Many names, many gains?
How regional diversity in Germany affects innovation

Anna Kremer[*]

Working Paper
June 28, 2024

Abstract

Meeting others with different backgrounds brings up new ideas. In this paper I show that this is relevant not only with respect to heterogeneous industries or nationalities but that regional differences matter too. Regions in a country vary in their traditions and culture. Cultural homogeneity within regions is mixed up by internal migration, that, like international migration, increases diversity of a place. In a novel approach I look at diversity in German municipality associations measured by different family names and investigate the effect it has on the number of generated patents. I use a unique data set from a phone book in 1996 and casualty lists from WWI. I find a significant positive relationship between diversity and generated patents and establish causality by using instrumental variables estimations with proximity to borders and building on the French occupation zone limiting the inflow of refugees. I show that diversity and openness of a place affect its economic performance positively in terms of innovation when referring to intra-country differences at the local level.

Keywords: cultural diversity, innovation, openness, phone book, patents, local level, Germany

Introduction

Family names tell, to some degree, a story about an individual's background. "Foreign" names point to a family immigration history more recently or long ago. "Domestic" names give hints, too, about where a family originally stems from because many names are specific for a region. When individuals or families move they also bring their names to a new place. Thus, earlier migration is reflected in the composition of family names to some degree. Migration itself is an important determinant of diversity because it increases the cultural heterogeneity of a place when migrants keep (parts of) their culture. This is illustrated vividly by the example of international migration. Earlier research has shown that diversity, measured by industry heterogeneity or on an inter-country level by different nationalities or places of birth, positively affects innovativeness and productivity of an economy. However, within a country cultural disparities arise, too. Usually they stem from a different history and tradition of the regions but also have to do with the natural environment, the economy and with the different upbringing of families. I hypothesize that not only inter-country differences are relevant for cultural diversity but that analogously also these intra-country, regional disparities advance innovativeness. To

[*]Technical University Dresden and ifo Institute Dresden, Einsteinstr. 3, 01069 Dresden, kremer@ifo.de

measure diversity within a country I make use of the spread of family names and prove my hypothesis by the example of Germany.

The distribution of family names proxies for family distribution (in the male line) in Germany quite well. This is due to the legal situation until 1976 that determined the last name of the husband as the family name. The legal circumstances changed; however, a large majority of couples still choose the husband's last name when they marry (only 3% decided for the wife's but 80% for the husband's name in 1996;[1] Rosar, 2020). How families historically spread across Germany is therefore reflected by their names in a phone book, which I use to account for diversity. A drawback of using a phone book to measure diversity stems from having common names that originated at different places and thus do not necessarily relate to a geographical distribution of the name (compare e.g. "Smith"). Still, many names are peculiar or predominant in a specific region and can account for movements of families. I use this insight and a data set containing all phone book entries from 1996 to determine where in Germany diversity of family names is high or low. For this I use a measure of deconcentration, a fractionalization and a Shannon index. I also regard the openness of municipalities throughout the past 100 years by comparing family names at places from the casualty list in World War I and the phone book entries from 1996.

The literature suggests that "interactions between diverse people [...] increases economic performance" (Möhlmann and Bakens, 2015, p. 235). This is based on the idea of Jacobs externalities – spillovers from firm and sector diversity (Jacobs, 1969; proven by Glaeser et al., 1992; Henderson, 1997) and also holds true for cultural diversity. Ottaviano and Peri (2006) provide evidence for this by showing that native workers have higher wages and rents in places where the share of foreign-born people is higher. Focusing on Germany, Niebuhr (2010) looks at a cross-section of German regions to study the effect of cultural diversity of the labour force on patent applications. She confirms that a positive spillover effect outweighs the costs of diversity. However, most studies focus on the effect of cultural diversity measured solely by nationalities.

I add to the literature by looking at whether diversity from intra-country differences, controlled for by family names, affects economic performance. This micro level has not been investigated yet but the reasoning is analogue to that of diversity in nationalities. In my work I estimate the effect of local diversity at the level of municipality associations, measured by (the change in) family names, on innovativeness, measured by patents. A higher amount of different family names is expected to be connected to a more dynamic composition of the municipality associations throughout history and thus a higher diversity. With this I build on the reasoning that relocation of people induces cultural diversity (Bakens et al., 2015) because culture varies over space and even small alterations, such as within Germany, shape people differently. From these different interacting backgrounds innovation is expected to arise. Additionally, firms in a more diverse environment have the possibility to choose the most capable employees from a larger pool which might also increase innovativeness.

My two hypotheses are:

- More diverse municipalities, in terms of family names, have a higher innovativeness.
- More open municipalities throughout the past 100 years, in terms of a changing pattern of family names, have a higher innovativeness.

The next section gives an overview on the existing work on how cultural diversity, measured at the inter-country level, affects economic performance. I then turn to explaining my empirical method by commenting on the theoretical model, the derived empirical regression, the measure of diversity and the data. Next, I explain my estimation and findings and consider the question of causality for which I employ proximity to country borders and terrain ruggedness as instruments for diversity (via remoteness and migration). Lastly, I perform some robustness checks before concluding.

[1] 17% kept their individual last names.

Literature review

According to Jacobs (1969), diversity helps to generate new knowledge by bringing together different proficiencies and exchanging ideas. The idea that diversity produces spillovers and then innovation, originally focuses on how different firms and sectors interact. Jacobs' theory is confirmed by the work of Glaeser et al. (1992) who show that it is rather spillovers from diversity than externalities from competition and specialization within an industry that advance innovation in agglomerations. For this, they relate cities' industry diversity and growth in the USA between 1956 and 1987. Likewise, Henderson (1997) shows for the US-American urban counties between 1977 and 1987 that diversity externalities between industries have a longer lasting effect on employment and seem to be dynamic.

After Jacobs' theory was proven successful with respect to industrial diversity, it also translated to a cultural definition of diversity: Culture determines how we live together and is defined as comprising features that are relatively homogeneous within but vary between groups. These include a common value system, historical traditions, religion and lifestyles (Bakens et al., 2015). Cultural groups are often delimited by geography. Thus, spatial distribution of people correlates with cultural dispersion to some extent. However, this relationship is not set in stone. Migration mixes it up and increases cultural diversity, when people relocate and keep (aspects of) their original culture (Bakens et al., 2015). This happens both on a inter- and intra-country level. The resulting cultural diversity can have a positive or negative effect on innovation and productivity. While heterogeneous workers might have to deal with communication barriers they can also exchange ideas stemming from their different backgrounds, generate knowledge and improve production.

Ottaviano and Peri (2006) provide evidence for a positive effect of cultural diversity on productivity by showing that native workers have higher wages and rents in places where the share of foreign-born is higher. They study 160 metropolitan areas in the USA between 1970-1990 and distinguish between places of birth for cultural diversity. With an IV method they aim to solve the endogeneity that migrants tend to go where industries and cities boom. They use the predicted change in national groups for each of the cities as an instrument, relying on the observation that nationalities tend to flock together. Their results are robust. Looking at a more recent time period Cooke and Kemeny (2017) confirm that diversity of birthplace is responsible for improvements in productivity in the USA between 1991 and 2008. They also show that more complex problem solving is enriched by diversity. Bellini et al. (2013) follow Ottaviano and Peri in their approach and regard how diversity, defined by country of birth, affects productivity in the NUTS3 regions of 12 European countries. They confirm a positive correlation and find further evidence for causation, too. Möhlmann and Bakens (2015) investigate on the same relationship with a two-step method. For firms in the Netherlands they firstly estimate total factor productivity by measuring the deviation to sector productivity. In a second step they investigate whether more diverse firms are more productive. Their results show that a higher share of foreigners in a firm has a negative effect on productivity. Diversity among these employees however is positively correlated to firm performance. The effect vanishes when including firm fixed effects. Ozgen et al. (2014) also look at firm level data from the Netherlands and Germany. They establish a connection between (parental) place of birth of employees and the number of new products in a firm. Their results suggest a modest positive influence of cultural diversity on innovativeness. They also subsume evidence from other studies in Europe, North America and New Zealand extensively. In this they emphasize the positive potential (productivity and innovation) and negative effects (communication barriers) of cultural diversity.

Regarding employers more closely, Nathan and Lee (2013) study 7600 firms in London and compare their cultural diversity with innovation, entrepreneurship and whereto sales are directed. They rely on firm micro data to define diversity as whether owners and partners are migrants or non-(white-)British. The authors find in their regression a small but significant diversity bonus: A higher diversity makes firms more innovative, they reach out better to international markets and migrants are more likely to be entrepreneurs. Evidence for spillovers from cultural diversity is also found in papers dealing with business management (Roberge and van Dick, 2010; Van Beers and Zand, 2014).

Focusing on Germany, Niebuhr (2010) estimated a cross-section of German regions to prove the effect of cultural diversity of the labour force on patent applications. She confirms that a positive spillover effect outweighs the costs of diversity. With respect to the endogeneity problem she argues that migration is often motivated by earlier migration from the same origin and thus uses latitude of region centre and lagged diversity of neighbouring regions as instruments in an IV regression. This approach is similar to Ottaviano and Peri (2006). Brunow and Stockinger (2015) also confirm the effect of cultural diversity on innovation at the firm level in Germany. They account for diversity by different nationalities while their dependent variables reflects different innovation outcomes (e.g. improvement or introduction of products).

Literature that specifically looks at how openness of societies affects innovation is scarce. However, the reasoning builds on the same theory and effects as for diversity. Moreover, openness and diversity are closely entangled because, expectedly, more open societies become more diverse.

Taken together, the literature emphasizes that cultural diversity has advantages (innovation and productivity) as well as disadvantages (communication barriers) for economic performance. Yet, it shows that more often than not, innovation is positively affected by it. I examine whether this also holds true on the municipality level in Germany. In contrast to existing work, I proxy for (intra-country) diversity by family names and their dispersion over time. A higher amount of different family names is expected to be connected to a more dynamic place. In this I build on the reasoning that relocation of people induces cultural diversity because culture varies over space and even little deviations, such as within Germany, shape people differently. From these differences, innovation is expected to arise. This might be due to people with different background sharing ideas or stems from having a larger pool of capable individuals from which firms can choose.

