnic diversity is much more relevant for patenting in different technological areas than for the patenting per se.

5 Sensitivity analysis

In this section, we examine whether the effects of labor diversity on patenting activity of firms differ between different occupational categories of workers, and across alternative diversity and innovation measures. All results are based on the full IV specifications described in the previous section.

Firstly, we calculate our diversity indexes for white- and blue-collar occupations separately. This is driven by the idea that diversity could play a different role for distinct occupational groups and consequently have diverse effects on firm innovation. In particular, we would expect that the beneficial effects of diverse problem-solving abilities and creativity would materialize more in terms of innovation for white-collar occupations compared to blue-collar occupations. The results of the effect of diversity indexes calculated separately for the two occupational groups on firm probability to innovate, number of patent applications and firm probability of applying for a patent in different technological areas are presented in the first two columns of Table 5, 6 and 7, respectively. Indeed, we find that workforce diversity is much more important for white-collar than for blue-collar occupations. The effect of ethnic diversity among the white-collar workers on probability to innovate and the number of patents is positive and statistically significant. On contrary, the effect of education and demographic diversity is insignificant for both white- and blue-collar occupations. Similar

[15] We have also investigated whether the effects of a particular dimension of diversity can be influenced by other forms of labor heterogeneity by inclusion of all possible interaction couples between the diversity indexes. Furthermore, driven by the hypothesis that there might be complementarities among different skills and demographic groups, in particular young and educated workers together with a more diverse workforce can stimulate innovation and creativity, we have augmented our models with interactions between diversity indexes and shares of highly skilled and younger workers. Nevertheless, none of the interactions turned out to be statistically significant. Figures showing marginal effects of the interactions are available from the authors upon request.

results are estimated for the number of patent applications and the extensive margins.

Futher, as a part of the sensitivity analysis we evaluate eventual variations in the effects of labor diversity when the diversity measure is differently computed. In particular, we use two alternative diversity indexes: the Shannon-Weaver entropy and the richness indexes. The entropy index is considered as one of the most profound and useful diversity indexes in biology (Maignan et al., 2003), whereas the richness index is defined as a number of categories observed for each dimension of interest (it does not account for the "evenness" dimension). Moreover, we include an Herfindhal index for the type of tertiary education (this index has now only 4 categories: engineering, natural sciences, social sciences and humanities) and the standard deviation for education and age. That allows us to compare the effects associated with the amount and types of higher education. Parameters on cardinal measures of diversity (standard deviation of years of education and age) are never significant: the richness component is the main driver.

Additional checks consist of excluding from the calculation of ethnic diversity alternative groups of foreigners (i.e. second generation immigrants), foreigners with tertiary education and foreigners speaking one of the language belonging to the germanic group. It is in fact plausible to expect that communication costs associated with ethnic diversity may increase after subtracting out foreigners who are likely to speak Danish or English. As cities have at same time a lot of immigrants and a high percentage of innovative firms, we drop Copenhagen (the only real agglomeration in Denmark) and environs in the analysis. Results from all these robustness checks are reported in columns 3 to 9 of Tables 5, 6, and 7 and do not qualitatively differ from the main results. Interestingly, the role of ethnic heterogeneity weakens once we exclude foreigners who probably speak English (highly educated workers) or Danish, confirming the idea that the communication costs and costs of "cross-cultural dealing" are likely to be more important when foreigners don't speak the same language. In fact, the ethnic diversity measure may capture both heterogeneity in a specific education (employees with same degree but coming from different universitary sistems may still present some degree of heterogeneity) and recruitment of talented workers with different origins.

Finally, as labor diversity has been computed at the firm level (weighting average of Herfindahl indexes computed at the workplace level), we evaluate how results change if multi-establishment enterprises are excluded from the sample. Restricting our attention to single workplaces, we check whether the relationship between workforce diversity and innovation is sensitive to the level of analysis or mainly driven by the inclusion of big companies. The last column of Table 5 and 7 report infomation on such a check : the interpretation of these findings does not significantly differ from that related to the main results. Table 6 does not include this robustness check as count models did not converge with the restricted sample.

6 Discussion and conclusions

In this paper we provide an overall assessment of the nexus between labor diversity and firms' patenting behavior. To the best of our knowledge, this study represents the first concrete attempt to formalize and generalize the relationship between labor diversity and innovation by using detailed information on firms' workforce composition.

Specifically, controlling for a large number of firm-specific characteristics, proxying for time-invariant unobservables, including reasonable measures of knowledge spillovers, adopting alternative categorizations for diversity and using proper instruments for the labor diversity dimensions of interest, we find robust evidence that ethnic diversity of the labor force is a fundamental source of innovation. That facilitates firms' patenting activity in several ways: (a) slightly increases their propensity to (apply for a) patent, (b) increases the overall number of patent applications and (c) enlarges the breadth of patenting technological fields. Being prudent in the quantification of ethnic heterogeneity effects on all these aspects of patenting activities, we find that a 10 percentage change in ethnic diversity increases the number of firms' patent applications by approximately 1.5 percent. The contribution of ethnic diversity in terms of general propensity to send at least one patent application in a given year is quite modest: a standard deviation change in its value turns to raise such a probability by 0.05 percent. However, the effect of educational diversity on extensive margins is large, a standard deviation change in skill diversity leads to a raise of about 7.3 percent in the firms' probability to apply for a patent in different technological areas. Thus, in order to widen the patent technological spectrum it seems to be fundamental to increase the heterogeneity in the workers' perspectives stemming from different cultural background. Regarding the results of education and demographic diversity on innovation, their effects typically vanishes when we include a the full set of controls or instrument such measures. Finally, we find that the beneficial effect of ethnic diversity on innovation materializes for both the white-collar and the blue-collar occupations: the effect is however larger for the former category of workers. These results support the hypothesis that diverse workers tend to have a wider pool of different experiences, knowledge bases and heuristics boosting their problem-solving capacities and creativity, which in turn facilitate innovations. In this regard, our findings are consistent with the theoretical frameworks proposed by Hong and Page (2001 and 2004) and Berliant and Fujita (2008).

The overall picture coming out from our empirical analysis seems to be particularly relevant not only for the design of firms' innovation strategies but also for public policies aimed at fostering innovation. Our results give an important insight into the technological process, a driver of productivity growth and hence of the economic growth. We find that an increase in firm labor diversity in terms of education and ethnicity has a positive effect on the firm innovation process. Thus, governmental policies aimed to encourage the employment of workers with different cultural backgrounds can be beneficial in terms of improvements in firms' patenting activities, increasing both private returns, directly, and social gains, through knowledge diffusion mechanisms. Nowadays, such policies might contribute to attract foreign and domestically less abundant skilled labor by supporting investments in human capital. That could be one of the determinants to invert the general decline in patenting activity recorded during the recent economic crisis among the OECD countries (OECD, 2009).

Acknowledgments

We thank Guglielmo Barone, Tor Eriksson, Hideo Owan, Pekka Ilmakunnas, Michael Rosholm, Chad Syverson and Måns Soderbom (alphabetical order) for helpful suggestions. In addition, we appreciate comments from participants at seminars organized by the Copenhagen Business School, University of Bergamo, Aarhus School of Business, University of Lausanne, and from participants at the following conferences: ESPE 2010, The 5th Nordic Summer Institute in Labor Economics, The 2010 Ratio Young Scientist Colloquium, ESEM 2010, CAED/COST 2010 in London, the 2010 International Symposium on Contemporary Labor Economics at WISE, Xiamen, and EALE 2011 in Paphos. The usual disclaimer applies.

