

Migration and Innovation

Does Cultural Diversity Matter for Regional R&D Activity?

Annekatrien Niebuhr

January 2007

Abstract

Recent theoretical research deals with economic costs and benefits of cultural diversity related to immigration. However, empirical evidence regarding the impact of cultural diversity on economic performance is still scarce. This paper investigates the significance of cultural diversity of the workforce on innovation output for a cross-section of German regions. The findings indicate that cultural diversity indeed affects innovative activity. The results suggest that differences in knowledge and capabilities of workers from diverse cultural backgrounds enhance performance of regional R&D sectors. However, education levels are also important. Diversity among highly qualified employees has the strongest impact on innovation output.

JEL classification: C21, O31, R11

Keywords: Cultural diversity, innovation, knowledge production function, Germany

IAB Nord
Regional Research Network of the Institute for Employment Research
Projensdorfer Straße 82
D-24106 Kiel
Germany

Phone: +49-(0)431-3395-3921

Fax: +49-(0)431-3395-9392

e-mail:annekatrin.niebuhr@iab.de,

Acknowledgements:

I would like to thank Christoph Grenzmann (Stifterverband für die Deutsche Wissenschaft) for the generous provision of regional R&D data and Andrea Stöckmann for excellent research assistance. Financial support from the Volkswagen Foundation is gratefully acknowledged as part of the Study Group on Migration and Integration “Diversity, Integration and the Economy”. I thank Herbert Brücker, Eckhardt Bode, Rüdiger Wapler and Stefan Fuchs for helpful comments and suggestions on an earlier version of this paper. I have also benefited from useful discussions with participants of the VW Study Group Meeting at HWWI, at the ERSA Congress 2006 and the annual meeting of the committee for regional science of the Verein für Socialpolitik 2006. The usual disclaimer applies.

1. Introduction

The significance of immigration of qualified workers will rapidly increase in the ageing European economies since demographic change will cause a decline and a sharp increase of the average age of the workforce. Foreign workers are already an important factor of the German economy. In 2004, almost 7% of all employees in Germany have foreign nationality. More than 100.000 highly skilled foreigners with a university degree work in Germany.¹ Zimmermann (2005) notes that in spite of the rising importance of migration, the issue is still controversial and the understanding of the effects of international labour mobility is rather limited. Research on the economic consequences of migration has mainly focused on labour market effects and, more precisely, on the question whether immigrants depress wages and increase unemployment of native workers. Many analyses stress substitution effects among native and foreign workers. However, taking into account that labour is not homogenous, the impact of immigration depends on whether migrants are skilled or unskilled and on labour market conditions in the host country.

The objective of this paper is to provide evidence on the impact of migration on innovation at the regional level, a subject that has not received much attention in the migration literature so far. Our analysis differs from many previous studies that focus on labour market effects of immigration. The second aspect that differentiates this analysis from other studies is that we do not restrict heterogeneity of labour to the level of education only. Due to their different cultural backgrounds, it is likely that migrants and native workers have fairly diverse abilities and knowledge. Thus, there might be skill complementarities between foreign workers and natives in addition to those among workers of different qualification levels. Presumably foreign and native workers of the same educational level are also imperfectly substitutable groups because of cultural differences. Fujita and Weber (2004) argue that cultural diversity of the labour force might be of special importance for R&D activity since the generation of new products and ideas heavily relies on individual talents and skills from diverse educational and cultural environments. Due to data restrictions, we define cultural diversity as diversity of workers' nationality rather than ethnicity or cultural background. Regionally differentiated information on country of origin of inhabitants and employees is not available in German official statistics.

The possibility that diversity can enhance productivity, innovation and growth has already been considered in the economic literature. However, most studies have concentrated on the impact of *economic* diversity rather than cultural or ethnic diversity. According to Jacobs (1969), diversity of geographically proximate industries promotes innovation and growth in cities. Glaeser et al. (1992) as well as Feldman and Audretsch (1999) provide corresponding empirical evidence for US cities. Romer (1990) highlights in his seminal endogenous growth model the significance of a variety of intermediate inputs for productivity. Empirical evidence

¹ According to recent estimates of the Federal Statistical Office, almost 20% of the population in Germany has a migration background. See Statistisches Bundesamt (2006).

provided by Anderson et al. (2005) suggests that creativity is greater in regions marked by more diverse employment bases, while Duranton and Puga (2001) investigate the role that a diversified urban environment plays in fostering innovation at the regional level.

While there is an emerging theoretical literature dealing with the economic effects of cultural diversity (e.g. Fujita and Weber 2004, Lazear 1999b, 2000), there are surprisingly few empirical studies within the field of economics. Theoretical models consider different costs and benefits of diversity and specify various linkages between diversity and economic performance. However, corresponding empirical work that can help determine whether positive or negative effects of cultural diversity prevail remains scarce. Until now, there has been mainly cross-country evidence, and studies focusing on growth and productivity effects in US regions (Easterly and Levine 1997; Ottaviano and Peri 2006). To the best of our knowledge, comprehensive empirical studies dealing with innovation and cultural diversity do not exist at all. Investigations that analyse the relationship between innovation input and output fail to take cultural diversity into account (e.g. Greunz 2003, Anselin et al. 1997, Bode 2004). The aim of this paper is to investigate the impact of cultural diversity on regional innovation in Germany. Therefore we extend the knowledge production framework to analyse whether a more diverse labour force, from a cultural point of view, fosters innovation due to production complementarities, or whether negative effects of diversity, e.g. language barriers, outweigh the benefits.

The rest of the paper is organised as follows. In section 2, the theoretical framework of the analysis is outlined. Production complementarities and costs associated with cultural diversity are discussed. The cross section and data sets applied in the empirical analysis are described in section 3. An important issue of the investigation concerns the measurement of cultural diversity. In section 4, we introduce the applied diversity indicator and provide some empirical evidence of cultural diversity in German regions. We employ the knowledge production function approach to investigate the impact of cultural diversity on regional innovation capacity. The corresponding regression model and some robustness issues are discussed in section 5. The regression results are presented in section 6. Conclusions follow.

2. Theoretical Framework

Ottaviano and Peri (2006) argue that skills of foreign workers might complement those of the native labour force. In their model of multicultural production, different cultural groups provide different services. Diversity has a positive impact on regional productivity. However, heterogeneity also hampers the exchange between different cultural groups: there are adverse productivity effects because of cultural distance. Other authors also recognise that there is a trade-off with respect to heterogeneity. Lazear (1999a, 2000) considers positive productivity effects of ethnic diversity, but there are also costs of diversity arising from barriers to

communication caused by different languages and cultures.² Thus according to the literature, there appears to be an optimal degree of diversity which is influenced by the nature of production. Some of the literature on this theme also examines the significance of institutions in this context (e.g. Easterly 2001). An important result of this research is that the implementation of growth enhancing effects of diversity may require a specific set of rules, or regulatory framework. Ottaviano and Peri (2006) emphasise the role of a core of shared norms (integration) that might constitute a prerequisite for realising the potential benefits of diversity.