In the next section I turn to the empirical method.

Method

Model

To take potential costs and advantages of diversity for innovativeness into account I rely on the model of Niebuhr (2010). It expects that excessive differences hinder exchange (e.g. via language barriers, misunderstandings are also possible with the same mother tongue) and thus innovation. Advantages on the other hand could arise from diversity via two channels: Firstly, if more diverse people meet, interactions are expected to be more productive and generate new knowledge. Secondly, from having a larger pool of more diverse people firms can choose the most capable employees who then generate more innovation.

$$Y_i = [1 - \tau(div_i)]^\alpha \sum_{o=1}^{n} (L_{oi})^\alpha$$

The common production function in place i is augmented by the effect of cultural diversity. Diversity div from culturally different origins, noted by o, increase output for a fixed number of workers and is defined as $div_i = \sum_{o=1}^{n} (L_{oi}/L_i)^\alpha$. With $\alpha \in (0,1)$ diversity increases (for a fixed n and L_i) if large groups are getting smaller and therefore groups are more evenly distributed. The costs of differences (τ), e.g. communication barriers, are an increasing function of diversity div and thus reduce the advantages of it, modeled by effectively limiting the number of inputs.

The elasticity of substitution between different origins is $\sigma = 1/(1-\alpha)$.

Whether (dis)advantages of cultural diversity prevail depends on this elasticity of substitution and on the costs of diversity as can be seen when looking at the marginal effect of diversity:

$$\frac{\delta Y_i}{\delta div_i} = L_i^{\alpha}\left[[1-\tau(div_i)]^{\alpha} - \alpha[1-\tau(div_i)]^{\alpha-1} * \frac{\delta\tau}{\delta div_i} * div_i\right]$$

Applying the diversity more concretely to intra-country diversity measured by last names $Div_i = \sum_{o=1}^{\bar{n}}(names_{oi}/entries_i)^{\alpha}$ is obtained with $\bar{n}$ being the number of different family names, $names_{oi}$ being number of people sharing family name o in i and $entries_i$ being the number of inhabitants (or phonebook entries) in i.

When the production function is multiplied by $(L_i/L_i)^{\alpha}$ to prevent from a merely quantitative effect of increasing labor supply, it can be rewritten as:

$$Y_i = L_i^{\alpha} div_i[1-\tau(div_i)]^{\alpha}$$

The derived basic regression model to determine the overall effect of diversity on innovativeness is

$$I_i = \beta_0 + \beta_1 * Div_i + \beta_2 * L + \beta_3 * C_i + \epsilon$$

The equation defines innovativeness as patents per employee and sets it in relation to diversity. It moreover refers to other differences of employees in L, here proxied by skill level and different national backgrounds. Additionally, it controls for urbanization (C).

There is a potential humb-shaped effect of diversity. I therefore control in one specification additionally for the square of diversity (Div_i^2) to see whether there is an optimal degree of diversity with respect to benefits and costs of it.

Measures of diversity

Within a country cultural disparities arise. Usually they stem from a different history and tradition of the regions but also from a dissimilar natural environment, economy, education and more. Additionally people are raised differently in different families. From these differences between people who meet innovation is expected to spark. This argumentation is analogue to that of Ottaviano and Peri (2006) who show that diversity in places of birth fosters productivity. The distribution of family names proxies for family distribution in Germany quite well. Families could not decide up until 1976 upon what name to choose as family name and even after that a large majority kept choosing the last name of the husband. How families historically spread across Germany in the male line is therefore reflected by the names in a phone book, which I use to account for diversity. This is pictured by the distribution of the family name "Wethmar" in Figure 1. The name originated in the region of Westfalia and many descendants still live around that area. Some however, have moved further and increased diversity elsewhere.

A drawback of using a phone book to measure diversity stems from having common names that originated at different places and thus do not necessarily relate to a geographical distribution of the name (compare e.g. "Müller" in Figure 1). Still, I consider these to be randomly enough distributed and many names in Germany are peculiar or predominant in a specific region and can account for movements of families.

To measure diversity I use three definitions. Most of the literature looking at diversity and innovation employs a fractionalisation or entropy index to account for birthplace diversity. These indices account not only for the number of different groups but additionally for the evenness of distribution of groups. The representation of groups is uneven if only one or two groups prevail while many more are present as (small) minorities. It is even if all groups are about equal in size. Most studies in the context of diversity and innovation employ these measures with respect to different nationalities, but Posch et al. (2023) also use a Shannon index for family names.

I make use of a fractionalisation index (1-Herfindahl index), relatively close to *div* in the model, that refers to the probability of two individuals, randomly drawn from the sample, to

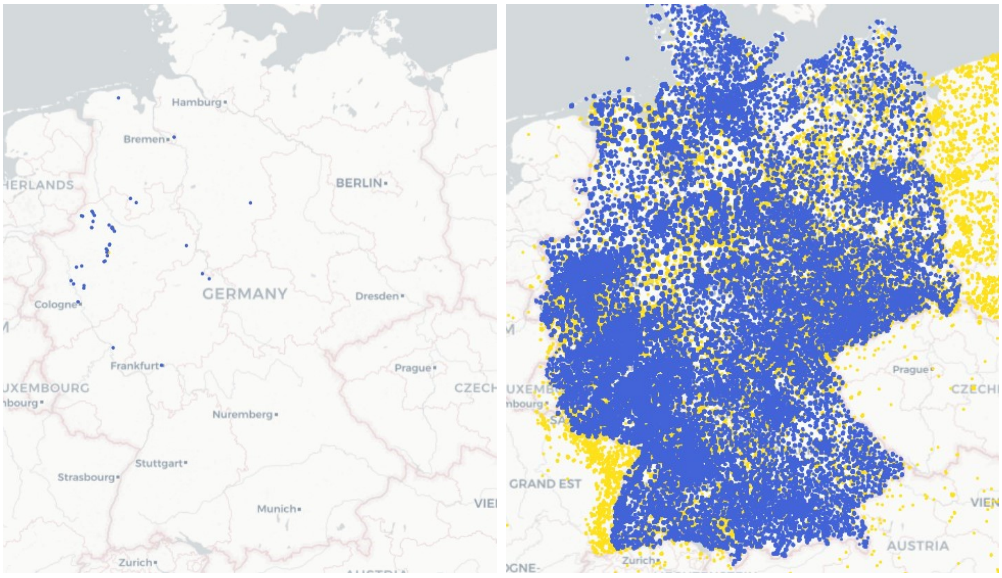

Figure 1: Distribution of family name "Wethmar" (left) and "Müller" (right). Yellow points refer to the distribution in 1890, blue ones to 1996.

share their family name n (in line with Ottaviano and Peri, 2006 and Niebuhr, 2010). This index puts relatively much weight to the largest group. It grows close to 1 when there is a large number of different names that are more or less evenly distributed (shared by the same amount of people in a municipality).

$$Fractionalization_i = 1 - [\sum_{o=1}^{\bar{n}} (names_{oi}/entries_i)^2]$$

Moreover I employ a Shannon index that also Posch et al. (2023) use. It puts more weight on rare family names compared to the fractionalisation index and also grows with a higher diversity.

$$Shannon_i = -[\sum_{o=1}^{\bar{n}} (names_{oi}/entries_i) * \log_2(names_{oi}/entries_i)]$$

Because family names, when compared to nationalities, define many more and much smaller groups that are usually all "minorities", the distributions deviate I also employ a more straightforward measure: I regard (de)concentration of family names. A comparison of the number of different family names with the total number of entries in the phone book in a municipality reveals how diverse a place is in terms of family names. This also relates to N_i in the theoretical model. I thus calculate for every place i

$$Deconcentration_i = \bar{n}_i/entries_i$$

Deconcentration varies between 0 and 1, a higher number referring to more diverse places. This measure however might suffer some bias in highly populated areas, because there are common family names of people who are unrelated. It is just more likely in a big city that people share family names by chance than in a small village. Thus diversity in very populated places might be underrated. I therefore run robustness regressions excluding municipalities with more than 0.75 quantile of inhabitants to make places more comparable.

Deconcentration correlates by about -0.04 with fractionalisation, with the Shannon index by -0.18. The fractionalisation and Shannon index correlate by about 0.47.

Furthermore, I make use of the distribution of family names in (approximately) 1890 to see how open places were throughout the past 100 years. For this I compare the development of the diversity measures between 1890 and 1996 and refer to this measure as openness. I construct openness for all three diversity measures. A positive openness implies a higher diversity in 1996 than in 1890.

$$Openness_i = Diversity\ measure_{i,1996} - Diversity\ measure_{i,1890}$$

Data

The distribution of names is provided by CompGen (Verein für Computergenealogie e.V.). The first dataset is based on casualty lists from WWI and proxies for the time about 1890 as it provides birth places of those that died. The second draws from a digitalized phone book that provides the distribution of names in 1996. I match geocoded names with today's municipality boundaries to have time-consistent areas. The data was georeferenced by a combination of postal codes and location names.[2]

Because private phone coverage in East Germany was only really advancing after reunification but had not finished in 1995 (Didczuneit, 2015) I restrict my sample to West Germany where the true distribution of names is expected to be reliably identified. This restricts external validity of the outcomes to West Germany. The total number of entries in West Germany is 26 millions with 1,123,716 different family names. This means that every person shares on average with 22 other their last name. In every place about 16,547 (mean of subsamples) different family names appear and every name accounts for 2.45 entries on average. Diversity, measured as (de)concentration, has a mean value of 0.5 (sd 0.1) names per inhabitant and is approximately normally distributed. The fractionalisation index ranges around .997 (sd 0.005) in a strongly right skewed distribution. The Shannon index has a mean of 10 (sd 1.3) and is approximately normally distributed.

Diversity of the municipality associations in 1996 and 1890 is displayed in Figures 2, 3 and 4.[3]

As for innovation, patent data is available from Deutsches Patent- und Markenamt and refers to the address of the innovator. Figure 5 shows the distribution of patents in West Germany averaged over the years 1995-1999 and for 1890.

