



Sorting out the impact of cultural
diversity on innovative firms.
An empirical analysis of
Dutch micro-data

Ceren Ozgen and Thomas de Graaff

Sorting out the impact of cultural diversity on innovative firms. An empirical analysis of Dutch micro-data[☆]

Ceren Ozgen^{a,b,1}, Thomas de Graaff^a

^a*Department of Spatial Economics, VU University Amsterdam, The Netherlands*

^b*Tinbergen Institute, Amsterdam, the Netherlands*

Abstract

An increasing amount of research in the migration literature shows a positive association between migrant diversity and firm productivity. However, the potential bias due to unobserved heterogeneity remains a challenge. In this paper we analyse the impact of cultural diversity on firm innovativeness, while using finite mixture modeling to control for observed and unobserved heterogeneity. Recent availability of microdata has enabled us to construct a linked employee-employer dataset through merging datasets on both workers and firms. We explore the possible ways of firm-level knowledge exchange among the employees with different cultural backgrounds and its impact on firms' product and process innovations. We find that workforce diversity is beneficial for innovativeness in capital-intensive sectors. It also positively impacts large firms that operate in high-level services, manufacturing, mining and R&D sectors, that are predominantly located in the non-urban areas in the Netherlands. In labour and land intensive sectors, the impact of cultural diversity on innovativeness is inconclusive.

Keywords: Cultural diversity, innovativeness, (un-)observed heterogeneity, finite mixture modeling, migration

JEL-classification: J15, J21, O31

[☆]This research is part of the Migrant Diversity and Regional Disparity in Europe (MIDI-REDIE) project, funded by the NORFACE research programme Migration in Europe—Social, Economic, Cultural and Policy Dynamics. We thank Raymond Florax, Maureen Lankhuizen, Peter Nijkamp, Jacques Poot, participants at the 2012 'Eureka' Seminar at VU University Amsterdam and at the 2012 'Workforce Heterogeneity and its Impact' Conference at IAB Nürnberg for helpful comments and suggestions.

Email addresses: c.ozgen@vu.nl (Ceren Ozgen), t.de.graaff@vu.nl (Thomas de Graaff)

¹Corresponding author. Address: Department of Spatial Economics, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands

Preprint submitted to Elsevier

March 28, 2013

1. Introduction

“The best way to have a good idea is to have a lot of ideas.”

The quote above is from Linus Pauling, a Nobel Laureate twice, and emphasizes the importance of having various ideas in creative success (as cited in [Uzzi and Dunlap, 2005](#), p.2). The economics of diversity literature stresses that it is not so much ideas that are common, but instead ideas that are different which are those most likely to make the change and spur knowledge accumulation. Ideas in the public domain help us to communicate, whereas exclusive ideas are important to bring in originality in collaborative work ([Berliant and Fujita, 2012](#)).² In this age, where non-rivalrous knowledge ([Romer, 1993](#)) is widespread and accessible more than ever, it is a challenge for firms to transform potentially new knowledge in productivity growth.

To explore the impact of these pervasive phenomena, we look specifically at within firm diversity. Although much information has become widely available, it is people and the interaction amongst them that makes the difference in the way circulated information and ideas are understood, interpreted and used. Country specific attributes increase heterogeneity amongst individuals, even though they might have similar educational backgrounds ([Mattoo et al., 2012](#)). Moreover, due to the dynamic nature of knowledge accumulation, individuals' location and migration over time and space may significantly impact their knowledge endowments. In an economic setting where labour is mobile, knowledge transfers do not necessarily need to be restricted by public mechanisms (i.e., schooling), but may be spurred as well by the influx of migrants and their ideas. Therefore, we focus on firms and their foreign employees from different countries of origin. By doing so, we assume that access to a localized diverse array of knowledge sets creates a firm specific advantage in knowledge production.

A number of recent studies, mostly from countries such as the USA, New Zealand and Germany, estimated the impact of diversity on innovation.³ The focus has been predominantly on the area level, where the cultural diversity within the region is used as a innovation and productivity enhancing indicator. Far fewer research projects explored

²[Lazear \(1999\)](#) made a similar argument in the context of the trade-off between culture and language.

³See, amongst others, [Ottaviano and Peri \(2006, 2005\)](#); [Maré et al. \(2009\)](#); [Kerr \(2009\)](#); [Niebuhr \(2010\)](#); [Hunt and Gauthier-loiselle \(2010\)](#); [Brunow and Blien \(2011\)](#); [Ozgen et al. \(2011\)](#).

the impact of migrant diversity on firms' outcomes. Here, there is some evidence that firms are able to transcend their skill limitations through employing relevant people whose experiences help to close the skill-gap firms face, and who bring in unique knowledge, that assists firms to be innovative.⁴

In this literature, observed and unobserved heterogeneity of firms remains a methodological challenge for measuring the impact of workforce diversity at the firms. Using various sub-samples of firms—for example, selections based on sector specialization, import-export behavior, and location choices—are used as a remedy for heterogeneous firms so far (see [Trax et al., 2012](#), for a recent example). A disadvantage of this approach is that a-priori it seems hard to identify the dimensions that define the firms' heterogeneity. Moreover, data are usually not available for some of the more crucial dimensions (e.g., managerial quality) leading to unobserved heterogeneity.

In this paper we propose an alternative approach in which we can endogenously segment firms in sub-samples and can simultaneously deal with unobserved heterogeneity that correlates between firms' innovativeness and diversity. In particular, we are interested in whether firms' innovativeness benefits from the cultural diversity of the firm's workforce. And if so, which types of firms actually benefit the most from cultural diversity?

To understand these knowledge spillovers, we focus on firms as the smallest micro units of production where new ideas can be transferred or created amongst employees. We exploit a firm level linked employee-employer micro-dataset obtained from Statistics Netherlands. We observe each firm every 2 years for 3 consecutive periods over 6 years: from 2000 to 2006. The data provides extensive information about the characteristics of firms, employees, and the location of both. To explore the underlying clusters of firms that exclusively profit from the diversity of the foreigners, we employ a multivariate finite mixture model.

The main advantage of finite mixture modeling within a regression context is that such models can endogeneously assign observations to groups or segments, where each segment is estimated separately ([McLachlan and Peel, 2000](#)). This is in contrast with traditional regression methods, which reflect the aggregate estimates for the whole sample,

⁴[Navon \(2009\)](#); [Nathan and Lee \(2012\)](#); [Ozgen et al. \(2011\)](#) and [Parrotta et al. \(2011\)](#) (the studies were conducted for Israel, the UK, the Netherlands and Denmark, respectively) are a few examples that studied within firm diversity of the workforce and its effect on firm-level innovations.

despite the fact that underlying data may be substantially heterogeneous. Therefore, the estimated parameters represent the predictions when the parameters are the same for all of the observations. Our results show that those approaches might be misleading. Indeed, with respect to the various employee, firm and regional characteristics, we find substantial heterogeneity within our dataset.

Moreover, this model addresses potential unobserved factors that may simultaneously influence the diversity of the firm's labour force and its productivity. We achieve this by adopting a simultaneous estimation procedure for two models, one of which explains firms' innovativeness and the other one which explains diversity. If unobserved heterogeneity affects both innovativeness and cultural diversity, then segmenting our heterogeneous multivariate sample into more homogeneous multivariate subsamples mitigates this problem.

