Migration and Innovation: The Impact of East German Inventors on West Germany's Technological Development*

Antonin Bergeaud[†] Max Deter[‡] Maria Greve[§] Michael Wyrwich[¶]

December 2024

Abstract

We investigate the causal relationship between inventor migration and regional innovation in the context of the large-scale migration shock from East to West Germany between World War II and the construction of the Berlin Wall in 1961. Leveraging a newly constructed, century-spanning dataset on German patents and inventors, along with an innovative identification strategy based on surname proximity, we trace the trajectories of East German inventors and quantify their impact on innovation in West Germany. Our findings demonstrate a significant and persistent boost to patenting activities in regions with higher inflows of East German inventors, predominantly driven by advancements in chemistry and physics. We further validate the robustness of our identification strategy against alternative plausible mechanisms. We show in particular that the effect is stronger than the one caused by the migration of other high skilled workers and scientists.
Keywords: Patents, Migration, Germany, Iron Curtain, Innovation.
JEL Codes: H10, N44, P20, D31.

*We thank Sebastian Braun, Davide Cantoni, Petra Moser, Hanna Schwank as well as seminar participants at Berlin, Dresden, Marseille, Paris, Regensburg, Strasbourg and Utrecht for the discussion.
[†]HEC Paris, CEP-LSE and CEPR, bergeaud@hec.fr
[‡]Berlin School of Economics, max.deter@gmail.com
[§]Utrecht University, m.greve@uu.nl
[¶]Groningen University, m.wyrwich@rug.nl

1 Introduction

The question of whether the migration of high-skilled workers boosts local innovation activity and productivity has been widely explored in the literature. From the theoretical perspective, endogenous growth models such as those proposed by Romer (1990) and Aghion and Howitt (1992) suggest that migration can significantly influence destination regions. This impact arises through an increased supply of skilled labor and the fostering of knowledge spillovers (Peters, 2022). Empirically, the characterization of such a causal impact poses notable challenges as surveyed by Lissoni and Miguelez (2024). In this paper, we contribute to this topic in the specific context of inventor migration from authoritarian East Germany to democratic West Germany after World War II (WWII). We use a novel identification strategy based on name similarity and report a sizable, positive and causal effect on the innovation rate at the county level.

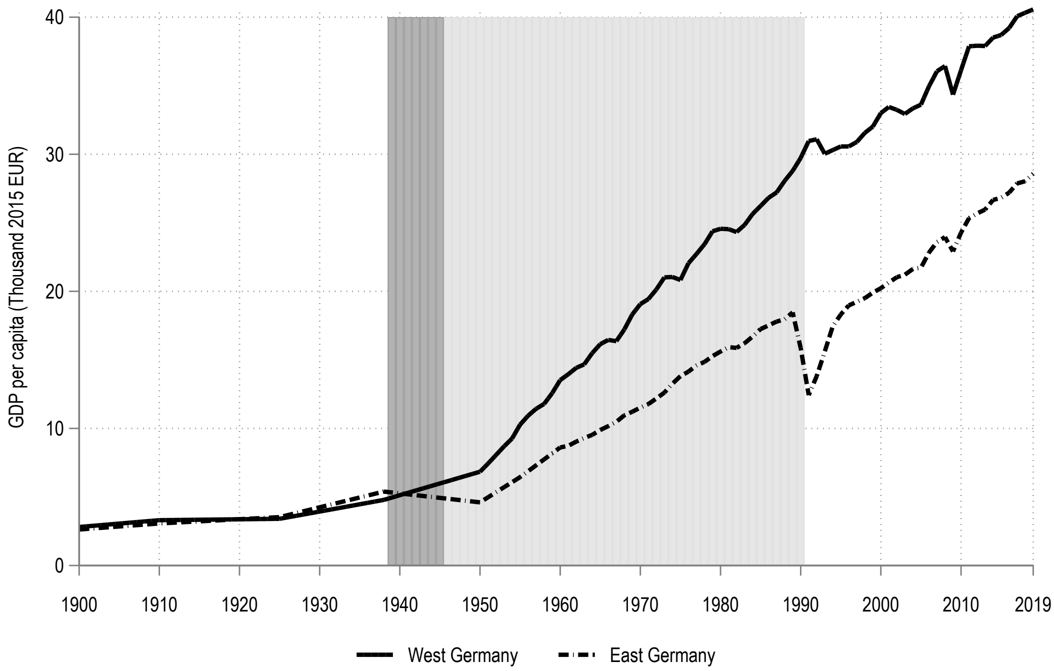
We focus on the unique case of Germany after WWII when the country was divided into democratic and market-based West Germany (*Federal Republic of Germany*, *FRG*) and authoritarian and socialist East Germany (*German Democratic Republic*, *GDR*). After both parts of the country had followed a parallel economic development until WWII, the division into East and West Germany profoundly shaped the country's economic history, more so than almost any other event. Figure I illustrates the stark differences in the long-term economic development between East and West Germany.[1] Until the building of the Wall that surrounded East Germany 1961-89 and made East-West-migration nearly impossible, almost three million people migrated to the West (one quarter of the East German population) motivated by both political and economic factors (Burchardi and Hassan, 2013; Fuchs-Schündeln and Schündeln, 2009). Among them, many members of the so-called *intelligence*[2] (doctors, engineers, scientists). This devaluation of human capital particularly incentivized professionals to leave. Many sought more prosperous opportunities in the West.

We use recently digitized data from two primary sources. First, information is extracted from the universe of West German patents provided by *PatentCity* (Bergeaud and Verluise, 2024). This novel dataset includes the precise location and name of every inventor based in Germany (before WWII) and in West Germany between 1950

[1]The free market economy of West Germany experienced the *Wirtschaftswunder* (economic miracle), benefiting significantly from U.S. economic aid and the Marshall Plan.

[2]The term *Intelligentsia*, initially coined in late 19th-century Russia, was later adopted in Soviet jargon to denote the upper echelon of human capital, often with a discriminatory undertone (Malia, 1960). In this context, the *Intelligentsia* referred to individuals with a university or technical school degree, represented by the compasses in the state emblem, with minimal wage disparity compared to ordinary workers.

FIGURE I. GDP per capita, 1900-2019



Notes: Figure I reports the GDP per capita in West and East Germany. Berlin is excluded. Dark gray-shaded area represents the WWII period (1939-1945) and light gray-shaded area represents the period of separation between 1949-1990 including occupation by the Allies between 1945-1949. Years 1900-1938 are based on Rosés and Wolf (2018). Years 1950-2019 are retrieved from the Maddison Project Database and the values are in 2015 Euros. Data are interpolated linearly during WWII.

and 1989, as well as details about the underlying technology. Second, we incorporate data on East German patentees during the GDR era, sourced from Hipp et al. (2022).[3] This provides us with over 700,000 patents filed between 1930 and 1979.

By looking at the descriptive evidence, we observe a significant gap in innovation activities post-1961 between West German regions that received an abnormally high number of East German inventors and those that did not. However, our objective is to establish a causal link between the inventor influx and regional innovation activities.

Quantifying the causal effect of migration on innovation has often been challenging, mainly due to the non-random settlement patterns of migrants in host regions. Migrants typically choose destinations based on economic factors and anticipated future performance rather than randomly. For instance, Donges and Streb (2024) demonstrate that the location choices of East German firms in West German counties are strongly correlated with the proximity to the East German border and local productivity levels. Since estimating an average treatment effect depends on the assumption

[3]Despite being a state-owned socialist economy, East Germany maintained a patent system aligned with the standards of the World Intellectual Property Organization. More details can be found in Hipp et al. (2022).

that destination choice is effectively random, conditional on observables, this selection bias complicates the estimation process (see Borjas, 2019 for a review).

To establish a causal relationship, our main contribution is the use of a novel instrumental variable to measure the propensity for migration from East Germany to specific West German counties. We assess the relative prevalence of surnames in each West German county and compare this with the list of East German inventors' surnames prior to the construction of the Berlin Wall. While the patent data allows us to identify the surnames of East German inventors, obtaining the complete geographical distribution of surnames is more challenging. To overcome this, we use a dataset of over 3 million missing WWI soldiers, which includes their names and birth locations (see A1 for an example of such a list). As with any instrumental variable, the relevance condition (i.e., surname proximity predicts future migration) and exclusion restriction (i.e., surname proximity is uncorrelated with future innovation performance) must hold. Our identification hinges on two key assumptions. First, surname proximity acts as a pull factor, reflecting shared family or cultural ties, which makes an inventor more likely to choose a county with a higher surname proximity, controlling for other observable economic and geographical factors. Second, surname proximity should not predict future economic or technological development at the local level. Thus, a potential threat to our identification is that the overrepresentation of certain family names, particularly prevalent among East German inventors, could signal characteristics that correlate with future innovation. We address these concerns in detail in Section 5 and provide empirical support for these assumptions. Specifically, we use various vintages of geocoded data from historical German address books to confirm that our innovation results are driven by surname proximity to inventors, rather than correlations with wealth or industrial development.

Our main OLS and IV results demonstrate that patentee migrants from East Germany significantly contributed to fostering innovation growth in West German counties. Specifically, a 1% per 100,000 inhabitants (based on the 1925's population, which is about 1 fourth of a standard deviation) increase in the number of East German migrant inventors between 1945 and 1961 led to an approximately 9% rise in patents filed from 1961 to 1979 — a substantial effect equivalent to the interquartile range (p75-p25).

Addressing potential identification threats, we first validate the representativeness of WWI casualty data using census and address book records, showing a strong correlation between the distribution of names in these sources. Second, we assess whether the East German inventor list might proxy for broader features correlated with innovation by constructing alternative instruments based on high-skilled occupations from over 10 million address book entries (incl. name, location, and occupation); these controls do not alter the first-stage results. Third, we evaluate whether inventor names could reflect specific technological trends, potentially confounding our findings. By estimating the instrument, migration, and innovation effects across eight distinct technology classes (county $\times$ technology panel), we confirm the robustness of the results. Furthermore, we control for technological proximity between East German inventors and West German counties, which might act as a pull factor, yet the results remain consistent and robust across all specifications.

Our paper contributes to an important economic literature that has emphasized that host regions gain from incoming migrants. While out-migration may severely affect the economic and scientific potential of the home country (Nunn, 2008; Akbulut-Yuksel and Yuksel, 2015; Waldinger, 2016; Becker, 2022), host countries usually benefit through increased economic activity and the gain in human capital. For example, the mass migration to the US in the 18^{th} and 19^{th} centuries or the migration of 8 million ethnic Germans to West Germany after WWII caused persistently higher incomes, industrialization, and innovation (Sequeira et al., 2017; Akcigit et al., 2017; Burchardi et al., 2020; Arkolakis et al., 2020; Peters, 2022). This occurs through the gain in knowledge and a higher population density that is beneficial to economic prosperity (Davis and Weinstein, 2002).[4]

In particular, high-skilled immigrants improve human capital and the stock of ideas in the host country (Kerr and Lincoln, 2010). Skilled-worker Huguenot immigration substantially increased productivity in Prussia (Hornung, 2014). Jewish émigrés from Nazi Germany significantly contributed to chemical innovation in the U.S. (Moser et al., 2014). Ferrucci (2020) shows that Soviet inventors that migrated to Germany following the collapse of the Soviet Union increased patent production in technology fields in which the Soviet Union was more active. However, the influx of high-skilled immigrants can also crowd-out existing structures. Borjas and Doran (2012) showed that the sudden influx of Soviet scientific personnel into the United States after 1991 led to a negative productivity effect on US mathematicians, an increased mobility, and no effect on overall scientific productivity in mathematics.

We make four contributions to the literature. First, the historical setting allows to exploit three large, sudden, unexpected yet well-delimited shocks - the separation of Germany into two wholly different countries, the East-West migration of a quarter

[4] Accordingly, Wyrwich (2020) as well as Ciccone and Nimczik (2022) exploit the policy regulation in their identification strategy that prevented refugees from settling in the French occupation zone after WWII, showing that refugees caused a higher population density in West Germany that coincides with long-term higher income and productivity.

of the population, and the erection of the Berlin Wall. In contrast to other migration studies, our set-up takes place between countries that were previously one, sharing the same language, culture, institutions, history, and economic development. This also allows us to exploit strong cultural and family ties in our instrument. Second, we are able to estimate the impact of inventor migration at the very local level, i.e., industries within counties, while previous studies often neglected the local dimension or restrict the analyses to specific technologies. We are thus able to distinguish between the aggregate effect of inventor migration (county level) and the setting where we expect the strongest impact (spillover effects within the same industry and location). Our third contribution is the establishment of a patent database that combines two unique sources on patents in East and West Germany. We identify 13,886 patents of 1,876 inventors that moved from East Germany to West German counties during the period 1945-1961. This is a unique source helping to better understand inventor migration during the Cold War and also helps identifying the (regional) impact of German division; a topic of great interest in Germany today much as the general impact of the Iron Curtain across Europe.

Our fourth main contribution is the development of a novel instrument to estimate the causal effect of high-skilled migration on innovation, based on surname proximity. This approach builds on the literature that uses historical surname registers to establish causal relationships between immigration and local innovation. Prior research indicates that communities tend to cluster socially and geographically based on shared traits such as kinship, ethnicity, and cultural heritage, which can influence interactions and information flows (Agrawal et al., 2008; Kerr, 2008). For example, Posch et al. (2024) show that increased immigration diversifies U.S. counties' social structures, as reflected in surname distributions, and enhances innovation. Extending this line of work, we construct a new dataset integrating records of WWI soldiers, East German inventor names, and innovation in West German counties. By measuring the name proximity between East German inventors and West German counties, we identify a pull factor for migration that is plausibly exogenous to local economic and technological conditions. While we remain agnostic about the precise mechanism captured by surname proximity, we leverage it to estimate the causal impact of inventor migration on innovation in West Germany.

The remaining of this paper is divided as follows. Section 2 presents some historical facts to motivate our approach. Section 3 presents the construction of the data and identification. Section 4 presents the results and discusses. Section 5 deals with identification threats, while Section 6 concludes.

2 Historical Background

2.1 Division and Economic Development

After WWII, Germany was divided in two steps. First, in 1945 the four Allied forces Great Britain, US, France, and the Soviet Union separated the country into four parts. Then, in 1949 the Soviet occupation zone became *East Germany*, i.e., the *German Democratic Republic* (GDR), and the other three parts became *West Germany*, i.e., the *Federal Republic of Germany* (FRG). East Germany was designated to become a role model for the socialist system by Soviet authorities.

