

Kremer, Anna

**Working Paper**

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

CEPIE Working Paper, No. 02/25

**Provided in Cooperation with:**

Technische Universität Dresden, Faculty of Business and Economics

*Suggested Citation:* Kremer, Anna (2025) : Many names, many gains? How local diversity in Germany affects innovation, CEPIE Working Paper, No. 02/25, Technische Universität Dresden, Center of Public and International Economics (CEPIE), Dresden,  
<https://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa2-1004928>

This Version is available at:

<https://hdl.handle.net/10419/333401>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**CEPIE Working Paper No. 02/25**

Center of Public and International Economics

**MANY NAMES, MANY GAINS?  
HOW LOCAL DIVERSITY IN GERMANY AFFECTS  
INNOVATION**

November 2025

Anna Kremer

Editors: Faculty of Business and Economics, Technische Universität Dresden.

This paper is published on the [Open Access Repository Qucosa](#).

The complete Working Paper Series can be found at the [CEPIE Homepage](#) | [EconStor](#) | [RePEc](#)



# Many Names, Many Gains?

## How Local Diversity in Germany Affects Innovation

**Anna Kremer\***

November 2025

### **Abstract**

Meeting others with different backgrounds brings up new ideas. This paper shows that this not only matters for a background in heterogeneous industries or nationalities, but that regional differences matter too. Regions within a country vary in their traditions and culture. Cultural homogeneity within regions becomes mixed due to internal migration, which, like international migration, increases the diversity of a place. In a novel approach, I look at diversity in German municipalities, measured by different family names, and investigate its effect on the number of generated patents. I use a unique dataset from a 1996 phonebook and casualty lists from WWI. There is a positive association between innovation and diversity when defined by the share of new names, a deconcentration measure, or a Shannon index. Causality is established by using instrumental variables estimations with historical borders. I show that intra-country diversity affects patenting positively and conclude that regional differences matter for economic outcomes.

*Keywords:* cultural diversity, innovation, family names, patents, local level, Germany

*JEL-Codes:* R11, O30, Z13

---

\*Corresponding author. Technische Universität Dresden, ifo Institut Dresden, Einsteinstr. 3, 01067 Dresden and Thünen Institut für Innovation und Wertschöpfung im ländlichen Raum, Bundesallee 64, 38116 Braunschweig. [anna.kremer@thuenen.de](mailto:anna.kremer@thuenen.de). This publication is an updated version of Kremer, A.: Names, diversity and innovation, CEPIE Working Paper 03/21.

# 1 Introduction

Family names tell a story about an individual’s background. “Foreign” names point to a family history of immigration, more recently or long ago. “Domestic” names give indications, too, about where a family originates because many names are specific to 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. Migration itself is an important determinant of diversity because it increases the cultural heterogeneity of a place if 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, positively affects the innovativeness and productivity of an economy. This is most likely connected to complementary skills and knowledge (Ottaviano and Peri, 2006). However, cultural disparities arise within a country as well. Usually, they stem from the regions’ different histories and traditions, but are also linked to the natural environment, the economy, and the different upbringings of families, including professional specializations. These differences might also lead to other knowledge, skills, and abilities; most likely on a smaller scale than in the international comparison.

This paper asks whether these differences between regions within a country are relevant to advancing innovativeness. To shed light on this issue, I employ a dataset on the distribution of family names in Germany to account for local diversity and examine the relationship with patents per employee in the municipality associations.

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, which determined that the husband’s name was the family name. The legal circumstances changed; however, a large majority of couples still choose the husband’s family name when they marry <sup>1</sup>. 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 locations and thus do not necessarily relate to a geographical distribution of the name (compare e.g. “Smith” or “Müller”). Still, many names are unique or predominant in a specific region and can account for movements of families. I use this insight and a data set containing all 1996 phone book entries to determine where the diversity of family names in Germany is high or low.

The literature suggests that “interactions between diverse people [...] increase 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 for cultural diversity. Ottaviano and Peri (2006) provide evidence 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 the cultural diversity of the labor force on patent applications. She confirms that a positive spillover effect outweighs the costs of diversity. However, most studies focus on the impact of cultural diversity measured solely by nationalities.

I add to the literature by looking at whether diversity from intra-country differences,

---

<sup>1</sup>Only 3% decided in favor of the wife’s name, whereas 80% opted for the husband’s name in 1996, and 17% kept their individual family names. (Rosar, 2020)

proxied by family names, affects economic performance. Posch et al. (2023) regard innovation in the USA and measure cultural diversity by family names, too. They show in an IV-Setting that higher diversity affects patenting positively, both in terms of quantity and quality, and at a regional or individual level. The proven mechanism behind this builds on individuals having different knowledge and being willing to interact. Because Posch et al. (2023) study diversity in an immigrant country, the historical USA, their family name measure also captures differences in the ancestral country. I add to their contribution by solely accounting for regional differences among the native population by restricting my sample to family names present in 1890 in Germany. I estimate the effect of local diversity at the level of municipality associations, measured by family names, on innovativeness, measured by patents. A higher number of family names is expected to be connected to a more dynamic composition of the municipality associations throughout history, thus resulting in a higher diversity. With this, I build on a reasoning analogous to the international case: that relocation of people induces cultural diversity (Bakens et al., 2015) because culture varies over space and even minor alterations, such as within Germany, could shape people differently. From these different interacting backgrounds, innovation is expected to arise via complementary skills, knowledge, and abilities, as well as better interactive problem-solving. Additionally, firms in a more diverse environment can choose the most capable employees from a larger pool, which might also increase innovativeness.

I find that local diversity affects innovation positively when defined by the share of new names that were not present in 1890, a deconcentration measure, or a Shannon index. Causality is established by using the historical borders as an instrument. Using different estimation specifications, adding more controls, and controlling for geographical spillovers shows that my findings are robust.

The following section provides an overview of the existing work regarding how cultural diversity, measured at the inter-country level, affects economic performance. I then explain the empirical method by commenting on the theoretical model, the measures of diversity, and the data. Next, I describe my estimation, findings, and how I tackle the endogeneity issue. For the latter, I employ historical, political borders as instruments for diversity. Lastly, I perform some robustness checks before concluding.

## 2 Literature Review

According to Jacobs (1969), diversity helps generate new knowledge and advance productivity by bringing together different expertise and exchanging ideas. Originally, diversity is thought of as differences between firms and sectors. This 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. They compare 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 more sustained effect on employment and seem dynamic.