Our main finding is that cultural diversity of the foreign workforce significantly increases the probability to innovate at the firm level. However, once observed and unobserved heterogeneity is accounted for, the effect is unequal across firms. It is predominantly capital-intensive firms with a focus on knowledge and high-tech production (e.g., the sectors research and development, computer and related and manufacturing) that actually benefit from the diversity of foreign labour. These firms are usually large, and mostly innovate in more than one type of innovation. Thus, this particular cluster of firms is able to utilize the diversity of their foreign employees. Because these firms are more capital than labour intensive, they are usually closely located to urban areas with their dense distribution of jobs and population, but not necessarily within these areas. Firm-level innovativeness benefitting from cultural diversity does not seem to be a particularly urban phenomenon.

The remainder of this paper reads as follows. The next section provides a concise theoretical exposition of the ways in which immigrants and cultural diversity among foreign born or second generation ethnic workers may influence innovation. Section 3 outlines the empirical model and the measurement of cultural diversity. Subsequently, Section 4 describes how the data were constructed, while Section 5 discusses the results of our analysis. Section 6 concludes.

2. Theoretical Background

Knowledge flows across any spatial scale have never been as exhaustive as nowadays. People all over the world can access various types of information and insights one could not have imagined just half a century ago. However, due to this immense inflow of codified information (Audretsch and Feldman, 2004), the uniqueness of the knowledge people possess is subject to erosion, which leads to a process of homogenization of knowledge. It is not only the information flow through formal channels (i.e., via official documents or the internet), but also through the mobility of the people transmitting tacit and unique knowledge, that knowledge gets spread. Theory indicates that firms indeed benefit from unique knowledge that people possess (Prat, 2002), but to what extent it is crucial for competitive firms to benefit from the flow of knowledge and diversity that people bring in to remain innovative, is still an empirical question.

Until very recently, neither the innovation nor the migration literature concentrated on the potential role of workforce diversity at the firm level. Moreover, human capital and economic growth theories also did not make a distinction between various particularities of skilled labour. Clearly, the shortage of attention partly stemmed from the unavailability of micro-level data. Another reason, however, might be the absence of recognition of the heterogeneity of employees. In reality, two people with identical educational backgrounds and experience may show different productivity levels.⁵ Part of this difference can be attributed to personal abilities and motivations, while the cultural background of people can also be an informative, albeit imperfect, proxy, for intrinsic productivity influenced by cultural background and working habits.

Most studies within the economics of diversity literature assume that knowledge production is a function of the public knowledge of employees as well as their private (and usually distinct) knowledge. Firms acquire this (distinct) knowledge by inter alia employing foreign workers in addition to cooperating internationally or participating in relevant networks. Diversity can also be enhanced by employing more women, older or younger workers, or disabled workers. Workers who are dissimilar relative to other workers are more likely to have distinct knowledge from others. This allows the workforce

⁵Since, as common practice in econometric specifications, all employees with similar education/skill levels have been included into similar categories.

to learn from each other, thus leading to increased productivity within the firm, despite some linguistic and cultural barriers (Lazear, 1999; Florax et al., 2005).⁶ In our case, the dissimilarity condition is satisfied by having employees from different birthplaces. Moreover, the employees are assumed to be able to communicate as required; a condition which we assume satisfied because workers are recruited for specific tasks. This implies that homogeneity of the workers in terms of their cultural background may slow down the innovation process. This is reflected in Agrawal et al. (2008), who find a negative interaction or substitution effect between co-location and co-ethnicity on the probability of a knowledge flow between inventors.

3. Methodology

Directly assessing the impact of cultural diversity on the innovativeness of firms is likely to be flawed because of possible unobserved factors, such as managerial quality, openness of the firm and position of the firm within its trade network. These factors might affect both cultural diversity and firms' innovativeness and thereby creating a possible source of bias. To address this, we adopt a multivariate finite mixture approach for two reasons: to address (*i*) the underlying heterogeneity of the impact of within-firm workforce diversity on the innovativeness of the firms in the Netherlands, and to control for (*ii*) unobserved heterogeneity.⁷

Our modelling approach has two important features. First, we allow for our sample to be a mixture of multiple subsamples (with varying characteristics), yet we do not know the proportions of these mixed distributions a-priori. We aim to segment the sample of firms through identifying more or less homogeneous segments (groups). We do so while for each segment predicting the parameters of the density function underlying the observed data. Such a finite mixture model addresses the underlying heterogeneity endogenously and allow us to cluster the firms according to the possible benefits they reap from the

⁶This line of reasoning is consistent with recent field experiments that show that there indeed is evidence of a (small) positive impact of cultural diversity on team performance (see Mitchell and Nicholas (2006) and Hoogendoorn and van Praag (2012) for empirical results).

⁷Finite mixture models are not widely used in economics, except in marketing (see, e.g., Jedidi et al., 1997; Wedel and Kamakura, 2000) and labour economics (e.g., Lancaster, 1990). Recently, probably due to the increased computing power and the availability of new software, finite mixture approaches have received more attention (see, e.g., Baum-Snow and Pavan, 2012; Lankhuizen et al., 2012).

diversity within the workforce and the other control variables. Segmenting heterogeneous firms into a smaller number of homogeneous subgroups thus enables us to identify the various firms' responses in terms of innovativeness to labour diversity.

Second, by incorporating a simultaneous estimation procedure, we are able to address potentially unobserved heterogeneity and possible sorting of a diverse workforce to innovative firms.⁸ Obviously, there can be numerous latent factors that simultaneously influence the sorting of a diverse workforce to firms, and increase, at the same time, the productivity of firms in terms of their innovativeness. To tackle this problem, we simultaneously model innovation and diversity, where the latter is used as well as a control variable for the former. The possible correlation between the models, denoting unobserved heterogeneity, is then identified through the probability mass each segment has for the parameters of the two models combined.

The main assumption of our approach so far is that cultural diversity within a firm has an impact on firms' innovativeness and *not* vice versa. There are, however, two possible causes for reverse causality. First, (skilled) foreign employees could be attracted to innovative firms. However, our assumption can be argued to be, at least partly, justified because we measure a composition effect and not a volume effect of foreigners: namely, when firms recruit a relatively large number of foreigners, prospective foreign employees will be attracted to the presence of colleagues of their own cultural background. Second, and more problematic, it could be the case that the innovative firm itself actively seeks appropriate talent from all over the world. Indeed, according to a recent Forbes report ([Forbes, 2011](#)), firms recognise the importance of diversity as a key driver of innovation—although improving gender diversity is generally seen as the most important issue. If firms actively promote cultural diversity to foster innovation, then reverse causality causes our estimate to be biased upwards. However, we expect this mechanism to apply only to large and international firms. Most firms recruit locally, or regionally, and take the ethnic composition of the labour force as given.

⁸The usual approach to account for unobserved heterogeneity within a panel is to use a fixed effects approach. In this paper, we choose not to because of two main reasons. First, inter temporal variation in our dataset is low, especially concerning the innovations and cultural diversity variables, which renders fixed effects estimation cumbersome. Moreover, a fixed effects approach does not necessarily remove all unobserved heterogeneity. Second, we are not only interested in the level effects (the various constants), but also in the variation of the slope parameters (or the various impacts cultural diversity has on innovativeness).

The analysis proceeds in a two-step procedure. First, we look at the impact of a diverse workforce, on varying innovation types, by using a simultaneous finite mixture procedure where the firms are endogenously segmented. To set this up as broadly as possible, we allow all our covariates, including the diversity measure, to vary.