There appears to be a link between the costs and benefits of diversity on the one hand and the concept of ethnic identity described in Constant et al. (2006) on the other hand. According to the authors, migrants start out from their ethnicity, i.e. permanent characteristics associated with the country of origin, and then develop their ethnic identity as they are exposed to the culture and values of the host country. Ethnic identity is defined as the balance between commitments with the host country and commitments with the country of origin. Constant et al. (2006) distinguish four states of ethnic identity: assimilation, integration, marginalization and separation. Assimilation seems to imply a strong decline of both costs and benefits of cultural diversity since it is characterised by a strong identification with the host country and conformity to the corresponding norms and codes. With respect to the economic effects of diversity, integration might be interpreted as the best state because it involves commitment to the host society but also a strong dedication to the culture of origin, thus still ensuring high benefits but relatively low costs of diversity. In contrast, in case migrants are primarily identified as marginalized or separated, cultural diversity may mainly entail high costs.

The benefits of diversity might be of particular importance in the R&D sector, whereas in industries specialized on more standardised forms of production the costs of a diverse labour force might easily outweigh the positive effects. Alesina and La Ferrara (2005) argue that cultural diversity may lead to innovation and creativity since it involves variety in abilities and knowledge. Fujita and Weber (2004) argue that knowledge production relies heavily on talents and skills of employees coming from a wide range of cultural backgrounds. The nature of R&D activity calls for interaction between different workers and a pooling of different ideas and abilities. Berliant and Fujita (2004) also refer to the significance of cultural diversity for knowledge creation and transfer. The heterogeneity of people is important for the creation of new ideas.

As outlined by Alesina and La Ferrara (2005), ethnic diversity can affect economic performance in different ways. Diversity might have a direct impact on economic outcomes via different preferences or by influencing individual strategies. Moreover, diversity might have an influence on the production process. Our analysis focuses on the latter approach.

² Costs of diversity might also be due to an inability to agree on common public goods and public policies. See Alesina and La Ferrara (2005).

Fujita and Weber (2004) consider a production function that includes diversity effects. They investigate the heterogeneity between the native population and immigrants that is associated with a production complementarity. In their model, the impact of diversity on the output Q of region i is as follows:

$$Q_i = (N_i^\gamma + I_i^\gamma)^{\frac{1}{\gamma}} \quad (1)$$

where N_i is the number of native workers and I_i is the immigrant work force. The parameter γ measures the strength of the production complementarity between workers with different cultural backgrounds. Fujita and Weber (2004) restrict the range of γ to non-negative values, more precisely $0 < \gamma < 1$. A negative value of γ implies an extremely strong complementarity effect such that output tends to zero as the labour force becomes more and more homogenous. In contrast, in case of $\gamma > 1$, cultural diversity has an unfavourable impact on production, indicating that the negative effects of diversity e.g. caused by communication barriers, are stronger than the benefits.

However, we cannot apply the production function proposed by Fujita and Weber (2004) since the focus of this analysis is on R&D activity and important determinants of regional knowledge production are missing in their approach. Moreover, the simple differentiation between migrants and natives proposed by Fujita and Weber is not appropriate given culturally diverse populations and marked differences between various migrant groups as regards their economic behaviour. Constant et al. (2006) argue that therefore migration research that treats immigrants as a homogenous group will become less important. In order to acknowledge differences between immigrants and natives as well as the diversity among immigrants, we choose a more general production function, similar to the one described in Alesina and LaFerrara (2005), as the starting point of our analysis:

$$KNOW_i = f(RD_i; DIV_i) \quad (2)$$

where $KNOW_i$ is knowledge output in region i , RD_i is R&D input and DIV_i is cultural diversity of the workforce. This function closely resembles the so-called knowledge production function introduced by Griliches (1979). The knowledge production function links knowledge output to R&D inputs. If $\partial KNOW_i / \partial DIV_i > 0$, diversity fosters regional innovation because the positive impact associated with the production complementarity outweighs negative effects linked to a labour force marked by more diverse cultural backgrounds. In contrast, $\partial KNOW_i / \partial DIV_i < 0$ implies that production complementarities are too weak to compensate for the negative effects associated with diversity. We check whether positive or negative effects dominate in the regression analysis.

3. Data

Point of departure of our empirical analysis is the knowledge production function given by equation (2) that links R&D input to R&D output, i.e. new products, processes and ideas. Thus, we first of all need adequate proxies for regional innovation and R&D input to investigate the impact of cultural diversity on knowledge production. Regional data on patent applications, used as a measure for knowledge output, and on R&D inputs in Germany are available on the county level (NUTS 3) and for planning regions (so-called Raumordnungsregionen) which comprise several counties linked by intense commuting. We have to restrict the analysis to planning regions due to some data restrictions for NUTS 3 regions. Overall, our cross section contains 95 regions. Furthermore, the analysis takes into account the region type. Starting from a classification based on a typology of settlement structure according to the criteria population density and size of the regional centre, we differentiate between agglomerated, urbanised and rural regions.³

Patent applications, applied as an indicator for innovative output of the region, comprise patents published by the German and the European patent office that have been assigned to the innovators' region of residence. As Bode (2004) notes, this approach avoids potential mismeasurement due to centralised patenting of multi-site companies. Annual patent data is available for the period 1995 to 2000.⁴ Information on R&D input was provided by the German Stifterverband. R&D data include R&D staff as well as R&D expenditure of commercial firms. The data come from a biannual census and are available for 1995, 1997 and 1999. However, we can only use data for 1997 and 1999 in our analysis. Data for 1995 is not compatible due to some changes in the delineation of regions. Thus, the investigation is restricted to a panel data set with only two observations in the time dimension. Finally, we include several explanatory variables in the regression model based on employment data provided by the German Federal Employment Agency. The employment statistic covers all employment subject to social security contributions.⁵ The information is given on the NUTS 3 level and refers to workplace location. We use employment data differentiated by nationality, educational level, branch, occupation and firm size in order to generate our diversity measure and several control variable that enter into the regression model.

4. Spatial Dimension of Cultural Diversity in Germany

Our indicator of cultural diversity is rooted in the literature on growth effects of ethnic fragmentation (e.g. Easterly and Levine 1997). In these studies, the probability that two randomly drawn individuals belong to two different groups is frequently applied as a measure

³ Four planning regions had to be merged due to restricted data availability. The classification has been developed by the Federal Office for Building and Regional Planning. For details see appendix A1 and http://www.bbr.bund.de/raumordnung/europa/download/spesp_indicator_description_may2000.pdf.

⁴ See Greif and Schmiedl (2002) for more detailed information on the patent data base.

⁵ Hence, civil servants and self-employed are not recorded in the employment statistic.

of fragmentation. The measure of diversity is calculated as 1 minus the Herfindahl index of concentration across groups:

$$DIV_{it} = 1 - \sum_{k=1}^K s_{ikt}^2 \quad (3)$$

where s_{ikt} is the share of employees with nationality k among all employees of region i in year t . Ottaviano and Peri (2006) note that this indicator accounts for both richness of the distribution (i.e. number of nationalities) and a relatively even distribution across nationalities. Thus, according to this measure, cultural diversity will increase if the number of nationalities rises or if the shares of different nationalities in employment converge. In this study, the indicator is based on regional employment data differentiated by educational level and nationality. We differentiate between 3 levels of education (no formal vocational qualification, completed apprenticeship, university degree) and 213 nationalities. Four different diversity indices are calculated: one aggregate measure which is based on total employment by nationality and three qualification-specific indices corresponding to the three levels of education mentioned above. By considering the cultural diversity of the labour force at different qualification levels we can check whether education matters, i.e. taking into account that it might be cultural diversity of highly qualified workers only that affects the process of innovation.