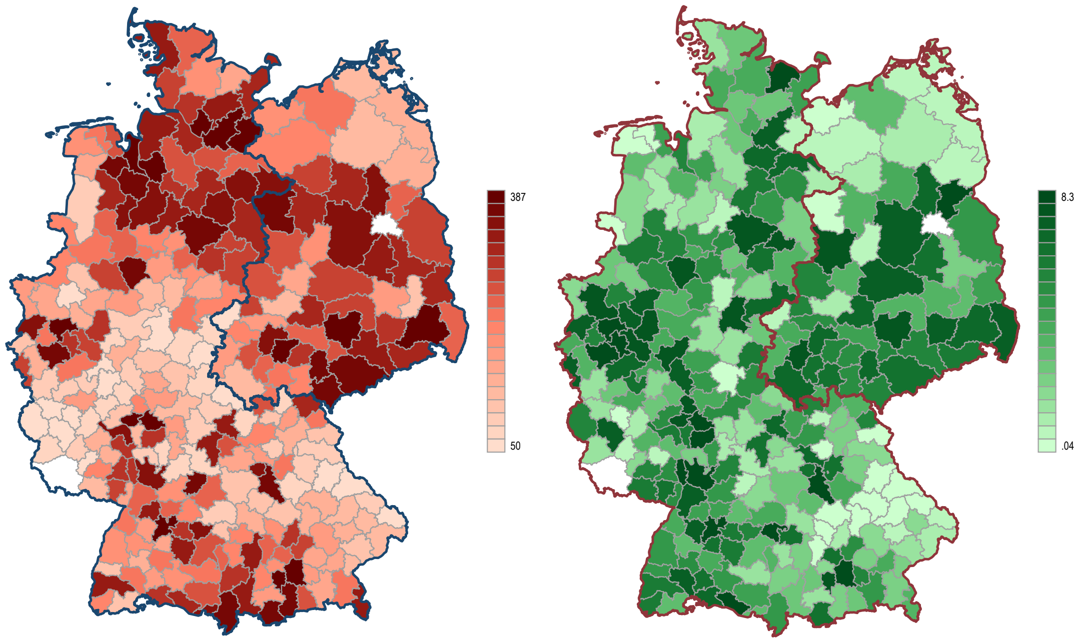
Before the separation, the two parts of the country underwent a relatively parallel economic development. Fritsch et al. (2023) show that patent intensity was very similar before WWII. In fact, before WWII, innovation was highly concentrated in Berlin and other locations with above-average innovation activities included South Saxony in the East, the Ruhr area, Munich, Stuttgart and Frankfurt to a lesser extent in the West. Other than this, inventive activities were relatively uniformly spread across the territory. In terms of economic development, Figure II shows that neither income per capita nor patenting per capita exhibited a clear disparity between East and West Germany.

The division into East and West has shaped the German economic history like hardly any other event. East Germany became an authoritarian socialist regime with a centrally planned economy. It had one of the most rigid systems of former communist states, with the single ruling party SED (Socialist Unity Party) and the Ministry of State Security (MfS), the so-called *Stasi*, repressing opposition by extensive observation, imprisonment, and psychological destruction (*Zersetzung*) (Rainer and Siedler, 2009; Hensel et al., 2009; Grashoff, 2006). On the other side of the Iron Curtain, the democratic, free market economy of West Germany experienced a surge in economic growth, the so-called *"Wirtschaftswunder"*, and benefited fully from economic aid provided by the US and the Marshall plan.

The (fear of) repression in East Germany together with new opportunities in West Germany implied a massive individual out-migration to West Germany. In addition, some highly innovative firms relocated to the West (Donges and Streb, 2023; Falck et al., 2014).[5] Reparations paid to the Soviet Union reduced the East German capital-

[5]Well-known examples include the German electrical engineering company Siemens or the producer of electrical equipment AEG (*"Allgemeine Elektricitäts-Gesellschaft AG"*) (Streb et al., 2006). Further prominent examples include car manufacturers Audi and BMW, or the optical instrument firm Carl Zeiss.

FIGURE II. Income and patent per capita before WWII



Notes: The Figure on the left-hand side reports the value of total income divided by population for each county, both taken in 1925. The Figure on the right-hand side reports the total number of patents filed between 1930 and 1939 by inventors in a given county, divided by population in 1925. Both figures use data from the Census in 1925 and from *PatentCity* Bergeaud and Verluise (2024). Total population in West Germany: 40.2m. Total population in East Germany (including all of Berlin): 19.4m. Counties are aggregated into 281 units to account for changing boundaries (see 3). Berlin and Saarland are excluded.

labor ratio in industry by about one third (Ritschl and Vonyó, 2014). Real output per worker in East Germany in 1954 was only two third of the West German level (see Figure I and Sleifer, 2006).

Today, two waves of mass migration (1945-61 and after 1989) together with four decades of socialism and oppression continue to pose serious gaps in the economic development of East and West Germany. Next to substantially lower wages, productivity, and population density, the region East Germany (excluding Berlin) registers currently only one fifth of the patents of the neighboring Bavaria with a similar population (Fuchs-Schündeln and Izem, 2012; The Economist, 2021). Only seven percent of Germany's most valued 500 companies (none is part of the *DAX*, the 40 major German companies at the Stock Exchange) and very few "hidden champions" have their headquarter in the East (The Economist, 2014).

2.2 East-West Migration

In total, three million people who had lived in the area that became part of the GDR after WWII fled to West Germany between 1945 and 1961.[6] This is about 25 percent of the population (Burchardi and Hassan, 2013; Becker et al., 2020). Figure A3 in the appendix shows Census based East-West migration in 1946, 1950, and 1961. Most East migrants went to Noth West Germany (Schleswig Holstein, Lower Saxony), while border regions generally received more migrants. Legal migration to West Germany was restricted. In 1952, East German authorities built a heavily protected border including a system of barbwire fences, strict border patrols, landmines, and watchtowers. In the following years, it was still possible to illegally cross the border via West Berlin until the unexpected building of the Berlin wall on August 13, 1961 by East German authorities. The building of the Wall made illegal out-migration virtually impossible until the fall of the Berlin wall in 1989. Legal out-migration remained heavily restricted and a relatively rare event. While in the month July 1961, more than 30,000 applied for admission in West Berlin, in 1962 and 1963 only 60 persons applied on average per month (Heidemeyer, 2019). The outflow from East to West decreased from 3 million 1945-61 to 625,000 individuals over the period 1962-88 (Bade, 2018).[7] Migrating East Germans were rather young and strongly positively selected in terms of their skill and occupational composition (Becker et al., 2020; Bauernschuster et al., 2012). In particular specialists, such as doctors, engineers, and scientists massively left the country (Hefele, 1998; Rytlewski and Opp de Hipt, 1987). East Germany lost about one third of their academics during this time.

There were three main incentives for high-skilled individuals to leave the country. First, East Germany's revolutionary upheavals in the economy and society targeted at members of the *intelligence*, which included expropriations, pressures of collectivization, and loss of privileges. Second, the lack of freedom and democracy. Third, the economic opportunities in West Germany compared to the stagnant development in the GDR (Geißler, 2014). The *Stasi* reported in 1959 to the SED that out-migration of "occupational groups of the *intelligence* (especially scientists, doctors, engineers, and technicians) increases substantially" (Ministry of State Security, 1959).[8] For exam-

[6]Although West Germany admitted in the beginning only one third of East German applicants (the quota increased later to 99%), the law determined that all German immigrants had the right to remain in the country, but rejected applicants with some restrictions on mobility. Note also that 400,000 individuals moved from West to East Germany in 1950-61, the majority were GDR refugees that migrated back to the East.

[7]East-West migrants until the 1980s consisted mainly of pensioners that could not contribute to the economic production in the socialist East (Bade, 2018).

[8]Another report of the same year states that main reasons for why doctors leave the country is West

ple, of all pre-war professors at the main universities (Berlin, Leipzig, Halle, Rostock, Greifswald, Jena) only 17 percent remained in their position until 1961.

Although academic and inventor migrants comprised only a small portion, the "Escape of the Mind from the GDR"[9] gained significant attention both in the East and the West. Due to the traditionally high levels of professional and geographical mobility within this demographic group, emigration was particularly accessible to them.

From the West German perspective, the economic integration of refugees from East Germany is generally seen as a "success story" (Baum, 1999) because immigrants were young, rather highly educated, spoke the same language, and integrated quickly. Moreover, West Germany urgently needed workers during their economic upswing after WWII. Anecdotal evidence suggests that the migration of scientists and inventors from East to West Germany contributed substantially to shaping West Germany's economic development. In the next section, we present some relevant evidence.

2.3 Historical Evidence on Patentee Migration and Innovation

The list of prolific migrant inventors from our patent data provides valuable anecdotal evidence of how East-West migration contributed to West German innovation. One example is Hellmut Bredereck (1904-1981), an organic chemist from East German Jena, was resettled to West German Baden-Württemberg in 1945 by the U.S. army. There, he established a company with 180 employees producing saccharin, became rector of the Technical University of Stuttgart, and served as president of the Society of German Chemists. His *Bredereck synthesis* further exemplifies the knowledge transfer from East to West (as an example, a patent document of Bredereck is shown in Table A4 in the appendix).[10]

German headhunting at medicine congresses as well as the hostility towards the *intelligence* and repression in the GDR (of State Security, 1959). Many members of the former academic, administrative, and economic elites experienced a hostile and often violent treatment of the new government that saw them as "leftovers" of a former capitalist society. Many engineers and professors lost their jobs and often experienced prison, lower occupational status, and (if they were lucky), could migrate to the West.

[9]Such is the title of a special edition from the Bulletin of the Press and Information Office of the Federal Government in West Germany, No. 199, 214/1960.

[10]Other examples in our dataset includes *August Karolus* (1893-1972), a physicist in television technology, emigrated from East German Leipzig after his research institute was destroyed during WWII. After a brief stop in Zurich, he settled in West German Freiburg, where he worked as a professor until his death in 1972. His expertise significantly contributed to the West's technological advancements in this field. *Alfred Kraus* (1899-1979), a chemist specializing in nitrocellulose, conducted research in East German Magdeburg before emigrating with his family in 1946. After a brief stay in England, he settled in West Germany, working in Osnabrück, Sythen, and Aschau until 1968, where he published over 40 scientific papers (Wikipedia, 2024). *Alfred Kuhn* (1895-1960), a chemist who worked

This is one example providing clear evidence of the critical role that East German inventors played in driving innovation and scientific progress in West Germany. This individual trajectory reflects a broader phenomenon: the migration of skilled inventors not only sustained their personal careers but also bolstered the innovation activities of their host regions.

Figure III suggests that aggregate impact on follow-up innovation was significant. In this Figure, we report the coefficients from estimating the following difference-in-difference model for every West German county k (228 counties) each year from 1919 to 1989 except 1940-1960:

$$\frac{Y_{k,t}}{pop_k} = \sum_{\tau=1919}^{1989} (\alpha_\tau M_k + \beta_\tau X_k) \mathbb{1}_{t=\tau} + \nu_k + \mu_t + \psi_{t,b(k)} + \varepsilon_{k,t} \quad \text{with} \quad \alpha_{1938} = 0 \tag{1}$$

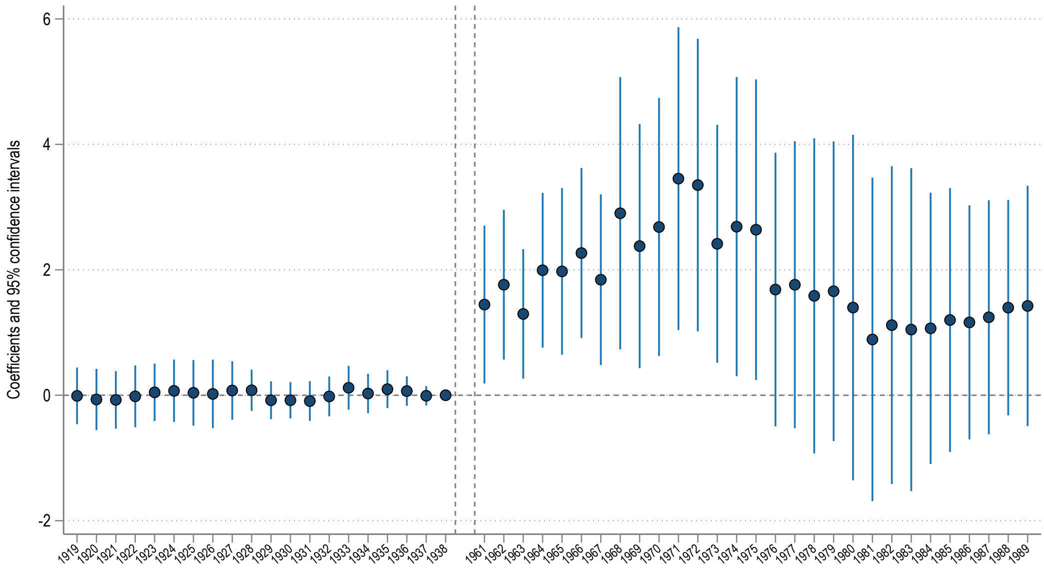
In equation (1), M_k is a treatment variable that is equal to 1 if the county is above the median in terms of the number of incoming migrant inventors from East Germany divided by 1961's population, X is a vector of county specific time unvarying variables: population and income taken in 1939 and the distance to the East German's border, all are taken in logarithm.[11] ν_k and μ_t are two set of fixed effects and $\psi_{t,b(k)}$ is a year-region fixed effects ($b(k)$ at the NUTS2 level which on average includes 8 counties). The coefficients of interest α_τ are plotted in Figure III and suggests that larger inflow of inventor migrants was associated with a larger number of patents per capita right after 1961 and for more than 15 years. The effect ultimately became less precise, although it continues to be positive, as migrant inventors likely retired and other regions progressively caught up. Nonetheless, the effect on the stock of knowledge is sizable and can have long-run effects.

However, the usual challenge emphasized by the literature on migration of innovation prevents us from using this result as an estimation of the average treatment effect. This is why we now turn to a less direct empirical identification.

on penicillin for the pharmaceutical company Madaus in East German Radebeul. After his company was expropriated by socialist authorities, Kuhn moved to West German Frankfurt in 1949 and continued working for Madaus, which also relocated to the West. He simultaneously built his own pharmaceutical business in West Germany.

[11] More details about the source of these data are given in Section 3.

FIGURE III. Difference-in-difference coefficients



Note: Figure III reports coefficients α_τ from an OLS estimation of model (1) along with the 95% confidence interval. The reference year is 1938 which is the last pre-war year. The model include county, year and year-region fixed effects. The standard errors are clustered on a NUTS2 level. The number of counties is 228.

