
Labor, Automation Innovation and Human Capital

Carsten Peter Feuerbaum

Kumulative Dissertation

zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
Doktor rerum politicarum (Dr. rer. pol.)

an der
Wirtschaftswissenschaftlichen Fakultät
der Katholischen Universität Eichstätt-Ingolstadt

vorgelegt von
Carsten Peter Feuerbaum
aus München

Ingolstadt, den 31. März 2020

Erstgutachter: Prof. Alexander Danzer, Ph.D.

Zweitgutachter: Prof. Dietmar Harhoff, Ph.D.

Acknowledgements

First and foremost, I am deeply thankful to my supervisor Alexander Danzer for his guidance, insightful suggestions, the productive cooperation on joint projects and his extraordinary strong support throughout all the stages of this thesis. I will miss our fruitful discussions during our countless car rides to Ingolstadt. Moreover, I am deeply grateful to my second advisor Dietmar Harhoff for inviting me to the Max Planck Institute for Innovation and Competition as well as for his continuous support, motivation, honest advice and critical comments that sharpened my research ideas. Another big thank goes to Simon Wiederhold for completing my dissertation committee and for giving me valuable feedback on my research.

I am truly grateful to my co-author and friend Fabian Gaessler for his helpfulness and mentoring throughout my master's degree and my PhD-studies. Furthermore, I would like to especially thank my co-authors Ludger Woessmann and Marc Piopiunik for the pleasant and productive collaboration from which I learned a lot.

The doctoral program Evidence-Based-Economics has provided me with cutting edge training and a supportive research environment. Within this context, a special thanks goes to Florian Englmaier and Joachim Winter. Besides, I am grateful for financial support from the Elite Network of Bavaria.

In addition, I would like to thank Stefan Pabst for providing me with archival documents from the Federal Employment Agency which helped me to construct a novel dataset of German guest workers.

A word of thank also goes to the European Economic Association, the Royal Economic Society, the German Academic Exchange Service, the Verein for Socialpolitik and proFOR for financial support which enabled me to present parts of this thesis at numerous international workshops and conferences. This thesis has greatly benefited from the comments received.

At KU Eichstätt-Ingolstadt, the Max Planck Institute for Innovation and Competition and the Evidence-Based-Economics Program, I have been lucky to be surrounded by many wonderful colleagues and friends who have enriched my life. I wish I had the space here to thank each by name. Among other things, I will miss the stimulating and cooperative atmosphere, the insightful lunches at Due and Oberbayern, the cheerful gatherings at the bridge, the relaxing dinners in Ingolstadt and the refreshment in the Eisbach. A word of gratitude goes to Svenja Friess, David Heller, Zhaoxin Pu and Cristina Rujan for their help with reviewing parts of this thesis.

Last but not least, I am utmost grateful to my family for their support and love. This thesis is dedicated to my mother and late father.

Statement of own Contribution

Chapter 2 "Labor Supply and Automation Innovation" (joint with Alexander M. Danzer and Fabian Gaessler) is based on a research idea of mine. The conception of the chapter, the literature work and the writing were a joint task. Data acquisition, data preparation and empirical analyses were largely carried out by me and partly by Fabian Gaessler.

Chapter 3 "Labor Recruitment and (Non-)Automation Innovation" is single-authored by the author of this dissertation.

Chapter 4 "Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children" (joint with Alexander M. Danzer, Marc Piopiunik and Ludger Woessmann) is based on a research idea by Alexander Danzer. The conception of the chapter, the literature work and the writing were a joint task. Data acquisition, data preparation and empirical analyses were largely carried out by the author of this dissertation.

I presented both Chapter 2 and Chapter 4 at numerous workshops and conferences. For this purpose, I raised several external travel funds (2* DAAD Travel Grants, 1 * EEA Travel Grant).