Pierpaolo Parrotta acknowledges the financial support from the Graduate School for Integration, Production and Welfare. Mariola Pytlikova gratefully acknowledges funding from the NORFACE Migration Programme "Migration: Integration, Impact and Interaction".

References

[1] Adsera, A., Pytlikova, M., 2010. The Role of Language in Shaping International Migration: Evidence from OECD Countries 1985-2006, Unpublished manuscript.

[2] Andersen, A.K., 2000. Commuting Areas in Denmark, AKF working paper.

[3] Audretsch, D.B., Feldman, M. P., 1996. R&D Spillovers and the Geography of Innovation and Production. American Economic Review, 86, 630-640.

[4] Adams, J.D., 1990. Fundamental Stocks of Knowledge and Productivity Growth, Journal of Political Economy, 98, 673-703.

[5] Adams, J.D., Jaffe, A., 1996. Bounding the Effects of R&D: An Investigation Using Linked Establishment and Firm Data Rand Journal of Economics, 98, 673-702.

[6] Anderson, R., Quigley, J. M., Wilhelmsson, M., 2005. Agglomeration and the spatial distribution of creativity, Papers in Regional Science, 83, 445–464.

[7] Bantel, K.A., Jackson, S. E., 1989. Top Management and Innovations in Banking: Does the Composition of the Top Team Make a Difference?, Strategic Management Journal, 10, 107-124.

[8] Basset-Jones, N., 2005. The Paradox of Diversity Management, Creativity and Innovation, Creativity and Innovation Management, 14, 169-175.

[9] Becker, G., 1957. The Economics of Discrimination. University of Chicago Press, Chicago.

[10] Berliant, M., Fujita, M., 2008. Knowledge Creation as a Square Dance on the Hilbert Cube. International Economic Review, 49, 1251–1295.

[11] Bloom, N., Van Reenen, J., 2002. Patents, real options and firm performance, Economic Journal, 112, 97-116.

[12] Blundell, R., Griffith, R., Van Reenen, J., 1995. Dynamic Count Data Models of Technological Innovation, Economic Journal, 105, 333-44.

[13] Blundell, R., Griffith, R., Van Reenen, J., 1999. Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms, The Review of Economic Studies, 66, 529-554.

[14] Blundell, R., Griffith, R., Windmeijer, F., 2002. Individual Effects and Dynamics in Count Data Models, Journal of Econometrics, 108, 113-131.

[15] Cohen, W.M., Levinthal, D.A., 1990. Absorptive Capacity: A New Perspective on Learning and Innovation, Administrative Science Quarterly, 5, 128-152.

[16] Deding, M., Filges, T., Van Ommeren, J., 2009. Spatial Mobility and Commuting: the Case of Two-Earner Households, Journal of Regional Science, 49, 113-147.

[17] Drach-Zahavy, A., Somech, A., 2001. Understanding Team Innovation: The Role of Team Processes and Structures, Group Dynamics: Theory, Research, and Practice, 5 (2), 111-123.

[18] European Commission, 2005. The Business Case for Diversity: Good Practices in the Workplace.

[19] Feldman, M.P., Audretsch, D.B., 1999. Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition. European Economic Review, 43, 409–429.

[20] Griliches, Z., 1990. Patent Statistics as Economic Indicators: A Survey, Journal of Economic Literature, 28, 1661-1707.

[21] Guiso, L., Sapienza, P., and Zingales, L., 2009. Cultural Biases in Economic Exchange?, Quarterly Journal of Economics, 124, 1095-1131.

[22] Harrison, D.A., Klein, K.J., 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations, Academy of Management Review, 32, 1199-1228.

[23] Hong, L., Page Scott E., 2001. Problem Solving by Heterogeneous Agents Journal of Economic Theory, 97 (1), 123-163.

[24] Hong, L., Page Scott, E., 2004. Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers, Proceedings of the National Academy of Sciences, 101, (46), 16385-16389.

[25] Horwitz, S.K., Horwitz, I.B., 2007. The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography, Journal of Management, 33, 987-1015.

[26] Jaffe, A.B., 1986. Technological Opportunity and Spillovers of R&D, American Economic Review, 76, 984-1001.

[27] Kaiser, U., Kongsted, H., Rønde, T., 2008. Labor Mobility and Patenting Activity, CAM working paper no. 16.

[28] Kerr, W., Lincoln, W., 2010. The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention, Journal of Labor Economics, 28, 473-508.

[29] Kelley, M.R., Helper, S., 1999. Firm Size and Capabilities, Regional Agglomeration, and the Adoption of New Technology, Economics of Innovation and New Technology, 8, 79–103.

[30] Knight, D., Pearce, C.L., Smith, K.G., Olian, J.D., Sims, H. P., Smith, K.A., Flood, P., 1999. Top management team diversity, group process, and strategic consensus, Journal of Strategic Management, 20, 445–465.

[31] Lazear, E.P., 1998. Personnel Economics for Managers. New York, John Wiley & Sons.

[32] Lazear, E.P., 1999. Globalisation and the Market for Team-Mates, The Economic Journal, 109, c15-c40.

[33] Maignan, C., Ottaviano, G., Pinelli, D., Rullani, F., 2003. Bio-Ecological Diversity vs. Socio-Economic Diversity: A Comparison of Existing Measures, Nota di Lavoro, Fondazione Eni Enrico Mattei.

[34] Montgomery, J.D., 1991. Social Networks and Labor Market Outcomes: Toward an Economic Analysis, American Economic Review, 81, 1408-1418.

[35] Munshi, K., 2003. Networks in the Modern Economy: Mexican Migrants in the US Labor Market, The Quarterly Journal of Economics, 118, 549-599.

[36] Niebuhr, A., 2010. Migration and innovation: Does cultural diversity matter for regional R&D activity?, Papers in Regional Science, 89, 563–585.

[37] OECD, 2009. Policy responses to the Economic crisis. Investing in Innovation for Long-Term Growth, Paris.

[38] Osborne, E., 2000. The Deceptively Simple Economics of Workplace Diversity, Journal of Labor Research, 21, 463-475.

[39] Parrotta, P., Pozzoli, D., Pytlikova, M., 2010. Does Labor Diversity Affect Firm Productivity? Aarhus School of Business Working Paper No. 10-12. Aarhus.

[40] Pedersen,J.P., Pytlikova, M., Smith, N., 2008. Selection and Network Effects - Migration Flows into OECD Countries 1990-2000, European Economic Review, 52 (7) ,1160-1186.

[41] Pitcher, P., Smith, A.D., 2001. Top Management Team Heterogeneity: Personality, Power, and Proxies, Organization Science, 12 (1), 1-18.

[42] Söllner, R., 2010. Human Capital Diversity and Product Innovation: A Micro-Level Analysis, Jena Economic Research Papers.

[43] Vuong, Q.H., 1984. Two-Stage Conditional Maximum Likelihood Estimation of Econometric Models, Social Science Working Paper, 538, California Institute of Technology.

[44] Wallsten, S.J., 2001. An Empirical Test of Geographic Knowledge Spillovers Using Geographic Information Systems and Firm-Level Data. Regional Science and Urban Economics, 31 (5), 571-599.

[45] Watson, W.E., Kumar, K., Michaelsen, L.K., 1993. Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups, The Academy of Management Journal, 36 (3), 590-602.