I identify places of innovation from the post codes of the patent inventors and times by the year of the effective European filing date of the innovation. The average number of patents per employee over the years 1995-1999 is 0.007 with a standard deviation of 0.026. The distribution is heavily left skewed. For patent data at the level of planning regions I rely on the OECD REGPAT database. However, data at a finer spatial scale is not available from there.

Data on control variables related to the regional employment structure stems from the Establishment History Panel of IAB (Ganzer et al., 2021). Skill level, share of foreign-born employees, participation of women and age groups are used as controls. Due to data censoring I lose around 6% of the observations. The sample of the baseline regression is seen in Figure 6. For the spatial structure, information on the urban type is gathered from the INKAR database (Bundesinstitut für Bau- Stadt- und Raumforschung (BBSR)). The reference category are large cities. I further lose 1% of the observations because this data source does not refer to the exact same geographical entities.

Descriptive statistics are in Table 1.

Moreover I rely on historical data taken from several sources. Patent data from the 19th century stems from Bergeaud and Verluise (2021) and De Rassenfosse et al. (2019), population data from Roesel (2022), information on universities from Fritsch and Wyrwich (2018) and data on control variables for Prussia in the 19th century from Becker et al. (2014). Descriptive statistics are in Table 2.

[2]The geocoding refers to post code and place name and is thus not entirely but acceptably enough accurate.

[3]White spaces mark the selection bias.