In contrast to most studies that are based on data for the US, we use employment instead of population data. The advantage of our measure is a closer connection to the production process. Moreover, nationality defines cultural identity of employees in the present analysis. Country of birth is the most widely used indicator in this context. However, information on country of birth is not available in the German employment statistic.⁶ Applying nationality to determine cultural identity has assets and drawbacks. Referring to nationality implies that naturalised citizens do not enter into the diversity measure as “foreign” persons. However, using country of origin as a definition of the foreign workforce implies that we do not consider people with a migration background born in Germany – unless we have information on the country of birth of the parents. Naturalised employees probably tend to be more successful with respect to qualification and labour market integration due to the terms of naturalisation in Germany (minimum duration of stay and language skills required). Therefore, our diversity measure might be imprecise especially with respect to the highly qualified labour force.

The share of foreign employees in Germany amounts to 7.1% in 2000.⁷ This corresponds with a value of the overall diversity measure of 0.136. Table 1 shows regions marked by relatively high and low diversity of the workforce, respectively. The group of regions with

⁶ The same applies to German population statistics.

⁷ This only refers to the labour force subject to social security contributions.

comparatively high cultural diversity almost solely consists of agglomerations and urbanised areas. Moreover, all of them are located in western Germany. The most diversified regions are Stuttgart, Munich and Rhein-Main (i.e. the Frankfurt area), highly agglomerated regions in the South-West of Germany.⁸ There are no cities from the northern part of the country among the leading regions. Lowest diversity measures arise for eastern German regions. Eastern Germany does poorly as regards diversity of their labour force, most notably some rural peripheral areas (Mecklenburger Seenplatte, Südwestsachsen, Vorpommern).

[Table 1 around here]

There are also distinct differences between the considered levels of education. Overall, diversity is highest among low-skilled employees who have no formal vocational qualification. The ranking of regions differs somewhat for the different qualification groups. However, the qualification-specific diversity measures are highly correlated. There are pronounced differences between eastern and western Germany for all diversity measures (see Figure 1), i.e. they pertain to all levels of education.⁹ We find the largest disparity between the East and West for diversity of low- and medium-skilled employees. Besides the disparities between eastern and western German regions, there are also marked differences among region types. The highest diversity indices are found in agglomerated regions irrespective of the qualification level, whereas rural areas on average exhibit relatively low cultural diversity of employment.

[Figure 1 around here]

5. Econometric Issues

5.1. Basic Specification

We apply the knowledge production function to investigate the impact of cultural diversity of the workforce on regional innovation. The knowledge production function links innovative output to R&D inputs. Since the number of patent applications is also affected by the size of the regional economy, we investigate the relationship between patents and R&D input in per capita terms. R&D staff¹⁰ and R&D expenditure per inhabitant are used as proxies for R&D activity. The basic regression model is given by:

$$\ln P_{it} = \alpha_0 + \alpha_1 \ln RD_{it-1} + \alpha_2 \ln DIV_{it} + \sum_{n=1}^N \beta_n CONTROL_{nit} + u_{it} \quad (4)$$

⁸ The evidence is in accordance with the findings provided by Ottaviano and Peri (2006). They find indices up to 0.58 for US cities based on population figures.

⁹ For maps showing the regional distribution of all qualification-specific diversity measures see appendix A3.

¹⁰ Data on R&D personnel is given in full-time equivalents.

where P_{it} is the number of patents per capita in region i and year t . RD_{it-1} is R&D personnel or R&D expenditure per capita in year $t-1$ and u_{it} is the error term. In order to appropriately model the relationship between R&D input and output, the input variable enters into the model with a time lag of one year. Patents as well as R&D input refer to data from firms only. With respect to the objective of the investigation, the most prominent variable is the diversity index DIV_{it} which is calculated according to equation (3). Separate models are estimated for diversity measures based on total employment and qualification-specific employment figures.

Furthermore, we expand the original knowledge production function by some control variables in order to avoid misspecification due to omitted variables. Controls comprise indicators for the sectoral composition of regional economies, more precisely the ratio of service to manufacturing employment in the region $STRUC_{it}$. The industry structure is considered because the propensity to patent is higher in manufacturing than in the service sector. Moreover, the innovative performance of regions might be influenced by the intensity of local university research. Therefore we also included the number of R&D staff at universities and polytechnics per inhabitant UNI_{it} as an explanatory variable. According to Bode (2004), the propensity to patent might also be affected by the size of firms. In order to capture corresponding effects two additional variables are considered: the employment shares of small (less than 20 employees) and large (500 or more employees) firms ($SMALL_{it}$, $LARGE_{it}$). As the innovation process in highly agglomerated areas may significantly differ from the process in rural peripheral regions, we take into account the region type as well ($REGTYPE_i$). Finally, an indicator for human-capital endowment of the region HC_{it} is included because human capital might foster the innovation process via facilitating knowledge spillovers. Human capital is measured by the share of highly skilled employees (university degree) in total employment. Inclusion of a human-capital variable also enables us to check whether diversity among highly qualified workers just works as an approximation of human-capital endowment of the region.

5.2. Robustness Checks

To investigate the robustness of our empirical results, a number of additional regression models are applied. Firstly, we have to consider potential effects of measurement errors and endogeneity of explanatory variables. The estimated effect of diversity on R&D output might be biased due to potential endogeneity of cultural diversity. We use diversity measures lagged by 5 years and a dummy variable differentiating between eastern and western German regions as instruments for contemporaneous diversity indices. These variables are highly correlated with contemporaneous diversity and unlikely to be affected by reverse causation. This applies especially to the East-West dummy as a pure geographic variable.

Secondly, fixed and random effects panel data models are applied so as to control for unobserved time-invariant explanatory variables:

$$\ln P_{it} = \alpha_0 + \alpha_1 \ln RD_{it-1} + \alpha_2 \ln DIV_{it} + \sum_{n=1}^N \beta_n CONTROL_{nit} + \eta_i + \lambda_t + v_{it} \quad (5)$$

where η_i denotes a region-specific effect, controlling for unobservable regional characteristics that are time-invariant. λ_t captures unobservable time effects and v_{it} is a white noise error term.

Evidence provided by Bode (2004) and Anselin et al. (1997) suggest that geographically bounded spillovers and spatial dependence are important for regional innovative activity. Therefore, we check for misspecification due to omitted spatial effects indicated by spatial autocorrelation in the regression residuals. Depending on the results of corresponding LM-tests, we might estimate spatial lag models or spatial error models. The spatial lag model is given by:

$$\ln P_{it} = \alpha_0 + \rho \sum_{j=1}^R w_{ij} \ln P_{jt} + \alpha_1 \ln RD_{it-1} + \alpha_2 \ln DIV_{it} + \dots + \eta_i + \lambda_t + v_{it} \quad (6)$$

Thus we extend the non-spatial model by a spatial lag of the dependent variable $\sum_{j=1}^R w_{ij} \ln P_{jt}$ where w_{ij} is the contiguity matrix. Taking into account the weighted sum of patent applications in neighbouring regions implies that spatial autocorrelation of the error term is caused by omission of some substantive form of spatial dependence caused by interaction among neighbouring regions. Geographic knowledge spillovers might result in interdependent innovation processes of adjacent R&D departments leading to spatial autocorrelation.