[46] Williams, K.Y., O'Reilly III. C.A, 1998. Demography and Diversity in Organizations: A Review of 40 Years of Research, In B.M. Staw & L.L. Cummings (Eds.), Research in Organizational Behavior, 20, 77-140.

[47] Zajac, E., Golden, B.R., Shortell, S.M., 1991. New Organizational Forms for Enhancing Innovation: The Case of Internal Corporate Joint Ventures, Management Science, Vol. 37 (2), 170-184.

Appendix 1: Measurement of ethnic diversity

1) The citizens in the different nationality groups are: **Danish**, Danish native including second generation immigrants; **North America and Oceania**, United States, Canada, Australia, New Zealand; **Central and South America**, Guatemala, Belize, Costa Rica, Honduras, Panama, El Salvador, Nicaragua, Venezuela, Ecuador, Peru, Bolivia, Chile, Argentina, Brazil; **Formerly Communist Countries**, Armenia, Belarus, Estonia, Georgia, Latvia, Lithuania, Moldova, Russia, Tajikistan, Ukraine, Bulgaria, Czech Republic, Hungary, Poland, Romania, Slovakia, Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Rep. of Macedonia, Montenegro, Serbia, and Slovenia; **Muslim Countries**, Afghanistan, Algeria, Arab Emirates, Azerbaijan, Bahrain, Bangladesh, Brunei Darussalem, Burkina Faso, Camoros, Chad, Djibouti, Egypt, Eritrea, Gambia, Guinea, Indonesia, Iran, Iraq, Jordan, Kazakstan, Kirgizstan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Malaysia, Maldives, Mali, Mauritania, Morocco, Nigeria, Oman, Pakistan, Palestine, Qatar, Saudi Arabia, Senegal, Sierra Leone, Somalia, Sudan, Syria, Tadzhikstan, Tunisia, Turkey, Turkmenistan, Uzbekistan, Yemen; **East Asia**, China, Hong Kong, Japan, Korea, Korea Dem. People's Rep. Of, Macao, Mongolia, Taiwan; **Asia,** all the other Asian countries non included in both East Asia and Muslim Countries categories and **Africa,** all the other African countries not included in the Muslim Country; **Western and Southern Europe**, all the other European countries not included in the Formerly Communist Countries category.

2) Using linguistic grouping: **Germanic West** (Antigua Barbuda, Aruba, Australia, Austria, Bahamas, Barbados, Belgium, Belize, Bermuda, Botswana, Brunei, Cameroon, Canada, Cook Islands, Dominica, Eritrea, Gambia, Germany, Ghana, Grenada, Guyana, Haiti, Ireland, Jamaica, Liberia, Liechtenstein, Luxemburg, Mauritius, Namibia, Netherlands, Netherlands Antilles, New Zealand, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and Grenadines, Seychelles, Sierra Leone, Solomon Islands, South Africa, St. Helena, Suriname, Switzerland, Trinidad and Tobago, Uganda, United Kingdom, United States, Zambia, Zimbabwe), **Germanic Nord** (Denmark, Iceland, Norway, Sweden), **Slavic West** (Czech Republic, Poland, Slovakia), **Slavic South (**Bosnia and Herzegovina, Croatia, Serbia, Slovenia)**, Slavic East** (Belarus, Georgia, Mongolia, Russian Federation, Ukraine)**, Baltic East** (Latvia, Lithuania)**, Finno-Permic** (Finland, Estonia), **Ugric** (Hungary)**, Romance** (Andorra, Angola, Argentina, Benin, Bolivia, Brazil, Burkina Faso, Cape Verde, Chile, Columbia, Costa Rica, Cote D'Ivoire, Cuba, Djibouti, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, France, French Guina, Gabon, Guadeloupe, Guatemala, Guinea, Guinea Bissau, Holy See, Honduras, Italy, Macau, Martinique, Mexico, Moldova, Mozambique, Nicaragua, Panama, Peru, Portugal, Puerto Rico, Reunion, Romania, San Marino, Sao Tome, Senegal, Spain, Uruguay, Venezuela), **Attic** (Cyprus, Greece)**, Turkic South** (Azerbaijan, Turkey, Turkmenistan), **Turkic West** (Kazakhstan, Kyrgystan)**, Turkic East** (Uzbekistan)**, Gheg** (Albania, Kosovo, Republic of Macedonia, Montenegro), **Semitic Central** (Algeria, Bahrain, Comoros, Chad, Egypt, Irak, Israel, Jordan, Kuwait, Lebanon, Lybian Arab Jamahiria, Malta, Mauritiania, Morocco, Oman, Qatar, Saudi Arabia, Sudan, Syrian Arab Republic, Tunisia, Yemen**,** United Arabs Emirates)**, Indo-Aryan (**Bangladesh, Fiji, India, Maldives, Nepal, Pakistan, Sri Lanka)**, Mon-Khmer East** (Cambodia)**, Semitic South** (Ethiopia)**, Malayo-Polynesian West** (Indonesia, Philippines)**, Malayo-Polynesian Central East** (Kiribati, Marshall Islands, Nauru, Samoa, Tonga)**, Iranian** (Afghanistan, Iran, Tajikistan)**, Betai** (Laos, Thailand)**, Malayic** (Malasya)**, Cushitic East** (Somalia)**, Viet-Muong** (Vietnam), **Volta-Congo** (Burundi, Congo, Kenya, Lesotho, Malawi, Nigeria, Rwanda, Swaziland, Tanzania, Togo)**, Barito** (Madagascar)**, Mande West** (Mali)**, Lolo-Burmese** (Burma)**, Chadic West** (Niger)**,**

Guarani (Paraguay), **Himalayish** (Buthan), **Armenian** (Armenia), **Sino Tibetan** (China, Hong Kong, Singapore, Taiwan), **Japonic** (Japan, Republic of Korea, Korea D.P.R.O.).

Appendix 2: External knowledge indexes

The main literature on agglomeration economies emphasizes the importance of firm's local environment, which may reflect information advantages, labor or other inputs pooling and further beneficial network effects aimed at alleviating the burden represented by fixed costs. A seminal contribution in this field is due to Audretsch and Feldman (1996), who find that industries characterized by elevated R&D intensity or particularly skilled labor forces present a greater degree of geographic concentration of production. Other relevant studies like Wallsten (2001) and Adams and Jaffe (1996) provide evidence of the geographic extent of knowledge spillovers by computing the distance in miles between each firm-pair. However, the geography is not the only dimension of the external knowledge. In fact, there exists at least another approach which focuses on the concept of technological proximity (Jaffe, 1986; Adams, 1990). Specifically, the idea that the technology developed by a firm can affect other firms, even though they are not geographically close or no transactions of goods occur between them, has led to the definition of technological proximity as closeness between firm-pairs' technological profiles.

Following both the cited approaches, we construct two indexes of knowledge spillovers. These are weighted sums of firms' codified knowledge proxied by the discounted stock of patent applications.[16] The weighting function for the first index refers to the geographical distance between pairs of workplaces' municipalities and is computed by using the firms' latitude and longitude coordinates (the address of their headquarters). Specifically, assuming a spherical earth of actual earth volume, this method allows us to measure the distance in kilometers between any pair of firms i and j .[17] The first knowledge spillover index is then computed as follows:

$$K\_geo_{it} = \frac{1}{e^{dist_{ij}}} \sum_{j \neq i}^{I} disc\_stock_{jt} \, .$$

The second index is instead based on the technological proximity. Following Adams (1990), we use the shares of differently skilled workers to define our alternative weighting function ψ_{ij} that is the uncentered

[16] See paragraph 4.2.