In the second stage, we analyze each segment looking at the firms' characteristics in each segment. Thus, we look at each firm's probability to 'belong' to more or less homogeneous segments and use the firms' characteristics to describe a posteriori these estimated segments. This enables us to discern between groups of firms which are similarly affected by diversity.

We describe our econometric model and diversity measure in subsections 3.1 and 3.2, respectively.

3.1. Econometric Model

Our aim is to model the impact of diversity on innovations while controlling for unobserved heterogeneity. To do so, we start with the following conceptual model:

$$f_P(P_{it}) = X_{it}\beta_{\delta_i} + f(D_{it})\gamma_{\delta_i} + \epsilon_{it} \quad (1a)$$

$$f_D(D_{it}) = Y_{it}\alpha_{\nu_i} + \mu_{it}, \quad (1b)$$

Where $f_P(P_{it})$ is a measure of the innovativeness of firm i at time t , X_{it} and Y_{it} denote firm specific characteristics at time t , $f_D(D_{it})$ is a measure of firm specific cultural diversity at time t —, i.e., our variable of interest—, β_{δ_i} , α_{ν_i} and γ_{δ_i} are firm specific parameters to be estimated, and ϵ_i and μ_i denote *i.i.d.* error terms.

As argued above, it is very likely that the relationship between firm innovation and cultural diversity is influenced by unobserved variables, which thus leads to unobserved heterogeneity bias. Here, unobserved heterogeneity is denoted by δ_i and ν_i and is assumed to vary over firms. Moreover, we assume that unobserved heterogeneity has an impact on all parameters (β_{δ_i} , γ_{δ_i} , and α_{ν_i}) and, finally, that there are multiple homogeneous subsamples with similar parameters.⁹ When δ_i and ν_i are uncorrelated, unobserved

⁹Note that this is a step further than the mass-point approach as advocated by, e.g., Heckman and Singer (1984) and Abbring and van den Berg (2003) where only the intercepts are allowed to vary and correlate between simultaneous models. Here, the entire parameter vectors is allowed to vary.

heterogeneity does not pose a problem for a direct estimation of eqn. (1a). However, when they are correlated, estimating eqn. (1a) without taking into account eqn. (1b) causes both β and γ parameters to be biased.

An example is the case of managerial quality. Suppose that there are two types of firms within the sample, those with a high managerial quality and those with a low managerial quality and that we do not have any information which firm belongs to which group. Moreover, firms with high managerial quality attract a culturally diverse workforce and are very likely to innovate. On the other hand, firms with low managerial quality do not have a culturally diverse workforce and are less likely to innovate. Thus, the *observed* high impact of cultural diversity on innovation should in fact be attributed to unobserved managerial quality.

The empirical strategy we employ to control for unobserved heterogeneity is to segment our sample into homogeneous groups of firms. In the above example this means that we would like to segment our sample into two groups of firms; those with high and those with low managerial quality. In reality, the dimensions along which segmentation takes place are of course unknown.

We proceed by integrating δ_i and ν_i out as follows.

$$f(P_{it}, D_{it}) = \int_{\delta} \int_{\nu} f_P(P_{it}) f_D(D_{it}) dG(\delta_i, \nu_i). \quad (2)$$

Finally, we use the assumption that δ_i and ν_i are homogeneous within (a finite number of) S subsets, and then eqn. (2) boils down to the following finite mixture specification:

$$f(P_{it}, D_{it} | X_{it}, Y_{it}, \alpha, \beta, \gamma) = \sum_{s=1}^S \pi_{is} [f_P(P_{it} | X_{it}, D_{it}, \beta_s, \gamma_s) \times f_D(D_{it} | Y_{it}, \alpha_s)], \quad (3)$$

where $\pi_{is} \geq 0$, $\sum_{s=1}^S \pi_{is} = 1$. For more details concerning estimation we refer to [Appendix A](#).

3.2. Measuring the Diversity of the Workforce

The literature offers various kinds of diversity measures (see, e.g., [Alesina and La Ferrara, 2005](#)). The measure used in our study incorporates two aspects of a non-homogeneous population. Firstly, we account for the relative share of each unique group in total population. Secondly, total variation in terms of richness of the population is considered (For a more detailed discussion about the index see [Ozgen et al. 2011](#)). The index is calculated as follows:

$$D_{it} = 1 - \sum_{j=1}^6 s_{jit}^2, \quad (4)$$

where s_{jit} is the share of the group j ($j=1, \dots, 6$) employed in firm i at time t . The index can get values between 0 and $5/6$ (in our case).¹⁰ A value of 0 refers to complete homogeneity and the index turns to its maximum value when no migrant in the same firm share a common birth place. Natives as a group are excluded, since it is our aim to measure the diversity among foreign employees. Once natives are included the diversity index becomes 96 percent correlated with share of foreigners, which is a crude measure of a firm's overall foreignness.

In addition to the diversity measure, other controls are included in the estimation. It is possible to group the covariates into three categories. These are variables on employee characteristics, firm-level controls and agglomeration variables to account for the regional features that may influence firms' location decisions as well as knowledge inflows from their external environment. All the estimations include the natural logarithm of firm size (employment) to take the firms' evolution into account. Moreover, the financial and personnel related obstacles that the firms encountered during the innovation process are controlled for. We expect that the firms facing such difficulties should be more likely to innovate since these are indications to grow and innovate further (for a discussion, see [Ozgen et al., 2011](#)). The last firm-level variable is openness of firms to change, and it appears to be a robust indicator as one of the soft factors that boost innovativeness in the

¹⁰ The squares of relative shares are summed over six large regions in the world, namely the EU15, other European countries, North America, South-east Asia, East Asia and Oceania, and Rest of the World. While creating these groups we tried to maximize the differences between groups and the similarities within groups. Thus, the grouping is broadly based on the cultural distances between countries. The Rest of the World category comprises various countries from which there is only a small number of immigrants in the Netherlands.

recent innovation literature ([Jensen et al., 2007](#)). Openness to change refers to whether the firms went through organizational changes with respect to third parties in the last 3 years.

The second set of variables relate to the employee characteristics. A firm's employee profile may make a distinctive contribution to its success to innovate. In addition, based on the needs of various sectors, the firms certainly differ on their search of employee qualities. To address potential heterogeneity, we consider the youthfulness, skills or qualities and the diversity of the firm's workforce. Since we specifically devote attention to tacit knowledge, ideas, and abilities that the foreign employees may possibly bring in, we construct our employee measures only with non-natives. Therefore, the youthfulness of the employees refer to relative share of foreign employees aged 25–45 to the total foreigners in a firm; and similarly, the high-skill intensity of foreign employees is the relative share of highly skilled foreigners among total foreigners in a firm. Finally, we include some regional controls to assess how firms' locational choices contribute to their innovation behaviour. Locating in dense urban economic agglomerations and the role of local competition appear as important indicators of innovations in the literature ([Carlino et al., 2007](#)).

Cultural diversity within the firm is modeled by 'exogenous' determinants, namely the foreign population per municipality and the number of second generation immigrants with both parents born abroad per firm. For each wave of the innovation surveys, the former is lagged for 6 years after a firm submitted the questionnaire back in the respective period. The migration literature shows that immigrants persistently follow each other over time. A municipality that was the destination of earlier generations of immigrants is likely to attract more newcomers for various purposes. Thus, firms that select their employees from a labor pool at their vicinity, are likely to encounter more foreigners. On the other hand, second generation foreigners have more information about the market conditions and firms' employment strategies. Therefore, they are more likely to self-select to firms renowned for employing foreign workers. However, since both of the parents of these second generation immigrants were born abroad, they should still carry the characteristics of their native cultures.