Jacobs' theory also translated to a cultural definition of diversity: Culture determines how we live together and is defined as the features that are relatively homogeneous within groups, but vary between them. These include a common value system, historical

traditions, religion, and lifestyles (Bakens et al., 2015). Cultural groups are often delimited by geography. Thus, the spatial distribution of people correlates with cultural dispersion to some extent. However, this relationship is not set in stone. Migration mixes this up and increases cultural diversity; when people relocate and retain (aspects of) their original culture (Bakens et al., 2015). This happens both on an inter- and intra-country level. The resulting cultural diversity can positively or negatively affect innovation and productivity. While heterogeneous workers might have to deal with, e.g., communication barriers, they can also exchange ideas from their different backgrounds, generate knowledge, and improve productivity.

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 people is higher. They expect this to be due to the complementarity of skills between workers. They study 160 metropolitan areas in the USA between 1970 and 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 each city’s predicted change in national groups as an instrument, relying on the observation that nationalities tend to flock together. Looking at a more recent period, Cooke and Kemeny (2017) confirm that birthplace diversity is responsible for productivity improvements 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. Posch et al. (2023) find a positive effect of cultural diversity on innovation as well, both in terms of quantity and quality, and at a regional or individual level. They measure diversity by using family names, which in their setting in the USA also includes differences in nationalities from immigration. Möhlmann and Bakens (2015) investigate the same relationship with a two-step method. For Dutch firms, they first estimate total factor productivity by measuring the deviation from sector productivity. In a second step, they investigate whether more diverse firms in terms of the ethnicity of their employees are more productive. Their results show that a higher share of foreigners in a firm hurts productivity. Diversity among these employees, however, is positively correlated with firm performance. The effect disappears when firm fixed effects are included. Ozgen et al. (2014) also look at firm-level data from the Netherlands and Germany. They establish a connection between (the 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 extensively subsume evidence from other studies in Europe, North America, and New Zealand. Here, they emphasize cultural diversity’s positive potential (productivity and innovation) and adverse effects (communication barriers).

Regarding employers more closely, Nathan and Lee (2013) study 7600 firms in London and compare their cultural diversity with innovation, entrepreneurship, and where sales are directed. They rely on firm microdata to define diversity as whether owners and partners are migrants or non-(white-)British. In their regression, the authors find a small but significant diversity bonus: Higher diversity makes firms more innovative, they are better at reaching out 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 labor force on patent applications. She confirms that a positive spillover effect outweighs the costs of diversity. Concerning the endogeneity problem, she argues that migration is often motivated by earlier migration from the same origin and thus uses the latitude of the region center and lagged diversity of neighboring regions as instruments in an IV regression. 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 reflect different innovation outcomes (e.g., improvement or introduction of products).

Recent literature looks at the channel of spillovers and suggests that it is skill (Docquier et al., 2020) and knowledge complementarity (Buchholz, 2021; Bernstein et al., 2022). Docquier et al. (2020) use panel data from the US from 1960 to 2010 and show that gains in diversity are more significant from highly-skilled immigrants. Moreover, they show that the impact is largest when immigrants come from countries that are either economically or culturally distant. Their evidence also points to skill complementarity being responsible for these outcomes. Buchholz (2021) investigates the mechanism of diversity spillovers by looking at US Metropolitan Statistical Areas with pseudo panels of workers and fixed effects regressions. He finds that it is both an effect of “interactive problem solving” and an “exposure” effect. The first refers to people pooling their knowledge as part of problem-solving, while the latter refers to becoming more productive by learning from others. Bernstein et al. (2022) look at immigrants in the US and their patenting activity employing their social security number. They find that 16% of inventors are immigrants and are responsible for an even larger number of patents. Using premature death as an instrument, they show that knowledge spillovers from immigrants to natives are greater than between natives.

I examine whether a positive relationship between diversity and innovation exists at the municipal level in Germany. In contrast with existing work, I proxy for (intra-country) diversity by family names and their dispersion over time. A greater number of family names is expected to be connected to a more dynamic place. In this, I build on the reasoning that relocation of people promotes cultural diversity because culture varies over space and minor deviations, such as within Germany, shape people differently.<sup>2</sup> They stem from the different upbringings of families, differing natural environments and economic factors, and traditions in different (historical) regions of Germany. They result in different mindsets and potentially in different abilities and skills. These differences are expected to give rise to innovation, e.g., by a collaboration of people with different mindsets to solve a problem together, or by people with different backgrounds sharing ideas.<sup>3</sup>

In the next section, I turn to the empirical method.

---

<sup>2</sup>Analogous to differences between countries, differences between regions can be smaller or greater but cannot be neglected for the case of Germany (compare e.g. Falck et al., 2012 or Kremer, 2022)

<sup>3</sup>Ciccone and Nimczik (2022) find that a higher share of expellees in Baden-Wuerttemberg led to a higher number of patents, also assuming knowledge spillovers drive this outcome.

### 3 Method

#### Model

To account for potential costs and advantages of diversity for innovativeness, I rely on the model of Niebuhr (2010). It expects excessive differences to hinder exchange (e.g., via language barriers, however, misunderstandings are also possible with the same mother tongue) and thus innovation. On the other hand, advantages 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, with a larger pool of more diverse people, firms can choose the most capable employees, who then create more innovation.

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

The effect of cultural diversity augments the common production function in place  $i$ . Diversity  $div$  from culturally different origins (names), noted by  $o$ , increases 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 get smaller and therefore groups are more evenly distributed. The costs of differences ( $\tau$ ), 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 (reflected by names) is  $\sigma = 1/(1 - \alpha)$ .