[17] We use the following formula $d_{ij} = 6378.7 * acos\{sin(lat_i/57.2958) * sin(lat_j/57.2958) + \; + cos(lat_i/57.2958) * cos(lat_j/57.2958) * cos(lon_j/57.2958 - lon_i/57.2958)\}$.

correlation:

$$\psi_{ij} = \frac{f_i f_j^{'}}{\left[\left(f_i f_i^{'}\right)\left(f_j f_j^{'}\right)\right]^{1/2}} .$$

The components of the generator vector f reflects firm's workforce composition in terms of skills using the disaggregated categorization as described in section 3.2. The second measure of knowledge spillover pool is therefore defined as

$$K\_tech_{it} = \psi_{ij} \sum_{j \neq i}^{I} disc\_stock_{jt} .$$

Thus, both K_geo_{it} and K_tech_{ij} contain weighting functions that might capture the so-called firm's absorptive capacity, which is the ability to identify and exploit the knowledge externally produced (Cohen and Levinthal, 1990).

Table 1: Descriptive statistics

		All sample			Non-patenting firms			Patenting firms		
Variables	Definition	Median	Mean	Sd	Median	Mean	Sd	Median	Mean	Sd
IDA Variables:										
males	men as a proportion of all employees	0.786	0.709	0.243	0.786	0.706	0.247	0.174	0.674	0.199
foreigners	non-danish employees as a proportion of all employees	0	0.042	0.086	0	0.423	0.494	1	0.750	0.433
age1	employees aged 15-28 as a proportion of all employees	0.304	0.325	0.172	0.304	0.325	0.173	0.263	0.280	0.127
age2	employees aged 29-36 as a proportion of all employees	0.255	0.259	0.120	0.250	0.257	0.121	0.296	0.300	0.090
age3	employees aged 37-47 as a proportion of all employees	0.200	0.204	0.109	0.200	0.204	0.110	0.222	0.219	0.079
age4	employees aged 47-65 as a proportion of all employees	0.251	0.212	0.124	0.252	0.178	0.15	0.232	0.162	0.067
skill1	employees with compulsory education as a proportion of all employees	0.176	0.271	0.129	0.164	0.272	0.128	0.201	0.238	0.123
skill2	employees with a secondary/ post-secondary education as a proportion of all employees	0.714	0.689	0.189	0.714	0.690	0.189	0.658	0.662	0.147
skill3	employees with a tertiary education as a proportion of all employees	0	0.040	0.099	0	0.038	0.097	0.043	0.100	0.137
tenure	average tenure	4.473	4.622	1.867	4.466	4.616	1.871	5.038	5.025	1.596
manager	managers as a proportion of all employees	0.018	0.045	0.064	0.016	0.045	0.064	0.037	0.052	0.059
middle manager	middle managers as a proportion of all employees	0.837	0.762	0.241	0.842	0.764	0.240	0.658	0.599	0.240
blue collars	blue collars as a proportion of all employees	0.140	0.236	0.348	0.140	0.234	0.348	0	0.384	0.486
size1	total number of employees (less than 50)	1	0.816	0.387	1	0.825	0.379	0	0.154	0.316
size2	total number of employees (50-100)	0	0.096	0.295	0	0.093	0.291	0	0.416	0.498
size3	total number of employees (more than 100)	0	0.087	0.281	0	0.080	0.272	0	0.056	0.324
Index Ethnic aggr	diversity index based on employees' nationality	0	0.101	0.204	0	0.098	0.201	0.499	0.373	0.206
Index Edu aggr	diversity index based on employees' education	0.444	0.415	0.115	0.444	0.415	0.115	0.500	0.478	0.107
Index demo aggr	diversity index based on employees' demographic characteristics	0.764	0.752	0.072	0.764	0.752	0.072	0.805	0.796	0.055
Index Ethnic disaggr	diversity index based on employees' spoken language	0	0.197	0.315	0	0.193	0.313	0.664	0.553	0.317
Index Edu disaggr	diversity index based on employees' education	0.583	0.578	0.134	0.582	0.577	0.133	0.706	0.714	0.104
Index Demo disaggr	diversity index based on employees' demographic characteristics	0.889	0.876	0.073	0.888	0.876	0.073	0.932	0.922	0.059
Accounting Variables:										
Patent applications	annual number of patent applications	0	0.029	0.628	0	0	0	0	0.829	3.142
capital	(1000 kr.)	11334.29	73542.36	841303.4	10864	57015.39	781429.8	77714.73	541278.6	2071364
foreign-ownership	1, if the firm is foreign owned	0	0.004	0.066	0	0.005	0.066	0	0.004	0.061
multi	1, if the firm is multi-establishment	0	0.093	0.260	0	0.093	0.291	0	0.298	0.457
exp	1, if the firm is exporting	1	0.506	0.499	0	0.488	0.499	1	0.874	0.331
geo_spillover	spillover variable based on the technological distance	40.1925	226.697	456.646	40.19252	228.2731	228.2731	1130.534	1063.769	362.0997
tech_spillover	spillover variable based on the geographical distance	1091.168	1031.535	345.931	1090.384	1030.382	345.2853	50.08433	182.6429	340.2594
N			102789			101655			1134	

Notes: : All workforce composition and accounting variables are expressed as time averages from 1995 to 2003. The industrial sectors included in the empirical analysis are the following: food, beverages and tobacco (4.05 %); textiles (2.24 %), wood products (6.68 %), chemicals (3.49 %), other non-metallic mineral products (1.50 %), basic metals (19.13 %), furniture (3.79 %), construction (22.40 %), wholesale trade (14.67 %), retail trade (9.02 %), post and telecommunications (0.27 %), financial intermediation (1.19 %) and business activities (11.02 %).

Table 2: The effects of labor diversity on firm probability to innovate. Main results.