In contrast, the spatial error model will be the appropriate specification if the misspecification is due to nuisance dependence. Anselin and Bera (1998) note that spatial autocorrelation in measurement errors or in variables that are otherwise not crucial to the model might entail spatial error dependence. The spatial error model may be expressed as:

$$\ln P_{it} = \alpha_0 + \alpha_1 \ln RD_{it-1} + \alpha_2 \ln DIV_{it} + \dots + \eta_i + \lambda_t + v_{it} \quad \text{and} \quad v_{it} = \lambda \sum_{j=1}^R w_{ij} v_{jt} + \varepsilon_{it} \quad (7)$$

Finally, we take into account that outlying observations might have a marked effect on the regression results. To address this issue we apply quantile regressions as introduced by Koenker and Basset (1978). The median regression corresponds to the least absolute deviation estimator and is a robust alternative to OLS. Quantile regressions minimise an objective function which is a weighted sum of absolute deviations:

$$\min_{\gamma} \left[\sum_{i: y_i \geq x'_i \gamma} \theta |y_i - x'_i \gamma| + \sum_{i: y_i < x'_i \gamma} (1 - \theta) |y_i - x'_i \gamma| \right] \quad (8)$$

Here y_i is the dependent variable and x_i is the vector of explanatory variables which is multiplied by the coefficient γ . The objective function can be interpreted as an asymmetric linear penalty function of deviations from predicted to actual patents per capita. An important special case is the median regression ($\theta = 0.5$). Since this regression puts less weight on outliers than OLS, it is a robust alternative. Minimising the distance to other quantiles than the median yields a family of coefficients and gives estimates for the marginal effects of a change in independent variables at different points of the conditional distribution (see Buchinsky 1998).

6. Regression Results and Discussion

Point of departure of the regression analysis is a basic pooled model including all control variables. The model is estimated with different versions of the focal explanatory variable, i.e. diversity measures based on total employment and employment at different levels of education. Table 2 shows the results of this basic model. The specifications in columns I to IV only differ with respect to the diversity measure included. In line with previous evidence on the knowledge production function, we get a highly significant impact of R&D expenditure on innovation output.¹¹ Furthermore, some control variables appear with significant coefficients, indicating that structural characteristics of the regions matter for innovative activity. The relative size of the industrial sector, importance of small firms as well as the settlement structure are associated with significant effects on the innovation output – at least in some specifications. According to the estimates, a specialisation of regions on the industrial as compared to the service sector tends to raise patents per capita. Furthermore, areas characterised by a relatively large share of small firms on average seem to perform better than other regions.

A positive effect is also associated with the region's human-capital endowment. However, the coefficient is only marginally significant at the 10% level in the models II and IV, i.e. the specifications including diversity among low- and high-skilled employees, respectively. The negative coefficient of the region-type variable implies that there are systematic differences between the innovation processes of metropolitan areas, urbanised and rural regions. More precisely, less densely populated regions, especially rural areas, are marked *ceteris paribus* by a lower productivity of R&D activity. This might point to some kind of positive agglomeration effect to be at work. In contrast, the findings indicate that university research has no important impact on innovation. The coefficient is insignificant in most specifications and negative.¹² Finally, the results point to an innovation-enhancing effect of cultural diversity of the workforce. The coefficient of the diversity measure is positive and highly

¹¹ Substituting R&D expenditure by R&D personnel does not change the results.

¹² The disappointing performance of university research might be caused by the fact our data set does not allow to focus on applied research at universities and institutes.

significant irrespective of the educational level considered. Further, the impact of diversity among highly educated employees is clearly stronger than the effect that is determined for low- and medium-skilled workers. Thus, the regression results indicate that cultural diversity is a factor which positively influences the process of knowledge creation, but the qualification level of labour also clearly matters in this context.

[Table 2 around here]

In a parsimonious specification, we delete university research because the variable is wrongly signed and mostly insignificant. Exclusion of the university research variable does not change the basic findings (see Table 3). In particular, the coefficients of all diversity measures remain positive and significant, although the effect of diversity among highly skilled workers declines somewhat. Evidence that diversity of employees with an university degree exerts the most pronounced influence of all education levels on innovation turns out to be a fairly robust result. However, as indicated by the tests for spatial autocorrelation, regional R&D activity is marked by some spatial interaction not captured by the regression model so far. The differences between the test statistics suggest that problems are caused by omission of some kind of substantive form of spatial dependence that rests upon knowledge spillovers between neighbouring regions.¹³ In order to check whether the identified impact of cultural diversity is affected by the omission of spatial dependence, we include a spatial lag of patent applications per capita in some specifications.

[Table 3 around here]

Before turning to the significance of spatial interaction, we check whether unobservable region-specific effects are important and adversely affect the estimates of the pooled model. Furthermore, we skip the model with the diversity measure based on total employment from now on and focus on the different qualification levels. According to the Breusch-Pagan tests (BP) and the F -tests displayed in Table 4, there are significant region-specific effects. However, the results of the random-effects model (columns I to III) are very similar to the estimates of the OLS regression of the pooled data. The coefficient of R&D expenditure slightly declines but is still highly significant. The impact of cultural diversity turns out to be very stable, the effect of diversity among highly skilled employees is even reinforced. However, the results change dramatically in the fixed-effects model (columns IV to VI). The coefficients of the diversity measures are insignificant, although still of the same sign at least for medium- and high-qualified employees. As regards the findings for R&D input, the result is even worse. We get a significant negative impact of R&D expenditure on patent applications.

¹³ Higher significance of LM lag tests suggests that the spatial lag model is the appropriate specification. The corresponding decision rule is proposed by Anselin and Florax (1995).

[Table 4 around here]

The problem of the fixed-effects model might be linked to the quality of the data on R&D input, i.e. survey data that may be affected by measurement errors. Johnston and DiNardo (1997) note that estimates may be biased towards zero due to mismeasurement of explanatory variables. The attenuation bias can be aggravated by fixed effects estimation, in particular if the explanatory variables are highly correlated across time, as is frequently the case when the time period between the two cross sections is small (see also Griliches and Hausman 1986). With respect to the data set used in the regression analysis, this applies to R&D expenditure per capita as well as to the diversity indices. Although there is a considerable variation across regions, there is much less variation in time. Because of the completely implausible implications and the outlined methodological problems of the fixed-effects specification, we focus on the random-effects model for the remainder of the robustness checks.

The results of the IV regressions suggest that endogeneity of cultural diversity is unlikely to be a major problem (see Table 5). The diversity measures are instrumented by the East-West dummy in the displayed specifications. The impact of cultural diversity on innovation output is even reinforced in some models. As regards the impact of spatial interaction, we do not arrive at robust results. Significance of the spatially lagged dependent variable is affected by the choice of the spatial weights matrix. Application of a binary contiguity matrix results in a significant positive effect of patent applications in neighbouring regions, whereas the corresponding coefficient is not significantly different from zero at the 10% level for a weight matrix based on inverse distance with a cut-off point. Altogether, the basic findings regarding the impact of cultural diversity are not changed in the spatial lag model. The use of the East-West dummy as an instrument for diversity yields very robust evidence.