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

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

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

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

#### Measures of Diversity

Cultural disparities arise within a country. They usually stem from a different history and tradition of the regions, but also from a dissimilar natural environment, economy, and education, among other things. Furthermore, people are raised differently in different families. I expect that when these people meet, the differences spark innovation. This argumentation is analogous 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. Until 1976, families could not decide what name to choose as a family name, and even after that, a large majority continued to select the husband's name. How families historically spread across Germany in the male line is therefore reflected by the

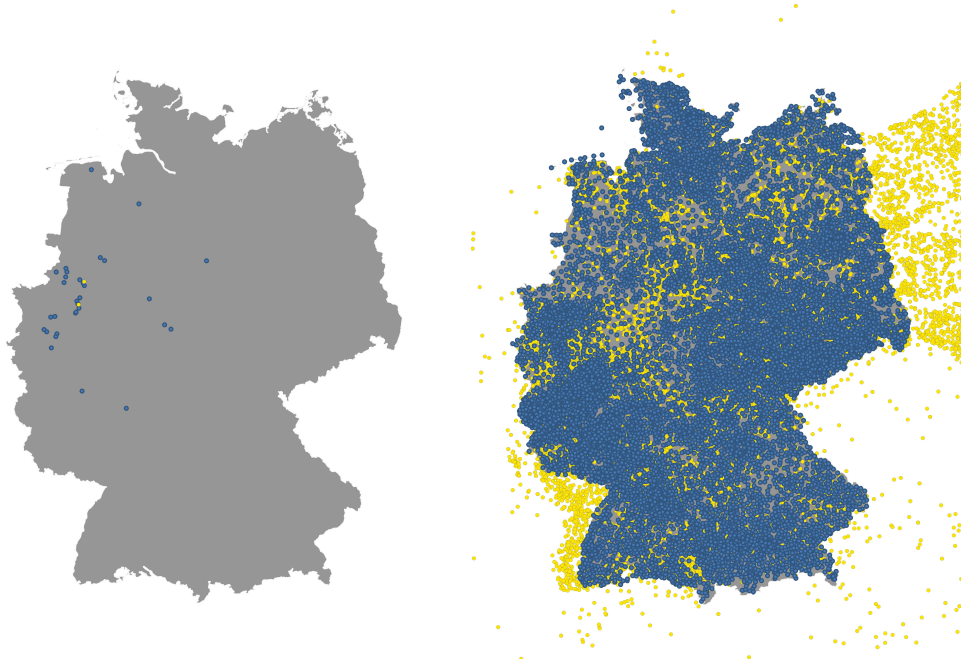


Figure 1: Distribution of family name “Wethmar” (left) and “Müller” (right). Yellow points refer to the distribution in 1890, blue ones to 1996.

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 Westphalia, 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 is that it has common names that originated in 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 distributed randomly enough not to harm my measure.<sup>4</sup> Moreover, many names in Germany are unique (the mean number is at 129 entries per family name, with the median at 21) or predominant in a specific region and can account for movements of families.

I am interested in intra-country differences in Germany and whether they affect innovativeness. Names of foreigners thus bias my measure. Therefore, I restrict the sample of family names to names that appeared in Germany in 1890. This excludes family names from, e.g., the so-called “guest workers” and double names from marriages. If, however, foreigners marry into a German family and receive the German name, they are still part of my sample.

To quantify diversity from the distribution of family names, I employ three measures that regard different dimensions: the share of new names, deconcentration, and the Shannon index. They are explained in the following.

<sup>4</sup>The relative shares of the name “Müller” across Germany is found in Figure 5 in the Appendix, as an example.

## Share of new names

The most straightforward way to quantify how diverse a place is is by looking at how many individuals came to it who were living elsewhere before. Therefore, as the first measure of diversity, I count how many people live in a place in 1996 with a family name that was not present in that area in 1890. Because I restrict my sample to names appearing in Germany in 1890, this measures how many people came to the place with a different regional background. I use the share of people with a name that was not present in 1890 over all inhabitants as my first diversity measure, and calculate it by

$$\text{Share new names}_i = \text{Entries new names}_i / \text{All entries}_i$$

A higher share of new names equals a more diverse area.

## Deconcentration

Next, I look at how many different family names are present in the place in 1996. For this, I divide the number of different family names by the number of inhabitants of every place and am left with the average number of family names per inhabitant. It is thus calculated by

$$\text{Deconcentration}_i = \text{Names}_i / \text{Entries}_i$$

I call this diversity measure “Deconcentration” as it looks at how many different family groups are present. A higher number of family names per inhabitant means a more diverse place.

## Shannon diversity

Last, I regard a more traditional measure of diversity, which is also employed by Posch et al. (2023). It stems from information theory and considers the number of different groups and the evenness of their distribution. It is calculated by

$$\text{Shannon Diversity}_i = -\left[\sum_{o=1}^{\bar{n}} (\text{Names}_{oi} / \text{Entries}_i) * \log_2(\text{Names}_{oi} / \text{Entries}_i)\right]$$

A higher value of the Shannon index reveals a more diverse municipality.

The correlations between all three measures are 0.76 between the share of new names and deconcentration, 0.14 between the share of new names and Shannon index, and -0.19 between deconcentration and the Shannon index.

## Data

The distribution of names is provided by CompGen (Verein für Computergenealogie e.V.). It draws from a digitized phone book that provides the distribution of names in 1996. I match geocoded names with today’s municipality boundaries. A combination of postal codes and location names georeferenced the data.<sup>5</sup> Additionally, I use data from casualty lists from World War I from the same source to see which names were present back then and where they appeared.

---

<sup>5</sup>The geocoding refers to postcode and place name and is thus not entirely accurate.



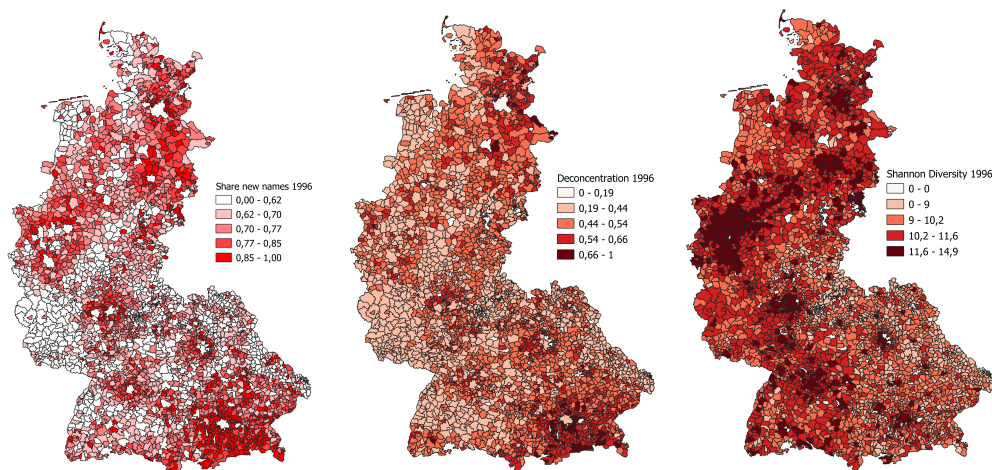


Figure 2: Diversity of family names

Since private phone coverage in East Germany only started to advance after reunification but had not finished in 1995 (Didczuneit, 2015), I restrict my sample to West Germany, where the proper distribution of names is expected to be reliably identified. The total number of entries in West Germany is 26 million, with 209,135 family names. The mean of the occurrence of family names is 128.6, with the median being at 21 entries per family name. This shows that large outliers (the maximum number of entries is 279,932 for a single family name, which is “Müller”) bias the mean strongly; many names are indeed attributable to only a couple of people, presumably related.