Table of Contents

1	Preface	1
2	Labor Supply and Automation Innovation	8
2.1	Introduction	8
2.2	Institutional Background: Germany’s Migration Placement Policy	11
2.3	Data	13
2.3.1	Automation Innovation	14
2.3.2	Ethnic German Inflows and Other Regional Data	16
2.3.3	Descriptive Statistics	17
2.4	Empirical Framework	18
2.5	Results	19
2.5.1	Main Results	19
2.5.2	Discussion of Mechanism	22
2.6	Robustness Analyses	27
2.6.1	Accounting for Patent Value	27
2.6.2	Alternative Measures of Automation Innovation	28
2.6.3	Alternative Estimation Model: Poisson Regression	28
2.6.4	Separate Analyses for Pre- vs. Post-Binding-Allocation Periods	28
2.6.5	Alternative Samples	30
2.7	Discussion and Conclusion	31
3	Labor Recruitment and (Non-)Automation Innovation	68
3.1	Introduction	68
3.2	The German Guest Worker Program	72
3.3	Data	74
3.3.1	Guest Worker Data	74
3.3.2	(Non-)Automation Innovation	75
3.3.3	Regional Data on the Size of the Manufacturing Sector	77
3.3.4	Unemployment	77
3.3.5	Summary Statistics	77
3.4	Empirical Model	78
3.5	The Association between Labor Recruitment and (Non-)Automation Innovation	79
3.5.1	Main Results	79
3.5.2	Dynamics of the Associations	80
3.5.3	Associations by Regional Labor Market Size	80
3.5.4	Results by Technology Area	81
3.6	Robustness Checks	82
3.6.1	Accounting for Pre-Existing Unemployment Rates	82
3.6.2	Excluding Specific Regions	82

3.6.3	Overlapping Observations Model	83
3.7	Conclusion	83
4	Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children	101
4.1	Introduction	101
4.2	Institutional Background on the German Guest-Worker Program	105
4.3	Data	107
4.3.1	Survey Data on Guest Workers and their Children	107
4.3.2	Ethnic Concentration	109
4.4	Empirical Model	111
4.4.1	Model Setup with Region and Ethnicity Fixed Effects	111
4.4.2	Balancing Test by Degree of Ethnic Concentration	112
4.5	The Effect of Ethnic Concentration on Immigrant Children's Language Proficiency and Educational Attainment	114
4.5.1	Main Results	114
4.5.2	Subgroup Analysis	115
4.6	Mediating Factors	116
4.6.1	Parental Proficiency in the Host-Country Language	116
4.6.2	Inter-Ethnic Contacts with Natives and Economic Conditions	118
4.7	Robustness	119
4.7.1	Measuring Ethnic Concentration by Ethnic Shares	119
4.7.2	Instrumenting Ethnic Concentration in 1985 by Ethnic Concentration in 1975	120
4.7.3	Measuring Ethnic Concentration with Census Data	120
4.7.4	Measuring Ethnic Concentration at the County Level	121
4.7.5	Accounting for Interview Mode	122
4.7.6	Investigating Return Migration	122
4.7.7	Investigating Regional Migration within Germany	123
4.7.8	Investigating Family Size	123
4.8	Conclusion	123
	Bibliography	152

List of Figures

2.1	West German states with allocation policy	13
2.2	Event study: the effect of the ethnic inflow rate on the level of patents . . .	21
2.3	Effect on automation innovation by different originators	25
2A-4	Ethnic German inflows by arrival year	41
2A-5	Occupations and demographics of incoming ethnic Germans	42
2A-6	Automation innovation across technology areas	43
2A-7	Automation keywords across technology areas	44
2A-8	The level of automation patents and non-automation patents	45
2A-9	The share of automation patents by technology area	45
2A-10	The number of low-skilled workers by technology area	46
2A-11	Share of automation patents across German regions	47
2A-12	Distance between focal patent and citing patent	48
3.1	Foreign workers in West Germany	88
3.2	Foreign workers in West Germany by occupation	91
3.3	West German labor office districts	92
3.4	Automation innovation across technology areas	93
3.5	Level of non-automation patents across regions	94
3.6	Level of automation patents across regions	95
3.7	The level of (non-)automation patents	96
3.8	Geographical overlap between labor office districts and counties	97
3.9	Association between labor recruitment and non-automation innovation - al- ternative lag structure	98
3.10	Association between labor recruitment and automation innovation - alter- native lag structure	99
3.11	Association between labor recruitment and non-automation innovation across technology areas	100
4.1	Ethnic concentrations across West Germany, 1985	125
4A-1	Ethnic concentrations across West Germany: census 1987	135
4A-2	County-level ethnic concentrations across West Germany: census 1987 . . .	136