3 Data and identification

3.1 Patents and inventors

We establish the relationship between migrating inventors and innovation by utilizing various sources of digitized and processed patent data from 1930 on. Our primary dataset is *PatentCity* (Bergeaud and Verluise, 2024), which were obtained by utilizing Natural Language Processing (NLP) techniques to digitize granted patent data retrieved from Germany's National Patent Office. Information include the specific date and technology of the granted patent, patentee's full name, and address. An example of such a patent document of formerly mentioned Hellmut Bredereck is in Figure A4. We supplement this dataset with a similar dataset that cover the universe of East German patents after 1950 from the *Office of Inventions and Patents* and curated by Hipp et al. (2022). Figure A2 in the appendix shows the regional distribution of patents over the period 1930-1970.

By combining both datasets, we were able to identify 720,534 patents (116,026 in East Germany) over the period 1930-1979 (the whole country before 1950 and West Germany from 1950 on). Inventors are identified using their full name as it appears in the patent and are localized using their addresses. Using the administrative identifier of the district in *PatentCity*, we can allocate each inventor to East or West Germany. Unfortunately, most patents in *PatentCity* do not contain detailed enough information to split between different part of Berlin and therefore it is impossible for us to know whether an inventor located in the capital belongs to the eastern or western part. We therefore used the additional details contained in (Hipp et al., 2022) to split inventors located in Berlin between the East and the West.[12]

During this period, we count 1,876 inventors who migrated from East to West Germany between 1945 and 1961 according to their residential address. These inventors filed altogether 13,886 patents. We identify such inventors by looking for patentees who filed a patent in East Germany before 1961 and then a subsequent application reporting an address in West Germany between 1945 and 1961 (and not again in East Germany). Figure IV displays the relative importance of patenting from migrant inventors in each district of West Germany based on our data. It becomes obvious that East German patent activity is more pronounced in border regions, but with a substantial level of variation.

Throughout the analysis, the initial unit of observation is a West German *Landkreis* (county). These correspond to NUTS administrative regions level 3 and this restriction is imposed by the geographical detail included in the patents. There are around 330 such regions in West Germany. However, due to border changes we merged some of these districts, in particular we integrated *Stadtkreis* (independent cities) into the district that contained them. This results in a dataset of 228 geographical units which are the final unit of observation in baseline our analysis.

Our main outcome is the number of patents filed between 1961 and 1979 at the county level. We retrieve this information from *PatentCity* and include every patent, regardless of the technology. In case of multiple inventors located in different counties, we assign a corresponding share in our count.

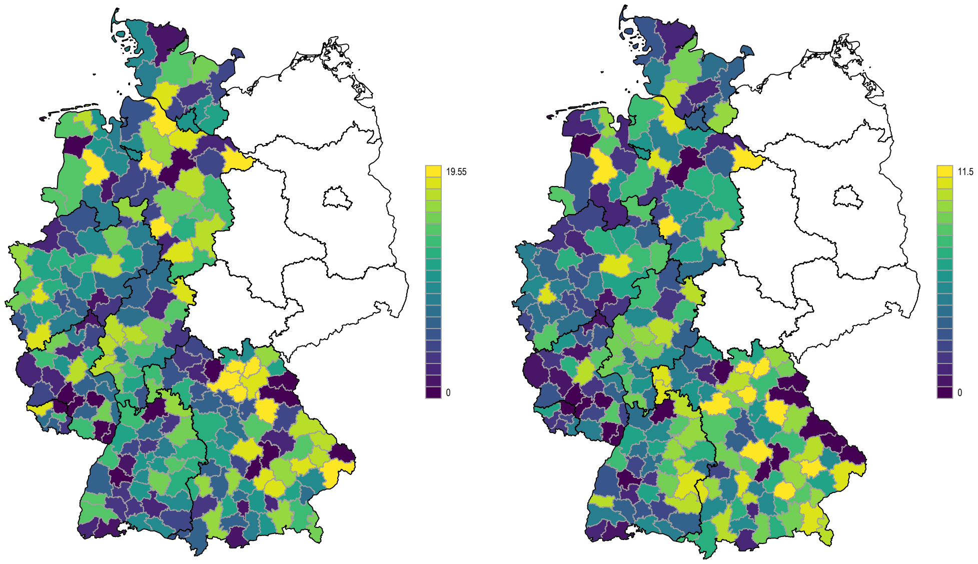
3.2 Census and further data

We augment our dataset with several information drawn from various vintages of the Census and the German Local Population Database. This allows us in particular to calculate various per capita indicators (Statistik des Deutschen Reichs, 1927; Roesel, 2022). We also use some further sources of historical data. The information regarding war-time destruction is sourced from the 1950 housing census *(Gebäude- und Wohnungszählung)* (Statistisches Bundesamt, 1956). Data on expellees is derived

[12]The availability of more precise address information allowed us to allocate inventors to specific quarters of the city, which were located either in East Berlin or West Berlin.

FIGURE IV. Innovation by East German migrant inventors

(a) Average share per year (b) Stock



Note: Figure 4(a) reports the average value of the number of patents filed yearly by "migrant" East German inventors over the total number of patents filed in the district over the period 1950-1970. Figure 4(b) reports the total number of patents filed by "migrant" East German inventors between 1950 and 1970 divided by the total number of patents filed in the district during the period 1934-1950.

from censuses conducted in 1950 and 1961.[13] The historical income data stems from the Statistics of the German Empire (Statistisches Reichsamt, 1929). These information will be useful to control for observable differences in socio-economic characteristics across counties.

3.3 Instrumental variable and identification assumptions

Given the endogeneity of the inflow of migrant inventors in a model explaining the number of patents after 1961, we build an instrument. Indeed, migrants sort endogenously into the most dynamic areas. In particular, inventors anticipate future economic benefits and migrate to cities and regions with the most favorable conditions for their research, such as research clusters and universities (Donges and Streb, 2023). This means that the correlation between migration of inventors and patents

[13]Expellees refer to Germans who originated from the former Eastern territories, which are now part of Poland and Russia. This group is subdivided into Sudeten Germans (*Sudetendeutsche*) who lived in Czechoslovakia before WWII, and *Volksdeutsche*, who were dispersed across various Central and Eastern European territories before 1945 (Wyrwich, 2020).

may be spurious.

The general idea is thus to measure a pull factor for East German inventors in each West German county that is unlikely to correlate with future innovation dynamics, holding all other observable factors constant. This instrument is constructed by examining the proximity between the set of East German inventors' names to the geographical distribution of names across regions in the West.

This argument relies on the assumption that, all else being equal, an inventor is more likely to migrate to a county where names similar to theirs are more prevalent than in other areas. Indeed, a significant driver of East-West migration of scientists and academics was the presence of personal connections in the West. Particularly in the immediate post-war years, family reunification in West Germany was the most common reason for out-migration, as many migrants had family ties there, which lowered social barriers to migration (Bispinck and van Melis, 2015; Heidemeyer, 2019).

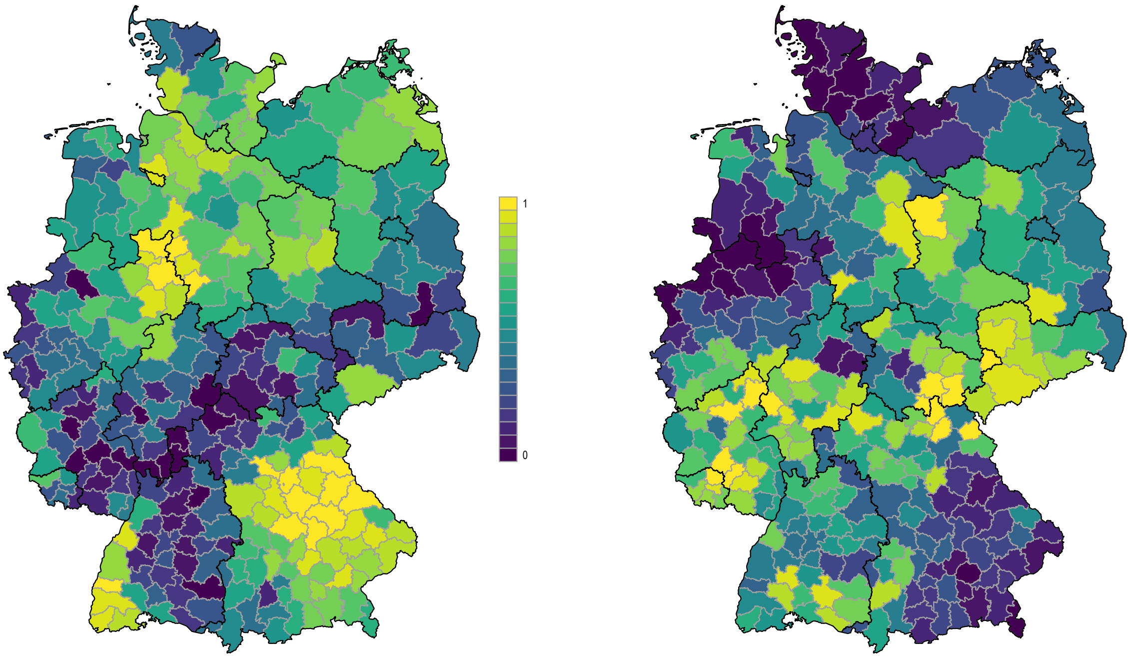
Proximity in names may also indicate cultural similarities between regions. A high concentration of people with similar names suggests cultural ties or specific cultural connections between these regions. Cultural ties are important for migration decisions (Guiso et al., 2009), as shown in various regional contexts (Falck et al., 2012a; Kremer, 2022). For example, dialect similarities in Germany have been found to align with religious similarities, reflecting historically rooted trade and migration patterns. Specific historical migration waves may also contribute to name similarity (Stoeckle, 2010); for anecdotal evidence, see Falck et al. (2012a).

The surname may also reflect an individual familial migration history, particularly for inventors who left East Germany after World War II. For example, someone with the last name 'Meier' in Central Germany, where this name was historically uncommon, might be a descendant of people from West German regions where 'Meier' was more prevalent. The inventor might be inclined to move back to areas where their family originated, not necessarily due to family ties, but because of a sense of cultural familiarity rooted in their family background. Similarly, an inventor with the surname 'Huber' living in Eastern Germany might be more likely to migrate to Bavaria, where the name is very common. Likewise, an inventor with a Slavic surname could be more likely to settle in the Ruhr area, where a significant proportion of the population has Polish surnames.

A fully geocoded database of individual Germans, including their full names, is unfortunately unavailable for the pre-1961 period. To measure the prevalence of each name, we therefore rely on surnames and birth locations drawn from the lists of missing soldiers in World War I. More specifically, we use digitized entries from the

FIGURE V. Name Density

(a) Meyer (b) Müller



Note: Map 5(a) shows for every county the relative density of the name 'Meyer' (as calculated in equation (2), where 1 means the highest prevalence of 'Meyer' in this county. Map 5(b) shows the same for the name 'Müller'.

Verlustliste,[14] i.e., lists of missing soldiers from World War I published in the *Verordnungsblatt* (the Army's Ordinance Sheet; see Figure A1 for an example). These lists represent official, personal disclosures by the Prussian government between 1914 and 1919, covering the entire German Empire's armed forces, including units from Prussia, Bavaria, Württemberg, Saxony, the Imperial Navy, and the Imperial Protection Force. They record soldiers who were killed, missing, wounded, in captivity, or released from captivity (De Juan et al., 2024). Published almost daily immediately after the war, these lists allow us to assume that the geographical distribution of missing soldiers' surnames serves as a reasonable proxy for the general distribution of names. We return to this issue in the following section.

Formally, the name density measure is calculated by

$$F_k(n) = \frac{\eta(n,k)}{\eta(n)} \tag{2}$$

[14]Data are available from nvk.genealogy.net.

where k is a West German county and n is a given name. $\eta(n,k)$ is the ratio of the number of occurrence of n over the total number of missing soldiers located in k and $\eta(n)$ is the same share but over the whole country (West Germany). The measure therefore shows the relative prevalence of a name for a given county. Figure V shows the dispersion of the name density measure for two popular surnames: Meyer and Müller.

To construct our instrument, we first calculate the Levenshtein distance between each name in the set of inventors located in East Germany during the period 1930-1961 (10,314 distinct names) and the list of names from the WWI missing soldiers (distinct 234,454 names). The Levenshtein distance measures the semantic proximity from the number of permutations needed to transform one word to another. Hence, for each region k in West Germany, we calculate the following distance:

$$I_k = \log\left(\sum_{n\in\mathcal{N}(k)}\sum_{n'\in\mathcal{N}} Lev(n,n')F_k(n)P(n')\right)$$

where $\mathcal{N}(k)$ is the set of names in county k from the soldier register and $\mathcal{N}$ is the set of all names in the list of East German inventors from 1930 to 1961. I_k is therefore a weighted average of the relative density using the semantic distance as a weight.[15] Thus, the more prevalent East German inventor names are in a given county k, the larger the value of I_k. We censor the value of $Lev(n,n')$ to 90, i.e. we set its value to 0 if the proximity between n and n' is lower than 90 over 100. We did that on the one hand to allow for some flexibility in how the names are spelled but at the same time to reduce the risk of aggregating over a large number of weakly correlated proximity. $P(n')$ is the number of patents filed by inventors with a name n' during the period 1930-1961 so that names that are more frequent among inventors, or more prolific inventors, are given a higher weight. Finally, we take the logarithm to interpret the effect multiplicatively.

From a technical point of view the instrument is somewhat related to the shift-share approach developed by Card (2001) and exploits the tendency of new immigrants to

[15]We use a standard transformation from the Levenshtein distance which consists in measuring:

$$Lev(n,n') = 100\left(1 - \frac{\tilde{Lev}(n,n')}{\max\left(Len(n),Len(n')\right)}\right)$$

where $\tilde{Lev}(n,n')$ is the baseline Levenshtein measuring the smallest number of permutations required to go from n to n'. This standardization transforms the Levensthein distance into a proximity measure which is equal to 100 if the two names are the same and to 0 if there is no common letter.

choose areas where previous immigrants of the same origin have settled in order to benefit from local co-ethnic networks. The approach suggested by Card (2001) has been extensively used in the immigration literature (Campo et al., 2022; Burchardi et al., 2020). Similarly to our approach, Falck et al. (2012b) estimate the effect of cultural ties between German regions on economic exchange. As a proxy for cultural ties they use the proximity of different dialects in the German language. For example, Dorner et al. (2016) show that West German regions with (historically determined) stronger social ties to the East attracted more East German inventors after 1989.

Our analysis relies on the Stable Unit Treatment Value Assumption (SUTVA), which requires that (i) the treatment effect of inventor migration on innovation in a given county is independent of migration to other counties, and (ii) the treatment effect is consistent across regions and not dependent on the migration mechanism. Potential violations, such as knowledge spillovers between counties or regional heterogeneity in innovation dynamics, could influence our results. We mitigate these risks by controlling for county-level pre-trends, regional fixed effects, and conducting robustness checks on technological proximity between names and regions.