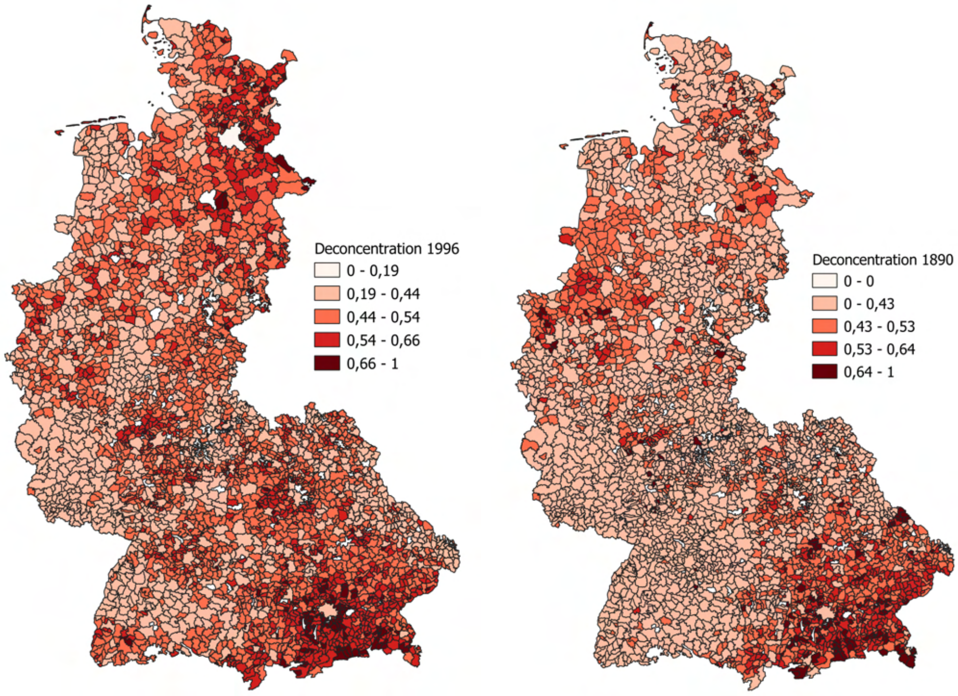

Figure 2: Deconcentration of family names

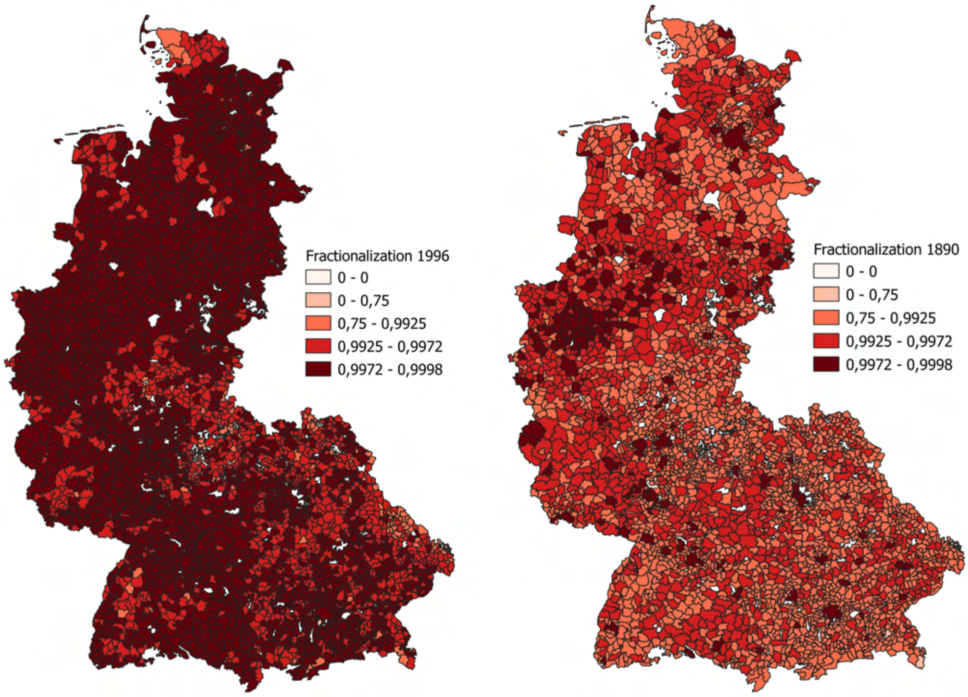

Figure 3: Fractionalization of family names

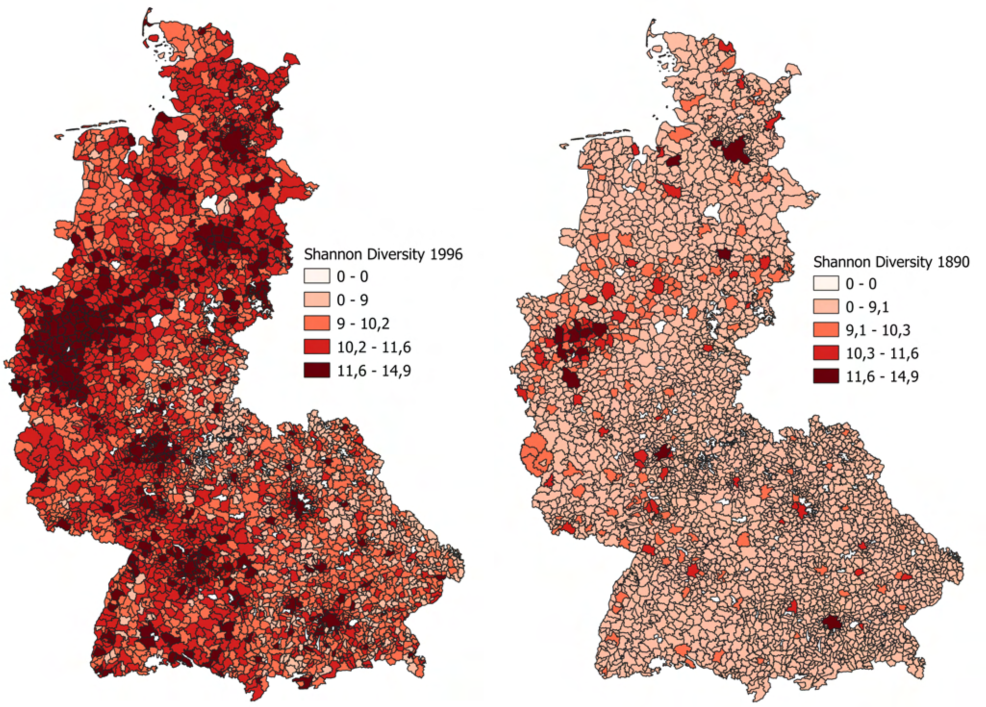

Figure 4: Shannon diversity of family names

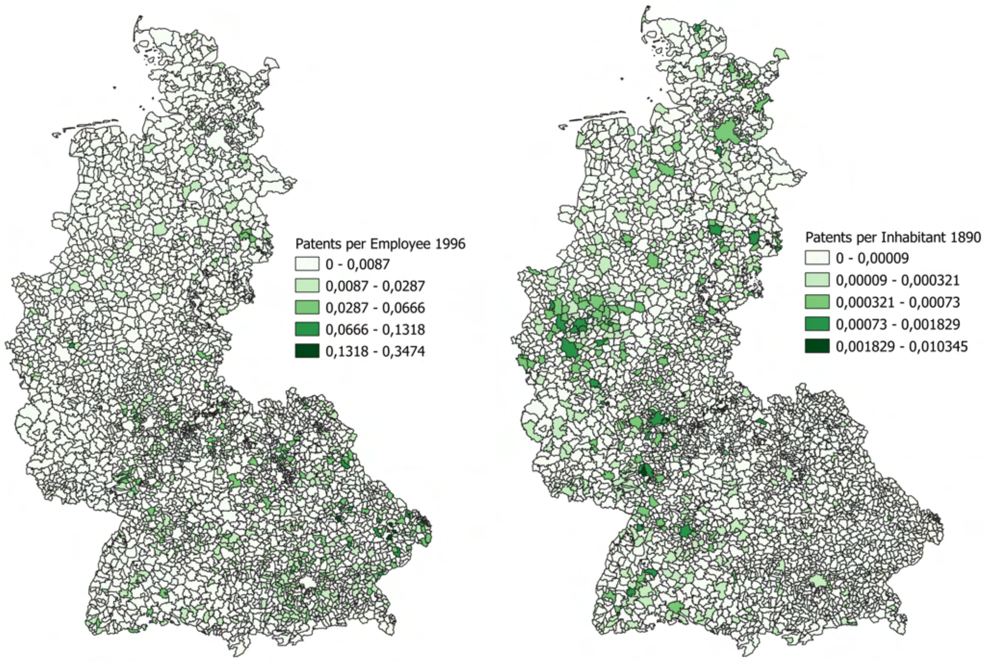

Figure 5: Patents per employee or inhabitant

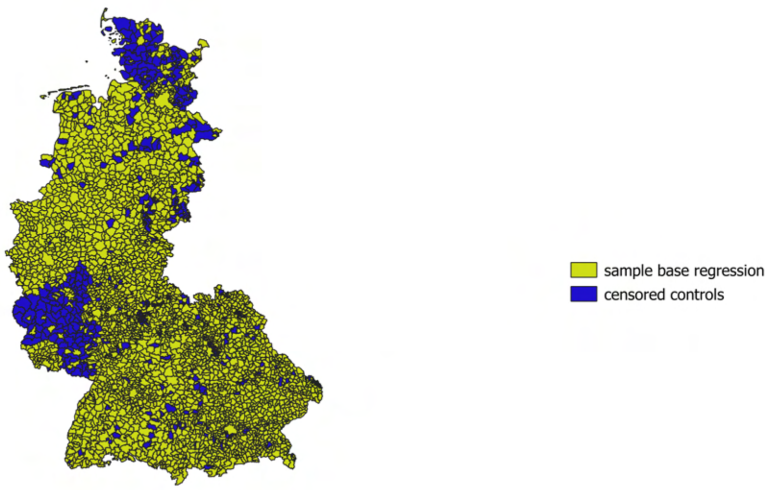

Figure 6: Sample with selection bias due to data censoring

Table 1: Descriptive statistics 1996

	count	mean	sd	min	max
Patents per employee	3161	.0065789	.0141616	0	.3474458
Deconcentration	3412	.4979376	.0949066	.1853195	1
Fractionalisation	3412	.9973773	.0048682	.75	.9997547
Shannon Diversity	3412	10.07775	1.276423	2	14.88683
Large Cities	69				
Medium sized cities	519				
Larger Towns	717				
Smaller Towns	932				
Rural areas	1160				
Share foreigners	3161	.0741151	.0476972	0	.3709796
Share high-skilled	3161	.0514364	.0358519	0	.527629
Share low-skilled	3161	.1782873	.0445727	.0513941	.4321005

Table 2: Descriptive statistics 19th century

	count	mean	sd	min	max
Patents per inhabitant in 1890	312	.0000294	.0000648	1.19e-06	.0007593
Deconcentration in 1890	312	.2933565	.0606719	.1639476	.6254072
Fractionalisation in 1890	312	.9973414	.0021572	.9799857	.9996426
Shannon diversity in 1890	312	10.04261	.8398909	6.968504	12.76263
Share of foreigners in 1871	312	.0079508	.013886	.0000477	.1476098
Share of illiterates in 1871	312	.0948179	.0943168	.006628	.5567229
Share of literates in 1871	312	.8889936	.0967909	.4265945	.9850926
Density in 1890	312	120.411	747.7274	.6551396	9870.489
Share of secondary sector in 1882	312	.1157501	.047821	.034972	.2789766
Share of self-employed in 1882	312	.1202061	.029809	.0598768	.2234438
Share of science-based professionals in 1882	312	.005302	.0047106	.0010704	.0414322

As the spatial unit I build on municipalities associations ("Gemeindeverbände") because they are more comparable across Germany than the more unevenly sized municipalities but still refer to a fine spatial level.

I drop the area of Grafling in Bavaria because it shows unreasonably three times higher numbers for patents per employees in the observation years. Moreover I drop Asperg and Reinhardwald because they both are outliers with respect to the fractionalization index. When looking at labour market regions I drop Kempten because patents per employees are 15 times higher than in the next productive region.

I estimate OLS regressions with patents transformed by inverse hyperbolic sine (IHS) and Poisson regressions to account for zero patents. In a robustness check I also use a SLX model. Because I estimate a cross-section but have panel data on patents and control variables in the estimations, I use averages of the latter to prevent measurement biases. All variables correlate by less than 50%, thus I expect no problematic multicollinearity.

Results

Diversity in 1996

Results of the base regression, connecting local diversity in 1996 with patents at that time, estimated by OLS, are found in Table 3. They are corroborated by Poisson estimations in Table 4.

There is a positive relation between all diversity measures and innovation. The association between deconcentration and innovation is significant. I can thus confirm my first hypothesis of a positive correlation between more diverse municipalities, in terms of family names, and a higher innovativeness. This is in line with the findings of Posch et al. (2023) who regard family names as diversity measures between 1850 and 1940 in the US and their effect on innovation. An increase of 10% in deconcentration of family names relates to around 20% more patents.

The controls point towards a peculiar, but German typical, result that less dense areas are more innovative than larger cities (see also Fritsch and Wyrwich, 2021).[4] While share of low-skilled employees has a negative effect in all estimations, the correlation between high-skilled employees and patents per employees remains unclear. The estimations moreover confirm a strong positive effect of the share of foreign-born employees for innovation and thus affirm the earlier literature on diversity referring to nationalities.

[4] The reference category for spatial type are large cities.

Table 3: OLS Estimation results (1) Deconcentration, (2) Fractionalisation, (3) Shannon index

	(1) Patents per employee (IHS)	(2) Patents per employee (IHS)	(3) Patents per employee (IHS)
Deconcentration	0.0113***		
	(0.00346)		
Fractionalisation		0.208	
		(0.164)	
Shannon Diversity			0.000146
			(0.000425)
Medium sized cities	0.000271	0.00199***	0.00220***
	(0.000625)	(0.000356)	(0.000783)
Larger Towns	0.00131	0.00378***	0.00403***
	(0.000817)	(0.000496)	(0.00125)
Smaller Towns	0.00289***	0.00583***	0.00605***
	(0.000964)	(0.000637)	(0.00156)
Rural areas	0.00610***	0.00984***	0.00983***
	(0.00110)	(0.000894)	(0.00170)
Share foreigners	0.0210***	0.0277***	0.0283***
	(0.00596)	(0.00537)	(0.00552)
Share high-skilled	-0.00275	0.000889	0.00165
	(0.00729)	(0.00698)	(0.00687)
Share low-skilled	-0.0149*	-0.0203**	-0.0212**
	(0.00876)	(0.00913)	(0.00937)
Constant	-0.000908	-0.205	0.000640
	(0.00174)	(0.165)	(0.00637)
Observations	3160	3160	3160
Adjusted R^2	0.039	0.037	0.036

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Poisson Estimation results (1) Deconcentration, (2) Fractionalisation, (3) Shannon index

	(1)	(2)	(3)
	patentssumperwork	patentssumperwork	patentssumperwork
patentssumperwork			
Deconcentration	1.502***		
	(0.457)		
Fractionalisation		23.45	
		(22.30)	
Shannon Diversity			0.0266
			(0.0601)
Medium sized cities	0.357***	0.593***	0.635***
	(0.123)	(0.102)	(0.144)
Larger Towns	0.643***	0.978***	1.035***
	(0.139)	(0.110)	(0.201)
Smaller Towns	0.919***	1.314***	1.376***
	(0.154)	(0.117)	(0.239)
Rural areas	1.326***	1.823***	1.870***
	(0.171)	(0.128)	(0.262)
Share foreigners	2.800***	3.741***	3.803***
	(0.783)	(0.677)	(0.685)
Share high-skilled	-0.0960	0.410	0.481
	(0.908)	(0.850)	(0.828)
Share low-skilled	-1.917	-2.724**	-2.804**
	(1.214)	(1.267)	(1.286)
Constant	-6.611***	-29.58	-6.498***
	(0.257)	(22.31)	(0.902)
Observations	3160	3160	3160
Adjusted R^2			

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Diversity in the 19th century

I also investigate the relationship between innovativeness and diversity in places in the late 19th century. For this I relate patent number per inhabitant in 1890 (extrapolated from the sum of patents between 1885 and 1895) to the distribution of family names around the same time. Since the geographic distribution is proxied for by family names and places of birth from casualty lists of World War I this is rather an approximation with respect to time. As control variables I use similar data as in the baseline regression: Share of non-Prussians replaces share of foreigners, share of illiterates proxies for low-skilled labor, literates proxy for high-skilled workers and density measures urbanisation of counties. I moreover control for the share of secondary sector, the share of self-employed and share of professionals in science-based industries (engineering and chemical industry). Data is available at the county level but only for the year 1871.[5] The Poisson regressions results are in Table 5 and show a positive relationship between diversity and innovation in the late 19th century that is even significant for the measures of deconcentration and the Shannon index. Despite the larger set of control variables the OLS regression suffers from an omitted variable bias.

Openness

Moreover I regard how openness is related to patents in 1996. Results are found in Table 6. I find a positive relationship for all three specifications that is only significant for the Shannon index. I thus cannot reject my second hypothesis that more open municipalities throughout the past 100 years, in terms of a changing pattern of family names, have a higher innovativeness. Poisson regressions give the same results. This shows that regions that had a higher inflow of migrants, and family names, are generating more patents today. This result, however, might be especially threatened by reverse causality that more innovative regions attracted more people from elsewhere.

Endogeneity

The association between diversity and innovativeness has a heavy threat of reverse causality. It seems reasonable that not only diversity enhances innovation but also that more people migrate to and thereby increase diversity of places with a high innovativeness. In these more productive places e.g. higher wages as a pull factor can be expected. I am trying to investigate on the mechanism by looking at instrumental variables regressions with two different instruments: the German country borders and the French occupation zone after World War II.

French occupation zone

As a first instrument I make use of the French occupation zone in Germany after World War II. This bases on the insight that 12 million displaced people from the formely Eastern regions of Germany that then belonged to Poland fled to the Soviet, British and American occupation zone but were not admitted to the French zone between August 1945 and 1949. Schumann (2014) shows that this significantly reduced flows to the French region in South Germany and that this affected later population patterns.