4. Data

We constructed a linked employee-employer dataset by using 3 different high-quality firm and individual level micro-data sources. Since the datasets used are at the micro-level, they were obtained under a confidentiality agreement with Statistics Netherlands. The study period extends over 6 years from 2000–2006. As a result of a 3 steps merging procedure, 888 firms can be observed every two years, over the 6 years. Hence, in total, we have a balanced panel of firms with 2,664 observations.

Altogether four different data sources are exploited:

- Community Innovation Survey (CIS). This dataset provides information on the innovation outputs, the innovation inputs and the obstacles to innovations at the firm level. Each period contains about 11,000 observations of firms. There is a 2-years timelag between when the questionnaires are sent to the firms, and when the responses are collected.
- Municipal registrations. This dataset provides information on the cultural background and the demographics of the employees. It registers all people living in the Netherlands and contains about 16 million observations.
- Tax registrations. This dataset provides information on all the employed tax payers and their work and firm characteristics. It registers about 10 million employees where an employee can be observed repeatedly during the year.
- National statistics. These are various statistics on agglomeration variables and are obtained at the municipal level.

The firms in our dataset are sampled from the Community Innovation Surveys, CIS 3.5 (2000–2002), CIS 4 (2002–2004), and CIS 4.5 (2004–2006) which provides the anchor of our empirical strategy. The CIS surveys defines two major types of innovations, namely product and process. The exact definitions of these innovation types are as follows: a product innovation is the market introduction of a new good or service or a significantly improved good or service with respect to its capabilities, such as improved software, user friendliness, components or sub-systems; a process innovation is the implementation of a new or significantly improved production process, distribution method, or support activity

for a firm's goods or services.¹¹ All the innovation variables are in binary form, thus if a firm reported an innovation at a particular period, the value of the variable is '1', and '0' otherwise.¹² The number of firms that answered all CIS questionnaires over the years decreases substantially. We, however, cannot track down why firms exited the sample.

In this study, a foreigner is defined as an employee who was not born in the Netherlands. The only time invariant identifier of an employee's background that is available to us is birthplace. Consequently, we proxy the employees' cultural background through their birthplaces. Acknowledging that culture is a multi-dimensional concept, and is shaped by many things in addition to the birthplace's unique characteristics and customs, the data do not allow us to control for further dimensions. Therefore, this study uses diversity of foreigners by birthplaces as an indication for cultural diversity within the firms. Unfortunately, our data do not include the entry time of the foreigners in the host country.

Table 1 gives the summary statistics of our data.

[INSERT TABLE 1 ABOUT HERE]

The firms that appear in our balanced panel are quite large; 85 percent of them have 100 or more employees. Therefore, the interpretation of the results should address large firms, and not necessarily small or medium-sized firms. More than 50 percent of the firms are innovative and many firms reported innovations in multiple categories, namely process or product innovations.

We have more than 1 million employees (1,127,210); 11 percent of which are born outside of the Netherlands. On average, there are 51 foreigners and 15 distinct birth places present in each firm. Foreigners are relatively low-skilled, whereas they are significantly younger than the native employees. However, we focus on the quality of the foreign stock among the foreign employees rather than their relative skill level to natives. In this respect, we measure the high-skill intensity of foreigners relative to the total foreign

¹¹Moreover, if a firm reported innovation in one of these categories and/or aborted an ongoing innovation effort for various reasons, the firm is coded as innovative in the CIS database. We do not use innovations as a whole as a dependent variable due to its vague definition. Using product or process innovation as dependent variables reveals qualitatively similar results to our main findings, and the use of overall innovativeness produces only marginally significant results for the impact of diversity (at the 7% level).

¹²Note that these variables are self-reported by the firm and not validated by Statistics Netherlands. Unfortunately, we cannot investigate whether this leads to a selection bias.

employees in each firm. Therefore, on average 13 percent of the foreigners are highly skilled, and 60 percent of them are aged between 25–45 year.

5. Findings

Our main interest is whether within firm diversity increases the probability to innovate, and if so, which firms actually benefit from it. We thus scrutinize whether it is possible to generalize the possible contribution of a diverse workforce for heterogenous firms. Subsection 5.1 first presents the results of the estimations whereas subsection 5.2 describes the firms’ composition over the various segments.

5.1. Estimation Results

Our econometric specification includes three sets of variables; variables related to firm characteristics, to employee characteristics, and finally control variables for economic agglomeration at different regional classifications. The diversity of the workforce is the variable of interest, and we test its impact on product and process innovations. In our data, innovations appear as a binary variable, and are therefore modeled with a logit model.¹³ The model that explains diversity is measured with a linear regression model.

Given the total number of observations and the computational limitations (specifically, due to the binary dependent variable for innovations), three segments performed as the best choice, and seem to sufficiently address the underlying heterogeneity of firms, once model (3) is estimated.¹⁴

[INSERT TABLE 2 ABOUT HERE]

Table 2 displays the ‘performance’ of our model, where the column ‘Size’ gives the number of firms actually assigned to each segment; where the assignment is based on the highest probability of a firm to belong to a particular segment. The column ‘Posterior’ reports the number of firms in each segment with a posterior probability exceeding 0.0001. The

¹³Our findings are robust when a linear probability model is applied instead of a binary logit model.

¹⁴To estimate equation (3) we use the ‘flexmix’ package in the free software environment ‘R’ (Leisch, 2004). The Akaike and Bayesian Information Criterion (AIC and BIC) indicate that three segments perform better than a smaller number of segments. As already noted, using a larger number of segments faces computational difficulties.

number of firms in column (2) exceeds the number of firms in (1), which indicates that some firms have membership probabilities exceeding the threshold for more than one segment. The values for the ratio in the third column indicate that there is some overlap for our data. Segment 2, and to a lesser extent segment 1, forms a group of observations that has distinct characteristics from the other observations. Segment 3 seems to be less distinct (and includes fewer number of observations). Overall, the results give additional evidence that using three segments adequately tackles the heterogeneity within this dataset.

[INSERT FIGURE 1 ABOUT HERE]

The same information is conveyed by the rootograms in Figure 1, which gives the posterior probabilities of membership for each segment.¹⁵ These rootograms indicate that, when many observations are close to 0 and 1, firms are well segmented. When there is a high probability mass in the middle, firms can belong to multiple segments with significant probabilities.

[INSERT TABLE 3 ABOUT HERE]

The overall results (displayed in Table 3) are in line with our theoretical expectations, the novelty of incorporating a finite mixture approach proves to be critical though. We observe that the theoretical expectations and the empirical findings of the economics of diversity literature so far appear to be valid only for a certain group of firms (which we label in our case segment 2) once unobserved heterogeneity is taken into account.