List of Tables

2.1	Summary statistics	33
2.2	The effect of the ethnic German Inflows on the share of automation innovation	34
2.3	The effect of the ethnic German inflows on the level of (non-)automation innovation	35
2.4	The effect of the ethnic German inflows on automation innovation - alternative lag structure	36
2.5	Effect on automation innovation across technology areas	37
2.6	Heterogeneous effects on automation innovation by labor market tightness .	38
2.7	Effect of ethnic German inflows on automation innovation – process and non-process innovation	39
2.8	Effect of ethnic German inflows on automation innovation – by labor market size 1991	40
2A-9	Analysis sample	49
2A-10	Regional characteristics	50
2A-11	The effect of the ethnic German inflows on the level of non-automation innovation - alternative lag structure	51
2A-12	Effect of ethnic German inflows on automation innovation – by firm size . .	52
2A-13	Effect of ethnic German inflows on automation innovation – by firm age . .	53
2A-14	Effect of ethnic German inflows on automation innovation – by firm regionality	54
2A-15	Effect on automation innovation (quality weighted)	55
2A-16	Effect of ethnic German inflows on automation innovation – alternative sets of automation keywords	56
2A-17	Effect of ethnic German inflows on automation innovation – poisson regressions	57
2A-18	Effect of ethnic German inflows on automation innovation – allocation period only	57
2A-19	Effect of ethnic German inflows on automation innovation – non-binding allocation period only	58
2A-20	Effect of ethnic German inflows on automation innovation – overlapping observations	58
2A-21	Effect of ethnic German inflows on automation innovation – exclusion of regions with low or high innovative capacity	59
2A-22	Effect of ethnic German inflows on automation innovation – exclusion of Gifhorn regions	60
2A-23	Effect of ethnic German inflows on automation innovation – winsorized inflow rate	61
2A-24	Effect of ethnic German inflows on automation innovation – exclusion of transition years	62
2A-25	Heterogeneous effects on automation innovation by labor market tightness (50 pct1)	63

2A-26	Heterogeneous effects on automation innovation by labor market tightness (90 pct1)	64
2A-27	Heterogeneous effects on non-automation innovation by labor market tightness	65
2A-28	Patent-level summary statistics	66
2A-29	Examples of automation patents	67
3.1	Summary statistics	85
3.2	Association between labor recruitment and (non-)automation innovation . .	86
3.3	Association between labor recruitment and non-automation innovation - alternative lag structure	86
3.4	Association between labor recruitment and automation innovation - alternative lag structure	87
3.5	Association between labor recruitment and automation innovation by regional labor market size	87
3.6	Association between labor recruitment and non-automation innovation by regional labor market size	88
3.7	Association between labor recruitment and (non-)automation innovation - controlling for unemployment rates	89
3.8	Association between labor recruitment and (non-)automation innovation - exclusion of regions with low or high innovative capacity	90
3.9	Association between labor recruitment and (non-)automation innovation - overlapping observations	90
4.1	Descriptive statistics by degree of ethnic concentration	126
4.2	Effect of ethnic concentration on host-country language proficiency	127
4.3	Effect of ethnic concentration on educational attainment	128
4.4	Subgroup analysis	129
4.5	Mediating factors - effect of ethnic concentration on host-country speaking proficiency	130
4.6	Mediating factors - effect of ethnic concentration on host-country writing proficiency	131
4.7	Mediating factors - effect of ethnic concentration on obtaining any school degree	132
4.8	Measuring ethnic concentration by share of own ethnicity in regional population	133
4.9	Instrumental-variable estimates using ethnic concentration in 1975	134
4A-1	Immigrants from guest-worker and other countries	137
4A-2	Individual-level variables	138
4A-3	Regional variables	142
4A-4	Ethnic concentration by ethnicity	142
4A-5	Occupation classes by ethnicity	143
4A-6	Balancing test using continuous share of ethnic concentration	144
4A-7	Subgroup analysis by ethnicity	145
4A-8	First-stage results using parents' leads and lags of language proficiency as instruments	146
4A-9	Ethnic concentration measured in 1987 census	147
4A-10	Ethnic concentration measured at county level (1987 census)	148
4A-11	Controlling for interview mode and return intention	149
4A-12	Predicting return migration by immigrant children between 1985 and 1995	150