I find negative significant first stage regressions in Tables 21, 22, 23. These corroborate the findings of Schumann (2014) that less migration happened towards areas of the French occupation zone during a time of many resettlements. Less people moved into the regions and thus diversity is reduced. The second stage results show positive but insignificant effects of diversity on innovation. Notice however that the French occupation zone was the smallest in Germany and the sample of the French occupation zone is further reduced because of data limitations in this study (see Figure 6). Therefore, estimates are not as precise with the lower

[5] Again, this an approximation with respect to time but people were less mobile in the 19th century.

Table 5: Poisson Estimation results (1) Deconcentration and (2) Fractionalization and (3)Shannon diversity

	(1) patents1890perinhab	(2) patents1890perinhab	(3) patents1890perinha
patents1890perinhab			
Deconcentration in 1890	3.544***		
	(1.120)		
Fractionalisation in 1890		193.5	
		(148.0)	
Shannon diversity in 1890			0.549***
			(0.140)
Share of foreigners in 1871	0.00293	3.568	4.549
	(6.078)	(6.100)	(5.495)
Share of illiterates in 1871	-2.121	-5.382	-4.393
	(9.049)	(8.534)	(8.359)
Share of literates in 1871	2.261	-0.125	1.065
	(8.655)	(8.270)	(8.281)
Density in 1890	0.000228***	0.000205***	0.000133***
	(0.0000251)	(0.0000309)	(0.0000337)
Share of secondary sector in 1882	13.06***	11.01***	8.851***
	(2.798)	(2.420)	(1.839)
Share of self-employed in 1882	0.750	3.211	5.327
	(3.228)	(3.651)	(3.298)
Share of science-based professionals in 1882	37.27***	39.16***	44.08***
	(10.81)	(10.86)	(9.531)
Constant	-15.59*	-205.3	-18.96**
	(8.967)	(150.4)	(9.123)
Observations	312	312	312
Adjusted R^2			

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: OLS Estimation results (1) Deconcentration, (2) Fractionalisation, (3) Shannon index

	(1) Patents per employee (IHS)	(2) Patents per employee (IHS)	(3) Patents per employee (IHS)
Openness Deconcentration	0.00216		
	(0.00268)		
Openness Fractionalization		0.00660	
		(0.00714)	
Openness Shannon index			0.000577**
			(0.000259)
Medium sized cities	0.00182***	0.00187***	0.00124***
	(0.000394)	(0.000362)	(0.000463)
Larger Towns	0.00341***	0.00349***	0.00295***
	(0.000561)	(0.000483)	(0.000523)
Smaller Towns	0.00530***	0.00537***	0.00499***
	(0.000654)	(0.000601)	(0.000600)
Rural areas	0.00895***	0.00898***	0.00879***
	(0.000841)	(0.000754)	(0.000746)
Share foreigners	0.0279***	0.0278***	0.0227***
	(0.00534)	(0.00552)	(0.00612)
Share high-skilled	0.00153	0.00129	-0.00185
	(0.00735)	(0.00709)	(0.00732)
Share low-skilled	-0.0212**	-0.0209**	-0.0173**
	(0.00883)	(0.00896)	(0.00867)
Constant	0.00268	0.00265	0.00134
	(0.00177)	(0.00170)	(0.00162)
Observations	3160	3160	3160
Adjusted R^2	0.036	0.036	0.038

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimation results IV (2SLS) regression, Being part of the French occupation zone as instrument

	(1) Patents per employee (IHS)	(2) Patents per employee (IHS)	(3) Patents per employee (IHS)
Deconcentration	0.0112		
	(0.00929)		
Fractionalisation		1.150	
		(0.979)	
Shannon Diversity			0.00255
			(0.00216)
Medium sized cities	-0.000399	0.000889*	0.00376
	(0.00108)	(0.000460)	(0.00258)
Larger Towns	0.000792	0.00333***	0.00812*
	(0.00159)	(0.000839)	(0.00474)
Smaller Towns	0.00196	0.00565***	0.0116*
	(0.00185)	(0.00151)	(0.00642)
Rural areas	0.00518**	0.0119***	0.0179**
	(0.00241)	(0.00356)	(0.00860)
Share foreigners	0.0159*	0.0176**	0.0168**
	(0.00887)	(0.00802)	(0.00846)
Share high-skilled	0.00102	-0.000434	-0.00232
	(0.00896)	(0.00985)	(0.0109)
Share low-skilled	-0.0110	-0.00921	-0.0103
	(0.0130)	(0.0140)	(0.0134)
(sum) plants	-8.37e-08	-0.000000179**	-0.000000367*
	(0.000000100)	(7.70e-08)	(0.000000197)
Constant	-0.000844	-1.146	-0.0294
	(0.00413)	(0.978)	(0.0280)
Observations	2479	2479	2479
Adjusted R^2	0.034	0.015	0.018

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

observation number. Still, the results however suggest that more locally diverse places are indeed more innovative.

Country borders

In order to estimate the effect more precisely, I apply a second instrument. I use distance to Germany's country borders as another instrument for diversity. As has been shown by Redding and Sturm (2008) market access matters also for population movements: Cities in West Germany located near the inner-German border experienced a significant drop in population growth during the division period. A similar, but most likely smaller, effect should hold true for all border regions. This market access effect is the first reason why less people moving towards border regions are expected and hence a lower diversity. Secondly, there is a strong effect of distance determining migration, already found by Ravenstein (1885) and confirmed for German internal migration by Parikh et al. (2003). Because border regions are geographically close to fewer regions this should also affect migration adversely. Again the inflow of people should be lower in border regions. I argue that therefore diversity should be diminished close to the German borders.

I am using dummies for 25 km, 50 km and 75 km distance to the (former) FRG[6] country border as an instrument.[7] The first stage regressions in Tables 24, 25, 26 show a negative relationship between the border dummies and diversity, being largest closest to the border.[8] The results of the second stage regressions are in Table 8 and show significant and positive effects. Because the limited market access implies also a lower number of firms that are able to innovate I control for number of plants.[9]

These results show that there is a causal effect of diversity on innovation.

Robustness

As robustness checks I perform several tests with respect to the regression specification, different definitions of diversity and geographical spillovers. All robustness checks are displayed in the OLS specification. Poisson regressions however corroborate the findings unless specified otherwise.

Different specifications

In order to prove robustness of the estimations I control for additional demographic factors of the work force in Table 11. For deconcentration and fractionalization I find a positive relation with number of generated patents. For the Shannon index it is negative in the OLS specification but positive when estimating a Poisson model. The share of women has a positive and partly even significant effect on the number of generated patents per employee. Employees between the ages of 20 and 24 years have a positive but insignificant effect.

In Table 9 I moreover control for the influence of the manufacturing sector which is expected to be responsible for a high number of patents. The sample size is severely reduced due to data censoring. A positive but insignificant relation between diversity and patents per employee is found. The effect of the share of manufacturing is insignificantly negative.

Additionally I check in Table 10 whether the effect of diversity on innovation is non-linear. Introducing the square of diversity shows that indeed the relationship has an inverted u-shape

[6] In line with Redding and Sturm (2008) only a few years after reunification there is still an adverse effect of the inner-German border expected.

[7] Redding and Sturm (2008) find a significant impact up to 75 km.

[8] Only for deconcentration there is a positive association, being numerically larger further from the border.

[9] Number of plants correlates with deconcentration by -0.29, with fractionalization by 0.19 and with the Shannon index by 0.45.

Table 8: Estimation results IV (2SLS) regression, FRG border dummies (25, 50, 75 km) as instrument

	(1)	(2)	(3)
	Patents per employee (IHS)	Patents per employee (IHS)	Patents per employee (IHS)
Deconcentration	0.0420*		
	(0.0229)		
Fractionalisation		2.992***	
		(0.783)	
Shannon Diversity			0.00563***
			(0.00166)
Medium sized cities	-0.00309	0.000898	0.00704***
	(0.00224)	(0.000577)	(0.00212)
Larger Towns	-0.00366	0.00413***	0.0141***
	(0.00344)	(0.000854)	(0.00369)
Smaller Towns	-0.00299	0.00777***	0.0196***
	(0.00412)	(0.00122)	(0.00472)
Rural areas	-0.000983	0.0167***	0.0284***
	(0.00470)	(0.00249)	(0.00613)
Share foreigners	-0.0000632	0.0136**	0.0139**
	(0.0173)	(0.00630)	(0.00616)
Share high-skilled	-0.0170	-0.0169*	-0.0198**
	(0.0149)	(0.00889)	(0.00938)
Share low-skilled	0.00306	-0.00357	-0.00817
	(0.0134)	(0.00907)	(0.00902)
(sum) plants	0.000000215	-0.000000223**	-0.000000661***
	(0.000000221)	(0.000000105)	(0.000000255)
Constant	-0.0118	-2.986***	-0.0672***
	(0.00781)	(0.782)	(0.0208)
Observations	3160	3160	3160
Adjusted R^2	0.013	-0.075	-0.017

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

for the Shannon index and deconcentration. This points towards an optimal degree of diversity for innovation that can be under- or overshot and is in line with Niebuhr (2010). For fractionalisation the outcome shows a contrary u-shaped relationship which is surprising.

Moreover, I drop all municipality associations with less than 100 phone book entries because they might be biased. Results are found in Table 12 and confirm the base findings. The same holds true if I drop places with less than 1000 phone book entries.

In Table 13 I split the sample from 1996 in order to make areas more comparable in size so especially the deconcentration measure should not be biased. The inhabitant distribution at municipal level is heavily right skewed. I therefore cut off at the 0.75 quantile (16668 inhabitants). Looking at solely the municipalities with less inhabitants confirms a significant positive significant association of deconcentration and patents. For fractionalisation and the Shannon index a positive insignificant association with innovation is found again. With a Poisson regression the Shannon estimation is insignificantly negative. In Table 14 I moreover split the sample from 1890 in order to only look at districts with less than the 0.75 quantile of inhabitants. Results are displayed for the Poisson estimation because again the OLS has a omitted variable bias. The estimations show positive and mostly significant relationships between diversity and innovation.