Diversity of the municipality associations in 1996 is displayed in Figure 2.

As for innovation, patent data is available from Deutsches Patent- und Markenamt and refers to the inventor’s address. Figure 3 shows the distribution of patents in West Germany averaged from 1995 to 1999. I identify places of innovation from the postcodes of the patent inventors and times by the year of the effective European filing date of the innovation. I average them over the years 1995 to 1999 to prevent measurement bias. The average number of patents per employee from 1995 to 1999 is 0.007 with a standard deviation of 0.026. The distribution is heavily left-skewed and displayed in Figure 3.

Data on control variables related to the regional employment structure stems from the Establishment History Panel of IAB (Ganzer et al., 2020). Skill level and share of foreign-born employees among employees are used as controls and averaged from 1995 to 1998 to prevent measurement bias. Due to data censoring, I lose around 6% of the observations. The sample of the main regression is seen in Figure 4. Information on the urban type with dummy variables for large cities, medium-sized cities, larger towns, smaller towns, and rural areas is gathered from the INKAR database (Bundesinstitut für Bau- Stadt- und Raumforschung (BBSR)).

Descriptive statistics are in Table 1.

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 estimate OLS and 2SLS regressions, FOR JOURNAL: as do Posch et al. (2023),

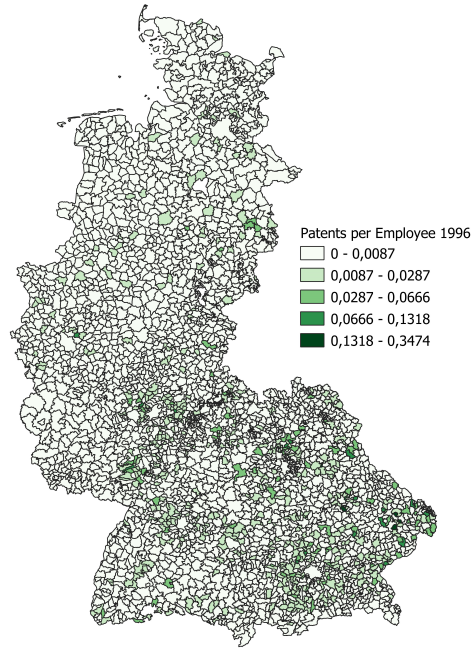


Figure 3: Patents per employee

Table 1: Descriptive statistics 1996

	count	mean	sd	min	max
Patents/Employee	3399	.0068502	.02487	0	1.204864
New names	3415	.7041405	.1390327	.202211	1
Deconcentration	3415	.4608311	.1021182	.1115982	1
Shannon Diversity	3415	9.746677	1.217753	0	13.82171
Share foreigners	3399	.0721459	.0473105	0	.3709796
Share high-skilled	3399	.0506658	.0350192	0	.527629
Share low-skilled	3399	.1781056	.0445319	.0513941	.501992
Large cities	69				
Medium sized cities	519				
Larger towns	717				
Smaller towns	932				
Rural areas	1160				

with patents transformed by inverse hyperbolic sine (IHS) to account for (a few) zero patents.

## 4 Results

In the following, I investigate whether a higher intra-country diversity, measured by family names, positively correlates with higher innovativeness. This is expected because people with different backgrounds may have complementary skills and/or knowledge; they exchange ideas more productively or solve problems better together because of other approaches.

I estimate OLS regressions of the diversity measures on innovativeness, measured by patents. I control for differences in employee structure, proxied by skill level and national background. Additionally, I control for the level of urbanization and use state fixed effects.

I find a positive and significant association between all three diversity measures and innovation in Table 2, Panel A.<sup>6</sup> This shows that there is likely to be a positive relationship between diversity and innovation in different dimensions: Share of new names shows that when additional families move into a place, this correlates with increased innovativeness. Likewise, a higher number of family names relative to number of inhabitants (Deconcentration) correlates positively with the number of patents. Lastly, the Shannon index reveals that a more diverse distribution of family names is connected to more innovation. However, the causality is unclear and is dealt with in the following section.

### Endogeneity

The association between diversity and innovativeness has a substantial risk of reverse causality. It seems reasonable that diversity not only enhances innovation, but also that more people migrate to and thereby increase the diversity of places with high innovativeness. In these more productive places, e.g., higher wages as a pull factor, are expected.

I am investigating the mechanism by looking at instrumental variables regressions with distance to historical borders. This is based on the idea that it is easier to move within political entities than across them. For example, it is now more costly to move between German federal states than within them, e.g., because children have to adapt to different school syllabi. From this follows that places closer to borders are less attractive because they are restricted on one side and distance impedes migration (Ravenstein, 1885 and Parikh et al., 2003). Thus, regions close to borders should be less diverse because they have fewer in- and outflows than more central locations.

I make use of historic maps that refer to the German states since 1820 and the Prussian provinces, the post-war occupation zones, and the federal states (see also Kremer, 2022). I construct the weighted distance over time for each municipality to the closest border between 1820 and 1996. I only regard the border distance since 1890 for the share of new names. I use the same instrument for all three diversity measures because I expect all dimensions of diversity to be affected by distance to historical

---

<sup>6</sup>I am not using count models since I transformed my dependent variable by normalizing it to patents per employee. Scatterplots of the relationship are found in Figure 6 in the Appendix.