[Table 5 around here]

Finally, we check whether outlying observations affect the estimates by applying quantile regressions. Table 6 shows the coefficients of the diversity measures only. The results are based on a specification that includes all variables considered in the models IV to VI in Table 5. The spatially lagged dependent variable is instrumented. Results are given for the median regression, i.e. the least absolute deviations estimator and the regressions minimising the weighted sum of deviations to 10th, 25th, 75th and 90th quantile. The coefficients of the median regression are rather similar to the previous findings, indicating that the effect of cultural diversity is not subject to serious bias caused by outliers. Furthermore, the estimates of the other quantile regressions reveal that diversity has a significant impact at almost all parts of the conditional distribution. Whereas the size of the effect seems to decline as we move towards the upper quantiles of the distribution for the low and medium qualification level, there is no such systematic change for diversity among highly skilled employees. Only in the upper part of the distribution does diversity exert no influence on innovation. This implies

that cultural diversity does not matter for over performing regions in terms of innovation success.

[Table 6 around here]

Altogether, the analysis provides evidence that cultural diversity matters for the productivity of R&D at the regional level. However, less convincing results emerge in case lagged diversity measures are employed as instruments. Some of the corresponding specifications give rise to insignificant coefficients of the diversity variables and in some cases even to changes of sign.¹⁴ Surprisingly, evidence on positive effects of diversity among low-skilled workers seems to be most robust in this context. Data problems are likely to play a prominent role with respect to these findings. In particular, it might be important that a relatively high proportion of highly qualified employees with migration background is not captured by our diversity indicator since it is based on employment data by nationality and there seems to be a significant positive correlation between the probability of naturalisation and educational achievement.¹⁵ The diversity measure for the highest educational level could therefore most notably be affected by measurement errors resulting in biased coefficient estimates. Up to now, there is no comprehensive information available on country of origin or migration background of employees in Germany.

7. Conclusions

The regression results indicate that cultural diversity might indeed matter for innovative activity at the regional level. The empirical evidence points to differences in knowledge and capabilities of workers from diverse cultural backgrounds that may enhance performance of regional R&D sectors. The benefits of diversity seem to outweigh the negative effects. But education matters as well in this context. The strongest impact on innovation output is found for diversity among highly qualified employees. This is a plausible result as we would expect especially cultural diversity of highly skilled labour to be of importance for the development of new ideas and products. Thus cultural diversity based on the immigrant labour force releases positive economic effects, in the present case on innovative activity. However, we need to keep in mind that our diversity measures rest upon employed migrants. Thus, the positive impact can only be attached to immigrants already integrated into the labour market.

Some theoretical literature on economic effects of cultural diversity stresses the significance of institutions in this context. An important result of this research is that the implementation of growth enhancing effects of diversity may require a specific set of rules, or regulatory framework. Our results, i.e. the significance of the educational level and the fact that our

¹⁴ The unreported regression results are available from the author upon request.

¹⁵ In fact, improvement of career prospects seems to be an important motive for naturalisation in Germany, see Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration (2005).

focus is on *employed* migrants, suggest that institutions and regulatory framework concerned with education and labour-market integration of immigrants play a particular role in realising the benefits of diversity for innovation activity.

As regards future research, measurement issues discussed above call for the provision of more and better data on the population and labour force with migration background. Data restrictions possibly affect the precision of our regression results. In particular, approximation of cultural diversity among highly qualified employees might be exposed to a serious downward bias because we cannot record naturalised persons who presumably tend to be the more economically successful among workers with migration background. This means that, assuming the same spatial distribution of naturalised and foreign employees, the impact identified for cultural diversity among highly skilled workers is likely to be subject to an upward bias. Thus, differences in economic effects of diversity at distinct educational levels might be smaller than implied by our regression results.

References

- Alesina, A. and E. La Ferrara (2005): Ethnic Diversity and Economic Performance, *Journal of Economic Literature* 43 (3): 762-800.
- Anderson, R., J.M. Quigley and M. Wilhelmsson (2005): Agglomeration and the spatial distribution of creativity, *Papers in Regional Science* 83 (3): 445-464.
- Anselin, L., A. Varga and Z. Acs (1997): Local Geographic Spillovers between University Research and High Technology Innovations, *Journal of Urban Economics* 42 (3): 422-448.
- Anselin, L. and A. Bera (1998): Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In: A. Ullah and D. Giles (Eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 237–289.
- Anselin, L. and R. Florax (1995): Small Sample Properties of Tests for Spatial Dependence in Regression Models. In: Anselin, L. and R. Florax (Eds.), *New Directions in Spatial Econometrics*, Berlin, Springer, pp. 21-74.
- Beaufragte der Bundesregierung für Migration, Flüchtlinge und Integration (2005): *Daten, Fakten, Trends. Einbürgerung*, Berlin.
- Berliant, M and M. Fujita (2004): Knowledge Creation as a Square Dance on the Hilbert Cube (Washington University, Department of Economics: <http://economics.wustl.edu/berliantfujitaworkingpaper2.pdf>).
- Bode, E. (2004): The spatial pattern of localized R&D spillovers: an empirical investigation for Germany, *Journal of Economic Geography* 4: 43-64.
- Buchinsky, M. (1998): Recent Advances in Quantile Regression Models. A Practical Guideline for Empirical Research, *Journal of Human Resources* 33: 88-126.
- Constant, A., L. Gataullina and K.F. Zimmermann (2006): Ethnosizing Immigrants. IZA Discussion Paper No. 2040.
- Duranton, G. and D. Puga (2001): Nursery cities: urban diversity, process innovation, and the life cycle of products, *American Economic Review* 91 (5): 1454-1477.
- Easterly, W. and R. Levine (1997): Africa's Growth Tragedy: Policies and Ethnic Divisions, *Quarterly Journal of Economics* 111 (4): 1203-1250.
- Feldman, M. and D. B. Audretsch (1999): Innovation in cities: Science-based Diversity, Specialization and Localized Competition, *European Economic Review* 43 (2): 409-29.
- Fujita, M. and S. Weber (2004): Strategic Immigration Policies and Welfare in Heterogeneous Countries, Fondazione Eni Enrico Mattei (FEEM) Working Paper No. 2.2004 (Milano: Nota di lavoro / Fondazione Eni Enrico Mattei).
- Glaeser, E.L., H.D. Kallal, J.A. Scheinkman and A. Shleifer (1992): Growth in Cities, *Journal of Political Economy* 100 (6): 1126-52.
- Greif, S. and D. Schmiedl (2002): *Patentatlas Deutschland - Ausgabe 2002. Dynamik und Strukturen der Erfindungstätigkeit*, München.
- Greunz, L. (2003): Geographically and Technologically Mediated Knowledge Spillovers between European Regions, *Annals of Regional Science* 37 (4): 657-80.