Distance measure of diversity

Another possibility of quantifying how diverse a place is, is by looking at the geographic family name distribution. When family names are clustered in one region but also singularly appear in another, it seems likely, that those singular people moved away and thus increased diversity in their new place. Therefore I regard the distance between all appearances of family names to the geographic mean of the same name. For each municipality association I then calculate the sum of distances relative to the number of entries in the phone book or casualty lists. When using the average distance of family names to their geographic means in 1996, I find a positive, significant association with innovativeness. Results are in Table 15. When looking at distance between names and geographic mean in 1890 a positive and significant relation of the measure with patents is found. This is estimated with a Poisson regression because the OLS suffers from OVB.

New names

As a further measure I look at the names that newly appear in a municipality in 1996 compared between to 1890. I relate these entries to the total number of entries to have a measure what share of families moved newly into a region. I find a positive significant association with patents. Results are in Table 16.

Spillovers between regions

Due to commutes between the municipalities, I also test my findings at the level of planning regions.[10]I adjust the diversity measure for this new level of unit of interest. This is due to the fact that the simple relation between number of different names to number of entries does not seem to be appropriate for such large regions. For this, consider an example: In a region there are ten municipalities. In total there are ten different family names in the area, each with ten families sharing this name. It seems to make a huge difference in diversity whether each municipality consists of a single family or whether last names are evenly distributed everywhere. People are more exposed to diversity in case of the latter because they might have less contact to other municipalities. However, the beforehand used measure of diversity would give the same diversity number for both different distributions. To prevent from this problem I take the mean of the municipality diversities in the planning region. Results are supporting my main finding

[10]Due to data availability of the RegPat data I do not estimate at labour market regions directly.

only for the specification with deconcentration. For the fractionalization and Shannon index I find a negative but insignificant association.[11] They are found in Table 17 and rely on only 37 (out of 74 West German) Planning regions due to data censoring. I explain the findings at the planning region level by the fact that I am not able to control for urban type. If I drop these controls from the base specification I get the same signs and significances.

I am moreover estimating at the district level: Here, I can further control for R& D expenditures and personnel (relating to Niebuhr (2010)). Data stems from Stifterverband. Results are in Table 18, estimated by Poisson because of OVB in the OLS specification. All relationships between diversity and innovation are insignificant. Deconcentration and fractionalization correlate negatively, the Shannon index positively. The effect of R&D expenses and personnel are insignificant but rather positive for expenses and negative for personnel.

I explain the findings at the planning region and district level that the interactions between people with different backgrounds that are responsible for increasing innovation happen at the local level.

For the same reason, because regions are not separated from each other but there are all kind of spillover effects I estimate a SLX model (Halleck Vega and Elhorst, 2015) at the municipality level in Table 19. The distance weighted model in Table 19 confirms my base findings of a positive correlation as does the adjacency weighted model in Table 20.

Conclusion

By measuring diversity with the distribution of family names in Germany I show that diversity enhances innovation also when defined as intra-country differences. The reasonening is analogue to that of knowledge spillovers between people with different nationalities. A more diverse family name distribution, in terms of deconcentration, by fractionalization indices or measured as openness by development over time, within a municipality association is correlated to a higher number of generated patents. This shows that cultural disparities within a country are relevant to economic outcomes. This result is robust with respect to several robustness tests. I confirm with two IV estimations building on geographical conditions that the association between diversity and innovativeness runs from the former to the latter.

Acknowledgements

Thanks to Annekatrin Niebuhr who invested time to discuss the research idea at the very beginning and to Jesper Zedlitz who provided me with an updated version of the distribution of family names. I am furthermore grateful to Wolfgang Dauth from IAB for data on the employment structure in the municipalities and Miquel-Àngel Garcia-López for the georeferenced maps of postal routes in Europe. I thank Georg Hirte and participants of the Brown Bag seminar at TU Dresden for their helpful comments. Moreover I am grateful to Matthias Hoelzlein and Daniel Sturm for proposing suitable instruments. Further ideas on improvements stem from the ERSA 2021, meetings of the DFG Netzwerk “Die Dynamik von Innovationssystemen”, the ifo Christmas Conference 2022, the BiB anniversary conference in 2023, the CDGE PhD workshop 2024 and the UEA meeting 2024.

I received fundings from a Cusanuswerk PhD scholarship.

[11] In the Poisson estimation the association with fractionalization is significant at 10%.

References

Bakens, J., Nijkamp, P., and Poot, J. (2015). E Pluribus Prosperitas: On cultural diversity and economic development. In Bakens, J., Nijkamp, P., and Poot, J., editors, *The Economics of Cultural Diversity*. Edward Elgar Publishing.

Becker, S. O., Cinnirella, F., Hornung, E., and Woessmann, L. (2014). ipehd—the ifo prussian economic history database. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 47(2):57–66.

Bellini, E., Ottaviano, G. I., Pinelli, D., and Prarolo, G. (2013). Cultural diversity and economic performance: Evidence from European regions. In Fischer, M. M., Thill, J.-C., van Dijk, J., and Westlund, H., editors, *Advances in Spatial Science*. Springer.

Bergeaud, A. and Verluise, C. (2021). One century of innovation in europe and the us. Technical report.

Brunow, S. and Stockinger, B. (2015). Establishments' cultural diversity and innovation: evidence from Germany. In Nijkamp, P., Poot, J., and Bakens, J., editors, *The economics of cultural diversity*. Edward Elgar Publishing.

Bundesinstitut für Bau- Stadt- und Raumforschung (BBSR) (2021). INKAR. Indikatoren und Karten zur Raum-und Siedlungsentwicklung. https://www.inkar.de/.

CompGen (Verein für Computergenealogie e.V.) (2021). Namensverbreitungskarte. https://nvk.genealogy.net/map.

Cooke, A. and Kemeny, T. (2017). Cities, immigrant diversity, and complex problem solving. *Research Policy*, 46(6):1175–1185.

De Rassenfosse, G., Kozak, J., and Seliger, F. (2019). Geocoding of worldwide patent data. *Scientific data*, 6(1):1–15.

Deutsches Patent- und Markenamt (2021). DPMA register. https://register.dpma.de/DPMAregister/pat/experte.

Didczuneit, V. (2015). "Endlich Telefon" - Der Aufbau des Telekommunikationsnetzes in den neuen Bundesländern. https://www.mfk-berlin.de/wp-content/uploads/Endlich-Telefon.pdf.

Fritsch, M. and Wyrwich, M. (2018). Regional knowledge, entrepreneurial culture, and innovative start-ups over time and space—an empirical investigation. *Small Business Economics*, 51:337–353.

Fritsch, M. and Wyrwich, M. (2021). Does successful innovation require large urban areas? germany as a counterexample. *Economic Geography*, 97(3):284–308.

Ganzer, A., Schmidtlein, L., Stegmaier, J., and Wolter, S. (2021). Establishment History Panel 1975-2019. FDZ-Datenreport. 16/2020 v2 (en). DOI: 10.5164/IAB.FDZD.2016.en.v2.

GarciaLópez, M. .-À. and ViladecansMarsal, E. (2023). The role of historic amenities in shaping cities. *Available at SSRN 4603290*.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 100(6):1126–1152.

Halleck Vega, S. and Elhorst, J. P. (2015). The slx model. *Journal of Regional Science*, 55(3):339–363.

Henderson, V. (1997). Externalities and Industrial Development. *Journal of Urban Economics*, 42(3):449–470.

Jacobs, J. (1969). *The economy of cities*. Random House, New York.

Möhlmann, J. and Bakens, J. (2015). Ethnic diversity and firm productivity in the Netherlands. In Nijkamp, P., Poot, J., and Bakens, J., editors, *The Economics of Cultural Diversity*. Edward Elgar Publishing.

Nathan, M. and Lee, N. (2013). Cultural diversity, innovation, and entrepreneurship: Firm-level evidence from London. *Economic Geography*, 89(4):367–394.

Niebuhr, A. (2010). Migration and innovation: Does cultural diversity matter for regional R & D activity? *Papers in Regional Science*, 89(3):563–585.

Ottaviano, G. I. and Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, 6(1):9–44.

Ozgen, C., Peters, C., Niebuhr, A., Nijkamp, P., and Poot, J. (2014). Does cultural diversity of migrant employees affect innovation? *International Migration Review*, 48:377–416.

Parikh, A., Van Leuvensteijn, M., et al. (2003). Internal migration in regions of germany: A panel data analysis. *Applied Economics Quarterly*, 49(2):173–192.

Posch, M., Schulz, J., and Henrich, J. (2023). Surname diversity, social ties and innovation. *Social Ties and Innovation (August 3, 2023)*.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London*, 48(2):167–235.

Redding, S. J. and Sturm, D. M. (2008). The costs of remoteness: Evidence from german division and reunification. *American Economic Review*, 98(5):1766–1797.

Riley, S. J., DeGloria, S. D., and Elliot, R. (1999). Index that quantifies topographic heterogeneity. *intermountain Journal of sciences*, 5(1-4):23–27.

Roberge, M.-É. and van Dick, R. (2010). Recognizing the benefits of diversity: When and how does diversity increase group performance? *Human Resource Management Review*, 20(4):295–308.

Roesel, F. (2022). The german local population database (gpop), 1871 to 2019. *Jahrbücher für Nationalökonomie und Statistik*.

Rosar, A. (2020). Ehenamenwahl in Deutschland. https://gfds.de/anne-rosar-ehenamen-in-deutschland/.

Schumann, A. (2014). Persistence of population shocks: Evidence from the occupation of west germany after world war ii. *American Economic Journal: Applied Economics*, 6(3):189–205.

Van Beers, C. and Zand, F. (2014). R & D cooperation, partner diversity, and innovation performance: An empirical analysis. *Journal of Product Innovation Management*, 31(2):292–312.