Column (1) in Table 3 gives the estimates for the whole sample for both the probability of product innovation (P_{it}) and the diversity (D_{it}). Clearly, cultural diversity has a large impact on product innovations, which is in line with, e.g., Niebuhr (2010) and Nathan and Lee (2012). If the diversity index increases with one standard deviation (which equals to 0.187), then the probability for the average firm to innovate increases with 24%. However, segmentation (columns 2–4) shows that this only accounts for a specific *type* of firms—the

¹⁵Firms with posterior probabilities smaller than 0.0001 are omitted. Usually many firms in each segment have posterior probabilities close to zero. To avoid having the high count in the corresponding bar obscure the information in other bars of the rootogram, these product groups are omitted (see, e.g., Leisch, 2004). Moreover, note that for comparison reasons the vertical axis corresponds to the square of the number of firms in each bar.

ones present in segment 2 (displayed in column 3). The probability for an average firm in segment 2 to innovate increases by 42% when the diversity index increases by one standard deviation (which equals to 0.085) of the diversity of the firms in segment 2. For the firms in segments 1 (column 2) and 3 (column 4) the diversity index has however no significant effect.

Table 3 shows that the innovativeness of the firms in segment 3 is driven by different inputs than those of the two other segments. Although segment 1 and segment 2 show similarities in terms of the drivers of innovation, the diversity index of foreign workers materializes to be beneficial for firms exclusively in the second segment. The diversity index is not significant for the two other segments. Consequently, not only the size of the predicted parameters are different, but also their statistical significance alter per segment.

The innovation literature is fairly convincing on the importance of firm size as a robust indicator of innovation. Our results, however, exhibit that the natural logarithm of the firm size is positive, yet not significant for all three segments (which may be due to the segmentation being driven partly by firmsize). For firms in segment 1 and 2, lack of personnel at most levels and costs up to a medium level are positive and significant inputs of innovativeness, with respect to those firms which did not report any obstacle. Firms that restructured their internal organization in relation to the third parties are significantly more likely to innovate at least at the 5 percent level in all segments. Therefore, the econometric specification where we accounted for firm's capacity, obstacles faced during innovations and institutional resources for its receptiveness from the outside resources play significant roles in determining its probability to innovate. However, the impacts are not even across the different subsamples; indeed, different set of firms perform differently with the similar inputs.

The high-skill intensity and the youthfulness of the foreign employment significantly increase the probability to innovate for firms in segment 2, whereas none of these inputs prove to be important for the firms in the third segment at any conventional significance levels. On the other hand, firms in segment 1 benefit from the high-skill intensity of the foreign workers at the five percent level. Eventually, this also shows that inherent differences exist between the sectors. For instance, the capital and physical inputs they utilize for innovation or the type of employees they benefit the most from vary considerably.

The last set of controls included in the specifications are local competition and density of the economic activity at the municipalities. In the Netherlands, functional regions¹⁶ are defined as meaningful economic regions based on daily commuting distances, therefore to measure the impact of local competition we used the number of firms per job at the NUTS 3 level. For the latter, following the literature on urban agglomeration economies, a density measure of number of firms per km² per municipality is employed to account for the role of input sharing, matching and external knowledge spillovers on innovations [Carlino et al. \(2007\)](#).

Diversity within firms itself is more difficult to explain. The size of the allochthonous population seems to have a small but only marginal negative impact while the number of foreigners with both parents born abroad seems to have a varying impact over the segments. For the first segment it is positive while for the third segment it is negative.

Because a direct interpretation of the differences between the three segments is not immediately obvious from the regression estimates, the next subsection deals with a more indepth description of each segment.

5.2. Description of the Segments

The FMM estimation procedure reveals three rather distinct clusters of firms in the Netherlands based on the impact of employee diversity on innovations. These clusters show significant variation in terms of firm, employee, and location characteristics. To highlight some of these distinct features, we proceed to describe the clusters more indepth.

The locations quotients which are calculated to show the clustering of the top 5 sectors in each segment are displayed in [Table 4](#).¹⁷ The location quotient is informative on the concentration of firms in particular sectors (regions) in a segment relative to their concentration in the overall distribution of economic sectors (in a region).

[INSERT TABLE 4 ABOUT HERE]

[INSERT TABLE 5 ABOUT HERE]

¹⁶At the national jargon these 40 so-called Corop regions correspond to the NUTS 3 division of Eurostat, for international comparison.

¹⁷The location quotient is calculated as follows: $\frac{F_{sr}/F_s}{F_r/F_n}$, where F_{sr} denotes the number of firms in segment s and region r , F_s the number of firms within segment s , F_r the number of firms within region r and F_n the total number of firms within our sample.

Tables 4–5 show that the composition of sectors in each segment is very different from each other. Segment 1 can be characterised by mostly labor-intensive industries and services sectors, such as low-skilled business services and the financial intermediation sector. The firms in this segment are quite large with a mean value of 362 employees, and the composition of the labour force is fairly diverse. The share of foreign employees who are 25–45 years old among total foreigners is about 63 percent, hence hosts predominantly young people. The second segment is a set of firms active in capital-intensive goods and knowledge production sectors, whilst having a considerably high share of highly-skilled labour. The sectors clustered within this segment are mostly manufacturing, R&D, mining & quarrying and computer & related.¹⁸ This is the segment that on average has the largest firm size of 550 employees, and has a very diverse workforce (diversity index equals to 0.687), although the relative share of foreigners is similar to those in other segments. The firms in this segment show by far the most innovations in all innovation categories, and it is the only segment that is positively impacted by the within-firm diversity of the employees. The third and last segment can be associated with land-intensive sectors such as agriculture & forestry, transport & communication and electricity, gas & water. In this segment, we observe a mixture of firms that are relatively smaller in mean firm size (compared to other segments), which equals to 156 employees, have a fairly homogenous workforce (diversity index only equals to 0.214) with on average 23 foreign employees in each firm. These sectors mostly need large land-plots to operate and produce. Amongst the other three, the firms in this segment appear to be the least innovative.

[INSERT FIGURE 2 ABOUT HERE]

Figure 2 shows the employment population ratio (2a) and the location quotients (2b–2d) of the regional distribution of firms over NUTS 3 regions in the Netherlands. There is clearly a spatial difference in the regional distribution of the three segments. The labor intensive-sectors that dominate segment 1 are positively correlated with regions with relative large amounts of both population and jobs, such as the Northern part of the Randstad area (the most urbanised Dutch regions), which includes the capital city

¹⁸The number one ranked sector in segment 2 is actually ‘other services’, however the number of firms in this category is only 3, so we therefore decided to leave this sector out of Table 4.

Amsterdam. The correlation between the location quotient and with total population, total employment and the employment-population ratio is 0.16 in all cases. In contrast, the capital intensive-sectors that dominate segment 2 have a smaller correlation with both population and employment. These areas can be found in the ‘intermediate zone’ (as labeled by [van Oort, 2004](#)), non-urban regions close to large urban centers. Its correlation with total population, total employment and the employment-population ratio is 0.09, 0.04, and -0.02 . Finally, the more land intensive-sectors of segment 3 are negatively correlated with both population and employment, with a correlation of -0.35 and -0.33 , respectively, and a correlation of -0.18 with the employment-population ratio. These regions can be predominantly found in the periphery of the Netherlands.

These findings are in line with those of [van Oort \(2004\)](#), who finds that, although it is definitely spatially clustered, innovation is not necessarily an urban phenomenon—at least for the Dutch case.

As can be seen from both [Tables 4–5](#) and [Figure 2](#), once the sample is decomposed into three segments, based on common innovative behavior and input characteristics of the firms, a clear segmentation of firms become visible. The first segment can be characterized by urban firms that are labour intensive. The second sector are more capital intensive firms with a more non-urban character (although close to urban centers). It is these firms that derive positive benefits from diversity. The third and last segment can be less clearly defined. However, it seems to be dominated by more land intensive firms.