Table 2: Municipality-level estimates of the effect of family name diversity on innovation

	(1)	(2)	(3)
	Patents/empl	Patents/empl	Patents/empl
<i>Panel A: OLS estimates</i>			
Share of new names	0.00795** (0.00389)		
Deconcentration		0.0120* (0.00699)	
Shannon Diversity			0.00000568 (0.000703)
<i>Panel B: Reduced form regressions</i>			
Distance to borders since 1820	0.0000658*** (0.0000158)		
Distance to borders since 1890		0.0000594*** (0.0000148)	
<i>Panel C: IV estimates</i>			
Share of new names	0.0914*** (0.0251)		
Deconcentration		0.104*** (0.0268)	
Shannon Diversity			0.0257*** (0.00821)
<i>Panel D: First stage regressions</i>			
	Share of new names	Deconcentration	Shannon Diversity
Distance to historical borders	0.000721*** (0.000107)	0.000632*** (7.07e-05)	0.00256*** (0.000532)
F-test of excluded instruments	45.55	79.94	23.09
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	3390	3390	3390

Standard errors in parentheses

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

borders. I find positive significant first-stage regressions for all three measures (Table 2, Panel D). Likewise, find Redding and Sturm (2008) that places closer to the inner-German border had lower population dynamics. The reduced form is also positively significant (Table 2, Panel B).

The second stage regressions in Table 2, Panel C show diversity's positive and significant effect on innovation in all three specifications.<sup>7</sup> I thus affirm my earlier finding that diversity has a positive causal effect on innovation. These results align with Posch et al. (2023) who similarly estimate the effect of family name diversity on innovation in the US by IV regressions. Their findings, however, also include the effect of cultural

<sup>7</sup>The quantitative difference between the numbers is due to estimating the average treatment effect in OLS and the local average treatment effect in IV.

diversity from immigration.

Balancing tests in Table 18 show that the instrument does not significantly affect either the number of plants or the number or the share of engineers and scientists, population density or unemployment, the total number of employees, or employees between 20 and 24 years old in reduced form regressions.

To prove the exclusion restriction, that the instrument affects diversity through migration, I model diversity in 1996. I use data from Roesel (2022) to proxy for population number in 1890 and model migration flows between regions depending on geographical and cultural distance<sup>8</sup>. Additionally, I make borders prohibitive for migration. I do not account for any other economic factors in the model and use the resulting new inflows into municipalities as an instrument for diversity. I confirm the positive significant effect of the share of new names and deconcentration in Table 3 and that diversity is related to innovation through migration. The instrument is, however, weak for the Shannon index.

Table 3: IV Regressions with modeled diversity

	(1)	(2)
	Patents/empl	Patents/empl
Share of new names	0.102*** (0.0312)	
Deconcentration		0.168*** (0.0494)
Controls: Urban Type	Yes	Yes
Controls: Labor Force	Yes	Yes
State fixed effects	Yes	Yes
Observations	3390	3390

Standard errors in parentheses

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

Diversity positively affects innovation via the following migration channel: A more dynamic population composition of the municipality associations throughout history leads to a larger number of different family names, thus higher diversity. This builds on the reasoning that relocation of people induces cultural diversity (Bakens et al., 2015) because culture varies over space and even minor alterations, such as within Germany, may shape people differently. These different interacting backgrounds and complementary skills and knowledge are expected to give rise to innovation. Recent literature has shown that these channels exist for the case of international diversity (Bernstein et al., 2022; Buchholz, 2021; Docquier et al., 2020).

## 5 Robustness

As robustness checks, I perform several tests concerning additional controls, the regression specification, and geographic spillovers.

<sup>8</sup>the latter stems from Falck et al. (2012)

## Additional Controls and Subsamples

In further estimations, I add more controls or estimate subsamples to prove the robustness of my findings:

In Table 4, I control for the manufacturing sector’s influence, which is expected to be responsible for many patents. The sample size is dramatically reduced due to data censoring, and the instrument is weak in this specification. Thus, OLS results are displayed to confirm the main specification. The effect of the share of manufacturing is positive and insignificant.

Similarly, I control for the workforce’s share of engineers and scientists, because they especially contribute to the generation of patents. The results in Table 5 mimic the main results and show positive significant effects of these workers in the first two specifications.

Additionally, I control whether a (technical) university is present in a place because, again, these institutions might especially contribute to innovativeness (compare Fritsch and Wyrwich, 2018). Results are displayed in Table 6 and show significant positive effects of diversity on patents. Positive impacts of university presence are found in the specification for the share of new names and deconcentration, negative ones for the Shannon diversity.

In Table 7, I also control each municipality association’s population size. The instrument is too weak for the Shannon index specification. The base findings for the share of new names and deconcentration are unharmed.

Moreover, I control geographical factors to ensure the effect is correctly identified. I use terrain ruggedness and the presence of waterways. Results are found in Table 8 and show no difference from the base specification.

In Table 9, I restrict the sample to towns and rural areas (1 and 2) and large and medium-sized cities and larger towns (3 and 4). These subsamples show that diversity has a positive but insignificant effect on innovation in more urbanized areas. On the other hand, the effect is positive and strongly significant in more rural areas. I expect this to be related to the fact that it matters more who you meet in less populated areas.

Lastly, I exclude prevalent names (such as “Müller”) because these might bias the results. If only regarding the distribution of less familiar family names (defined by less than 192 entries in Germany, referring to the 90th percentile), my results in Table 10 are robust.

## Different Specifications

To prove the methodological soundness of my results, I run further specifications:

I drop all municipality associations with fewer than 1000 phone book entries since they may be biased due to their small size. Results are found in Table 11 and show similar effects to the base results. Again, in the Shannon specification, the instrument is too weak.

If I drop the federal states of Bremen and Hamburg, which are city states and thus might bias the outcomes, it does not change anything.

Moreover, I estimate a Poisson regression in Table 12 and can show that I find an outcome similar to the OLS for share of new names and deconcentration. For the Shannon index, the effect is negative but insignificant.

I also estimate the extensive margin of patents. Results are found in Table 13 and show only a positive effect of the Shannon index. However, only 120 municipalities are without patents, a small subsample of 3.5%.

Table 14 shows the outcomes when the level of patents per employee is used instead of the inverse hyperbolic sine. They are similar to the base findings.

When controlling for the number of foreigners, high- and low-skilled as well as women and young employees, instead of using their shares, Table 15 shows positive and significant effects of diversity on innovation in all three columns.

## Geographic Spillovers

Because spillover effects are suspected due to people not always working where they live and meeting in the surroundings, I estimate SLX models both distance- and adjacency-weighted. Results are found in Tables 16 and 17. The OLS results are displayed for the share of new names and Shannon diversity because the IV is weakly instrumented. I find a positive significant effect of diversity on innovation. The spatial weight does not show a clear pattern.