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)	Model (9)	Model (10)	Model (11)	Model (12)
	Probit	**Probit**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**	**Probit**	**Probit**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**
Index Ethnic	0.02258***	0.00069**	0.00266**	0.00282**			0.01012***	0.00027**	0.00217*	0.00271**		
	(0.00161)	(0.00026)	(0.00097)	(0.00097)			(0.00103)	(0.00012)	(0.00115)	(0.00119)		
Index Edu	0.01842***	0.00105	-0.00162		-0.00048		0.03101***	0.00404***	0.00168		0.00240	
	(0.00405)	(0.00104)	(0.00394)		(0.00376)		(0.00243)	(0.00091)	(0.00161)		(0.00162)	
Index Demo	0.04182***	0.00105	0.00315			0.00470	0.02424***	-0.00004	0.00553			0.00691
	(0.00585)	(0.00148)	(0.00381)			(0.00366)	(0.00412)	(0.00125)	(0.00350)			(0.00359)
Log(K)		0.00055***	0.00055***	0.00055***	0.00055***	0.00054***		0.00047***	0.00046***	0.00048***	0.00047***	0.00047***
		(0.00010)	(0.00010)	(0.00010)	(0.00010)	(0.00010)		(0.00010)	(0.00010)	(0.00010)	(0.00010)	(0.00010)
Log(L)		0.00045**	0.00043**	0.00044**	0.00048**	0.00047**		0.00046***	0.00045**	0.00045**	0.00048***	0.00047***
		(0.00016)	(0.00015)	(0.00015)	(0.00016)	(0.00016)		(0.00014)	(0.00014)	(0.00014)	(0.00014)	(0.00014)
Log(fixed effects)		0.00601***	0.00599***	0.00602***	0.00603***	0.00596***		0.00535***	0.00533***	0.00551***	0.00535***	0.00540***
		(0.00142)	(0.00142)	(0.00143)	(0.00143)	(0.00141)		(0.00133)	(0.00131)	(0.00135)	(0.00133)	(0.00132)
age1		0.00079	0.00113	0.00080	0.00072	0.00111		0.00031	0.00108	0.00074	0.00046	0.00131*
		(0.00073)	(0.00084)	(0.00073)	(0.00073)	(0.00080)		(0.00066)	(0.00076)	(0.00066)	(0.00066)	(0.00075)
age2		0.00223**	0.00230**	0.00217**	0.00209**	0.00232**		0.00163**	0.00191**	0.00194**	0.00163**	0.00194**
		(0.00079)	(0.00080)	(0.00077)	(0.00076)	(0.00080)		(0.00069)	(0.00072)	(0.00070)	(0.00069)	(0.00070)
age3		0.00242**	0.00241**	0.00254**	0.00242**	0.00221**		0.00208**	0.00199**	0.00241**	0.00212**	0.00190**
		(0.00101)	(0.00102)	(0.00101)	(0.00100)	(0.00100)		(0.00093)	(0.00093)	(0.00094)	(0.00092)	(0.00092)
males		-0.00047	-0.00017	-0.00047	-0.00068	-0.00014		-0.00025	0.00025	-0.00042	-0.00042	0.00009
		(0.00048)	(0.00058)	(0.00042)	(0.00044)	(0.00058)		(0.00043)	(0.00054)	(0.00039)	(0.00042)	(0.00052)
foreigners		-0.00000	-0.00000	-0.00000	-0.00000	-0.00000		-0.00037	-0.00274	-0.00324*	0.00017	0.00028
		(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)		(0.00098)	(0.00174)	(0.00182)	(0.00083)	(0.00082)
exp		0.00096***	0.00093***	0.00095***	0.00100***	0.00094***		0.00085***	0.00082***	0.00092***	0.00089***	0.00088***
		(0.00021)	(0.00021)	(0.00020)	(0.00021)	(0.00021)		(0.00020)	(0.00020)	(0.00020)	(0.00021)	(0.00020)
skill1		0.00266**	0.00187	0.00252***	0.00196	0.00220***		0.00191**	0.00195**	0.00192**	0.00185**	0.00187**
		(0.00087)	(0.00171)	(0.00067)	(0.00163)	(0.00064)		(0.00062)	(0.00063)	(0.00063)	(0.00062)	(0.00062)
skill2		0.00666***	0.00763***	0.00697***	0.00725***	0.00712***		0.00465***	0.00537***	0.00572***	0.00513***	0.00626***
		(0.00138)	(0.00200)	(0.00136)	(0.00194)	(0.00138)		(0.00123)	(0.00130)	(0.00126)	(0.00129)	(0.00132)
manager		0.00099	0.00120	0.00123	0.00095	0.00087		0.00129	0.00133	0.00139	0.00117	0.00095
		(0.00100)	(0.00101)	(0.00101)	(0.00100)	(0.00099)		(0.00090)	(0.00092)	(0.00094)	(0.00090)	(0.00090)
middle manager		-0.00059	-0.00043	-0.00049	-0.00058	-0.00055		0.00019	-0.00003	-0.00034	-0.00010	-0.00045
		(0.00044)	(0.00044)	(0.00044)	(0.00044)	(0.00044)		(0.00041)	(0.00050)	(0.00041)	(0.00050)	(0.00039)
tenure		-0.00013**	-0.00012**	-0.00011**	-0.00013**	-0.00014***		-0.00011**	-0.00013**	-0.00012**	-0.00012**	-0.00012**
		(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)		(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
multi		-0.00013	-0.00012	-0.00013	-0.00018	-0.00015		-0.00008	-0.00005	-0.00009	-0.00011	-0.00009
		(0.00016)	(0.00016)	(0.00016)	(0.00015)	(0.00016)		(0.00015)	(0.00016)	(0.00015)	(0.00015)	(0.00015)
geo_spillover		0.00000	0.00000	0.00000	0.00000	0.00000		-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
		(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)		(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
tech_spillover		0.00001**	0.00001**	0.00001**	0.00001**	0.00001**		0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
		(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)		(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Industry/size/year dummies	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes
N	103025	97976	97976	97976	97976	97976	103025	97976	97976	93527	97976	97976
pseudo R2	0.138	0.422	0.423	0.425	0.423	0.422	0.192	0.425	0.426	0.426	0.425	0.425

Notes: The dependent variable in all estimations is the probability to have at least one patent application. Marginal effects reported. Model1-Model6: diversity based on the aggregate specification. Model7-Model12: diversity based on the detailed specification. Model3-Model6 and Model9-Model12 report results from IV estimation. Significance levels: ***1%, **5%, *10%. Standard errors clustered at the firm level.

Table 3: The effects of labor diversity on firm patent applications.