3.4 Threat to identifications

Before showing the main results, we discuss the main threat to identification and briefly describe how we addressed them. The corresponding tests, when applicable, will be presented in Section 5.

Selection The first potential issue with our instrument is that the measure of exposure to East German inventors may not accurately reflect actual exposure levels, as the prevalence of a name in our dataset may not correspond to its true local prevalence. There are two primary reasons for this. First, the probability of showing up in the *Verlustliste*, conditional on deployment to the front lines, is not random. For instance, wealthier households may have leveraged political influence to avoid conscription early in the war or were more likely to belong to officers, who faced lower fatality risks. Second, the probability of being sent to the front line itself was not random. Scientists and industrialists, for example, were often kept behind the lines to contribute to the war effort.[16] Consequently, names common among wealthier fam-

[16]However, historical evidence suggests that no groups were substantially exempted from conscription. The legal option of substitution (paying someone to serve in your place) from the 19^{th} century was abolished. Thus, universal conscription led to a diverse pool of recruits, including skilled workers, professionals, and students from all levels of society. Moreover, many soldiers came from the edu-

ilies or more frequently used in scientific occupations might be underrepresented in our fatality list, which we used to measure the density of names in each county.

Although these issues are valid concerns, they would only affect our estimates under two conditions. First, if they systematically biased the instrument towards counties that experienced greater economic and technological development post-1961 and second if these names are indeed overrepresented in the list of inventors. In addition, the bias is likely to play against our positive effect, as under these conditions, our instrument will underestimate the proximity between a dynamic county and the list of inventors.

We present evidence suggesting that this mechanism is unlikely to apply in our context. In Section 5, we conduct validation tests. Firstly, we demonstrate a high correlation (88%) between population share in the 1910 census and the corresponding share based on missing soldiers at the county level, indicating that the missing soldier lists are numerically representative. This correlation rises to 97% when excluding two outlier counties, *Kreisfreie Stadt Hamburg* and *Kreis Pinneberg*, without altering our results. Secondly, we address potential bias from mismeasurement of name density among missing soldiers. We show that names with higher mismeasurement are not prevalent in the East-German inventor list or in occupations typically associated with economic and innovation development. To confirm this, we analyze early 20$^{\text{th}}$-century address books (details in Section 5.1), which include extensive occupational data. Additional methodological details are provided in Section 5.3.[17]

Predictive Power of Name Density One potential challenge arises from the concern that the distribution of surnames may not be sufficiently dispersed to provide enough statistical power for our instrument. However, even among relatively common surnames, there is substantial regional variation across German-speaking areas (Kunze, 2003). Although Müller (derived from the occupation of miller) is the most common surname in the German-speaking world, other surnames predominate in specific regions. For example, Schmidt (derived from the occupation of smith) is more prevalent in eastern Low-German-speaking areas.[18] Similarly, the surname Meyer (referring to

cated middle class and professional backgrounds (university students, lawyers, doctors, etc.) because of a strong tradition of civic and national duty (Bessel, 1993).

[17]Further information about these address books is available in Section 5.1. This dataset offers an alternative list of German names, along with occupations and locations, covering the 20$^{\text{th}}$ century. However, due to inconsistent geographic coverage, it was not used to construct our baseline instrument, favoring the WW1 casualty register instead.

[18]The spelling of Schmidt also varies by counties (e.g., Schmidt, Schmitz, Schmitt, Schmid, Schmied).

a farm administrator)[19] is especially common in Low-German-speaking regions like Lower Saxony, where it often surpasses Müller in frequency. However, in Central Germany, Meyer is less common, as Hof(f)mann was the typical occupational name for a farm administrator in this area.[20] We illustrate this pattern of regional name density for two of the most common German names Meyer and Müller in Figure V. These examples highlight how regional variations in agricultural and socioeconomic practices have shaped the diversity of surnames across Germany. This significant heterogeneity in surname prevalence, even among common names, gives us confidence that the instrument is sufficiently dispersed to produce robust statistical power.

Exclusion restrictions As with any instrument, the exclusion restriction requires careful examination. In our context, this implies that the composition of surnames in a county, relative to the national distribution of surnames, being more similar to that of inventors based in East Germany before 1961, should not be correlated with the dependent variable—the number of patents filed by all inventors in the county after 1961. We identify two potential backdoors that might invalidate this assumption.

First, it is possible that proximity to a list of inventor surnames could simply reflect the scientific potential of a given county. This could occur if these surnames were not fundamentally different from those more prevalent in skilled occupations. In particular, surnames that are historically associated with certain ethnic or religious groups, or other socio-economic factors, might have been over-represented among workers in occupations requiring more education or scientific training. Such surnames may also be linked to families who, due to their historical or cultural background, were more likely to pursue careers in science, technology, or other highly skilled fields.

Second, there is a concern that some of the inventor surnames on our list may belong to families that were prominent in industrial or entrepreneurial sectors. If certain surnames are historically linked to famous industrialist families, this could introduce bias, as such families may have influenced both the local economy and the likelihood of technological innovation in these regions. In the regression, we can control for economic and technological characteristics before the treatment, so these threats would only be relevant if they would impact the local dynamics of innovation *after* 1961. In any case, we will use the address books to calculate different instruments based on

[19]The name also appears in various forms, such as Maier or Meyer, depending on the region.

[20]In contrast, Bauer (meaning farmer) is the most common surname in Bavaria, which is part of the eastern Upper-German-speaking region. Additionally, certain rare surnames are regionally concentrated, such as Huber (denoting a large landholding farmer), which is common in southern Bavaria, excluding Munich.

alternative lists of names that are associated with scientific and skilled activities. We show that our instrument, while being correlated with these alternative instruments, is still significantly positively associated with future patenting activities even when controlling for these other instruments. More details are given in Section 5.4.

Technology Finally, a remaining concern, although related, is that the positive correlation between patents and our instrument may simply reflect technology-specific trends. If the surnames of inventors are heavily concentrated within a specific technology, and if this technology experienced a boom after 1961, our results could be driven by this trend rather than by the migration of inventors. To address this concern, we first demonstrate that the positive impact of migration holds in a model where the dependent variable is defined at the technology-county level, and the flow of migrant inventors (and corresponding instruments) is measured at the same level. In addition, we construct a measure of the technological proximity between each county and East-Germany based on the set of patents filed before 1961. This proximity is clearly a pull factor for East-German inventors and could potentially be correlated with a surname proximity if names are indicative of a specific industry. We thus include it as a control variable in our baseline model.

4 Results

4.1 Baseline estimation

Our main dataset is made up of counties (*Kreise*) located in West Germany. As explained in Section 3, we merged some observations in order to have time-consistent boundaries, resulting in 228 areas. For each of these observations, we observe the yearly number of patents filed by inventors residing in the county, the number of incoming migrant inventors from East Germany, and a number of control variables. We can also construct our instrument as defined in the previous section. Basic descriptive statistics are given in Table I.

From this, we estimate the following model:

$$Y_k = \beta M_k + \gamma X_k + \psi_{b(k)} + \varepsilon_k, \tag{3}$$

where Y_k is our measure of innovative activities after 1961, measured, for example, using the number of patents filed from 1961 to 1979 per capita or in logarithms. M_k is the total number of migrant inventors coming from East Germany to county k, standardized by population in 1925. X_k is a vector of control variables that includes distance to the East German border, total income in 1925, population in 1939 (both in logarithms), the share of the manufacturing sector in 1939, war destruction, and the number of incoming expellees from the Eastern territories immediately after the war. These controls are similar to the ones used in equation (1) and correspond to important pull factors that would explain a higher incoming flow of East German migrants. Finally, $\psi_{b(k)}$ is a region-specific fixed effect that captures the average dynamic of innovation at a broader local level and ensures that the identification comes from comparing similar counties within the same region. To account for potential spatial correlation in the error term ε_k, we cluster standard errors at the regional level $b(k)$.

We first estimate this model without using our instrument. The results are presented in Table II. Columns 1 and 4 use the logarithm of the number of patents filed in county k between 1961 and 1979 as the dependent variable Y_k, columns 2 and 5 uses the number of patents divided by population in 1939. We use 1939 in order to avoid possible correlation of measurement errors with the main regressor, which is standardized by population recorded in 1925. Columns 3 and 6 consider a count model and uses Poisson Quasi Maximum Likelihood Estimation (PQMLE).

In all cases, the estimation of β is positive and statistically significant. In our preferred model in column 4, which controls in particular for the stock of patents before 1950 as a measure of the innovative intensity of the county, the coefficient on the inflow of migrant inventors should be interpreted as a semi-elasticity: increasing the number of incoming inventors from East Germany per thousand inhabitants by one standard deviation (0.042, see Table I would result in an additional 27% patents filed over the whole 1961-1979 period ($6.340 \times 0.042 = 0.26628$).

However, endogeneity of migration with respect to future innovation prevents us from interpreting this magnitude as the true value of β, given that using OLS or PQMLE is likely to yield biased results. There are many reasons for such endogeneity. An upward bias can arise from positive selection and reverse causality: more innovative counties may attract a larger number of migrant inventors, causing M_k to be correlated with unobserved determinants of Y_k. We expect the control for the initial patent stock to account partly from this and indeed including this variable in the model reduces the effect by a quarter. Conversely, a downward bias may occur due to negative selection if migrant inventors settle in counties that are less innovative, or from measurement error in M_k, both leading to an attenuation of the estimated coefficient β. To address these concerns, we now use our instrument for M_k.

TABLE I. Descriptive statistics of main variables

Variable	Mean	Standard Dev.
Number of patents filed 1961-1979	716	1416
Migrants inventors (1950-1961) / 1925 pop ($\times$1000)	0.041	0.042
Population in 1939	172,529	227,223
Minimal distance to East Germany (in km)	127	89
Total income in 1925 (in thousands of Reichsmark)	6914	5101
Manufacturing share of employment in 1939	0.308	0.102
War destruction (share of houses destroyed in WWII)	0.163	0.163
Number of incoming expellees per capita in 1950	0.19	0.09

Notes: More details on how each variable is constructed is give in Section 3.

TABLE II. Correlation between migration of inventors and subsequent innovation

	(1) OLS (log)	(2) OLS (per capita)	(3) Poisson	(4) OLS (log)	(5) OLS (per capita)	(6) Poisson
Migrant inventors	8.365***	0.073***	8.448***	6.340***	0.065***	6.736***
	(1.239)	(0.019)	(1.150)	(1.420)	(0.020)	(1.315)
Distance (log)	0.099	0.000	-0.046	0.121	0.000	-0.001
	(0.107)	(0.001)	(0.104)	(0.122)	(0.001)	(0.111)
Population 1939 (log)	0.981***	-0.000	0.982***	0.670***	-0.002*	0.721***
	(0.109)	(0.000)	(0.084)	(0.117)	(0.001)	(0.132)
Total income 1925 (log)	-0.036	-0.001	-0.005	-0.072	-0.001*	-0.006
	(0.080)	(0.001)	(0.100)	(0.086)	(0.001)	(0.096)
Manufacturing share 1939	2.588***	0.007**	1.870***	2.105***	0.005*	1.538***
	(0.634)	(0.003)	(0.432)	(0.619)	(0.003)	(0.499)
War-time destruction	0.596	0.001	-0.151	0.466	0.000	-0.279
	(0.686)	(0.003)	(0.391)	(0.682)	(0.003)	(0.345)
Expellees from Eastern territories	1.862	0.001	0.875	1.772	0.001	0.981
	(1.224)	(0.003)	(1.023)	(1.094)	(0.004)	(0.987)
Pre-1950 patent stock (log)				0.241***	0.001*	0.188***
				(0.082)	(0.000)	(0.072)
R-squared	0.790	0.543		0.804	0.554	
N	228	228	228	228	228	228

Notes: This table presents various estimation of model (3). The unit of observation is a West German county. The dependent variable is the number of patents filed from 1961 to 1979 by inventors residing in the county. Columns 1 and 4 take the dependent variable in log and columns 2 and 5 standardize it by 1925's population. These four columns presents estimates of the coefficients using the OLS. Columns 3 and 6 use a Poisson Quasi Maximum Likelihood estimation of the coefficients where the dependent variable is taken in level. Migrant inventors correspond to the number of inventors with a patent filed in East Germany between 1930 and 1960 that migrated to West Germany per thousand of inhabitant, measured using population in 1925. Definition of control variables are given in Table B1. All models include region (NUTS2) fixed effects. Standard errors are clustered by region.

4.2 First Stage

I_k has been built from name proximity and should in principle not be directly related to future innovation development, except through its impact on M_k. Before turning to a two stage least square estimation of equation (3), we first show that this instrument indeed is relevant in the sense that it is positively associated with larger incoming migrant inventors conditional on covariates. We thus estimate the first stage equation:

$$M_k = \phi I_k + \delta X_k + \psi_{b(k)} + \eta_k. \tag{4}$$

Results are presented in Table III where we progressively saturate the model with more control variables, and confirm that our instrument I_k is conditionally positively correlated with the number of migrant inventors, even when controlling for broader regional fixed effects.

TABLE III. First stage model: correlation between name proximity instrument and inventor migrations

	(1)	(2)	(3)	(4)
I_k	0.031**	0.027**	0.044***	0.039***
	(0.014)	(0.012)	(0.008)	(0.008)
Distance (log)		-0.008*	-0.001	0.001
		(0.004)	(0.005)	(0.005)
Population 1939 (log)		0.004	0.011	-0.019***
		(0.006)	(0.006)	(0.006)
Total income 1925 (log)		0.001	-0.009	-0.010
		(0.007)	(0.009)	(0.007)
Manufacturing share 1939		0.148***	0.100***	0.039
		(0.031)	(0.034)	(0.030)
War-time destruction		-0.004	0.012	-0.003
		(0.015)	(0.032)	(0.025)
Expellees from Eastern territories		0.005	0.013	0.004
		(0.036)	(0.038)	(0.042)
Pre-1950 patent stock (log)				0.021***
				(0.004)
Fixed Effects				
Region			✓	✓
R-squared	0.025	0.164	0.284	0.414
N	228	228	228	228

Notes: This table presents various estimation of model (4). The unit of observation is a West German county. The dependent variable is the number of inventors with a patent filed in East Germany between 1930 and 1960 that migrated to West Germany per thousand of inhabitant, measured using population in 1925 (M_k). Definition of control variables are given in Table B1. Models of columns 3 and 4 include region (NUTS2) fixed effects. Standard errors are clustered by region.

The result presented in column 4 suggests that an increase of I_k by 1 standard deviation (0.23) corresponds to an additional 0.00897 inventors per thousand inhabitants, which is about a quarter of the interquartile range. With an elasticity of the distance taken from column 2 and equal to 0.008, this suggests that such an increase is equivalent to reducing the geographical distance to East Germany by a factor 3 ($\exp(0.23 \times 0.039/0.008)$).