- Griliches, Z. (1979): Issues in Assessing the Contribution of R&D to Productivity Growth, *Bell Journal of Economics* 10 (1): 92-116.
- Griliches, Z. and J. Hausman (1986): Errors in Variables in Panel Data, *Journal of Econometrics* 31: 93-118.
- Jacobs, J. (1969): *The Economy of Cities*, New York, Random House.
- Johnston, J. and J. DiNardo (1997): *Econometric Methods*, McGraw-Hill.
- Koenker, R. and G. Bassett (1978): Regression Quantiles, *Econometrica* 46: 33-50.
- Lazear, E.P. (1999a): Culture and Language, *Journal of Political Economy* 107 (6): 95-126.
- Lazear, E.P. (1999b): Globalisation and the Market for Team-Mates, *The Economic Journal* 109: C15-C40.
- Lazear, E.P. (2000): Diversity and Immigration, in: G. J. Borjas (ed.), *Issues in the Economics of Immigration*, Chicago: University of Chicago Press, pp. 117-42.
- Ottaviano, G.I.P. and G. Peri (2006): The economic value of cultural diversity: evidence from US cities, *Journal of Economic Geography* 6 (1): 9-44.
- Romer, P.M. (1990): Endogenous Technological Change, *Journal of Political Economy* 98 (2): S71-S102.
- Schürmann C. and A. Talaat (2000): *Towards a European Peripherality Index: Final Report*, Berichte aus dem Institut für Raumplanung, No. 53, Dortmund
- Statistisches Bundesamt (2006): *Leben in Deutschland. Haushalte, Familien und Gesundheit - Ergebnisse des Mikrozensus 2005*, Wiesbaden.
- Zimmermann, K.F. (2005): European Labour Mobility: Challenges and Potentials, *The Economist* 153 (4): 425-450.

Appendix

A1. Cross section and region types

Type	Spatial categories	Size of the regional centre (number of inhabitants)	Population density (inhabitants per km ²)
1 Agglomerated regions			
	Highly agglomerated with large centre	> 300.000	> 300
	Agglomerated with large centre	> 300.000	150 up to 300
2 Urbanised regions			
	Urbanised with large centre	< 300.000 or > 300.000	> 150 (and a centre with < 300.000 inhabitants) or 100 up to 150 (and a centre with > 300.000 inhabitants)
	Urbanised without large centre	< 300.000	100 up to 150
3 Rural regions			
	Low population density and centre	> 125.000	< 100
	Low population density without centre	< 125.000	< 100

A2. Data

R&D data from Stifterverband für die Deutsche Wissenschaft on NUTS 2 and NUTS 3 level

- R&D personnel 1997, 1999
- R&D expenditure 1997, 1999

Patent data from Patentatlas Deutschland - edition 2002 on NUTS 3 level

- Patent applications 1995 - 2000

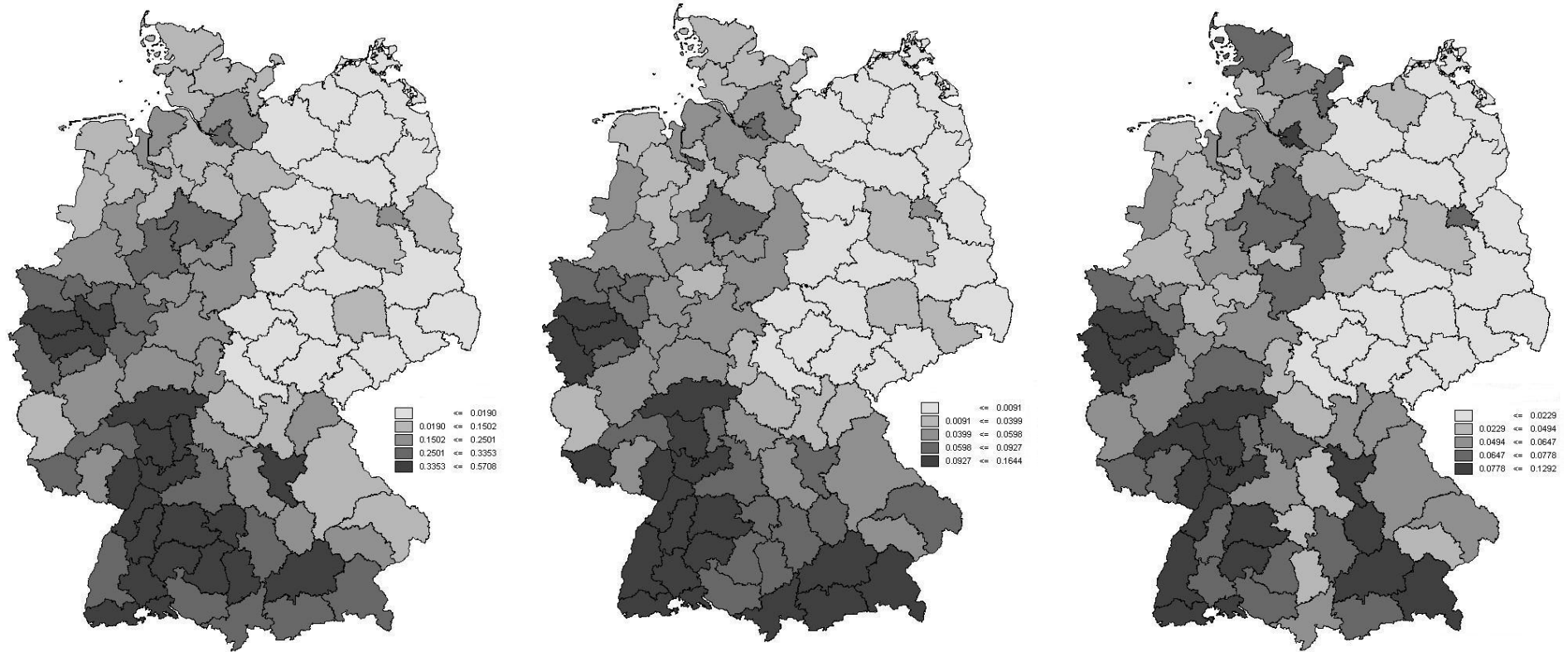
Employment data from the German Federal Employment Agency on NUTS 3 level

- Employment by nationality, qualification level and occupation 1993, 1995, 1998, 2000

Distance and travel time

- Interregional travel time bases on estimates for NUTS 3 regions by IRPUD Dortmund (Schürmann and Talaat 2000). Travel time for planning regions was generated by calculating weighted averages of NUTS 3 data.