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)	Model (9)	Model (10)	Model (11)	Model (12)
	Poisson	Poisson	Poisson (IV)	Poisson (IV)	Poisson (IV)	Poisson (IV)	Poisson	Poisson	Poisson (IV)	Poisson (IV)	Poisson (IV)	Poisson (IV)
Index Ethnic	0.48115***	0.07572**	0.15251**	0.14168**			0.54240***	0.06835**	0.20673**	0.23283**		
	(0.04152)	(0.02562)	(0.07402)	(0.07017)			(0.05235)	(0.03294)	(0.10196)	(0.10616)		
Index Edu	2.37860***	0.91495*	-2.41377		-2.69811		5.97907***	2.37100***	0.93471		1.13124	
	(0.48204)	(0.56572)	(2.37654)		(2.15371)		(0.59530)	(0.62668)	(0.90999)		(0.89126)	
Index Demo	8.79827***	0.10902	-1.65693			-2.37123	11.48962***	-0.09175	0.33787			1.24195
	(1.47977)	(1.44370)	(2.84554)			(2.48879)	(2.14927)	(1.48847)	(3.20575)			(3.13482)
Log(K)		5.13789***	5.22192***	5.13664***	5.20850***	5.08145***		4.93562***	4.93290***	4.95866***	4.93126***	4.97260***
		(0.60471)	(0.59774)	(0.61075)	(0.59961)	(0.60742)		(0.61851)	(0.61743)	(0.61424)	(0.61429)	(0.61905)
Log(L)		0.83986**	0.84720**	0.81652**	0.91050**	0.91581**		0.96090**	0.92641**	0.92034**	0.98015**	0.95729**
		(0.36785)	(0.36738)	(0.36582)	(0.36353)	(0.35934)		(0.35250)	(0.34438)	(0.34941)	(0.34777)	(0.34099)
Discounted stock of applications		-0.00014	-0.00014	-0.00010	-0.00024	-0.00021		-0.00006	-0.00013	-0.00019	-0.00016	-0.00022
		(0.00038)	(0.00038)	(0.00038)	(0.00038)	(0.00037)		(0.00038)	(0.00041)	(0.00039)	(0.00041)	(0.00040)
Log(fixed effects)		0.00342	0.00344	0.00330	0.00379*	0.00382*		0.00310	0.00337	0.00357	0.00343	0.00362
		(0.00229)	(0.00222)	(0.00227)	(0.00220)	(0.00223)		(0.00227)	(0.00244)	(0.00234)	(0.00239)	(0.00238)
Fixed effect dummy		0.05409***	0.05332***	0.05369***	0.05301***	0.05346***		0.05335***	0.05330***	0.05330***	0.05323***	0.05327***
		(0.00581)	(0.00579)	(0.00576)	(0.00579)	(0.00579)		(0.00582)	(0.00587)	(0.00588)	(0.00585)	(0.00587)
age1		0.18680	0.20681	0.19697	0.25515	0.11956		0.16123	0.23961	0.25809	0.21491	0.28081
		(0.23997)	(0.27548)	(0.24127)	(0.24296)	(0.24666)		(0.23785)	(0.24651)	(0.23682)	(0.23797)	(0.23734)
age2		0.21909	0.19276	0.23211	0.23732	0.19123		0.19164	0.24987	0.27462	0.23081	0.27345
		(0.27515)	(0.27861)	(0.27308)	(0.26544)	(0.27544)		(0.26907)	(0.27877)	(0.27550)	(0.27425)	(0.26947)
age3		0.28949	0.33926	0.29910	0.31785	0.33040		0.29628	0.34242	0.36597*	0.33271	0.33554
		(0.22752)	(0.21676)	(0.22633)	(0.21447)	(0.21834)		(0.22338)	(0.22241)	(0.22360)	(0.21851)	(0.21672)
males		0.05643	-0.31801	-0.00096	-0.19496	-0.28737		0.28395	0.19850	0.07635	0.13696	0.12131
		(0.53450)	(0.49134)	(0.41860)	(0.34573)	(0.49161)		(0.52814)	(0.50646)	(0.40491)	(0.44018)	(0.49062)
foreigners		-0.08525	-0.08765	-0.08316	-0.08528	-0.08749		0.01253	-0.02539	-0.02768	0.05064	0.06030*
		(0.06546)	(0.06284)	(0.06347)	(0.06254)	(0.06270)		(0.03517)	(0.07043)	(0.06908)	(0.03435)	(0.03175)
exp		0.52206***	0.55964***	0.52798***	0.56319***	0.54909***		0.51055***	0.53439***	0.55165***	0.54384***	0.55839***
		(0.11581)	(0.11980)	(0.11689)	(0.12071)	(0.11780)		(0.11323)	(0.11337)	(0.11432)	(0.11325)	(0.11332)
skill1		1.76808**	-0.26923	1.36318**	-0.53601	1.17806**		0.86503**	0.83850**	0.82499**	0.80831*	0.82118**
		(0.65893)	(1.61313)	(0.45323)	(1.48118)	(0.44059)		(0.41243)	(0.42138)	(0.41343)	(0.41458)	(0.41566)
skill2		0.16098***	0.27888**	0.19634***	0.29432***	0.19588***		0.12450***	0.15478***	0.17533***	0.15594***	0.18633***
		(0.03407)	(0.09217)	(0.02943)	(0.08331)	(0.02952)		(0.03255)	(0.03762)	(0.02860)	(0.03741)	(0.02930)
manager		-0.01169	0.00523	-0.00088	-0.00463	-0.00805		0.00808	0.00536	0.00111	-0.00185	-0.00881
		(0.03899)	(0.03645)	(0.03862)	(0.03676)	(0.03764)		(0.03836)	(0.03833)	(0.03838)	(0.03813)	(0.03719)
middle manager		-0.34684	-0.19535	-0.25467	-0.19296	-0.33304		0.07243	-0.17496	-0.36729	-0.19765	-0.41335
		(0.26734)	(0.26958)	(0.27218)	(0.26488)	(0.27982)		(0.29182)	(0.34885)	(0.28845)	(0.34680)	(0.29176)
tenure		-0.33242	-0.23732	-0.29018	-0.30125	-0.30591		-0.30569	-0.32792	-0.33508	-0.33752	-0.36754
		(0.26523)	(0.28340)	(0.27208)	(0.26626)	(0.27034)		(0.27698)	(0.30216)	(0.28360)	(0.27341)	(0.29695)
multi		-0.00259	-0.00569	-0.00119	-0.00910	-0.00699		0.00385	0.00526	0.00390	0.00223	0.00135
		(0.01941)	(0.01813)	(0.01884)	(0.01841)	(0.01944)		(0.01975)	(0.01912)	(0.01888)	(0.01952)	(0.01954)
geo_spillover		0.53659	0.43005	0.68790	0.39735	0.54939		0.87239	0.81209	0.65691	0.79920	0.63429
		(0.57836)	(0.57920)	(0.56229)	(0.57774)	(0.55393)		(0.55387)	(0.55404)	(0.50878)	(0.55612)	(0.51021)
tech_spillover		0.00331	0.00736	0.00694	0.00208	0.00005		0.02457	0.02360	0.02215	0.01971	0.01733
		(0.04263)	(0.04326)	(0.04317)	(0.04210)	(0.04136)		(0.03898)	(0.04034)	(0.04070)	(0.03872)	(0.03920)
Industry/size/year dummies	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes
N	103025	97976	97976	97976	97976	97976	103025	97976	97976	93527	97976	97976
chi2	173.0	55680.5	51514.1	57270.4	53456.9	53578.2	184.4	47358.8	47794.8	46576.3	47199.0	48025.8

Notes: The dependent variable in all estimations is the number of patent applications. Elasticities reported. Model1-Model6: diversity based on the aggregate specification. Model7-Model12: diversity based on the detailed specification. Model3-Model6 and Model9-Model12 report results from IV estimation. Significance levels: ***1%, **5%, *10%. Standard errors clustered at the firm level. Poisson (IV): standard errors are bootstrapped using a sequential two step bootstrapping procedure with 200 replications.

Table 4: The effects of labor diversity on the probability of applying in different technological areas.