## 6 Conclusion

By measuring diversity with the distribution of family names in Germany, I show that diversity enhances innovation when defined as intra-country differences. A more diverse family name distribution, in terms of deconcentration 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 to several robustness tests. I confirm with an IV estimation that the association between diversity and innovativeness runs from the former to the latter.

My findings indicate that cultural disparities within a country are relevant to economic outcomes.

## 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 georeferenced maps of European postal routes. I thank Georg Hirte, Marcel Thum, Dimitria Freitas, Christian Lessmann, Valentin Lindlacher, and participants of the Brown Bag seminar at TU Dresden for their helpful comments. Further ideas on improvements stem from the ERSA conference 2021, meetings of the DFG Netzwerk “Die Dynamik von Innovationssystemen” 2022-2024, the ifo Christmas Conference 2022, the BiB anniversary conference in 2023, the CDGE PhD workshop 2024, the meeting of

the Urban Economic Association in 2024, the ERSA conference 2024, and the meeting of Verein für Socialpolitik in 2024.

I received funding from a Cusanuswerk PhD scholarship.

## **Data availability statement**

The data supporting this study’s findings stems from CompGen (Verein für Computergenealogie e.V.) (2021) and is available upon request. Further data is openly available from INKAR (<https://www.inkar.de/>) of Bundesinstitut für Bau- Stadt- und Raumforschung (BBSR), OECD RegPat Database (Maraut et al., 2008) or was provided by IAB from Ganzer et al. (2020). Data for the instrumental variables stems from Kremer (2022).



## 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.
- 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.
- Bernstein, S., Diamond, R., Jiranaphawiboon, A., McQuade, T., and Pousada, B. (2022). The contribution of high-skilled immigrants to innovation in the united states. Technical report, National Bureau of Economic Research.
- 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.
- Buchholz, M. (2021). Immigrant diversity, integration and worker productivity: uncovering the mechanisms behind 'diversity spillover' effects. *Journal of Economic Geography*, 21(2):261–285.
- Bundesinstitut für Bau- Stadt- und Raumforschung (BBSR) (2021). INKAR. Indikatoren und Karten zur Raum-und Siedlungsentwicklung. <https://www.inkar.de/>.
- Ciccone, A. and Nimczik, J. (2022). The long-run effects of immigration: Evidence across a barrier to refugee settlement.
- 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.
- Deutsches Patent- und Markenamt (2021). DPMA register. <https://register.dpma.de/DPMAregister/pat/experte>.
- Dideczneit, V. (2015). "Endlich Telefon" - Der Aufbau des Telekommunikationsnetzes in den neuen Bundesländern. <https://www.mfk-berlin.de/wp-content/uploads/Endlich-Telefon.pdf>.
- Docquier, F., Turati, R., Valette, J., and Vasilakis, C. (2020). Birthplace diversity and economic growth: Evidence from the us states in the post-world war ii period. *Journal of Economic Geography*, 20(2):321–354.
- Falck, O., Heblich, S., Lameli, A., and Südekum, J. (2012). Dialects, cultural identity, and economic exchange. *Journal of urban economics*, 72(2-3):225–239.
- 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.
- Ganzer, A., Schmidtlein, L., Stegmaier, J., and Wolter, S. (2020). Betriebs-Historik-Panel (BHP) – Version 7519 v1. Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 100(6):1126–1152.
- 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.

- Kremer, A. (2022). Home is where the history is: How today’s migration in germany is shaped by regional identity. *Journal of regional science*, 62(2):604–638.
- Maraut, S., Dernis, H., Webb, C., Spiezia, V., and Guellec, D. (2008). The oecd regpat database: a presentation.
- 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.
- 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/>.
- 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