A3. Regional disparities in cultural diversity in Germany (low, medium, high skilled employment), 2000



Tables and Figures

Table 1: Cultural diversity in German planning regions, 2000

Region	Diversity index			
	Total	Low skilled	High skilled	
Düsseldorf	0.200	0.401	Hochrhein-Bodensee	0.095
Köln	0.200	0.368	Neckar-Alb	0.096
Hochrhein-Bodensee	0.213	0.422	Südlicher Oberrhein	0.098
Neckar-Alb	0.219	0.419	Mittlerer Oberrhein	0.100
Mittlerer Oberrhein	0.219	0.423	Starkenburger	0.103
Starkenburger	0.219	0.457	Südostoberbayern	0.105
Nordschwarzwald	0.221	0.447	Unterer Neckar	0.112
Rhein-Main	0.244	0.485	Rhein-Main	0.116
München	0.254	0.496	München	0.127
Stuttgart	0.290	0.571	Aachen	0.129
Mecklenb. Seenplatte	0.006	0.007	Dessau	0.007
Altmark	0.006	0.008	Südwestsachsen	0.008
Nordthüringen	0.008	0.011	Mecklenb. Seenplatte	0.009
Vorpommern	0.008	0.007	Altmark	0.010
Südwestsachsen	0.009	0.016	Oberlausitz-Niederschles.	0.010
Oberlausitz-Niederschles.	0.010	0.012	Uckermark-Barnim	0.011
Chemnitz-Erzgebirge	0.010	0.015	Westmecklenburg	0.011
Ostthüringen	0.010	0.015	Nordthüringen	0.011
Westmecklenburg	0.010	0.013	Vorpommern	0.013
Magdeburg	0.011	0.015	Südthüringen	0.014
Agglomerated regions	0.169	0.355		0.083
Urbanised regions	0.101	0.235		0.052
Rural regions	0.069	0.144		0.039
East Germany	0.035	0.074		0.029
West Germany	0.161	0.323		0.086
Germany	0.136	0.293		0.072

Source: German Employment statistic, own calculations.

Figure 1: Regional disparities in cultural diversity in Germany (total employment), 2000

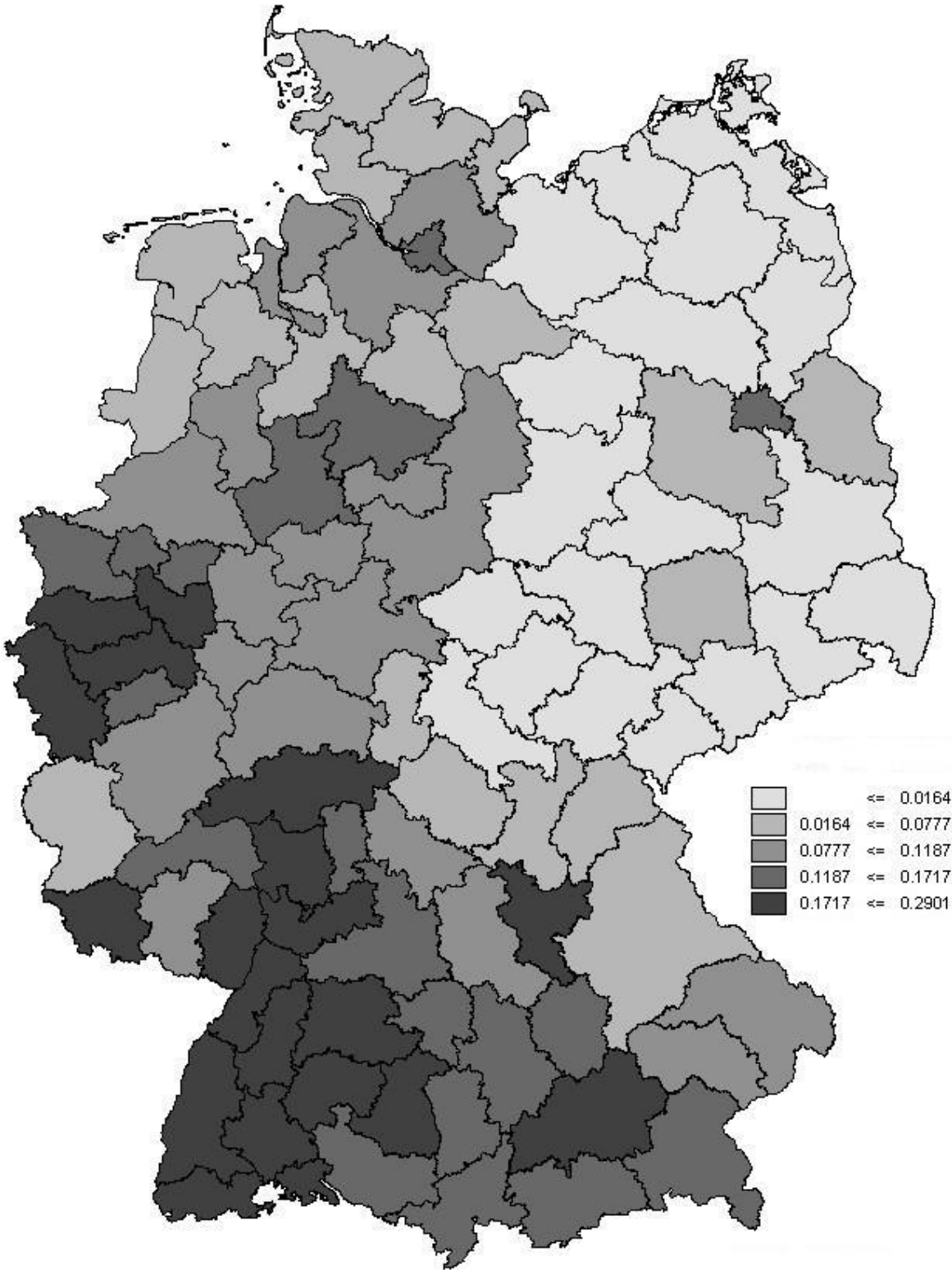


Table 2: Regression results – OLS pooled

Dependent variable	ln(patents per capita)			
	I	II	III	IV
Cons	2.77** (2.60)	2.96** (2.75)	1.99 (1.88)	3.62** (3.16)
lnRD _{it-1}	0.39** (8.15)	0.38** (7.97)	0.42** (8.57)	0.38** (7.20)
ln(UNI _{it})	-0.03 (1.58)	-0.03 (1.86)	-0.02 (1.14)	-0.04* (2.35)
ln(DIV _{it}) total	0.31** (5.99)			
ln(DIV _{it}) low		0.28** (6.14)		
ln(DIV _{it}) medium			0.27** (5.10)	
ln(DIV _{it}) high				0.43** (5.91)
ln(STRUC _{it})	0.57** (6.52)	0.55** (6.25)	0.56** (6.25)	0.71** (7.92)
ln(HC _{it})	0.22 (1.39)	0.29 (1.79)	0.03 (0.22)	0.28 (1.69)
ln(SMALL _{it})	0.93** (2.92)	1.02** (3.09)	0.86** (2.74)	0.82* (2.54)
ln(LARGE _{it})	0.19 (1.44)	0.21 (1.67)	0.19 (1.42)	0.15 (1.09)
REGTYPE _i	-0.05 (0.89)	-0.04 (0.69)	-0.08 (1.55)	-0.09 (1.80)
Adj. R ²	0.86	0.86	0.85	0.86
Observations	190	190	190	190

Notes: *t*-statistics are based upon White's heteroscedasticity-adjusted standard errors.