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (4)	Model (5)	Model (6)	Model (10)	Model (11)	Model (12)
	Probit	**Probit**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**	**Probit**	**Probit**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**	**Probit (IV)**
Index Ethnic	0.20253***	0.08572**	0.37377**	0.36025**			0.16594***	0.02485**	0.47286**	0.50231**		
	(0.05932)	(0.03612)	(0.16393)	(0.16358)			(0.04700)	(0.06121)	(0.23043)	(0.22723)		
Index Edu	0.66098***	1.12806**	0.31624		0.23833		0.90935***	0.50233**	0.02664		0.15989	
	(0.15334)	(0.34577)	(0.77091)		(0.75456)		(0.15177)	(0.22325)	(0.37167)		(0.37332)	
Index Demo	0.35108	0.12846	-0.43509			0.02358	0.54070*	0.57128*	0.48273			0.78830
	(0.30547)	(0.33855)	(0.80866)			(0.78969)	(0.30485)	(0.31785)	(0.80312)			(0.79251)
Log(K)		0.03205**	0.03319**	0.03335**	0.03243**	0.03272**		0.02936**	0.02774**	0.02786**	0.02770**	0.02714**
		(0.01200)	(0.01206)	(0.01208)	(0.01230)	(0.01236)		(0.01202)	(0.01191)	(0.01193)	(0.01226)	(0.01223)
Log(L)		0.02087	0.01918	0.01886	0.02406	0.02424		0.02431	0.02423	0.02509	0.03035	0.02860
		(0.02322)	(0.02300)	(0.02316)	(0.02343)	(0.02326)		(0.02342)	(0.02318)	(0.02348)	(0.02395)	(0.02353)
Log(fixed effects)		0.25801***	0.25321***	0.25241***	0.25299***	0.25236***		0.26798***	0.27038***	0.27106***	0.26707***	0.26812***
		(0.04677)	(0.04719)	(0.04708)	(0.04726)	(0.04731)		(0.04611)	(0.04582)	(0.04624)	(0.04615)	(0.04630)
age1		0.57099**	0.55412**	0.59831**	0.57784**	0.58840**		0.60236**	0.65887**	0.62246**	0.58157**	0.65943**
		(0.20239)	(0.20872)	(0.19973)	(0.20421)	(0.20318)		(0.20638)	(0.21332)	(0.20765)	(0.20909)	(0.21343)
age2		0.69057***	0.65777***	0.67916***	0.67822***	0.67367***		0.69681***	0.73910***	0.72480***	0.69007***	0.72066***
		(0.18933)	(0.18817)	(0.17843)	(0.18050)	(0.19061)		(0.18443)	(0.18862)	(0.18449)	(0.18103)	(0.18618)
age3		0.46901*	0.52787**	0.50504**	0.48515*	0.48803*		0.46175*	0.53973**	0.57719**	0.53074**	0.48097*
		(0.26024)	(0.26460)	(0.25519)	(0.25525)	(0.26330)		(0.25982)	(0.27151)	(0.25659)	(0.25459)	(0.26838)
males		0.10024	0.02564	0.06271	0.04616	0.04008		0.16285	0.14392	0.09172	0.06709	0.13774
		(0.10108)	(0.12765)	(0.08351)	(0.09141)	(0.13071)		(0.10346)	(0.12965)	(0.08742)	(0.09873)	(0.12652)
foreigners		0.00002	0.00001	0.00002	0.00001	0.00001		0.06526	-0.42656	-0.46270	0.18494	0.20729
		(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)		(0.30281)	(0.41045)	(0.41129)	(0.27068)	(0.26642)
exp		0.01074	0.01356	0.01258	0.01692	0.01848		0.02691	0.03053	0.03416	0.03878	0.03502
		(0.04073)	(0.04008)	(0.04059)	(0.03997)	(0.03941)		(0.04119)	(0.03996)	(0.03906)	(0.03808)	(0.03887)
skill1		0.79780**	0.47688	0.35795**	0.42024	0.31874**		0.11090	0.09946	0.07893	0.06605	0.10061
		(0.25391)	(0.35628)	(0.16147)	(0.35312)	(0.16009)		(0.14380)	(0.13847)	(0.13960)	(0.14207)	(0.13895)
skill2		0.14635	0.36830	0.50944**	0.46647	0.55330**		0.32775*	0.37779	0.33875*	0.37414*	0.50329**
		(0.21803)	(0.35683)	(0.18830)	(0.34526)	(0.18897)		(0.20464)	(0.23813)	(0.18613)	(0.22925)	(0.19197)
manager		0.37396	0.46451*	0.46036*	0.36865	0.37083		0.48560*	0.54175*	0.55292**	0.46037*	0.43626
		(0.26490)	(0.26373)	(0.26392)	(0.26585)	(0.26499)		(0.27414)	(0.28052)	(0.27899)	(0.27776)	(0.27718)
middle manager		0.06083	0.06925	0.08629	0.07769	0.08272		0.14020	0.08072	0.06670	0.06691	0.06056
		(0.09549)	(0.09595)	(0.09432)	(0.09566)	(0.09839)		(0.10277)	(0.11797)	(0.09764)	(0.11843)	(0.09843)
tenure		-0.01008	-0.00659	-0.00681	-0.00934	-0.00912		-0.01048	-0.00976	-0.00912	-0.00978	-0.01104
		(0.00849)	(0.00878)	(0.00864)	(0.00864)	(0.00867)		(0.00912)	(0.00934)	(0.00928)	(0.00908)	(0.00922)
multi		0.00077	-0.00062	0.00024	-0.00619	-0.00739		0.00247	0.00379	0.00063	-0.00600	-0.00208
		(0.03505)	(0.03454)	(0.03458)	(0.03419)	(0.03386)		(0.03667)	(0.03593)	(0.03645)	(0.03620)	(0.03559)
copatent		0.05318	0.05353	0.05404	0.05765*	0.05801*		0.06712*	0.06543*	0.06574*	0.06686*	0.06619*
		(0.03481)	(0.03493)	(0.03505)	(0.03518)	(0.03537)		(0.03676)	(0.03722)	(0.03739)	(0.03685)	(0.03676)
geo_spillover		0.00006	0.00006	0.00006	0.00006	0.00006		0.00006	0.00005	0.00006	0.00006	0.00007
		(0.00011)	(0.00010)	(0.00010)	(0.00011)	(0.00010)		(0.00011)	(0.00011)	(0.00011)	(0.00011)	(0.00011)
tech_spillover		0.00003	0.00003	0.00003	0.00002	0.00002		0.00002	0.00002	0.00002	0.00002	0.00002
		(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)		(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Industry/size/year dummies	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes
N	1090	1037	1037	1037	1037	1037	1090	1037	1037	1104	1104	1104
pseudo R2	0.068	0.330	0.335	0.335	0.331	0.330	0.101	0.321	0.327	0.326	0.322	0.323

Notes: The dependent variable in all estimations is the probability of applying a patent in different technological areas. Marginal effects reported. Model7-Model12: diversity based on the detailed specification. Model3-Model6 and Model9-Model12 report results from IV estimation. Significance levels: ***1%, **5%, 10*%. Standard errors clustered at the firm level.

Table 5: The effects of labor diversity on firm probability to innovate. Robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Occupation specific diversity		**Shannon entropy index**	**Richness**	**Edu and demo diversity as sd**	**2nd gen. Imm. as natives**	**Germanic group as natives**	**University graduates as natives**	**Copehagen is excluded**	**Mono-establishment firms**
	White collar	*Blue collar*								
index_fore_aggr	0.00689**	0.00263**	0.00092**	0.00373***	0.00383**	0.00011*		0.00023*	0.00248**	0.00301**
	(0.00242)	(0.00104)	(0.00033)	(0.00045)	(0.00118)	(0.00006)		(0.00010)	(0.00104)	(0.00108)
index_edu_aggr	-0.00191	-0.00222	-0.00091	-0.00305	0.02355	-0.00233		-0.00233	-0.00231	0.00029
	(0.00578)	(0.00378)	(0.00150)	(0.00079)	(0.01519)	(0.00390)		(0.00387)	(0.00424)	(0.00405)
Sd(years of education)					-0.00944					
					(0.00579)					
index_demo_aggr	0.00039	0.00249	0.00022	-0.00012		0.00541		0.00523	0.00253	0.00024
	(0.00449)	(0.00341)	(0.00024)	(0.00030)		(0.00383)		(0.00379)	(0.00413)	(0.00377)
Sd(age)					0.00045					
					(0.00031)					
Male					-0.00108					
					(0.00065)					
N	97976	97976	97976	97976	97976	97976		97976	91561	83970
pseudo R2	0.423	0.423	0.424	0.388	0.426	0.422		0.422	0.417	0.369
index_fore_disaggr	0.00515**	0.00222	0.00828**	0.00224***	0.00295**	0.00017**	0.00025**	0.00023*	0.00211**	0.00237**
	(0.00217)	(0.00160)	(0.00419)	(0.00033)	(0.00149)	(0.00006)	(0.00009)	(0.00011)	(0.00106)	(0.00113)
index_edu_disaggr	-0.00064	-0.00021	0.00021	0.00029	0.01370	0.00214	0.00217	0.00206	0.00085	0.00212
	(0.00213)	(0.00191)	(0.00218)	(0.00032)	(0.01054)	(0.00160)	(0.00164)	(0.00159)	(0.00175)	(0.00165)
Sd(years of education)					-0.00486					
					(0.00465)					
index_demo_disaggr	0.00374	0.00599	0.00341	0.00022		0.00642	0.00645*	0.00630	0.00540	0.00398
	(0.00322)	(0.00385)	(0.00201)	(0.00032)		(0.00352)	(0.00360)	(0.00348)	(0.00382)	(0.00351)
Sd(age)					0.00085					
					(0.00049)					
Male					0.00024					
					(0.00047)					
N	97976	97976	97976	97976	97976	97976	97976	97976	91561	83970
pseudo R2	0.423	0.423	0.425	0.382	0.427	0.426	0.425	0.426	0.419	0.373