4.3 The causal impact of migration on innovation

We now turn to the full estimation of the model using the instrument I_k, several alternative specifications are presented in Table IV. Columns 1 to 4 estimate the direct effect of the instrument of the dependent variable (the reduced form, whether in log or per capita), and columns 5 to 8 estimate the model using a two-stage least squares (2SLS) procedures. In all cases, the coefficient of interest is positive. Every model using the IV rejects the null hypothesis of the underidentification test and the first stage is relevant as measured by the first stage F-stat (see also Table III).

TABLE IV. Full model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I_k	0.441*	0.002*	0.366*	0.001**				
	(0.237)	(0.001)	(0.189)	(0.001)				
Migrant Inventors					10.077*	0.038**	9.262*	0.032**
					(5.042)	(0.017)	(4.944)	(0.014)
Distance (log)	0.088	0.000	0.127	0.000	0.102	0.000	0.119	0.000
	(0.096)	(0.001)	(0.131)	(0.001)	(0.113)	(0.001)	(0.127)	(0.001)
Population 1939 (log)	1.070***	0.000	0.553***	-0.003***	0.962***	-0.000	0.727***	-0.002***
	(0.129)	(0.001)	(0.119)	(0.001)	(0.130)	(0.000)	(0.104)	(0.001)
Total income 1925 (log)	-0.111	-0.001	-0.139	-0.001	-0.022	-0.001	-0.043	-0.001
	(0.110)	(0.001)	(0.099)	(0.001)	(0.103)	(0.001)	(0.115)	(0.001)
Manufacturing share 1939	3.418***	0.013***	2.347***	0.007**	2.410***	0.009**	1.987***	0.006**
	(0.748)	(0.004)	(0.671)	(0.003)	(0.636)	(0.004)	(0.563)	(0.003)
War-time destruction	0.674	0.001	0.420	0.000	0.558	0.001	0.447	0.000
	(0.756)	(0.003)	(0.690)	(0.003)	(0.669)	(0.003)	(0.674)	(0.003)
Expellees from Eastern territories	1.992	0.002	1.828*	0.001	1.860	0.001	1.793	0.001
	(1.194)	(0.003)	(1.037)	(0.004)	(1.234)	(0.003)	(1.149)	(0.003)
Pre-1950 patent stock (log)			0.375***	0.002***			0.177	0.001**
			(0.079)	(0.000)			(0.117)	(0.000)
F statistic (first stage)					32.50	32.50	24.30	24.30
R-squared	0.740	0.275	0.781	0.378	0.681	0.370	0.696	0.395
N	228	228	228	228	228	228	228	228

Notes: Columns 1-4 of this Table present regression results from estimating the model (3) replacing M_k by the instrument I_k (reduced form model). Columns 5-8 estimate model (3) using the 2SLS procedure where M_k is instrumented by I_k. The unit of observation is a West German county. The dependent variable is the number of patents filed from 1961 to 1979 by inventors residing in the county. Columns 1, 3, 5 and 7 take the dependent variable in log and columns 2, 4, 6 and 8 standardize it by 1939's population. Definition of control variables are given in Table B1. All models include region (NUTS2) fixed effects. Standard errors are clustered by region. In columns 5-8, the F-stat of the first stage is obtained using Kleibergen and Paap (2006) as calculated by Baum et al. (2002).

The positive and significant coefficients indicate a causal positive impact of the migration of East German inventors on local innovation in West German counties. These coefficients are meant to correspond to a local average treatment effects (LATE), identified for those inventors whose migration decisions were influenced by surname similarity between East and West German regions. Their magnitudes suggest that such migrations indeed significantly boost local innovation: increasing the number of migrant inventors by 1% per 100,000 inhabitants (based on the 1925's population which is about 1 fourth of a standard deviation) increases the number of patents by about 9% based on the model estimated in column 7, and by 0.32 per thousand of inhabitants (population of 1939) using the model presented in column 8. This roughly corresponds to the interquartile range (p75-p25) so the effect is indeed sizable. However, in the presence of many covariates, the LATE interpretation of the 2SLS estimator has recently been called into question (Blandhol et al., 2022; Mogstad and Torgovitsky, 2024). Consequently, we rely primarily on the reduced form models presented in columns (3) and (4) as our preferred estimates. Although the corresponding coefficient on I_k in these reduced form models cannot be directly linked to the magnitude of migration, our main contribution is to provide evidence of a causal, positive effect of migrations on innovation. As long as the instrument remains valid, these columns demonstrate precisely that causal link.

4.4 Additional results

4.4.1 Results by technology

The previous results can be decomposed by technology. Specifically, we split patents into eight categories based on the first character of their IPC classification.[21] We then replicate our main regression separately for each technology. Table V presents the results using the reduced form specification from column (3) of Table IV, showing that our effect is driven by Chemistry, Physics, and Electricity. Although it is difficult to draw a universal conclusion from this finding, note that these three domains tend to be 'science driven' technologies, where cutting-edge research and deep technical knowledge play a greater role than in many more applied or process-orientated IPC classes. The underlying knowledge may also be more transferable and less dependent on specific industrial processes. For example, Moser et al. (2014) documents how the immigration of Jewish chemistry professors to the United States before WWII significantly advanced the field in their host country.

4.4.2 Different time horizons

We now turn to examining different time horizons for the dependent variable. Specifically, we divide the post-1961 period into smaller windows, capturing patents filed during 1961–1965, 1965–1970, 1970–1975, and 1975–1979. The results, presented in Figure VI, are based on the reduced form specification of column 3 from Table IV, where the inflow of migrant inventors (M_k) is replaced with the instrumental variable (I_k).

[21] A: Necessities, B: Manufacturing and Logistics, C: Chemistry, D: Textiles, E: Constructions, F: Engineering, G: Physics, H: Electricity.

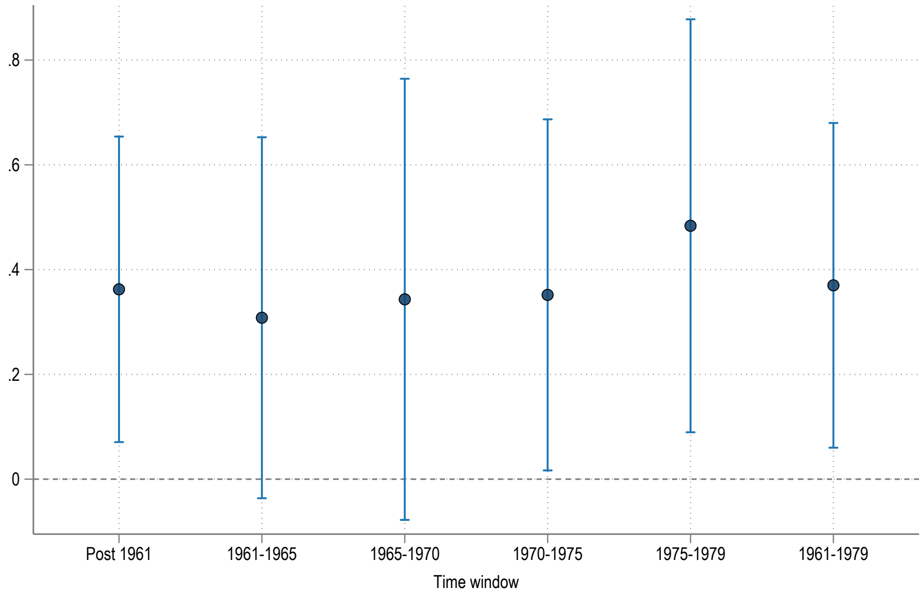
TABLE V. Regression by technology

IPC code	(1) A	(2) B	(3) C	(4) D	(5) E	(6) F	(7) G	(8) H
I_k	-0.071	0.251	0.509**	0.165	-0.235	0.223	0.503***	0.570**
	(0.164)	(0.152)	(0.195)	(0.208)	(0.177)	(0.171)	(0.148)	(0.266)
Distance (log)	0.214	0.252*	-0.018	-0.028	0.175	0.285**	0.182	0.145
	(0.141)	(0.123)	(0.198)	(0.196)	(0.112)	(0.113)	(0.130)	(0.108)
Population 1939 (log)	0.279*	0.228	0.369	0.348*	0.611***	0.331**	0.187	0.351**
	(0.159)	(0.173)	(0.233)	(0.202)	(0.158)	(0.159)	(0.200)	(0.165)
Total income 1925 (log)	0.074	-0.046	-0.108	-0.038	-0.161	-0.045	0.135	-0.086
	(0.117)	(0.147)	(0.189)	(0.156)	(0.129)	(0.144)	(0.184)	(0.175)
Manufacturing share 1939	0.928	2.208***	1.514	2.715***	1.499*	2.009*	2.729***	3.057***
	(0.773)	(0.700)	(1.087)	(0.908)	(0.816)	(0.983)	(0.745)	(0.871)
War-time destruction	-0.434	0.278	0.977	-0.362	-0.144	0.207	0.732	0.551
	(0.668)	(0.663)	(0.594)	(0.903)	(0.467)	(0.872)	(0.665)	(0.697)
Expellees from Eastern territories	1.056	1.722	2.722**	-1.391	-0.701	0.836	3.057***	1.923
	(1.284)	(1.540)	(1.274)	(1.344)	(1.411)	(1.162)	(0.902)	(2.048)
Pre-950 patent stock (log)	0.568***	0.558***	0.533***	0.395***	0.329***	0.458***	0.523***	0.501***
	(0.112)	(0.101)	(0.090)	(0.096)	(0.083)	(0.083)	(0.098)	(0.070)
R-squared	0.677	0.766	0.667	0.649	0.658	0.728	0.720	0.653
N	228	228	228	228	228	228	228	228

Notes: Each column of this Table replicates column 3 of Table IV but restrict to patents in a specific technology code. Definition of control variables are given in Table B1. All models include region (NUTS2) fixed effects. Standard errors are clustered by region. Technology codes follows one digit IPC and patents are matched to their most common IPC. A: Necessities, B: Manufacturing and Logistics, C: Chemistry, D: Textiles, E: Constructions, F: Engineering, G: Physics, H: Electricity

Figure VI demonstrates that the point estimates of the coefficient β become significant about 10 years after the last migrants in our sample arrived in West Germany. The effect becomes most pronounced in the later period, 1975–1979, suggesting that the influence of migrant inventors on innovation may take time to materialize.

FIGURE VI. Different time horizons



Note: Point estimate and 90% confident interval of coefficient β from estimating a reduced form version of the model 3 where M_k is replaced by the instrument I_k. The model is the same as in column 3 of Table IV but the dependent variable consider the sum of patents filed in k across different time windows. All models include region (NUTS2) fixed effects. Standard errors are clustered by region.

5 Dealing with identification threats

5.1 Address Books

As discussed in Section 3.4, the identification assumption underlined by our instrumental variable strategy is that the name of inventors is not a proxy for anything that would be correlated with future innovation dynamics. We will therefore construct several alternative pools of names reflecting different occupations around the timing of our treatment. For this, we make use of a unique database that contains more than 10.5 million address book entries in Germany, including name, location, and occupation. Data were provided by the *Verein für Computegenealogie* which used NLP techniques to digitize address books in Germany (cities and countryside) from the 18th century until the 1960s (Moeller and Fertig, 2023). We restrict attention to the years after 1900 and locations in present German boundaries.

From this new dataset, we select the 100 most frequent occupations. The occupations here should be considered very broadly as the information that would appear in the address book. Hence, this includes widows, landowners, or pensioners in addition to more standard occupations that cover various skill levels.

It is worth mentioning that while the coverage of the address book is fairly high, both geographically and across occupations, with more than 6.8 million entries between

1900 and 1960, not all current counties are included, and the number of distinct observations remains lower than with the list of casualties from WW1 which motivated our choice to use the latter source to build our baseline instrument and the former for robustness checks. The correlation between the share of names in the two datasets is around 65% and does not change over time from 1900 to 1960.

5.2 Randomization

Before addressing the threats discussed in Section 3.4, we first use the distribution of surnames in the Address Books to construct placebo instruments based on randomly drawn names, rather than on East German inventors. This randomization allows us to test whether our main result is driven by the migration-specific mechanism or whether it is an artifact of broader cultural or structural patterns embedded in the name distribution.

Specifically, let $N(n)$ be the probability weight of name n in East Germany taken from the Address Books. We draw a random set of K names from the full distribution, where K matches the number of East German inventors. Using this randomly selected set of surnames, denoted by $\mathcal{N}^i$, we construct a placebo instrument I_k^i for each county k as follows:

$$I_k^i = \log\left(\sum_{n\in\mathcal{N}(k)}\sum_{n'\in\mathcal{N}^i}\left(Lev(n,n')\cdot F_k(n)\cdot N(n',i)\right)\right),$$

where: $\mathcal{N}(k)$ is the set of surnames present in county k, $\mathcal{N}^i$ is the i-th random sample of K surnames drawn from the full distribution $N(n)$, $N(n',i)$ is the probability weight of the surname n' in the i-th random draw, $Lev(n,n')$ is the same name similarity measure defined earlier, $F_k(n)$ is the relative frequency of name n in county k.

Intuitively, this placebo instrument emphasizes how strongly the distribution of names in West Germany correlates with a random and representative set of East German surnames from the Address Books.[22]

We repeat this procedure 1000 times ($i = 1,\ldots,1000$). This approach serves two main purposes. First, by comparing our baseline IV coefficient from the main regression (3) to the distribution of coefficients obtained using I_k^i as an instrument for M_k, we conduct randomization-based inference. These random instruments will capture some

[22]We restrict to Address Books published between 1930 and 1961 and only keep surnames appearing at least five times.

of the predicting power of I_k as they keep the distribution of West German name unchanged. Yet their impact on future local innovation should be significantly different from what we find in our baseline estimates as they do not consider the list of inventors but a representative random set of East German name.[23]

Figure A5(a) in the appendix reports the distribution of coefficients from the 1000 randomizations. We can see that the average coefficient is around 6.5 and that our baseline coefficient of 9.3 (see Column 7 of Table IV) is above the maximal value obtained through these randomizations. Figure A5(b) does the same, but for the reduced-form equation (column 3 of Table IV) and similarly shows that the baseline coefficient is significantly higher than those obtained following the randomization.

5.3 Selection

Our baseline instrument I_k is based on the premise that the relative prevalence of a name $F_k(n)$ measured from the list of WW1 casualties is representative of the actual population, as explained in Section 3.4. As a first simple exercise, we show that the geographical coverage is consistent with the information drawn from the Census. To do so, we measure the correlation between the share of population of each county in the 1910 Census and the corresponding share among missing soldiers. We find a strong correlation of 88% which increases to 97% once we exclude two outliers around the city of Hamburg: Kreisfreie Stadt Hamburg and Kreis Pinneberg which are respectively under and over-represented in the WW1 casualties data. Table B2 in Appendix B shows that our main results hold when we exclude these two observations.