4A-13 Predicting regional migration within Germany by immigrant children between 1985 and 1995	150
4A-14 Predicting the presence of children in the household	151

Chapter 1

Preface

Automation technologies and human capital are fundamental drivers of economic growth. First, automation technologies allow for the more efficient production of goods and services. In the face of rapidly advancing automation technologies, there are increasing concerns about their implications on employment as these technologies potentially enable capital to be substituted for labor (e.g., Acemoglu *et al.*, 2020). Second, as human capital is the stock of skills and knowledge of the labor force, it allows the factor labor to be more productive and generate more ideas (Mincer, 1984). With the rise of knowledge-based economies, human capital has become of central importance (Powell and Snellman, 2004).

The motivation of this thesis is to advance the literature by providing novel insights into the determinants of automation innovation and human capital. For this purpose, this thesis exploits immigration episodes, uses empirical methods for causal inference and introduces new sources of data on immigrants and automation innovation. Chapter 2 and Chapter 3 each examine the relationship between labor endowments and automation innovation. Chapter 4 examines the effects of ethnic concentration on immigrant childrens' acquisition of human capital.

The three chapters have methodological and thematic commonalities. The focus of the empirical analyses lies on regions in Germany. This setting is particularly well-suited to address the research questions as Germany offers two large-scale immigration waves with highly informative placement policies: the placement of labor migrants during the German Guest Worker Program (1955 - 1973) and the allocation of ethnic Germans from the former Soviet Union to regions in the 1990s and 2000s. Most importantly, these allocation policies allow to circumvent the potential bias from self-selection of immigrants into regions. Immigrants would otherwise select into regions, potentially taking into account ethnic networks, living costs, employment opportunities or wages (see e.g., Edin *et al.*, 2003; Albert and Monras, 2017; Hunt, 1992). Furthermore, these large immigration waves offer rich variation to study the role of labor endowments and ethnic concentration. Finally, Germany is among the most innovative countries in the world, as evidenced by its total patent count, patents per research personnel and patents relative to GDP (Seeni and Brown, 2015; WIPO, 2009). Immigration might matter particularly for automation innovation in countries with a high

innovation capacity.

Focusing on labor endowments induced by immigration, Chapter 2 and Chapter 3 relate to the classic economic question on the role of factor endowments for technological change (e.g., Acemoglu, 2002, 2007; Hanlon, 2015; Kiley, 1999). In his seminal work *Theory of Wages*, Sir John Richard Hicks proposed that “*a change of the factors of production is itself a spur to [...] innovation of a particular kind - directed to economizing the use of a factor which has become relatively expensive*”. Against this backdrop, economic theory has been suggesting for a long time that labor endowments may discourage *labor-saving* innovation (see e.g., Hicks, 1932; Habakkuk, 1962; Acemoglu, 2010). Despite the economic and societal relevance, empirical tests on labor supply are scarce and remain inconclusive: San (2019) finds a positive relationship between low-skilled labor shortages and invention activities in the US farming industry. In contrast, Doran and Yoon (2019) find a positive relationship between low-skilled worker inflows and the overall rate of innovation. The scarcity of empirical tests on labor supply might be explained by the unavailability of systematic data on automation innovation and a lack of suitable variation in labor endowments.

Attempting to fill this gap in the literature, Chapter 2 and Chapter 3 construct novel measures of automation innovation across a broad spectrum of technologies. Since the bibliographic information of patents does not tell whether a patent covers an automation invention, these chapters utilize the invention descriptions of patents to measure automation innovation. For this purpose, both chapters employ processing methods from text mining and keyword searches to define the set of automation patents.¹ The used automation keywords have been shown to be highly indicative of automation innovation in invention descriptions (Mann and Püttmann, 2018). Chapter 2 and 3 expand on the prior immigration literature by introducing novel measures of automation innovation. Existing studies have largely focused on the relationship between immigration and the overall level of innovation (e.g., Hornung, 2014; Hunt and Gauthier-Loiselle, 2010; Moser *et al.*, 2014).