## Selection bias

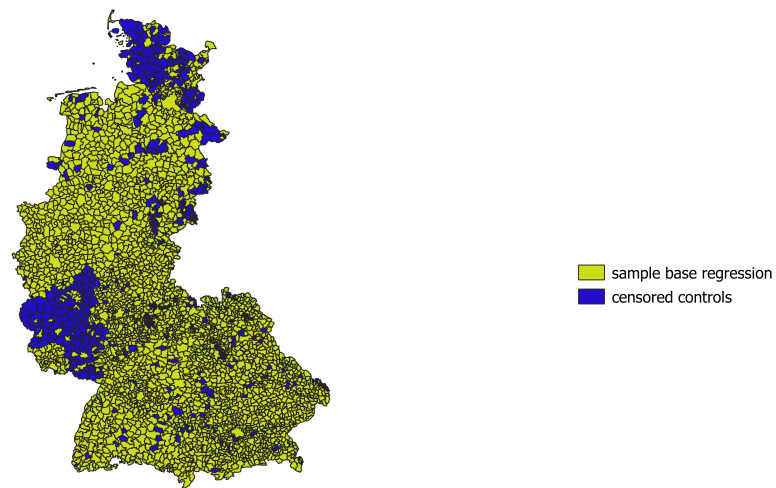


Figure 4: Sample with selection bias due to data censoring

Share of Müllers

Scatterplots

Robustness tests results

Balancing tests

Modeled diversity

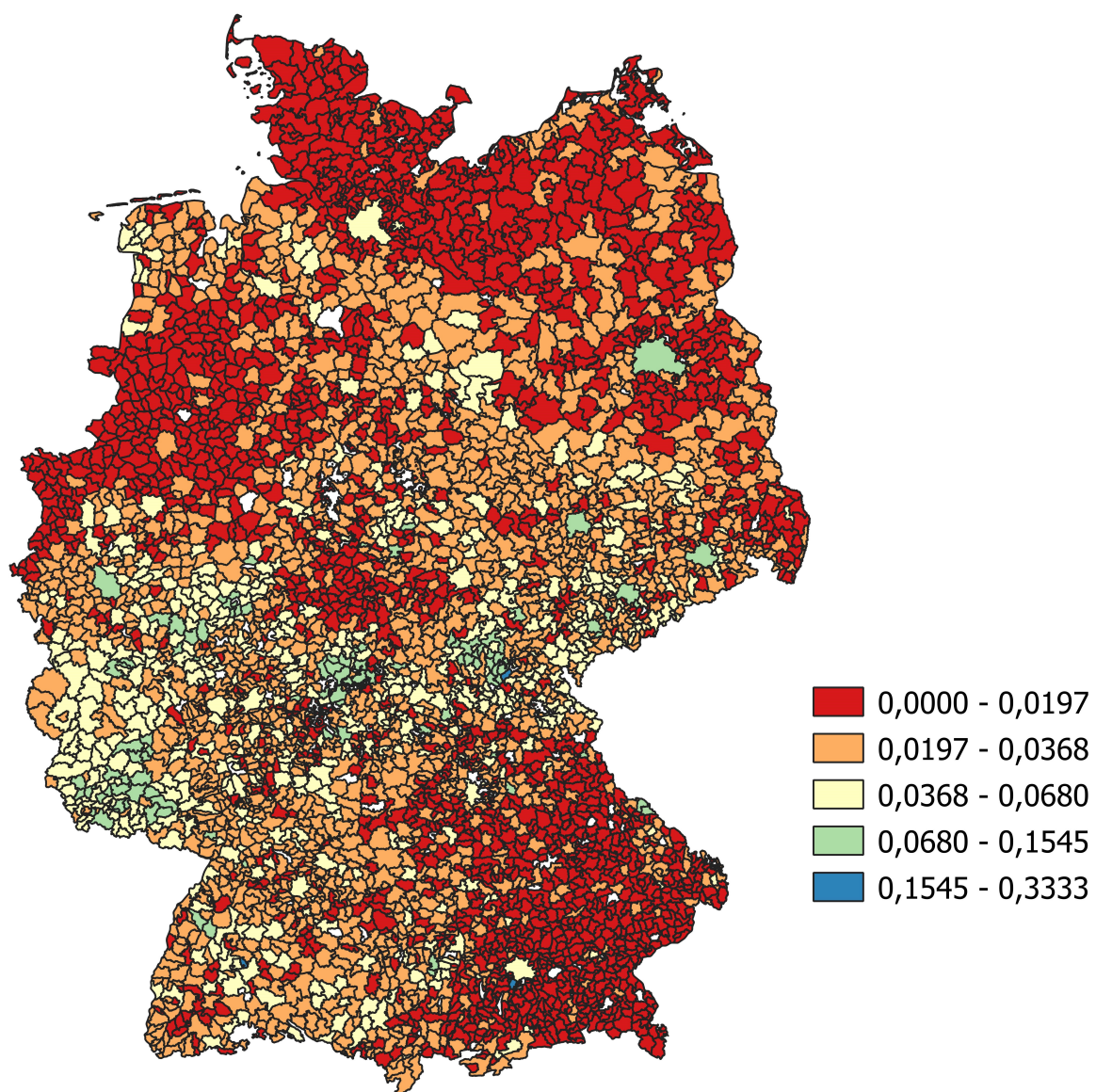


Figure 5: The shares of the family name “Müller” in each municipality show no distinct geographic pattern but appears rather randomly.

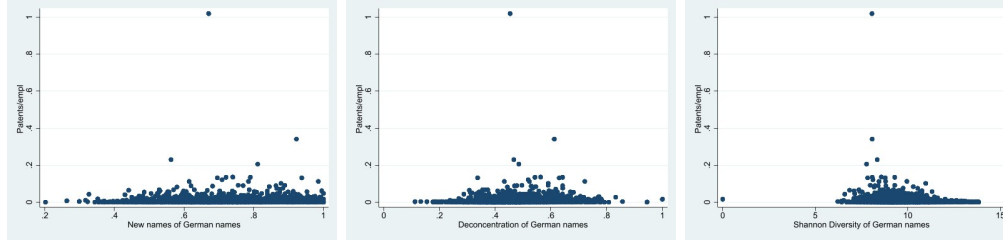


Figure 6: Scatterplots between diversity and  $IHS(patents/employee)$ . If I exclude outliers  $> 0.2$  patents/employee in the regressions, signs or significances of results are stable.

Table 4: Estimation results when Controlling for Manufacturing: OLS regressions

	(1)	(2)	(3)
	Patents/empl	Patents/empl	Patents/empl
New names	0.0105* (0.00585)		
Deconcentration		0.0185** (0.00919)	
Shannon Diversity			0.000906 (0.000941)
Share manufacturing	0.00288 (0.00288)	0.00275 (0.00304)	0.00461 (0.00305)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	837	837	837

Standard errors in parentheses

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

Table 5: Estimation results when Controlling for Share of Engineers and Science: IV (2SLS) regression, Proximity to historical borders as instrument

	(1)	(2)	(3)
	Patents/empl	Patents/empl	Patents/empl
New names	0.0902*** (0.0247)		
Deconcentration		0.103*** (0.0264)	
Shannon Diversity			0.0261*** (0.00833)
Share engineers and science	0.00000181*** (0.000000644)	0.00000153*** (0.000000517)	-0.00000174** (0.000000691)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	3390	3390	3390

Standard errors in parentheses

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

Table 6: Estimation results when Controlling for (Technical) Universities: IV (2SLS) regression, Proximity to historical borders as instrument

	(1)	(2)	(3)
	Patents/empl	Patents/empl	Patents/empl
New names	0.0970*** (0.0252)		
Deconcentration		0.111*** (0.0268)	
Shannon Diversity			0.0257*** (0.00754)
University before 1900	0.00384 (0.00286)	0.00212 (0.00221)	-0.00623 (0.00447)
Technical University before 1900	0.0168*** (0.00554)	0.0111*** (0.00408)	-0.00292 (0.00358)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	3388	3388	3388

Standard errors in parentheses

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

Table 7: Estimation results when Controlling for Population size: IV (2SLS) regressions, Proximity to historical borders as instrument

	(1)	(2)
	Patents/empl	Patents/empl
New names	0.0883*** (0.0240)	
Deconcentration		0.101*** (0.0257)
Population size	4.44e-08*** (1.47e-08)	4.28e-08*** (1.37e-08)
Controls: Urban Type	Yes	Yes
Controls: Labor Force	Yes	Yes
State fixed effects	Yes	Yes
Observations	3390	3390

Standard errors in parentheses

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

Table 8: Estimation results when Controlling for Terrain Ruggedness (TRI) and Waterways: IV (2SLS) regressions, Proximity to historical borders as instrument

	(1)	(2)	(3)
	Patents/empl	Patents/empl	Patents/empl
New names	0.0938*** (0.0274)		
TRI_mean	0.000554* (0.000287)	0.000491* (0.000278)	0.000865** (0.000349)
TRI_stdev	-0.000860** (0.000386)	-0.000795** (0.000376)	-0.00143*** (0.000500)
LENGTH	1.73e-08** (7.07e-09)	1.46e-08*** (5.46e-09)	-1.67e-08** (8.45e-09)
Deconcentration		0.103*** (0.0277)	
Shannon Diversity			0.0241*** (0.00777)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	3390	3390	3390