Although the numbers may match quantitatively, there is still the possibility that the distributions of names are not representative of the actual population. In particular, if names are informative about the nature of occupation, political power, or wealth, and if this dimension would be predictive of the likelihood of being missed or killed during WW1, this would mean that our measure of the prevalence of names will not adequately reflect the reality. Conceptually, this would mean that the true local prevalence of a given name $\tilde{F}(k)$ would only be measured with a noise:

$$F_k(n) = \tilde{F}_k(n) + \varepsilon_k(n)$$

Hence our instrument will only be observable up to the sum of $\varepsilon_k(n)$ weighted by

[23] Statistically, the distribution of East German inventor names is not orthogonal to the overall distribution of names from the Address books. For this reason, we expect I_k to be correlated to I_k^i for all i (and in particular to the average value across all i).

how name n (or relatively similar names) is frequent in the list of inventors. In our estimation, the bias of the coefficient will therefore be positive and meaningful if the names where the prevalence is over-estimated are also the ones that are correlated with the unobserved determinants of Y_k and that are particularly frequent in the set of names of East German inventors. Intuitively, we would expect the bias to go in the other direction. The first measurement error will create an attenuation bias even more that the variance of $\varepsilon_k(n)$ is large. Second, if anything we would think that inventors have names that are more closely related to a set of names of individuals that were less likely to be sent to the battlefield. This would mean that the correlation between $\varepsilon_k(n)$ and the weights $Lev(n, n')P(n')$ will be negative, effectively biasing the coefficient downward.

While this is indeed impossible to check formally, we can use the address books to estimate $\varepsilon_k(n)$ as the difference between our value of $F_k(n)$ and a similar value measured in the address book. We do it using all vintages from 1900 to 1930, for each county and name and then evaluate the correlation between

$$\sum_k \varepsilon_k(n) \text{ and } \sum_{n' \in \mathcal{N}} Lev(n, n')P(n).$$

The resulting correlation is very small and not significantly different from 0 which suggests that the proximity of a name to the list of East German inventors name is independent from the value of the mismeasurement.[24]

5.4 Exclusion restrictions

So far, we have discussed potential issues that arise from the measurement of $F_k(n)$ in our instrument. An additional issue could arise from the other components, that is, from the list of East German inventor names. As discussed in section 3.4, it is possible that this list of names may in fact be a proxy for some general features that will be correlated with innovation. Here again, we use the address book to construct alternative instruments that replace the set of names $\mathcal{N}$ and the corresponding weights $P(n)$ with alternative sets and weights measuring: (1) people working in scientific or technical occupations and (2) people working in occupations that are likely to be associated with higher income. We use information provided in the address books for the years

[24] Alternatively, we can use the address book to look at the average value of ε across occupations. We found that the 5 occupations with the highest value of ε in absolute values are: Widow, Seamstress, Train driver, Housewife, and Stenotypist, which are occupations were indeed more preserved from being sent to the battlefield during WW1, but are unlikely to be correlated with future innovation development and hence would bias our estimation.

1930-1960. More precisely, we define an entry to belong to category (1) if the type of task is labeled "komplexe Spezialistentätigkeiten" or "hoch komplexe Tätigkeiten" (Highly Complex Tasks and Complex Specialists Tasks) or if the occupation contains the word "Ingenieur" or "Professor". We consider an entry to belong to category (2) if it is associated with an industry defined as "Corporate organization, accounting, law and administration" or "Owners, proprietors and tenants" to which we add all entries labeled as "Führungskräfte" (managers). These two definitions are very large and, as a result, we find that, respectively, 15 and 11% of the address book entries match one of the two categories. The corresponding measures, denoted I_k^{sci} and I_k^w are positively correlated with I_k with a coefficient of around 40% and we include them as control in our baseline reduced form specification (specifically these presented in columns 3 and 4 of Table IV). The results are shown in Table VI. Columns 1 and 2 use the baseline specification without any additional control (they replicate columns 3 and 4 of Table IV). Columns 3 and 4 add a control for I_k^w, columns 5 and 6 for I_k^{sci} and columns 7 and 8 both I_k^w and I_k^{sci}. In these horse-race models, the coefficient associated with I_k remains positive and significant.

TABLE VI. Controlling for name proximity with scientific and wealthy occupations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I_k	0.370*	0.001**	0.336*	0.001**	0.438**	0.001**	0.406**	0.001**
	(0.188)	(0.001)	(0.173)	(0.001)	(0.180)	(0.001)	(0.170)	(0.000)
I_k^w			0.099	0.000			0.139	0.000
			(0.118)	(0.000)			(0.122)	(0.000)
I_k^{sci}					-0.152	-0.000	-0.189	-0.000
					(0.124)	(0.001)	(0.120)	(0.001)
Distance (log)	0.127	0.000	0.133	0.000	0.119	0.000	0.125	0.000
	(0.131)	(0.001)	(0.130)	(0.001)	(0.128)	(0.001)	(0.127)	(0.001)
Population 1939 (log)	0.554***	-0.003***	0.549***	-0.003***	0.572***	-0.003***	0.569***	-0.003***
	(0.119)	(0.001)	(0.123)	(0.001)	(0.127)	(0.001)	(0.131)	(0.001)
Total income 1925 (log)	-0.139	-0.001	-0.149	-0.001	-0.140	-0.001	-0.154	-0.001
	(0.099)	(0.001)	(0.098)	(0.001)	(0.104)	(0.001)	(0.105)	(0.001)
Manufacturing share 1939	2.347***	0.007**	2.234***	0.007**	2.328***	0.007**	2.165***	0.007**
	(0.671)	(0.003)	(0.644)	(0.003)	(0.679)	(0.003)	(0.643)	(0.003)
War-time destruction	0.419	0.000	0.424	0.000	0.364	0.000	0.357	0.000
	(0.689)	(0.003)	(0.690)	(0.003)	(0.730)	(0.003)	(0.734)	(0.003)
Expellees from Eastern territories	1.835*	0.001	1.800*	0.001	1.848*	0.001	1.802	0.001
	(1.035)	(0.004)	(1.049)	(0.004)	(1.036)	(0.004)	(1.062)	(0.004)
Pre-950 patent stock (log)	0.374***	0.002***	0.372***	0.002***	0.377***	0.002***	0.376***	0.002***
	(0.079)	(0.000)	(0.080)	(0.001)	(0.078)	(0.000)	(0.080)	(0.001)
R-squared	0.781	0.378	0.781	0.375	0.782	0.375	0.782	0.373
N	228	228	228	228	228	228	228	228

Notes: Columns 1 and 2 of this Table replicates columns 3 and 4 of Table IV. Other columns control for I^w and I^{sci} which are defined in Section 5.4. The unit of observation is a West German county. The dependent variable is the number of patents filed from 1961 to 1979 by inventors residing in the county. Columns 1, 3, 5 and 7 take the dependent variable in log and columns 2, 4, 6 and 8 standardize it by 1925's population. Definition of control variables are given in Table B1. All models include region (NUTS2) fixed effects. Standard errors are clustered by region.

5.5 Technological trend

A more specific concerns regarding the name of inventors is that they would be indicative of a specific technology. Such argument could be warranted by the fact that the names are often related to some occupations as explained in Section 3.4. A high proximity with a given county could then reflect the fact that the corresponding sector is overrepresented in that county, in which case we could mostly be picking up a technological trend with our specification.

To address these concerns, we use the fact that each patent is associated with a technological code, and in particular, we use the first letter of the IPC class, which splits patents into 8 categories, as in Section 4.4.1. For each of these categories, we can construct an instrument I_k^c, the number of migrants M_k^c, and the dependent variable Y_k^c. Each of these is calculated exactly the same way but restricts the sample of patents to those belonging to category c. We therefore have a panel at the county $\times$ technology level which allows to estimate the following model:

$$Y_k^c = \beta I_k^c + \nu_k + \mu_c + \psi_{b(k),c} + \varepsilon_{k,c}.$$

This model has the advantage of allowing us to control for county and technology fixed effects separately (ν_k and μ_c), and we also include a regional-technology specific set of dummy variables $\psi_{b(k),c}$. The identification of the parameter β is now done within county across technology, controlling for regional trends, or within technology across county. Table VII reports the results where the dependent variable is the logarithm of the number of patents. Table B3 in the Appendix shows the corresponding results using patents per capita. We can see that the coefficient associated with I_k in the reduced form model is positive and significant, except in the most saturated model (column 5) where we include both county and region-technology fixed effects. This means that the predictive power of name proximity to future patent development remains positive even with different sources of identifications and even when important source of variations are completely absorbed by fixed effects. In particular, the fact that column 4 reports a positive coefficient even when controlling for a set of technology and county fixed effects suggests that it is unlikely that our effect would be driven by a technological trend that would be correlated with I_k.

Alternatively, one can construct for each name a measure of the technological proximity to any West German county based on the cosine similarity between the vector of technology share of all patents filed between 1930 and 1961. This metric was introduced by Jaffe (1986) and is frequently used in patent analysis. It is formally defined as:

$$DC(n,k) = \frac{v_n.v_k}{\sqrt{||v_n||^2||v_k||^2}}$$

where the coordinate c of vector v_k is given by the share of patents in technological class c among all patents filed in k, a set denoted P_k

$$v_{k,c} = \frac{1}{|P_k|} \sum_{p \in P_k} \mathbb{1}_{c(p)=c} \text{ for } c = 1 \cdots C$$

and conversely for $v_{n,c}$ for patents filed by East German inventors named n. Aggregating this metric, we can construct a measure of technological proximity of each region to East German inventor which we denote P_k and is conceptually different from I_k. We construct two alternative measures of technological proximity, the first one uses the same 1 digit classification of technologies and categorizes patents into 8 groups. The second one uses a more detailed classification into 117 categories using the first 3 characters of the IPC classification. Such technological proximity can certainly serve as a pull factor, and if we think that names proximity merely capture this technological proximity, then our baseline estimation could be wrongly attributing an effect of migration on patenting that is in fact coming from a technology-specific boom after 1961, for a technology that is especially prevalent in East-Germany. To address this issue, we augment the models presented in Table IV and add a control for P_k. Results are presented in Table VIII and show that P_k, while being positively correlated with the number of incoming inventor migrants (column 1), does not capture all the variance explained by I_k when included as a control variable.

TABLE VII. Regression at the county-technology level

	(1)	(2)	(3)	(4)	(5)
I_k	0.533**	0.496**	0.856***	0.561*	0.458
	(0.213)	(0.209)	(0.090)	(0.314)	(0.341)
Distance (log)	-0.112	-0.119			
	(0.195)	(0.198)			
Population 1939 (log)	0.654***	0.654***			
	(0.164)	(0.164)			
Total income 1925 (log)	-0.265**	-0.260**			
	(0.118)	(0.117)			
Manufacturing share 1939	2.507***	2.530***			
	(0.844)	(0.847)			
War-time destruction	-0.174	-0.150			
	(1.107)	(1.106)			
Expellees from Eastern territories	-2.155	-2.065			
	(3.178)	(3.197)			
Pre-950 patent stock (log)	0.367***	0.368***			
	(0.097)	(0.098)			
Fixed Effects					
Technology	✓			✓	
County			✓	✓	✓
Region	✓		✓	✓	
Region-Technology		✓			✓
R-squared	0.456	0.461	0.427	0.529	0.536
N	1809	1809	1809	1809	1809

Notes: The unit of observation is a pair of West German county and 1 digit IPC technology (8 groups). The dependent variable is the number of patents filed in the technology from 1961 to 1979 by inventors residing in the county (in log). Definition of control variables are given in Table B1. Different models include different set of fixed effects shown in the bottom panel of the Table. Standard errors are clustered by region (NUTS2).

TABLE VIII. Controlling for technological proximity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I_k	0.044***	0.371*	0.001*			0.382*	
	(0.008)	(0.199)	(0.001)			(0.197)	
Migrant Inventors				9.378*	0.033**		9.699*
				(5.084)	(0.016)		(5.199)
Technological Proximity	0.046**	0.786*	0.005**	0.313	0.003*	0.535	0.555
	(0.021)	(0.409)	(0.002)	(0.399)	(0.002)	(0.470)	(0.491)
Distance (log)	0.000	0.150	0.000	0.128	0.000	0.131	0.123
	(0.004)	(0.133)	(0.001)	(0.130)	(0.001)	(0.131)	(0.130)
Population 1939 (log)	0.009	0.509***	-0.003***	0.711***	-0.002***	0.537***	0.717***
	(0.006)	(0.124)	(0.001)	(0.110)	(0.001)	(0.124)	(0.104)
Total income 1925 (log)	-0.011	-0.179*	-0.001	-0.058	-0.001	-0.152	-0.051
	(0.010)	(0.094)	(0.001)	(0.123)	(0.001)	(0.094)	(0.113)
Manufacturing share 1939	0.109***	2.484***	0.008**	2.037***	0.006**	2.304***	1.925***
	(0.034)	(0.686)	(0.003)	(0.564)	(0.003)	(0.662)	(0.554)
War-time destruction	0.023	0.617	0.001	0.526	0.001	0.444	0.473
	(0.031)	(0.682)	(0.003)	(0.688)	(0.003)	(0.680)	(0.660)
Expellees from Eastern territories	0.020	1.933*	0.002	1.833	0.001	1.788	1.743
	(0.037)	(1.072)	(0.004)	(1.157)	(0.003)	(1.067)	(1.195)
Pre-950 patent stock (log)		0.378***	0.002***	0.176	0.001**	0.376***	0.169
		(0.077)	(0.000)	(0.119)	(0.000)	(0.079)	(0.119)
R-squared	0.297	0.785	0.390	0.695	0.402	0.781	0.694
N	228	228	228	228	228	228	228

Notes: The unit of observation is a West German county. The dependent variable is the number of patents filed in the technology from 1961 to 1979 by inventors residing in the county (in log for columns 2, 4, 6 and 7 and standardized by 1939's population in columns 3 and 5) except for column 1 which uses the number of migrant inventors divided by population in 1925. Columns 1, 2, 3, 4, 6 uses an OLS estimator and colums 5 and 7 instruments the number of migrant inventors per capita by I_k and uses a 2SLS estimation strategy. Technological proximity is measured using the first 3 characters of the IPC in columns 1 to 5 and the first character in columns 6 and 7. Definition of control variables are given in Table B1. All models include region (NUTS2) fixed effects. Standard errors are clustered by region.