Notes: The dependent variable in all estimations is the probability to have at least one patent application. Marginal effects reported. All regressions are estimated with the IV approach and include all the firm specific characteristics, year and three-digit industry dummies. Significance levels: ***1%, **5%, *10%. Standard errors clustered at the firm level.

Table 6: The effects of labor diversity on firm patents. Robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Occupation specific diversity		**Shannon entropy index**	**Richness**	**Edu and demo diversity as sd**	**2nd gen. Imm. as natives**	**Germanic group as natives**	**University graduates as natives**	**Copehagen is excluded**
	White collar	*Blue collar*							
index_fore_aggr	0.20655**	0.16282*	0.73902**	1.13326***	0.40523***	0.00248**		0.19666**	0.46958***
	(0.08572)	(0.08989)	(0.35446)	(0.16599)	(0.09620)	(0.00088)		(0.06006)	(0.08613)
index_edu_aggr	-3.42664	-4.13509	-7.57788	-2.83950	-4.06066	-2.56163		-2.60852	-1.65049
	(2.69407)	(2.54483)	(4.01640)	(1.75373)	(1.99070)	(2.33162)		(2.34655)	(1.29744)
Sd(years of education)									
index_demo_aggr	-0.38105	-2.57837	-0.85381	0.84853	3.98048	-0.74046		-1.02748	0.79522
	(3.15709)	(2.69090)	(1.08903)	(1.16158)	(2.50242)	(2.73825)		(2.73289)	(2.49926)
Sd(age)					-0.56736				
					(0.97602)				
Male					-0.07360				
					(0.34096)				
N	97976	97976	97976	97976	97976	97976		97976	91561
pseudo R2	47157.4	48067.4	49242.2	5603.9	4395.9	53204.8		52262.6	5331.3
index_fore_disaggr	0.33620**	0.11239	0.52259**	0.05739**	0.83791***	0.00452**	0.03327*	0.18542**	0.78419***
	(0.16758)	(0.21649)	(0.19826)	(0.01975)	(0.16621)	(0.00159)	(0.01419)	(0.06808)	(0.14807)
index_edu_disaggr	-0.27752	-0.59977	-0.59075	0.47504	0.75629	1.07781	0.93566	1.06254	0.74427
	(1.03574)	(0.87961)	(0.64881)	(0.82797)	(1.54937)	(0.89016)	(0.90347)	(0.88562)	(0.88114)
Sd(years of education)					0.69131				
					(2.18772)				
index_demo_disaggr	-2.37148	-1.72161	-0.13367	3.43535		0.74481	0.61254	0.28825	0.83860
	(2.21932)	(3.61201)	(1.16488)	(1.81449)		(3.12195)	(3.16525)	(3.11454)	(3.34457)
Sd(age)					-0.02135				
					(1.03504)				
Male					0.14771				
					(0.39855)				
N	97976	97976	97976	97976	97976	97976	97976	97976	91561
pseudo R2	51395.3	46137.9	56488.1	5171.8	4286.8	49861.5	50423.7	48431.8	5738.5

Notes: The dependent variable in all estimations is the probability to have at least one patent application. Elasticities reported. All regressions are estimated with the IV approach and include all the firm specific characteristics, year and three-digit industry dummies. Significance levels: ***1%, **5%, *10%. Standard errors clustered at the firm level.

Table 7: The effects of labor diversity on the probability of applying in different technological areas. Robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Occupation specific diversity		**Shannon entropy index**	**Richness**	**Edu and demo diversity as sd**	**2nd gen. Imm. as natives**	**Germanic group as natives**	**University graduates as natives**	**Copehagen is excluded**	**Mono-establishment firms**
	White collar	*Blue collar*								
index_fore_aggr	0.96282**	0.48811**	0.00062**	0.10157**	0.35329**	0.01073		0.04169	0.39599**	0.17908*
	(0.46780)	(0.19450)	(0.00027)	(0.04179)	(0.12549)	(0.01036)		(0.04468)	(0.15680)	(0.07151)
index_edu_aggr	-0.05940	0.05626	-0.00137	-0.15659**	3.27028	0.28235		0.27678	0.27616	0.64617
	(1.23363)	(0.73684)	(0.00122)	(0.07414)	(3.33870)	(0.76899)		(0.77167)	(0.73320)	(0.58210)
Sd(years of education)					-1.16065					
					(1.25988)					
index_demo_aggr	-0.52883	-0.72934	0.00027	-0.05698		-0.08767		-0.14616	-0.42290	0.02795
	(1.05541)	(0.76887)	(0.00020)	(0.06368)		(0.80169)		(0.79359)	(0.77422)	(0.66495)
Sd(age)					-0.01216					
					(0.06971)					
Male					-0.03074					
					(0.13437)					
N	1037	1037	1037	1037	1037	1037		1037	1037	753
pseudo R2	0.333	0.325	0.416	0.309	0.426	0.331		0.331	0.417	0.369
index_fore_disaggr	1.02540*	0.49244	0.00493**	0.06125*	0.44625**	0.00566	0.10345*	0.01578	0.46884**	0.09548*
	(0.68433)	(0.32132)	(0.00142)	(0.03766)	(0.20623)	(0.02076)	(0.06253)	(0.04739)	(0.22131)	(0.04184)
index_edu_disaggr	-0.50067	-0.25844	0.00007	-0.06176	0.15329	0.10087	0.03323	0.10546	0.01493	0.10738
	(0.46485)	(0.39456)	(0.00173)	(0.04004)	(2.91797)	(0.37101)	(0.36890)	(0.37112)	(0.36296)	(0.38236)
Sd(years of education)					0.00290					
					(1.28083)					
index_demo_disaggr	0.35143	0.39576	0.00317	0.05710		0.75549	0.76212	0.72946	0.41049	0.25341
	(0.77419)	(0.85835)	(0.00169)	(0.06366)		(0.79400)	(0.79058)	(0.79423)	(0.79274)	(0.73213)
Sd(age)					0.04304					
					(0.09353)					
Male					0.09675					
					(0.11457)					
N	1037	1037	1037	1037	1037	1037	1037	1037	1037	753
pseudo R2	0.326	0.320	0.418	0.313	0.427	0.323	0.327	0.323	0.372	0.288

Notes: The dependent variable in all estimations is the probability to have at least one patent application. Marginal effects reported. All regressions are estimated with the IV approach and include all the firm specific characteristics, year and three-digit industry dummies. Significance levels: ***1%, **5%, *10%. Standard errors clustered at the firm level.

Figure 1: Commuting areas,1995, Denmark.