Relating these regional measures of automation innovation to lagged labor endowments, Chapter 2 and Chapter 3 employ a pure spatial approach. This approach captures the *total* innovation effect of labor endowments at the region level taking complementarities across regional skill groups into account (Dustmann *et al.*, 2016b).² To control for unobserved factors, Chapter 2 and Chapter 3 employ a rich set of fixed effects. First, region fixed effects control for systematic time-invariant differences across regions such as the general innovative capacity or time-constant sectoral specialisation. Second, time fixed effects account for year-specific shocks that affect all regions in the same way such as general trends in innovation.

¹In this regard, both chapters contribute to recent research on the classification of patents (see e.g., Dechezleprêtre *et al.*, 2019; Webb, 2019).

²This pure spatial approach has usually been applied to study the effects of immigration on labor market outcomes (e.g., Boustan *et al.*, 2010).

Chapter 2 “Labor Supply and Automation Innovation” (joint with Alexander M. Danzer and Fabian Gaessler) advances on the prior literature by providing the first evidence of the causal effect of regional labor supply on automation innovation by firms. Our analysis takes advantage of the quasi-experimental placement of ethnic Germans across regions. Following the collapse of the Soviet Union, approximately 2.49 million ethnic Germans entered Germany between 1990 and 2006. The majority of incoming ethnic Germans was low-educated and worked in manual occupations prior to migration. To prevent ethnic enclaves, most German states introduced an allocation policy, which became binding in the middle of our study period from 1992 to 2006. Note that earlier studies have documented the exogeneity of regional inflows of ethnic Germans with respect to regional conditions of the labor market (Glitz, 2012), crime (Piopiunik and Ruhose, 2017), or the capacity to innovate (Jahn and Steinhardt, 2016). This chapter analyzes the effect of the labor supply shocks on automation innovation in a difference-in-difference framework by comparing automation innovation in region-year pairs with differential labor supply shocks. Using high-quality data from administrative sources, we are able to control for a rich set of regional characteristics including detailed skill and occupation groups.

Our analysis shows that the exogenous labor supply shock of ethnic Germans had a negative effect on automation innovation activities in the respective region. The impact on automation innovation is concentrated in the first and second year after the labor supply shock, in mechanical engineering and among corporate patent applicants. In contrast to that, we do not find any significant effects of labor inflows on the level of non-automation innovation. Within the context of Germany’s heavy labor market regulations and inflexible wages, our analysis highlights the role of search costs as a potential channel: the substitution between workers and automation innovation is much stronger in regions where labor is scarce compared to regions plagued by unemployment. This findings is in line with labor inflows relaxing labor supply shortages in tight markets.

Relating to the literature on the role of factor endowments for technology adoption (e.g., Zeira, 1998; Clemens *et al.*, 2018; Lewis, 2011; Monras, 2019), this chapter shows that firms do not only *adopt* but also *develop* these technologies in response to changes in labor endowments. Furthermore, our results provide suggestive evidence that internal as well as external demand drives these R&D investments. Our findings on these regional effects of labor supply on automation innovation are, in addition, consistent with a geographical bias of buyer-supplier relationships (Bernard *et al.*, 2019) and technology transfer (Almeida and Kogut, 1999; Audretsch and Feldman, 1996).

Finally, Chapter 2 contributes to a nascent literature investigating the relationship between labor abundance and the direction of innovation (Doran and Yoon, 2019; San, 2019). While all these studies exploit variation in labor inputs resulting from the *supply* of the labor market, research exploiting variation in labor inputs resulting from active *labor recruitment* has remained scarce. The final Chapter 3 of this dissertations aims to be a first step in filling this gap.