6 Conclusion

This study provides robust evidence that the migration of skilled inventors from East to West Germany significantly contributed to innovation activities in the destination regions during the post-World War II era. By constructing a novel dataset on patenting in Germany since 1930 and leveraging quasi-random variation in surname proximity to gauge the propensity to migrate, we identify a causal, positive impact of inventor migration on regional patenting activity. Our findings indicate that an inflow of East German inventor migrants led to a subsequent increase in patents filed, which partly explains the enduring technological gap between East and West Germany observed today.

We demonstrate the robustness of these results through an array of checks, and show that name-based proximity is a reliable predictor of future local technological development. This paper thus contributes to the broader literature on migration and innovation, while highlighting the unique historical context of Cold War Germany. Because the migrants in this setting spoke the same language, shared the same culture, and came from a similarly developed region, confounding factors often present in cross-border migrations are substantially reduced.

However, current data constraints prevent us from exploring the individual trajectories of migrant inventors or tracing how their ideas may have influenced local economic development in the longer run. As newly available census datasets emerge, future research may uncover more granular evidence on the specific mechanisms through which such migrations spur local innovation and growth.

References

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

Agrawal, Ajay, Devesh Kapur, and John McHale, "How do spatial and social proximity influence knowledge flows? Evidence from patent data," *Journal of urban economics*, 2008, *64* (2), 258–269.

Akbulut-Yuksel, Mevlude and Mutlu Yuksel, "The long-term direct and external effects of Jewish Expulsions in Nazi Germany," *American Economic Journal: Economic Policy*, 2015, *7* (3), 58–85.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas, "Immigration and the rise of American ingenuity," *American Economic Review*, 2017, *107* (5), 327–31.

Arkolakis, Costas, Sun Kyoung Lee, and Michael Peters, "European immigrants and the United States' rise to the technological frontier," 2020. Mimeo Yale.

Bade, Klaus J, "Enzyklopädie Migration in Europa vom 17. Jahrhundert bis zur Gegenwart [2007/2011]," *Historical Social Research/Historische Sozialforschung. Supplement*, 2018, (30), 293–305.

Bauernschuster, Stefan, Oliver Falck, Robert Gold, and Stephan Heblich, "The shadows of the socialist past: Lack of self-reliance hinders entrepreneurship," *European Journal of Political Economy*, 2012, *28* (4), 485–497.

Baum, Christopher F, Mark E Schaffer, and Steven Stillman, "IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation," Statistical Software Components, Boston College Department of Economics April 2002.

Baum, K.–H., *Die Integration von Flüchtlingen und Übersiedlern in die Bundesrepublik Deutschland* 1999.

Becker, Sascha O, "Forced displacement in history: Some recent research," *Australian Economic History Review*, 2022.

_, Lukas Mergele, and Ludger Woessmann, "The separation and reunification of Germany: Rethinking a natural experiment interpretation of the enduring effects of Communism," *Journal of Economic Perspectives*, 2020, *34* (2), 143–171.

Bergeaud, Antonin and Cyril Verluise, "A new dataset to study a century of innovation in Europe and in the US," *Research Policy*, 2024, *53* (1), 104903.

Bessel, Richard, *Germany after the first world war*, Oxford University Press, 1993.

Bispinck, Henrik and Damian van Melis, *" Republikflucht": Flucht und Abwanderung aus der SBZ/DDR 1945 bis 1961. Veröffentlichungen zur SBZ-/DDR-Forschung im Institut für Zeitgeschichte. Mit einer Einleitung von Damian van Melis*, Walter de Gruyter GmbH & Co KG, 2015.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, "When is TSLS actually late?," Working Paper w29709, National Bureau of Economic Research Cambridge, MA 2022.

Borjas, George J, "Immigration and Economic Growth," Working Paper 25836, National Bureau of Economic Research May 2019.

— **and Kirk B Doran**, "The collapse of the Soviet Union and the productivity of American mathematicians," *The Quarterly Journal of Economics*, 2012, *127* (3), 1143–1203.

Burchardi, Konrad B and Tarek A Hassan, "The economic impact of social ties: Evidence from German reunification," *The Quarterly Journal of Economics*, 2013, *128* (3), 1219–1271.

— **, Thomas Chaney, Tarek Alexander Hassan, Lisa Tarquinio, and Stephen J Terry**, "Immigration, innovation, and growth," *National Bureau of Economic Research*, 2020.

Campo, Francesco, Mariapia Mendola, Andrea Morrison, and Gianmarco Ottaviano, "Talents and Cultures: Immigrant Inventors and Ethnic Diversity in the Age of Mass Migration," *Journal of the European Economic Association*, 2022, *20* (5), 1971–2012.

Card, David, "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Ciccone, Antonio and Jan Nimczik, "The long-run effects of immigration: Evidence across a barrier to refugee settlement," 2022.

Davis, Donald R and David E Weinstein, "Bones, bombs, and break points: The geography of economic activity," *American Economic Review*, 2002, *92* (5), 1269–1289.

Donges, Alexander and Jochen Streb, "How the West was Settled. The Location Choice of East German Companies Migrating to West Germany after World War II," *The Location Choice of East German Companies Migrating to West Germany after World War II (August 11, 2023)*, 2023.

— **and** — , "How the West was Settled: The Location Choice of East German Companies Migrating to West Germany after World War II," in "Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024: Upcoming Labor Market Challenges" ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg 2024.

Dorner, Matthias, Dietmar Harhoff, Tina Hinz, Karin Hoisl, and Stefan Bender, "Social ties for labor market access-Lessons from the migration of East German inventors," *CEPR Discussion Paper No. DP11601*, 2016.

Falck, Oliver, Robert Gold, and Stephan Heblich, "E-lections: Voting Behavior and the Internet," *American Economic Review*, 2014, *104* (7), 2238–65.

— **, Stephan Heblich, Alfred Lameli, and Jens Südekum**, "Dialects, cultural identity, and economic exchange," *Journal of urban economics*, 2012, *72* (2-3), 225–239.

— **,** — **,** — **, and** — , "Dialects, cultural identity, and economic exchange," *Journal of urban economics*, 2012, *72* (2-3), 225–239.

Ferrucci, Edoardo, "Migration, innovation and technological diversion: German patenting after the collapse of the Soviet Union," *Research Policy*, 2020, *49* (9), 104057.

Fritsch, Michael, Maria Greve, and Michael Wyrwich, "Shades of a socialist legacy? Innovation activity in East and West Germany 1877-2014," 2023.

Fuchs-Schündeln, Nicola and Matthias Schündeln, "Who stays, who goes, who returns? East–West migration within Germany since reunification 1," *Economics of Transition*, 2009, *17* (4), 703–738.

— **and Rima Izem**, "Explaining the low labor productivity in East Germany–A spatial analysis," *Journal of Comparative Economics*, 2012, *40* (1), 1–21.

Geißler, Rainer, *Die Sozialstruktur Deutschlands*, Springer-Verlag, 2014.

Grashoff, Udo, *In einem Anfall von Depression... Selbsttötungen in der DDR*, Ch. Links Verlag, Berlin, 2006.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Cultural biases in economic exchange?," *The quarterly journal of economics*, 2009, *124* (3), 1095–1131.

Hefele, Peter, *Die Verlagerung von Industrie-und Dienstleistungsunternehmen aus der SBZ/DDR nach Westdeutschland: unter besonderer Berücksichtigung Bayerns (1945-1961)*, Vol. 4, Franz Steiner Verlag, 1998.

Heidemeyer, Helge, "Flucht und Zuwanderung aus der Sowjetischen Besatzungszone und der DDR in die Bundesrepublik," in "Migration und Integration in Deutschland nach 1945," De Gruyter, 2019.

Hensel, Jürgen, Łukasz Kamiński, Krzysztof Persak, and Jens Gieseke, *Handbuch der kommunistischen Geheimdienste in Osteuropa 1944-1991*, Ruprecht Gmbh & Company, 2009.

Hipp, Ann, Michael Fritsch, Maria Greve, Jutta Günther, Marcel Lange, Christian Liutik, Beate Pfeifer, Mariia Shkolnykova, and Michael Wyrwich, "Comprehensive Patent Data of the German Democratic Republic 1949–1990," *Jahrbücher für Nationalökonomie und Statistik*, 2022, *0* (0).

Hornung, Erik, "Immigration and the diffusion of technology: The Huguenot diaspora in Prussia," *American Economic Review*, 2014, *104* (1), 84–122.

Jaffe, Adam B, "Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value," Working Paper w1815, National Bureau of Economic Research 1986.

Juan, Alexander De, Felix Haass, Carlo Koos, Sascha Riaz, and Thomas Tichelbaecker, "War and Nationalism: How WW1 battle deaths fueled civilians' support for the Nazi Party," *American Political Science Review*, 2024, *118* (1), 144–162.

Kerr, William R, "Ethnic scientific communities and international technology diffusion," *The Review of Economics and Statistics*, 2008, *90* (3), 518–537.

— **and William F Lincoln**, "The supply side of innovation: H-1B visa reforms and US ethnic invention," *Journal of Labor Economics*, 2010, *28* (3), 473–508.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of econometrics*, 2006, *133* (1), 97–126.

Kremer, Anna, "Home is where the history is: How today's migration in Germany is shaped by regional identity," *Journal of regional science*, 2022, *62* (2), 604–638.

Kunze, Konrad, "dtv-Atlas Namenkunde: Vor-und Familiennamen im deutschen Sprachgebiet," *(No Title)*, 2003.

Lissoni, Francesco and Ernest Miguelez, "Migration and innovation: Learning from patent and inventor data," *Journal of Economic Perspectives*, 2024, *38* (1), 27–54.

Malia, Martin, “What Is the Intelligentsia?,” *Daedalus*, 1960, *89* (3), 441–458.

Ministry of State Security, “Entwicklung der Republikflucht im Jahre 1958,” 1959.

Moeller, K. and G. Fertig, “Daten der Citizen Science für die Wissenschaft? Kontrollierte Vokabulare als Herausforderung und Chance für die Auswertung und Qualitätsverbesserung von Massendaten. Die crowdbasierte Datensammlung von Adressbucheinträgen des Vereins für Computergenealogie im Qualitätstest,” in “Althage, Melanie / Dröge, Martin / Hiltmann, Torsten / Prinz, Claudia (Hrsg.), Digitale Methoden in der geschichtswissenschaftlichen Praxis: Fachliche Transformationen und ihre epistemologischen Konsequenzen: Konferenzbeiträge der Digital History 2023” 2023, pp. 213–241.

Mogstad, Magne and Alexander Torgovitsky, “Instrumental variables with unobserved heterogeneity in treatment effects,” in “Handbook of Labor Economics,” Vol. 5, Elsevier, 2024, pp. 1–114.

Moser, Petra, Alessandra Voena, and Fabian Waldinger, “German Jewish émigrés and US invention,” *American Economic Review*, 2014, *104* (10), 3222–55.

Nunn, Nathan, “The long-term effects of Africa’s slave trades,” *The Quarterly Journal of Economics*, 2008, *123* (1), 139–176.

of State Security, Ministry, “Reaktion von Angehörigen der medizinischen Intelligenz auf Kommuniqué,” 1959.

Peters, Michael, “Market size and spatial growth—evidence from Germany’s postwar population expulsions,” *Econometrica*, 2022, *90* (5), 2357–2396.

Posch, Max, Jonathan Schulz, and Joseph Henrich, “How Social Structure Drives Innovation,” 2024. Unpublished.

Rainer, Helmut and Thomas Siedler, “Does democracy foster trust?,” *Journal of Comparative Economics*, 2009, *37* (2), 251–269.

Ritschl, Albrecht and Tamás Vonyó, “The roots of economic failure: what explains East Germany’s falling behind between 1945 and 1950?,” *European Review of Economic History*, 2014, *18* (2), 166–184.

Roesel, Felix, “The German Local Population Database (GPOP), 1871 to 2019,” *Jahrbücher für Nationalökonomie und Statistik*, 2022.

Romer, Paul M, “Endogenous technological change,” *Journal of political Economy*, 1990, *98* (5, Part 2), S71–S102.

Rosés, Joan Ramón and Nikolaus Wolf, *The economic development of Europe’s regions: A quantitative history since 1900*, Routledge, 2018.

Rytlewski, Ralf and Manfred Opp de Hipt, “Die Deutsche Demokratische Republik in Zahlen 1945/49-1980: ein sozialgeschichtliches Arbeitsbuch,” *(No Title)*, 1987.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian, “Migrants and the making of America: The short-and long-run effects of immigration during the age of mass migration,” 2017.

Sleifer, Jaap, *Planning Ahead and Falling Behind: The East German Economy in Comparison with West Germany 1936-2002*, Berlin: Akademie Verlag, 2006.

Statistik des Deutschen Reichs, *Volks-, Berufs- und Betriebszählung vom 16. Juni 1925: Die berufliche und soziale Gliederung der Bevölkerung in den Ländern und Landesteilen*, Vol. 403–405, Berlin: Reimar Hobbing, 1927.

Statistisches Bundesamt, "Gebäude- und Wohnungszählung in der Bundesrepublik Deutschland," 1956. Statistics regarding housing and buildings from the 1950 census.

Statistisches Reichsamt, *Einkommen- und Körperschaftsteuerveranlagung für 1925* Statistik des Deutschen Reichs (Band 348), Berlin, Germany: Verlag von Reimar Hobbing, 1929.

Stoeckle, Philipp, "Subjektive Dialektgrenzen im alemannischen Dreiländereck," *perceptual dialectology–Neue Wege der Dialektologie. Berlin, New York: de Gruyter*, 2010, pp. 291–315.

Streb, Jochen, Jörg Baten, and Shuxi Yin, "Technological and geographical knowledge spillover in the German empire 1877–1918," *The Economic History Review*, 2006, *59* (2), 347–373.

The Economist, ""Germans still don't agree on what reunification meant"," https://www.economist.com/europe/2019/10/31/germans-still-dont-agree-on-what-reunification-meant 2014. Accessed: 29 March 2021.

_, ""Parts of the country are desperate for more people"," https://www.economist.com/special-report/2021/09/20/parts-of-the-country-are-desperate-for-more-people 2021. Accessed: 1 February 2022.

Waldinger, Fabian, "Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge," *Review of Economics and Statistics*, 2016, *98* (5), 811–831.