6. Conclusion

This paper analyzed the impact of cultural diversity on firm innovativeness, while using finite mixture modeling to control for observed and unobserved heterogeneity. In terms of observed heterogeneity, we find that only a specific set of firms actually benefits from cultural diversity: namely, large firms that operate in high-level services, manufacturing, mining and R&D sectors, which are predominantly found in the non-urban areas in the Netherlands. The effect itself for this subset is large. For a one standard deviation increase in our diversity index, the probability to innovate increases with 42%.

This result is in line with previous empirical research, but adds that there is only a positive effect of cultural diversity for the type of firms mentioned above. Firms that

benefit from cultural diversity (in terms of their innovativeness) are large (probably mature) firms in capital and R&D intensive sectors and are not necessarily located within urban areas. Indeed, it seems that those firms are more likely to be located in less dense areas, though perhaps still reasonably close to urban areas. This is opposite to what [Carlino et al. \(2007\)](#) finds, but in line with [Henderson \(2007\)](#) who finds that wages and rents are too high in the densest areas for R&D firms. Moreover, R&D firms' interactions with other firms may not be intensive enough to seek the proximity of other R&D firms or high rent and wage areas in urban centres.

The innovativeness of labor intensive, urban and more service oriented firms and of more land intensive firms operating in the periphery does not benefit from cultural diversity. These firms are obviously less innovative than capital intensive firms, but even so, diversity does not seem have an impact. If firm diversity coincides with regional diversity, then this implies that the diversity within larger cities is less important than the regional diversity surrounding those cities. However, to what extent regional and firm diversity coincide is subject for further research.

Obviously, there are some limitations to this study. The first is a computational one. Innovation is measured by a dichotomous variable, which leaves little variation for this variable. Therefore, three segments is the best we can give given the structure of our data. However, it is well conceivable, that there are more possible subsets of firms that have distinct innovation behavior. On the other hand, most applications of finite mixture models end up with no more than 5 segments (optimized according to information criteria). But clearly, a continuous measure for innovation would be desirable.

The second one is the definition of cultural diversity by birthplace. It is quite likely that cultural diversity entails more than ethnic diversity. Constructing a precise definition (and implementation) of cultural diversity falls however outside the scope of the paper, but definitely require further research.

Third and finally there is the possibility of reverse causality. It is perhaps not so much that immigrants are drawn to innovative firms, but that innovative firms actively seek worldwide for the best employees, which would then lead to higher cultural diversity within the firm. If so, then our current estimate is biased upwards.

Appendix A. Segmenting Firms with Finite Mixture Modeling

This paper uses the following finite mixture approach to segment firms into three segments (we follow here the notation of [Leisch, 2004](#)). In total we have 2,664 observations belonging to 888 firms. Thus, we have 888 firms that we want to segment, where firm i consists of N_i observations. Assume that observations on $f(P_{it})$ arise from a population that is a mixture of S segments in proportions π_1, \dots, π_S , where we do not know in advance from which segment observations on $f(P_{it}, D_{it})$ arise. Then, the conditional density function of $f(P_{it}, D_{it})$ can be decomposed into its various segments as follows:

$$f((P_{it}, D_{it})|X_{it}, Y_{it}, \alpha, \beta, \gamma) = \sum_{s=1}^S \pi_s f_s((P_{it}, D_{it})|X_{it}, Y_{it}, \alpha_s, \beta_s, \gamma_s), \quad (\text{A.1})$$

where $\pi_s \geq 0$, $\sum_{s=1}^S \pi_s = 1$, X_{it} is the matrix of variables that measures various firm specific characteristics, Y_{it} is the matrix of variables explaining cultural diversity within firm i and α_s , β_s , and γ_s are the vectors of parameters specific for each segment s .

The log-likelihood of (A.1) is estimated by applying the expectation maximization (EM) algorithm of [Dempster et al. \(1977\)](#). The first step is the expectation (E) step, which computes the expected value of the complete log-likelihood function with respect to the segments s . This is given by:

$$\ln L = \sum_{i=1}^N \sum_{n=1}^{N_i} \ln f((P_{it}, D_{it})|X_{it}, Y_{it}, \alpha, \beta, \gamma), \quad (\text{A.2})$$

where $\sum_{i=1}^I N_i/N = 1$. The posterior probability that firm i belongs to segment s is given by:

$$\hat{\pi}_{is} = \frac{\pi_s \prod_{n=1}^{N_i} f_s((P_{it}, D_{it})|X_{it}, Y_{it}, \alpha_s, \beta_s, \gamma_s)}{\sum_{s=1}^S \pi_s \prod_{n=1}^{N_i} (f_s((P_{it}, D_{it})|X_{it}, Y_{it}, \alpha_s, \beta_s, \gamma_s))}. \quad (\text{A.3})$$

We can now derive the probability of segment s which can be inserted in (A.1) as:

$$\hat{\pi}_s = \frac{1}{I} \sum_{p=1}^I \hat{\pi}_{ps}. \quad (\text{A.4})$$

Thus, $\hat{\pi}_s$ are estimated using current values of the model parameters and can be inserted in (A.1).

In the maximization (M) step, the expected value of the complete log-likelihood function (A.1) is maximized with respect to the model parameters using the posterior probabilities as weights. This maximization step is performed sequentially (see [van Dijk et al., 2007](#)) as follows:

$$\max_{\delta_s} \sum_{i=1}^I \hat{\pi}_s \ln f_s((P_{it}, D_{it}) | X_{it}, Y_{it}, \alpha_s, \beta_s, \gamma_s), \quad (\text{A.5})$$

Both steps E and M are now iteratively applied until convergence occurs ([Leisch, 2004](#)).

References

- Abbring, J., van den Berg, G., 2003. The Non-Parametric Identification of Treatment Effect in Duration Models. *Econometrica* 71, 1491–1517.
- Agrawal, A., Kapur, D., McHale, J., Sep. 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics* 64 (2), 258–269.
- Alesina, A., La Ferrara, E., 2005. Ethnic diversity and economic performance. *Journal of economic literature* 43 (3), 762–800.
- Audretsch, D. B., Feldman, M. P., 2004. Knowledge spillovers and the geography of innovation. In: Henderson, J. V., Thisse, J. F. (Eds.), *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, Amsterdam, Ch. 61, pp. 2714–2739.
- Baum-Snow, N., Pavan, R., 2012. Understanding the City Size Wage Gap. *Review of Economic Studies* 79 (1), 88–127.
- Berliant, M., Fujita, M., Jul. 2012. Culture and diversity in knowledge creation. *Regional Science and Urban Economics* 42 (4), 648–662.
- Brunow, S., Blien, U., 2011. Effects of cultural diversity on individual establishments. Norface migration, discussion paper no. 2011-6.
- Carlino, G. A., Chatterjee, S., Hunt, R. M., May 2007. Urban density and the rate of invention. *Journal of Urban Economics* 61 (3), 389–419.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39 (1), 1–38.
- Florax, R., de Graaff, T., Waldorf, B., 2005. A Spatial Economic Perspective on Language Acquisition: Segregation, Networking and Assimilation of Immigrants. *Environment and Planning A* 37, 1877–1897.
- Forbes, 2011. Global diversity and inclusion: fostering innovation through a diverse workforce. *Forbes insights*, 1–20.