Standard errors in parentheses

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

Table 9: Estimation results for (1)-(2) Urbanized regions and (3)-(4) Rural areas: IV (2SLS) regressions, Proximity to historical borders as instrument

	(1)	(2)	(3)	(4)
	Patents/empl	Patents/empl	Patents/empl	Patents/empl
New names	0.108*** (0.0308)		0.0172 (0.0115)	
Deconcentration		0.110*** (0.0284)		0.0277 (0.0178)
Controls: Urban Type	Yes	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
Observations	2802	2802	1306	1306

Standard errors in parentheses

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



Table 10: Estimation results for less common names only: IV (2SLS) regression, Proximity to historical borders as instrument

	(1)	(2)
	Patents/empl	Patents/empl
Deconcentration	0.0901*** (0.0238)	
Shannon Diversity		0.0186*** (0.00551)
Controls: Urban Type	Yes	Yes
Controls: Labor Force	Yes	Yes
State fixed effects	Yes	Yes
Observations	3389	3389
Adjusted $R^2$	-0.076	-0.292
Standard errors in parentheses		
* $p < 0.1$ , ** $p < 0.05$ , *** $p < 0.01$		

Table 11: Estimation results Excluding places with less than 1000 obs: IV (2SLS) regressions, Proximity to historical borders as instrument

	(1)	(2)
	Patents/empl	Patents/empl
New names	0.0783*** (0.0233)	
Deconcentration		0.0934*** (0.0246)
Controls: Urban Type	Yes	Yes
Controls: Labor Force	Yes	Yes
State fixed effects	Yes	Yes
Observations	3071	3071
Standard errors in parentheses		
* $p < 0.1$ , ** $p < 0.05$ , *** $p < 0.01$		

Table 12: Estimation results Poisson estimation

	(1)	(2)	(3)
	Patents/employee	Patents/employee	Patents/employee
New names	0.832* (0.486)		
Deconcentration		0.948 (0.816)	
Shannon Diversity			-0.0862 (0.0732)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	3390	3390	3390

Standard errors in parentheses

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

Table 13: Estimation results Extensive margin

	(1)	(2)	(3)
	extensive	extensive	extensive
New names	-1.175 (0.946)		
Deconcentration		-0.204 (1.035)	
Shannon Diversity			0.568*** (0.139)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	1927	1927	1927

Standard errors in parentheses

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

Table 14: Estimation results in Levels: IV (2SLS) regressions, Proximity to historical borders as instrument

	(1)	(2)	(3)
	patentssumperwork	patentssumperwork	patentssumperwork
New names	0.0907*** (0.0254)		
Deconcentration		0.103*** (0.0271)	
Shannon Diversity			0.0256*** (0.00826)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	3390	3390	3390

Standard errors in parentheses

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

Table 15: Estimation results with Labor force measured in numbers: IV (2SLS) regression, Proximity to historical borders as instrument

	(1)	(2)	(3)
	Patents/empl	Patents/empl	Patents/empl
New names	0.0699*** (0.0170)		
Deconcentration		0.0844*** (0.0198)	
Shannon Diversity			0.0186*** (0.00493)
Controls: Urban Type	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Observations	3390	3390	3390

Standard errors in parentheses

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

Table 16: Estimation results for SLX model, distance weighted: IV regressions

	(1)	(2)	(3)
	ih_s_patentssumperwork	ih_s_patentssumperwork	ih_s_patentssumperwork
New names	0.00335** (0.00147)		
new_weighted	0.00336 (0.00693)		
Deconcentration		0.115*** (0.0361)	
deconc_weighted		-0.156*** (0.0527)	
Shannon Diversity			0.0000479 (0.0000923)
shannon_weighted			-0.00124 (0.000811)
Controls and State fe	Yes	Yes	Yes
Observations	3454	3454	3454
Adjusted $R^2$	0.021	-0.431	0.020

Standard errors in parentheses

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

Table 17: Estimation results for SLX model, adjacency weighted: IV regressions

	(1)	(2)	(3)
	ih_s_patentssumperwork	ih_s_patentssumperwork	ih_s_patentssumperwork
New names	0.00292** (0.00141)		
new_weighted	0.00406*** (0.00113)		
Deconcentration		0.0899*** (0.0266)	
deconc_weighted		-0.0182*** (0.00700)	
(mean) shannon			0.0000620 (0.0000919)
shannon_weighted			0.000353*** (0.000103)
Controls and State fe	Yes	Yes	Yes
Observations	3454	3454	3454
Adjusted $R^2$	0.023	-0.256	0.021

Standard errors in parentheses

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

Table 18: Balancing tests IV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	No of plants	No engineers/scient.	Share engineers/scient.	Population density	Unemployment	No of employees	Share 20-24 yo
Dist. Hist. Borders	0.710 (0.699)	0.452 (0.458)	0.467 (0.476)	0.147 (0.216)	1.815 (1.281)	9.120 (11.54)	0.00000312 (0.0000176)
Controls: Urban Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: Labor Force	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3390	3390	3390	3390	3390	3390	3390
Adjusted $R^2$	0.650	0.480	0.478	0.579	0.699	0.632	0.227

Standard errors in parentheses

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

Table 19: Reduced form with modeled diversity

	(1)
	Patents/empl
(sum) flows_x_y	0.000000598* (0.000000361)
Medium sized cities	0.00163*** (0.000504)
Larger Towns	0.00293*** (0.000674)
Smaller Towns	0.00404*** (0.000776)
Rural areas	0.00776*** (0.000839)
Share foreigners	0.0336*** (0.0128)
Share high-skilled	-0.0177 (0.0170)
Share low-skilled	-0.0419** (0.0171)
DE1	0 (.)
DE2	-0.180* (0.110)
DE5	-0.00259*** (0.000692)
DE6	-0.00196*** (0.000612)
DE7	-0.00177** (0.000770)
DE9	-0.00279*** (0.000736)
DEA	-0.00142** (0.000578)
DEB	-0.000194 (0.00107)
DEC	-0.00401*** (0.000583)
DEF	-0.00460*** (0.00122)
Constant	0.00868** (0.00361)
Observations	3390
Adjusted $R^2$	0.024

Standard errors in parentheses

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