** significant at the 0.01 level, * significant at the 0.05 level.

Table 3: Parsimonious specification - OLS pooled

Dependent variable	ln(patents per capita)			
	I	II	III	IV
Cons	2.24* (2.01)	2.33* (2.07)	1.63 (1.50)	2.65* (2.23)
lnRD _{it-1}	0.40** (8.14)	0.40** (8.04)	0.43** (8.52)	0.40** (7.51)
ln(DIV _{it}) total	0.29** (5.43)			
ln(DIV _{it}) low		0.27** (5.59)		
ln(DIV _{it}) medium			0.26** (4.69)	
ln(DIV _{it}) high				0.38** (4.94)
ln(STRUC _{it})	0.58** (6.58)	0.56** (6.35)	0.57** (6.31)	0.71** (7.51)
ln(HC _{it})	0.11 (0.70)	0.15 (0.98)	-0.04 (0.29)	0.09 (0.57)
ln(SMALL _{it})	0.95** (2.90)	1.03** (3.03)	0.87** (2.74)	0.86* (2.55)
ln(LARGE _{it})	0.17 (1.32)	0.19 (1.51)	0.18 (1.33)	0.14 (1.03)
REGTYPE _i	-0.05 (1.02)	-0.04 (0.86)	-0.08 (1.50)	-0.10* (1.98)
Adj. R ²	0.86	0.86	0.85	0.85
Observations	190	190	190	190
Moran's I	3.10**	3.09**	3.02**	3.29**
LM error	7.07**	7.06**	6.61**	8.15**
Robust LM error	1.27	1.43	0.62	0.89
LM lag	8.46**	7.83**	10.1**	13.2**
Robust LM lag	2.66	2.20	4.08*	5.93*

Notes: *t*-statistics are based upon White's heteroscedasticity-adjusted standard errors.

** significant at the 0.05 level, * significant at the 0.10 level.

Test on spatial autocorrelation were conducted with different weight matrices in order to check robustness. The results presented in the table are based on a binary contiguity matrix.

Table 4: Robustness analysis – Random effects and fixed effects

Dependent variable	ln(patents per capita)					
	Random effects			Fixed effects		
	I	II	III	IV	V	VI
Cons	3.46** (2.98)	3.09** (2.65)	4.49** (3.63)	5.61* (2.50)	5.68* (2.54)	6.17** (2.62)
lnRD _{it-1}	0.23** (4.18)	0.25** (4.43)	0.22** (3.92)	-0.19* (2.08)	-0.19* (2.09)	-0.18* (2.06)
ln(DIV _{it}) low	0.29** (5.24)			-0.02 (0.10)		
ln(DIV _{it}) medium		0.28** (4.55)			0.06 (0.32)	
ln(DIV _{it}) high			0.44** (5.35)			0.12 (0.70)
ln(STRUC _{it})	0.75** (6.18)	0.79** (6.39)	0.94** (7.77)	1.23 (1.95)	1.24 (1.98)	1.25* (2.00)
ln(HC _{it})	0.28 (1.45)	0.10 (0.53)	0.29 (1.47)	-0.59 (0.95)	-0.64 (1.00)	-0.52 (0.83)
ln(SMALL _{it})	0.34 (1.39)	0.22 (0.89)	0.31 (1.26)	0.08 (0.28)	0.07 (0.25)	0.08 (0.29)
ln(LARGE _{it})	0.16 (1.15)	0.16 (1.11)	0.14 (1.00)	0.82* (2.58)	0.81* (2.55)	0.81* (2.57)
REGTYPE _i	-0.05 (0.74)	-0.10 (1.43)	-0.10 (1.51)			
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.86	0.86	0.86	0.36	0.40	0.44
Observations	190	190	190	190	190	190
BP	39.9**	40.1**	43.3**			
F(94, 89)				8.81**	9.11**	9.10**

Notes: ** significant at the 0.01 level, * significant at the 0.05 level.

Table 5: Robustness analysis – Instrument variables and spatial lag

Dependent variable	ln(patents per capita)								
	IV estimation			Spatial lag model			Spatial lag model		
	2SLS			G2SLS ^{a)}			G2SLS ^{b)}		
				Binary contiguity matrix			Inverse distance, cut-off point 150 km		
	I	II	III	IV	V	VI	VII	VIII	IX
Cons	3.64** (3.05)	3.49** (2.88)	5.36** (3.86)	2.75*(2.48)	2.27* (2.08)	4.00** (3.35)	3.53** (3.06)	2.74* (2.34)	4.98** (3.57)
W_lnP _{it}				0.02** (2.89)	0.02** (3.39)	0.02** (3.96)	0.10 (0.48)	0.18 (0.95)	0.24 (1.19)
lnRD _{it-1}	0.22** (3.89)	0.24** (4.08)	0.19** (3.06)	0.26** (7.42)	0.28** (5.37)	0.22** (3.96)	0.23** (3.94)	0.29** (5.31)	0.20** (3.23)
ln(DIV _{it}) low	0.31** (4.81)			0.27** (5.34)			0.30** (4.36)		
ln(DIV _{it}) medium		0.37** (4.81)			0.28** (5.16)			0.34** (4.65)	
ln(DIV _{it}) high			0.54** (4.75)			0.42** (5.44)			0.50** (4.36)
ln(STRUC _{it})	0.75** (6.08)	0.74** (6.12)	0.96** (7.82)	0.58** (4.93)	0.57** (4.90)	0.77** (6.26)	0.73** (5.82)	0.65** (5.77)	0.91** (7.12)
ln(HC _{it})	0.32 (1.56)	0.20 (1.01)	0.42 (1.92)	0.23 (1.27)	0.06 (0.38)	0.26 (1.44)	0.32 (1.54)	0.15 (0.82)	0.41 (1.86)
ln(SMALL _{it})	0.34 (1.39)	0.23 (0.90)	0.30 (1.24)	0.46 (1.85)	0.37 (1.48)	0.33 (1.36)	0.36 (1.46)	0.41 (1.51)	0.36 (1.45)
ln(LARGE _{it})	0.14 (1.00)	0.06 (0.44)	0.07 (0.47)	0.16 (1.28)	0.14 (1.07)	0.14 (1.08)	0.14 (0.99)	0.05 (0.39)	0.07 (0.47)
REGTYPE _i	-0.04 (0.60)	-0.07 (1.02)	-0.08 (1.17)	-0.04 (0.65)	-0.07 (1.24)	-0.09 (1.35)	-0.04 (0.50)	-0.05 (0.86)	-0.06 (0.86)
Random effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.86	0.86	0.86	0.87	0.88	0.87	0.86	0.87	0.87
Observations	190	190	190	190	190	190	190	190	190

Notes: ** significant at the 0.01 level, * significant at the 0.05 level.

a) Spatial lag of patents per capita instrumented.

b) Spatial lag of patents per capita and diversity measures instrumented.

Table 6: Robustness analysis – Quantile regressions

	10th	25th	50th	75th	90th
$\ln(DIV_{it})$ low	0.34** (4.95)	0.34** (12.0)	0.31** (6.89)	0.25** (5.26)	0.15 (0.96)
$\ln(DIV_{it})$ medium	0.38** (5.26)	0.34** (7.53)	0.28** (4.93)	0.22** (2.85)	0.19 (1.06)
$\ln(DIV_{it})$ high	0.43** (3.58)	0.28** (3.26)	0.40** (3.75)	0.45** (7.32)	0.35 (1.66)

Notes: ** significant at the 0.01 level, * significant at the 0.05 level.
t-ratios in parentheses are based on standard errors bootstrapped with 200 replications.