Appendix

Robustness tests results

Table 9: OLS Estimation results with manufacturing sector

	(1) Patents per employee (IHS)	(2) Patents per employee (IHS)	(3) Patents per employee (IHS)
Deconcentration	0.000479		
	(0.0168)		
Fractionalisation		0.0934	
		(0.718)	
Shannon Diversity			0.000627
			(0.00166)
Medium sized cities	0.00256	0.00266	0.00395
	(0.00461)	(0.00205)	(0.00368)
Larger Towns	0.00459	0.00476*	0.00656
	(0.00607)	(0.00282)	(0.00521)
Smaller Towns	0.00773	0.00799**	0.0102
	(0.00670)	(0.00318)	(0.00632)
Rural areas	0.0152*	0.0157***	0.0184**
	(0.00824)	(0.00525)	(0.00824)
Share foreigners	0.0317	0.0316	0.0304
	(0.0375)	(0.0326)	(0.0335)
Share high-skilled	-0.0104	-0.0108	-0.0123
	(0.0300)	(0.0294)	(0.0303)
Share low-skilled	0.0119	0.0129	0.0138
	(0.0426)	(0.0374)	(0.0371)
mean_Verarb	-0.00734	-0.00742	-0.00708
	(0.00710)	(0.00726)	(0.00705)
Constant	-0.000157	-0.0935	-0.00880
	(0.00777)	(0.718)	(0.0229)
Observations	344	344	344
Adjusted R^2	0.045	0.045	0.045

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: OLS Estimation results u-shaped diversity

	(1) Patents per employee (IHS)	(2) Patents per employee (IHS)	(3) Patents per employee (IHS)
Deconcentration	0.0122		
	(0.0183)		
Fractionalisation		-22.55	
		(56.43)	
Shannon Diversity			-0.000491
			(0.00317)
conc2	-0.000854		
	(0.0180)		
Medium sized cities	0.000247	0.00203***	0.00252**
	(0.000752)	(0.000371)	(0.00107)
Larger Towns	0.00129	0.00389***	0.00444***
	(0.000918)	(0.000570)	(0.00140)
Smaller Towns	0.00287***	0.00600***	0.00648***
	(0.00106)	(0.000775)	(0.00147)
Rural areas	0.00607***	0.0101***	0.0102***
	(0.00116)	(0.000933)	(0.00147)
Share foreigners	0.0210***	0.0273***	0.0283***
	(0.00595)	(0.00542)	(0.00553)
Share high-skilled	-0.00273	0.000508	0.00150
	(0.00730)	(0.00684)	(0.00698)
Share low-skilled	-0.0150*	-0.0200**	-0.0212**
	(0.00876)	(0.00938)	(0.00938)
frac2		11.46	
		(28.39)	
shannon2			0.0000324
			(0.000146)
Constant	-0.00110	11.10	0.00336
	(0.00429)	(28.04)	(0.0170)
Observations	3160	3160	3160
Adjusted R^2	0.039	0.036	0.036

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: OLS Estimation results more controls

	(1) Patents per employee (IHS)	(2) Patents per employee (IHS)	(3) Patents per employee (IHS)
Deconcentration	0.0100***		
	(0.00384)		
Fractionalisation		0.167	
		(0.162)	
Shannon Diversity			-0.00000852
			(0.000425)
Medium sized cities	0.000429	0.00196***	0.00193**
	(0.000694)	(0.000363)	(0.000762)
Larger Towns	0.00153	0.00373***	0.00361***
	(0.000932)	(0.000501)	(0.00122)
Smaller Towns	0.00316***	0.00577***	0.00555***
	(0.00112)	(0.000639)	(0.00150)
Rural areas	0.00648***	0.00982***	0.00930***
	(0.00126)	(0.000900)	(0.00167)
Share foreigners	0.0225***	0.0288***	0.0297***
	(0.00654)	(0.00558)	(0.00580)
Share high-skilled	-0.000948	0.00190	0.00285
	(0.00727)	(0.00702)	(0.00693)
Share low-skilled	-0.0164*	-0.0217**	-0.0229**
	(0.00876)	(0.00908)	(0.00940)
mean_women	0.00525	0.00687*	0.00735*
	(0.00439)	(0.00404)	(0.00403)
mean_20_29	0.00954	0.00920	0.00894
	(0.00868)	(0.00865)	(0.00891)
Constant	-0.00495	-0.169	-0.00236
	(0.00353)	(0.162)	(0.00760)
Observations	3160	3160	3160
Adjusted R^2	0.040	0.038	0.037

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: OLS Estimation results with observations with less than 100 entries dropped

	(1)	(2)	(3)
	ihs_patentssumperwork	ihs_patentssumperwork	ihs_patentssumperwork
concentration	0.0115***		
	(0.00350)		
StadtLand2	0.000232	0.00199***	0.00219***
	(0.000631)	(0.000357)	(0.000792)
StadtLand3	0.00127	0.00378***	0.00401***
	(0.000822)	(0.000496)	(0.00126)
StadtLand4	0.00283***	0.00583***	0.00602***
	(0.000973)	(0.000638)	(0.00158)
StadtLand5	0.00602***	0.00984***	0.00978***
	(0.00111)	(0.000897)	(0.00173)
mean_foreign	0.0208***	0.0277***	0.0284***
	(0.00598)	(0.00537)	(0.00552)
mean_hq	-0.00293	0.000887	0.00166
	(0.00730)	(0.00698)	(0.00687)
mean_gq	-0.0148*	-0.0203**	-0.0212**
	(0.00876)	(0.00913)	(0.00938)
frac		0.206	
		(0.166)	
shannon			0.000136
			(0.000431)
_cons	-0.000965	-0.203	0.000775
	(0.00174)	(0.166)	(0.00645)
N	3159	3159	3159
adj. R^2	0.040	0.036	0.036

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: OLS Estimation results (1)-(3) less than 16669 inhabitants (0.75 quantile)

	(1)	(2)	(3)
	Patents per employee (IHS)	Patents per employee (IHS)	Patents per employee (IHS)
Deconcentration	0.0116***		
	(0.00416)		
Fractionalisation		0.189	
		(0.170)	
Shannon Diversity			0.0000257
			(0.000579)
Einwohnerdichte 1997	0.00000357***	0.00000490***	0.00000497***
	(0.00000121)	(0.00000128)	(0.00000135)
Rural areas	0.00401***	0.0104***	0.00999***
	(0.000715)	(0.00102)	(0.00116)
Share foreigners	0.0132*	0.0195***	0.0204***
	(0.00772)	(0.00695)	(0.00699)
Share high-skilled	-0.0112	-0.00594	-0.00458
	(0.00939)	(0.00899)	(0.00879)
Share low-skilled	-0.0129	-0.0181*	-0.0192*
	(0.00995)	(0.0104)	(0.0107)
Larger Towns		0.00402***	0.00400***
		(0.000687)	(0.000719)
Smaller Towns		0.00614***	0.00604***
		(0.000771)	(0.000907)
Constant	0.000981	-0.187	0.00127
	(0.00243)	(0.170)	(0.00710)
Observations	2382	2382	2382
Adjusted R^2	0.023	0.022	0.022

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Poisson Estimation results (1)-(3) less than 62779 inhabitants (0.75 quantile)

	(1) patents1890perinhab	(2) patents1890perinhab	(3) patents1890perinhab
patents1890perinhab			
Deconcentration in 1890	3.702***		
	(1.216)		
Fractionalisation in 1890		269.8***	
		(95.35)	
Shannon diversity in 1890			0.523***
			(0.109)
Share of foreigners in 1871	-7.859	-2.847	0.769
	(7.982)	(5.953)	(4.461)
Share of illiterates in 1871	15.05	10.68	8.840
	(9.875)	(8.556)	(7.557)
Share of literates in 1871	16.97*	13.41*	11.39
	(9.031)	(7.852)	(7.116)
Density in 1890	0.000259***	0.000224***	0.000185***
	(0.0000491)	(0.0000374)	(0.0000330)
Share of secondary sector in 1882	16.16***	13.92***	11.90***
	(3.133)	(2.341)	(1.688)
Share of self-employed in 1882	7.392*	9.957***	9.894***
	(3.928)	(3.528)	(2.885)
Share of science-based professionals in 1882	42.99***	43.92***	48.33***
	(12.57)	(11.40)	(10.04)
Constant	-31.57***	-296.0***	-30.03***
	(9.367)	(94.73)	(6.967)
Observations	237	237	237
Adjusted R^2			

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Distance between family names and their geographic means as diversity measure (1) in 1996, (2) in 1890, OLS estimates

	(1) Patents per employee (IHS)	(2) patents1890perinhab
main		
Distance measure	0.00000922**	
	(0.00000446)	
Medium sized cities	0.00199***	
	(0.000359)	
Larger Towns	0.00372***	
	(0.000471)	
Smaller Towns	0.00562***	
	(0.000579)	
Rural areas	0.00914***	
	(0.000765)	
Share foreigners	0.0255***	
	(0.00602)	
Share high-skilled	0.00160	
	(0.00709)	
Share low-skilled	-0.0212**	
	(0.00906)	
Distance measure		0.00437***
		(0.000757)
Share of foreigners in 1871		2.633
		(2.050)
Share of illiterates in 1871		-13.05**
		(6.196)
Share of literates in 1871		-7.353
		(5.992)
Density in 1890		8069.7***
		(1066.6)
Share of agriculture in 1882		-2082.8
		(1769.5)
Share of secondary sector in 1882		2.065
		(1.501)
Share of self-employed in 1882		-0.496
		(1.798)
Share of science-based professionals in 1882		28.29***
		(8.213)
University before 1900		0.978***
		(0.195)
Technical University before 1900		1.525**
		(0.659)
Constant	0.00103	-4.127
	(0.00190)	(5.832)
Observations	3160	1404
Adjusted R^2	0.038	

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: OLS Estimation results with share of new names

	(1)
	ihs_patentssumperwork
Share of new names	0.00807***
	(0.00257)
StadtLand2	0.000659
	(0.000552)
StadtLand3	0.00223***
	(0.000633)
StadtLand4	0.00415***
	(0.000712)
StadtLand5	0.00769***
	(0.000767)
mean_foreign	0.0209***
	(0.00609)
mean_hq	-0.00171
	(0.00723)
mean_gq	-0.0155*
	(0.00869)
Constant	-0.00231
	(0.00191)
Observations	3160
Adjusted R^2	0.040