- Heckman, J. J., Singer, B., 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52 (2), 271–320.
- Henderson, V. J., Jul. 2007. Understanding knowledge spillovers. *Regional Science and Urban Economics* 37 (4), 497–508.
- Hoogendoorn, S., van Praag, M., 2012. Ethnic Diversity and Team Performance: A Field Experiment. Tinbergen Institute Discussion Paper, TI 2012-068/3.
- Hunt, J., Gauthier-loiselle, M., 2010. How Much Does Immigration Boost Innovation? *American Economic Journal: Macroeconomics* 2 (410), 31–56.
- Jedidi, K., Jagpal, H. S., Desarbo, W. S., 1997. Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity. *Marketing Science* 16 (1), 39–59.
- Jensen, M. B., Johnson, B., Lorenz, E., Lundvall, B. A., Jun. 2007. Forms of knowledge and modes of innovation. *Research Policy* 36 (5), 680–693.
- Kerr, W. R., 2009. Breakthrough inventions and migrating clusters of innovation. NBER Working Paper No. 15443, Cambridge.
- Lancaster, T., 1990. *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge.
- Lankhuizen, M. B. M., de Graaff, T., de Groot, H. L. F., 2012. Product Heterogeneity , Intangible Barriers and Distance Decay : The effect of multiple dimensions of distance on trade across different product categories. Tinbergen Institute Discussion Paper.
- Lazear, E. P., 1999. Culture and Language. *Journal of Political Economy* 106 (6), S95–S126.
- Leisch, F., 2004. Flexmix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software* 11 (8), 1–18.
- Maré, D. C., Fabling, R., Stillman, S., Maré, D., 2009. Immigration and Innovation. Motu working paper (May).
- Mattoo, A., Neagu, I. C., Özden, c., Sep. 2012. Performance of skilled migrants in the U.S.: A dynamic approach. *Regional Science and Urban Economics* 42 (5), 829–843.
- McLachlan, M., Peel, D., 2000. *Finite mixture models*. John Wiley & Sons Ltd., New York.
- Mitchell, R., Nicholas, S., 2006. Knowledge Creation in Groups : The Value of Cognitive Diversity , Transactive Memory and Open-mindedness Norms. *Electronic Journal of Knowledge Management* 4 (1), 67–74.
- Nathan, M., Lee, N., 2012. Cultural diversity , innovation and entrepreneurship: Firm-level evidence from London. Working paper, London school of economics, London.
- Navon, G., 2009. Human Capital Spillovers in the Workplace: Labor Diversity and Productivity. Munich Personal RePEc Archive, Munich.
- Niebuhr, A., Jan. 2010. Migration and innovation: Does cultural diversity matter for regional R&D activity? *Papers in Regional Science* 89 (3), 563–585.
- Ottaviano, G. I., Peri, G., Sep. 2005. Cities and cultures. *Journal of Urban Economics* 58 (2), 304–337.
- Ottaviano, G. I., Peri, G., Jun. 2006. The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography* 6 (1), 9–44.

- Ozgen, C., Poot, J., Nijkamp, P., 2011. The Impact of Cultural Diversity on Innovation : Evidence from Dutch Firm-Level Data. IZA Discussion Paper no. 6000.
- Parrotta, P., Pozzoli, D., Pytlikova, M., 2011. Does Labor Diversity affect Firm Productivity? Does Labor Diversity affect Firm Productivity? Norface migration, discussion paper no. 2011-22.
- Prat, A., Jul. 2002. Should a team be homogeneous? *European Economic Review* 46 (7), 1187–1207.
- Romer, P. M., Dec. 1993. Idea gaps and object gaps in economic development. *Journal of Monetary Economics* 32 (3), 543–573.
- Trax, M., Brunow, S., Suedekum, J., 2012. Cultural diversity and plant level productivity. Duisburg: Mercator School of Management, University of Duisburg-Essen.
- Uzzi, B., Dunlap, S., 2005. How to Build Your Network. *Harvard Business Review* 1, 1–9.
- van Dijk, B., Franses, P. H., Paap, R., van Dijk, D. J. C., 2007. Modeling regional house prices. *Econometric Institute Report*, 2007-55, Erasmus University Rotterdam.
- van Oort, F. G., 2004. *Urban Growth and Innovation. Spatially Bound Externalities in the Netherlands*. Ashgate, Aldershot.
- Wedel, M., Kamakura, W., 2000. *Market Segmentation: Conceptual and Methodological Foundations*. Kluwer Academic Publishers, Boston.

Figures

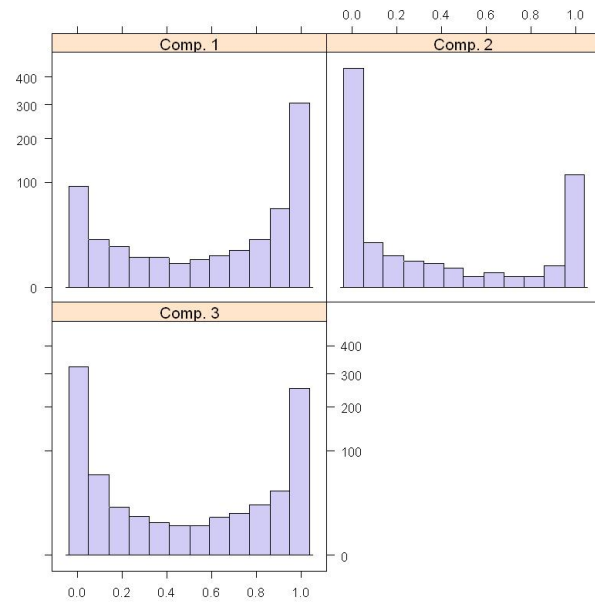
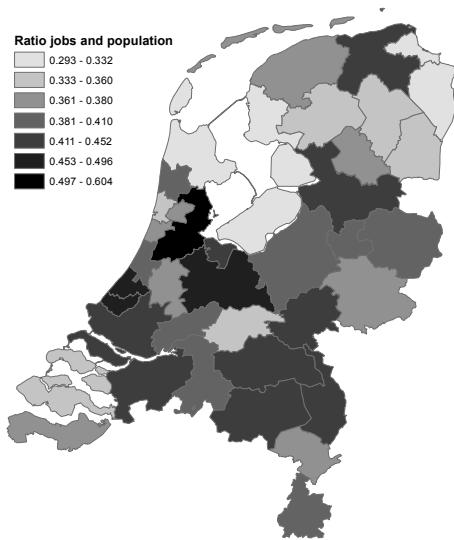
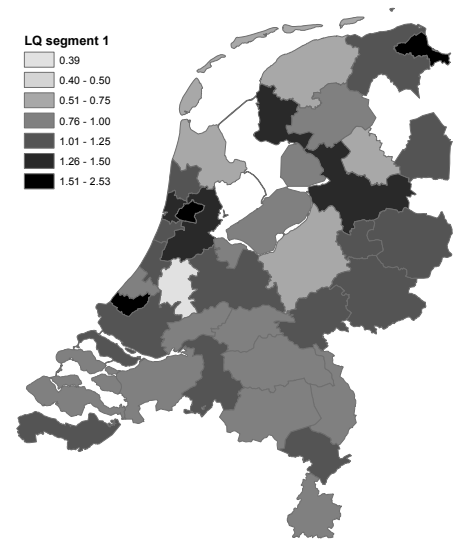