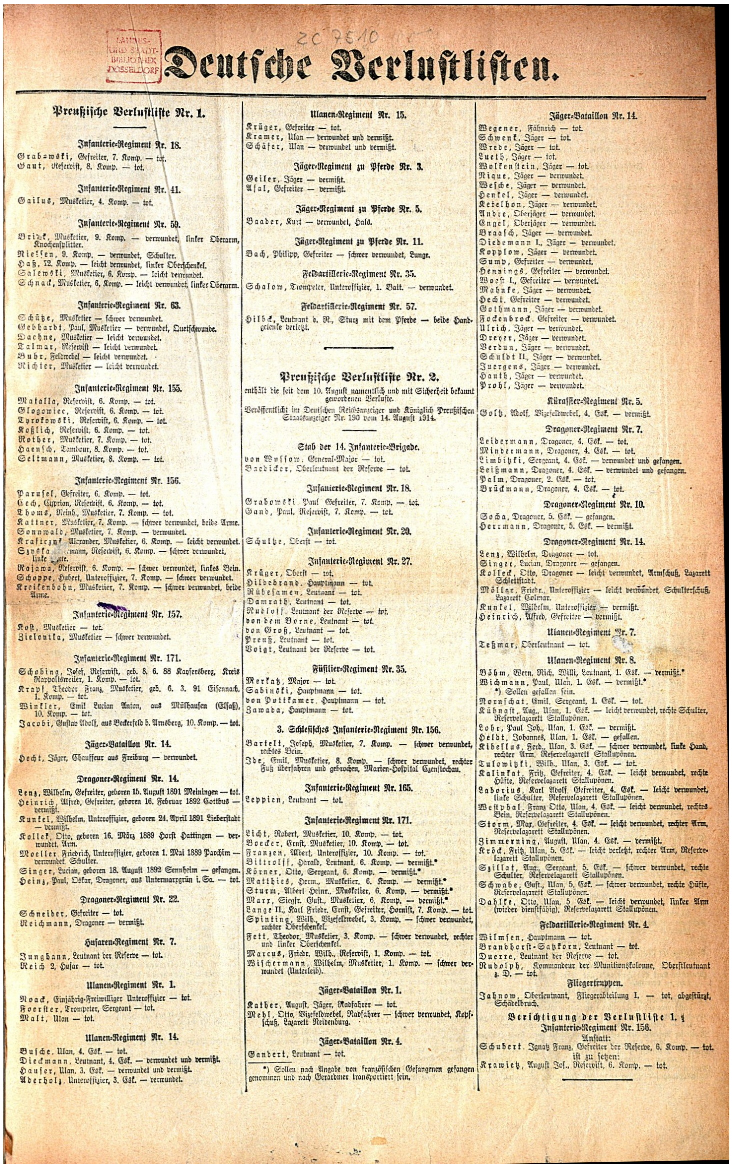
Wikipedia, "Alfred Kraus (Chemiker)," 2024. https://de.wikipedia.org/wiki/Alfred_Kraus_(Chemiker), Accessed: 10 December 2024.

Wyrwich, Michael, "Migration restrictions and long-term regional development: evidence from large-scale expulsions of Germans after World War II," *Journal of Economic Geography*, 2020, *20* (2), 481–507.

Appendix (for online publication)

A Additional Figures

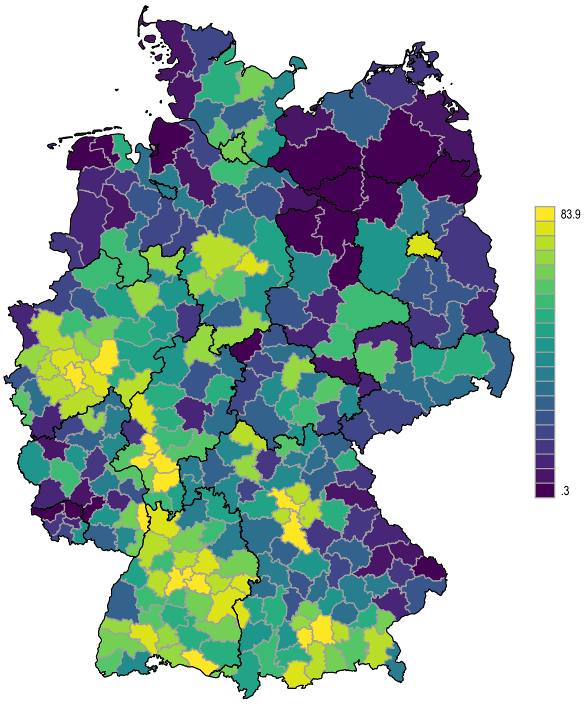
FIGURE A1. An example list of missing soldiers



Deutsche Verlustlisten.

Note: First edition of the German *Verlustliste* (List of Missing Soldiers from World War 2), Source: Wiki Genealogy

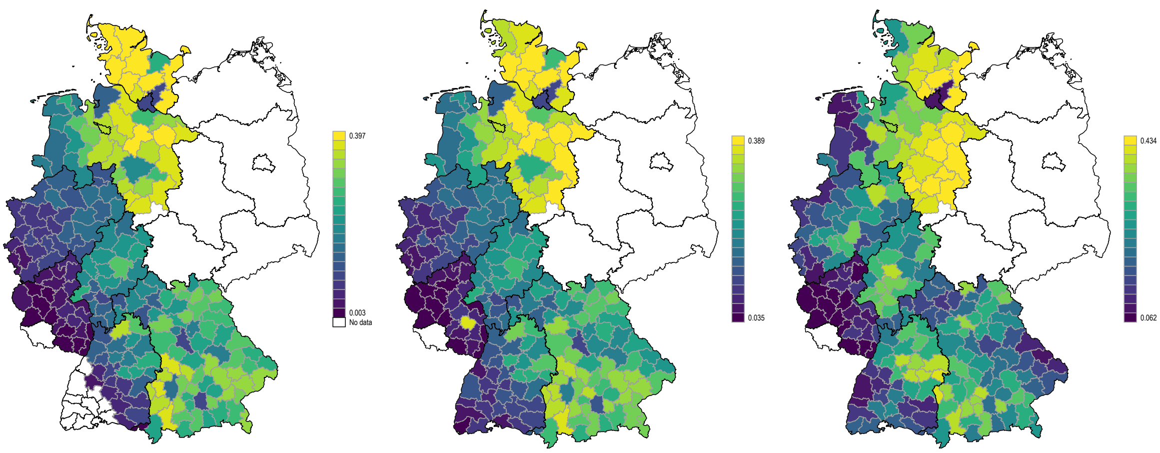
FIGURE A2. Patents per thousand inhabitants by districts



Note: Total number of patents filed between 1935 and 1970 allocated to the district of residence of the inventors. Population is taken in 1939. Source: *PatentCity* and Census.

FIGURE A3. East-West Migration in three distinct years

(a) 1946 (b) 1950 (c) 1961



Note: Number of East German migrants over total population in West German districts. 234 districts. Source: Census Data

FIGURE A4. Example of a Patent Document

BUNDESREPUBLIK DEUTSCHLAND

AUSGEGEBEN AM 16. MAI 1957

DEUTSCHES PATENTAMT

PATENTSCHRIFT

Nr. 963 995

KLASSE 39c GRUPPE 25 01

INTERNAT. KLASSE C 08 f

H 22507 IVb/39c

Dr. Hellmut Bredereck und Dr. Fritz Rochlitz, Stuttgart
sind als Erfinder genannt worden

W. C. Heraeus Gesellschaft mit beschränkter Haftung, Hanau/M.,
und Deutsche Gold- und Silber-Scheideanstalt vormals Roessler,
Frankfurt/M.

Verfahren zur Herstellung von Polymerisationsprodukten

Patentiert im Gebiet der Bundesrepublik Deutschland vom 24. Dezember 1954 an
Patentanmeldung bekanntgemacht am 29. November 1956
Patenterteilung bekanntgemacht am 2. Mai 1957

Als Polymerisationsbeschleuniger zur Polymerisation ungesättigter organischer Verbindungen bei niedriger Temperatur hat man bereits tertiäre Amine, organische Sulfinsäuren, α-Oxy- und α-Aminosulfone und Mercaptane vorgeschlagen.

Es wurde nun gefunden, daß als Polymerisationsbeschleuniger bei der Herstellung von Polymerisationsprodukten aus polymerisierbaren organischen Verbindungen mit einer doppelt gebundenen endständigen Methylengruppe α-substituierte Thioäther der allgemeinen Formel

$$R - S - CH_2X$$

besonders geeignet sind, wobei R einen organischen, insbesondere aromatischen Kohlenwasserstoffrest und X die Reste OH oder NR'R'' bedeutet, während R' bzw. R'' hierbei Wasserstoff oder Kohlenwasserstoffreste darstellen, die gegebenenfalls zu einem Ring geschlossen sind.

Besondere Fortschritte bringen Verbindungen der allgemeinen Formel

$$(R - S - CH_2)_2NR''',$$

wobei R''' Wasserstoff oder einen Kohlenwasserstoffrest bedeutet.

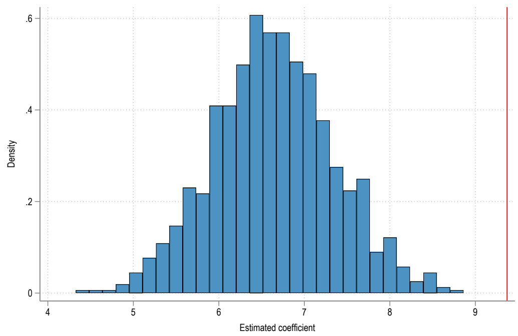
Polymerisierbare organische Verbindungen mit endständiger Methylengruppe, die nach dem Verfahren der Erfindung polymerisiert werden können, sind beispielsweise: ungesättigte aliphatische Kohlenwasserstoffe, z. B. Butadien, Vinylacetylen oder deren

709 514/288

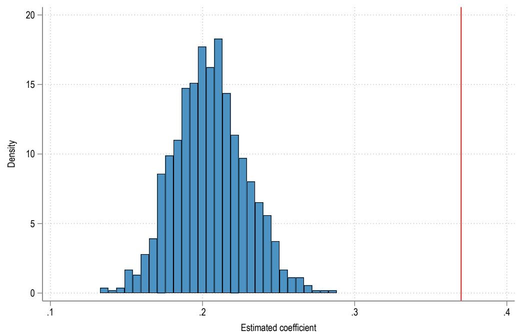
Note: The document shows the patent document of Hellmut Bredereck, a patent inventor who also shows up in our data.

FIGURE A5. Distribution of coefficients when instrument is randomized

(a) 2SLS estimation



(b) Reduced Form estimation



Note: Distribution of coefficients β from the estimation of model (3) when the instrument I_k is replaced by I_k^i constructed from randomizing the set of inventor names using the Address Books as explained in Section 5.2. Figure 5(a) and Figure 5(b) respectively use the specifications presented in columns 7 and columns 3 of Table IV. The vertical red line represents the coefficient when the actual instrument I_k is used and correpsond to the point estimates in these columns.

B Additional Tables

TABLE B1. Variable Descriptions

Variable	Description
Number of patents filed (1961–1979)	Total number of patents filed in the county between 1961 and 1979.
Migrant inventors (1945–1961)	Number of inventors who migrated from East to West Germany during 1945–1961, standardized by population in 1925.
Population (1939)	Total county population in 1939 (based on the Census 1939).
Distance to East Germany (log)	Logarithmic measure of the minimum distance from the county to the East German border (in km).
Total income (1925)	Total income in a county in 1925, measured in thousands of Reichsmark (based on the Census 1925).
Manufacturing share (1939)	Share of manufacturing employment in total employment in 1939 (based on the Census 1939).
War destruction (1945)	Share of houses destroyed during World War II, measured for the year 1945 using the 1950 housing census.
Expellees (1950)	Number of expellees from Eastern territories in 1950, standardized by population.
Pre-1950 patent stock (log)	Logarithm of the stock of patents filed in the county before 1950.
Instrument (I_k)	Surname proximity between East German inventors and West German counties, based on Levenshtein distance.

TABLE B2. Robustness check: excluding two counties

	(1)	**(2)**	**(3)**	**(4)**
Migrant Inventor	9.262*	0.032**	8.012*	0.029**
	(4.944)	(0.014)	(4.320)	(0.014)
Distance (log)	0.119	0.000	0.110	0.000
	(0.127)	(0.001)	(0.127)	(0.001)
Population 1939 (log)	0.727***	-0.002***	0.730***	-0.002***
	(0.104)	(0.001)	(0.102)	(0.001)
Total income 1925 (log)	-0.043	-0.001	-0.087	-0.001
	(0.115)	(0.001)	(0.104)	(0.001)
Manufacturing share 1939	1.987***	0.006**	1.933***	0.006**
	(0.563)	(0.003)	(0.546)	(0.003)
War-time destruction	0.447	0.000	0.418	0.000
	(0.674)	(0.003)	(0.683)	(0.003)
Expellees from Eastern territories	1.793	0.001	1.788	0.001
	(1.149)	(0.003)	(1.275)	(0.004)
Pre-1950 patent stock (log)	0.177	0.001**	0.202*	0.001***
	(0.117)	(0.000)	(0.106)	(0.000)
F statistic (first stage)	24.30	24.30	24.86	24.86
R-squared	0.696	0.395	0.697	0.384
N	228	228	226	226

Notes: Columns 1 and 2 of this Table replicates columns 7 and 8 of Table IV. Columns 3 and 4 do the same but exclude two observations: Kreisfreie Stadt Hamburg and Kreis Pinneberg (kreis number 2000 and 1056). Columns 1 and 3 take the dependent variable in log and columns 2 and 4 standardize it by 1925's population. Definition of control variables are given in Table B1. All models include region (NUTS2) fixed effects. Standard errors are clustered by region. The F-stat of the first stage is obtained using Kleibergen and Paap (2006) as calculated by Baum et al. (2002).

TABLE B3. Regression at the county-technology level - robustness

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
I_k	0.176**	0.159**	0.336***	0.135	0.084
	(0.078)	(0.075)	(0.042)	(0.132)	(0.124)
Distance (log)	0.040	0.040			
	(0.086)	(0.086)			
Population 1939 (log)	-0.325***	-0.325***			
	(0.081)	(0.081)			
Total income 1925 (log)	-0.151	-0.151			
	(0.104)	(0.104)			
Manufacturing share 1939	0.892**	0.892**			
	(0.376)	(0.374)			
War-time destruction	0.008	0.012			
	(0.361)	(0.362)			
Expellees from Eastern territories	0.131	0.127			
	(0.494)	(0.492)			
Pre-950 patent stock (log)	0.244***	0.244***			
	(0.061)	(0.061)			
Fixed Effects					
Technology	✓			✓	
County			✓	✓	✓
Region	✓		✓	✓	
Region-Technology		✓			✓
R-squared	0.250	0.292	0.350	0.392	0.447
N	1824	1824	1824	1824	1824

Notes: The unit of observation is a pair of West German county and 1 digit IPC technology (8 groups). The dependent variable is the number of patents filed in the technology from 1961 to 1979 by inventors residing in the county divided by population in 1939 (in thousand of inhabitants). Definition of control variables are given in Table B1. Different models include different set of fixed effects shown in the bottom panel of the Table. Standard errors are clustered by region (NUTS2).