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: OLS Estimation results for the planning regions

	(1)	(2)	(3)
	Patents per employee (IHS)	Patents per employee (IHS)	Patents per employee (IH
Adj. Deconcentration	0.00692**		
	(0.00297)		
Adj. Fractionalisation		-0.0315	
		(0.0196)	
Adj. Shannon Diversity			-0.000331
			(0.000225)
Density	-0.000000108	-0.000000946***	-0.000000314
	(0.000000412)	(0.000000255)	(0.000000516)
Share foreigners	0.0208***	0.0291***	0.0279***
	(0.00586)	(0.00679)	(0.00645)
Share high-skilled	0.00166	-0.000109	0.00104
	(0.00241)	(0.00258)	(0.00256)
Share low-skilled	-0.00121	-0.00390	-0.00292
	(0.00280)	(0.00365)	(0.00365)
Constant	-0.00223	0.0336*	0.00503*
	(0.00204)	(0.0194)	(0.00259)
Observations	37	37	37
Adjusted R^2	0.646	0.580	0.597

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Poisson Estimation results at district level, controlling for R& D and state fixed effects

	(1) patentssumperwork	(2) patentssumperwork	(3) patentssumperwork
patentssumperwork			
Deconcentration	-0.0486		
	(0.0497)		
Fractionalisation		-195.5	
		(289.8)	
Shannon Diversity			0.134
			(0.140)
Density	-0.000565***	-0.000519***	-0.000566***
	(0.0000908)	(0.0000884)	(0.0000921)
Share foreigners	7.115***	8.548***	6.680***
	(2.288)	(2.142)	(2.148)
Share high-skilled	1.622**	1.736**	1.564**
	(0.740)	(0.729)	(0.786)
Share low-skilled	1.910*	1.433	2.152*
	(1.117)	(1.014)	(1.121)
R&D expenses	0.00000161	0.00000177	0.00000209
	(0.00000169)	(0.00000157)	(0.00000162)
R&D personnel	-0.000181	-0.000209	-0.000247
	(0.000212)	(0.000196)	(0.000204)
Constant	-7.864***	187.5	-9.737***
	(0.751)	(289.5)	(1.843)
state fixed effects	yes	yes	yes
Observations	249	249	249
Adjusted R^2			

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 19: Estimation results SLX model with distance at municipality level

	(1)	(2)	(3)
	(patentssumperwork)/.1911319	(patentssumperwork)/.1911319	(patentssumperwork)
(mean) concentration	0.275***		
	(0.103)		
(mean) frac		0.405	
		(0.303)	
frac_w		-22.36*	
		(11.77)	
(mean) shannon			0.0277***
			(0.00681)
shannon_w			-0.111***
			(0.0246)
conc_w	-1.526***		
	(0.542)		
StadtLand2	-0.0446	-0.0383	-0.0215
	(0.0462)	(0.0461)	(0.0458)
StadtLand3	-0.0805	-0.0802	-0.0584
	(0.0632)	(0.0631)	(0.0627)
StadtLand4	-0.116	-0.118	-0.111
	(0.0759)	(0.0759)	(0.0756)
StadtLand5	-0.165*	-0.175**	-0.181**
	(0.0888)	(0.0883)	(0.0884)
ln_mean_foreign	0.214***	0.210***	0.209***
	(0.0170)	(0.0161)	(0.0160)
ln_mean_hq	0.144***	0.148***	0.155***
	(0.0240)	(0.0236)	(0.0235)
mean_gq	-0.627**	-0.606**	-0.599**
	(0.265)	(0.263)	(0.260)
ln_mean_plants	-0.0471**	-0.0602***	-0.0760***
	(0.0215)	(0.0206)	(0.0213)
ln_mean_Betriebsgr	-0.378***	-0.372***	-0.389***
	(0.0266)	(0.0266)	(0.0268)
Constant	-0.760**	20.60*	-0.379
	(0.334)	(11.74)	(0.325)
distance weighted	yes	yes	yes
Observations	3166	3166	3166
Adjusted R^2	0.160	0.158	0.166

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 20: Estimation results SLX model with adjacency at municipality level

	(1)	(2)	(3)
	(patentssumperwork)/.1911319	(patentssumperwork)/.1911319	(patentssumperwork)
(mean) concentration	0.128		
	(0.101)		
(mean) frac		0.410	
		(0.301)	
frac_wa		0.684***	
		(0.199)	
(mean) shannon			0.0254***
			(0.00669)
shannon_wa			0.0136**
			(0.00654)
conc_wa	0.152		
	(0.126)		
StadtLand2	-0.0510	-0.0403	-0.0364
	(0.0462)	(0.0461)	(0.0463)
StadtLand3	-0.0921	-0.0775	-0.0648
	(0.0631)	(0.0631)	(0.0632)
StadtLand4	-0.134*	-0.118	-0.0991
	(0.0759)	(0.0758)	(0.0759)
StadtLand5	-0.193**	-0.178**	-0.146
	(0.0884)	(0.0882)	(0.0886)
ln_mean_foreign	0.200***	0.208***	0.203***
	(0.0169)	(0.0161)	(0.0163)
ln_mean_hq	0.138***	0.145***	0.134***
	(0.0240)	(0.0235)	(0.0237)
mean_gq	-0.536**	-0.651**	-0.590**
	(0.264)	(0.261)	(0.259)
ln_mean_plants	-0.0569***	-0.0635***	-0.0885***
	(0.0213)	(0.0206)	(0.0213)
ln_mean_Betriebsgr	-0.361***	-0.368***	-0.358***
	(0.0267)	(0.0265)	(0.0268)
Constant	-1.514***	-2.363***	-1.627***
	(0.240)	(0.414)	(0.228)
adjacency weighted	yes	yes	yes
Observations	3166	3166	3166
Adjusted R^2	0.158	0.161	0.162

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 21: First stage IV regression French occupation zone

	(1)
VARIABLES	Deconcentration
French occupation zone	-0.0649***
	(0.00634)
Medium sized cities	0.102***
	(0.00947)
Larger Towns	0.155***
	(0.0105)
Smaller Towns	0.186***
	(0.0109)
Rural areas	0.224***
	(0.0115)
Share foreigners	0.756***
	(0.0370)
Share high-skilled	0.479***
	(0.0743)
Share low-skilled	-0.516***
	(0.0403)
(sum) plants	-8.92e-06***
	(1.74e-06)
Constant	0.355***
	(0.0144)
Observations	2,479

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

First stage regression results

Table 22: First stage IV regression French occupation zone

	(1)
VARIABLES	Fractionalisation
French occupation zone	-0.000631***
	(0.000126)
Medium sized cities	-0.000123*
	(7.20e-05)
Larger Towns	-0.000702***
	(8.28e-05)
Smaller Towns	-0.00141***
	(9.25e-05)
Rural areas	-0.00363***
	(0.000126)
Share foreigners	0.00584***
	(0.000773)
Share high-skilled	0.00592***
	(0.00107)
Share low-skilled	-0.00660***
	(0.00118)
(sum) plants	-4.21e-09
	(8.33e-09)
Constant	0.999***
	(0.000239)
Observations	2,479

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 23: First stage IV regression French occupation zone

VARIABLES	(1) Shannon Diversity
French occupation zone	-0.285***
	(0.0433)
Medium sized cities	-1.181***
	(0.103)
Larger Towns	-2.198***
	(0.112)
Smaller Towns	-2.951***
	(0.115)
Rural areas	-4.004***
	(0.118)
Share foreigners	2.945***
	(0.258)
Share high-skilled	3.411***
	(0.499)
Share low-skilled	-2.543***
	(0.296)
(sum) plants	7.20e-05***
	(1.99e-05)
Constant	12.78***
	(0.134)
Observations	2,479

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24: First stage IV regression Proximity to Borders

VARIABLES	(1) Deconcentration
dummy_FRGborder25	0.00150
	(0.00371)
dummy_FRGborder50	0.0226***
	(0.00384)
dummy_FRGborder75	0.0370***
	(0.00365)
Medium sized cities	0.0959***
	(0.0105)
Larger Towns	0.147***
	(0.0118)
Smaller Towns	0.178***
	(0.0121)
Rural areas	0.217***
	(0.0124)
Share foreigners	0.716***
	(0.0327)
Share high-skilled	0.449***
	(0.0639)
Share low-skilled	-0.567***
	(0.0353)
(sum) plants	-9.91e-06***
	(2.10e-06)
Constant	0.355***
	(0.0147)
Observations	3,160

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25: First stage IV regression Proximity to Borders

VARIABLES	(1) Fractionalisation
dummy_FRGborder25	-0.000766***
	(9.39e-05)
dummy_FRGborder50	-0.000352***
	(8.44e-05)
dummy_FRGborder75	-0.000293***
	(7.03e-05)
Medium sized cities	-7.88e-05
	(8.97e-05)
Larger Towns	-0.000627***
	(9.95e-05)
Smaller Towns	-0.00125***
	(0.000106)
Rural areas	-0.00304***
	(0.000126)
Share foreigners	0.00452***
	(0.000677)
Share high-skilled	0.00579***
	(0.00102)
Share low-skilled	-0.00661***
	(0.00103)
(sum) plants	9.72e-09
	(1.15e-08)
Constant	1.000***
	(0.000237)
Observations	3,160

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 26: First stage IV regression Proximity to Borders

	(1)
VARIABLES	Shannon Diversity
dummy_FRGborder25	-0.327***
	(0.0281)
dummy_FRGborder50	-0.212***
	(0.0294)
dummy_FRGborder75	-0.188***
	(0.0292)
Medium sized cities	-1.131***
	(0.122)
Larger Towns	-2.100***
	(0.134)
Smaller Towns	-2.769***
	(0.138)
Rural areas	-3.674***
	(0.141)
Share foreigners	2.345***
	(0.250)
Share high-skilled	3.705***
	(0.535)
Share low-skilled	-2.667***
	(0.282)
(sum) plants	8.38e-05***
	(2.41e-05)
Constant	12.87***
	(0.155)
Observations	3,160

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1