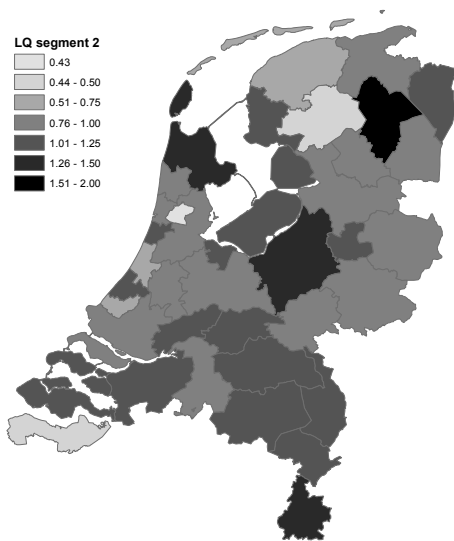
Figure 1: Rootograms of segments (with posterior probabilities larger than 0.0001)



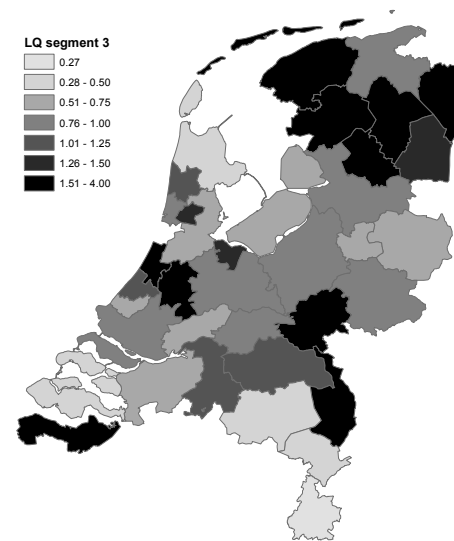
(a) Employment-population ratio



(b) LQ firms in segment 1



(c) LQ firms in segment 2



(d) LQ firms in segment 3

Figure 2: Employment-population ratio and the spatial distribution of firms by 3 segments at the NUTS 3 level

Tables

Table 1: Descriptive statistics^a ($N = 2,664$)

Variables	Mean	Std. dev.
Firm innovated	0.525	0.499
Firm innovated products	0.401	0.491
Firm innovated processes	0.379	0.485
Firm size (total employment)	423	773
Firm's openness to change	0.161	0.367
Obstacles: Lack of personnel	0.581	0.858
Obstacles: Cost	0.576	0.913
Number of firms per job (Nuts 3 level)	0.104	0.0202
Number of firms per municipality	59	54
Diversity index	0.566	0.187
High-skill intensity of foreign employment ^b	0.134	0.217
Youthfulness of foreign employment ^c	0.596	0.225
Foreign population per municipality	41702	83454
Number of 2 nd generation foreigners in a firm (both parents born abroad)	10	38

^a Due to the confidentiality agreement with Statistics Netherlands, minimum and maximum values of the variables cannot be displayed.

^b Share of highly-skilled foreigners in total foreign employment per firm.

^c Share of foreigners aged 25–45 in total foreign employment per firm.

Table 2: Summary statistics of the finite mixture modeling results

Segment	Size (1)	Posterior (2)	Ratio = $\frac{(1)}{(2)}$
1	1,053	2,376	0.443
2	1,251	1,707	0.733
3	360	1,803	0.200
Total	2,664	5,886	0.534

Table 3: Results for product innovation and diversity

Probability of product innovation (P_{it})	(1)		(2)		(3)		(4)	
	Whole sample		Segment 1		Segment 2		Segment 3	
	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.
Diversity index (excl natives)	1.149***	(0.264)	1.342	(0.994)	4.079**	(1.499)	-0.387	(0.706)
Log firm size	-0.013	(0.047)	-0.053	(0.089)	-0.011	(0.067)	0.146	(0.191)
Obstacles: Lack of personnel (low level)	0.732***	(0.123)	0.613**	(0.209)	0.846***	(0.186)	0.725	(0.396)
Obstacles: Lack of personnel (med. level)	1.048***	(0.142)	0.841***	(0.239)	1.072***	(0.215)	1.684***	(0.485)
Obstacles: Lack of personnel (high level)	1.052***	(0.231)	0.658	(0.407)	1.518***	(0.375)	0.815	(1.046)
Obstacles: Cost (low level)	0.741***	(0.132)	0.580*	(0.231)	0.772***	(0.196)	1.238**	(0.459)
Obstacles: Cost (medium level)	0.766***	(0.139)	0.872***	(0.241)	0.653**	(0.210)	0.693	(0.450)
Obstacles: Cost (high level)	0.386*	(0.195)	0.141	(0.325)	0.442	(0.308)	0.904	(0.556)
Openness to change	0.872***	(0.116)	0.893***	(0.202)	0.911***	(0.175)	0.845*	(0.393)
Number of firms per jobs (Nuts3 level)	-1.313	(2.607)	-0.973	(5.119)	0.763	(4.157)	-10.567	(10.330)
ln(firms/km2) per municipality	-0.243***	(0.052)	-0.279**	(0.102)	-0.198*	(0.085)	-0.305	(0.209)
High-skill intensity of foreign employment	1.018***	(0.201)	0.935*	(0.382)	0.966***	(0.281)	1.357	(0.743)
Youthfulness of foreign employment	0.367	(0.200)	0.342	(0.381)	1.007*	(0.394)	-0.059	(0.370)
Period 2002-2004	-0.325*	(0.127)	-0.112	(0.212)	-0.412*	(0.193)	-0.545	(0.393)
Period 2004-2006	-0.491***	(0.131)	-0.290	(0.224)	-0.588**	(0.200)	-0.609	(0.423)
Constant	-0.781	(0.466)	-0.641	(1.077)	-3.502*	(1.362)	0.118	(1.764)
Diversity (D_{it})								
	(1)		(2)		(3)		(4)	
	Whole sample		Segment 1		Segment 2		Segment 3	
	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.
Log foreign population per municipality	0.004	(0.003)	-0.002	(0.002)	-0.002	(0.001)	-0.004	(0.008)
Number of 2nd generation foreigners with both parents born abroad	0.0003**	(0.0001)	0.0001	(0.0001)	-0.0001	(0.0002)	-0.010***	(0.002)
Constant	0.525***	(0.022)	0.557***	(0.024)	0.712***	(0.011)	0.238***	(0.074)
N	2,664		1,053		1,251		360	

Significance levels : * : 5% ** : 1% *** : 0.1%

Table 4: Location quotients of top 5 sectors in each segment

Segment 1	LQ	Segment 2	LQ	Segment 3	LQ
Low-skilled business services	1.7	Manufacturing	1.6	Agriculture & forestry	2.6
Textile clothes& leather	1.6	R&D	1.5	Transport & communication	2.1
Financial intermediation	1.5	Mining & quarrying	1.4	Real estate & renting machinery	2
Environmental	1.4	Computer & related	1.3	Construction	1.9
Retail trade	1.3	Machinery & equipment	1.2	Electricity, gas & water	1.8

Table 5: Firm size distribution for each segment

Segment 1	%	Segment 2	%	Segment 3	%
≤ 50	0.04	≤ 50	0.01	≤ 50	0.17
50–100	0.11	50–100	0.06	50–100	0.19
100–250	0.51	100–250	0.44	100–250	0.54
> 250	0.34	> 250	0.49	> 250	0.1
N	1,053	N	1,251	N	360