Chapter 3 "Labor Recruitment and (Non-)Automation Innovation" provides the first evidence for the regional association between low-skilled labor recruitment and (non-)automation innovation. For this purpose, it investigates 1) the *substitutability* of labor recruitment and *automation* innovation and 2) the *complementarity* of labor recruitment and *non-automation* innovation. While the former is motivated by the above mentioned economic theory on labor abundance and labor-saving innovation (see e.g., Hicks, 1932), the latter relies on economic theory suggesting that labor abundance might encourage innovation that is *labor-complementary* (see e.g., Kremer, 1993; Acemoglu, 2010). Furthermore, labor recruitment may allow firms to grow, to extend the scope of production and to enter new markets: the increased flexibility of production potentially induces firms to undertake exploratory R&D investments leading to novel products or processes. Chapter 3 utilizes the massive demand-driven allocation of labor migrants to regions during the German Guest Worker Program to isolate the regional labor recruitment by firms. The recruitment process arguably circumvents immigrant self-selection into regions as the initial job location of incoming guest workers depended on firms' labor recruitment and preferences of guest workers for regions were not considered. Lasting from 1955 to 1973, the goal of the recruitment program was to reduce labor shortages in the booming post-war Germany. The majority of guest workers were low-educated and employed in manual occupations.

A distinguishing feature of this chapter is that it introduces several novel sources of data. First, this chapter introduces a new dataset on the universe of foreign workers at the regional level. I collect high-quality and longitudinal data over the period 1964-1973 from archival documents published by the Federal Employment Agency. Importantly, the rich data on the foreign workforce by country-of-origin allow to distinguish between workers from countries with and without guest worker treaties. Using these data, this chapter can exploit the substantial regional variation in the increase in the stock of guest workers across regions and years. Second, this chapter introduces novel regional data on the level of non-automation patents and automation patents. It corresponds to the first spatial data on innovation during the period of the German Guest Worker Program. I construct these innovation outcomes using historic patent data from the German Patent and Trademark Office. Finally, using modern methods of geographical science, this chapter adds regional information from the 1961 full population census on the size of the manufacturing sector. To the best of my knowledge, this chapter presents the first analysis on the German Guest Worker Program at a regional level.

The results of this chapter, while descriptive, point towards towards labor recruitment having different associations with automation innovation and non-automation innovation. First, my results indicate a significant (insignificant) degree of substitution between labor recruitment and automation innovation in regions with large (small) labor markets. This heterogeneity might be explained by larger labor markets capturing a greater part of the total innovation impact of labor recruitment. Prior research shows that the adoption of

production technologies is a key mechanism to absorb regional labor supply shocks (see e.g., Lewis, 2011; Monras, 2019). Larger regions might capture stronger demand effects for products related to automation innovation within regional borders as buyer-supplier relationships tend to be regionally clustered (Bernard *et al.*, 2019). Second, this chapter finds a significant and positive relationship between the stock of guest workers and non-automation innovation. Models with different lag structures suggest that the association materializes in the second and third year after the labor recruitment. The positive association between labor recruitment and non-automation innovation tends to be concentrated in textiles, paper, metallurgy, transportation and performing operations.

This exploratory chapter cannot attribute causality to the estimated associations because the recruitment of guest workers might be endogenous to innovation. While the analysis on the lag structures suggests that reverse causality is an unlikely explanation for the found patterns, time-variant unobserved factors that influence both innovation and labor recruitment would bias my estimates. Within this context, there has been very little prior research on the endogeneity between regional labor recruitment and (non-)automation innovation. Notwithstanding the limitations of the descriptive analysis, this exploratory chapter can be considered as a first step towards estimating the impact of labor recruitment on (non-) automation innovation.

This chapter's combination of an highly informative low-skilled labor recruitment program, the novel data on the guest workers and the new measures of (non-)automation innovation advances the existing literature on the relationship between labor recruitment and innovation (see e.g., Kerr and Lincoln, 2010; Doran *et al.*, 2014; Dimmock *et al.*, 2019). In a related context, this chapter adds to the literature on immigration and innovation (e.g., Hornung, 2014; Hunt and Gauthier-Loiselle, 2010). Both literature strands have predominantly focused on the role of high-skilled immigration.

A comparison of the results from Chapter 2 and Chapter 3 shows that they share a key finding: the relationship of labor endowments with automation innovation is significantly different compared to the one with non-automation innovation. Furthermore, the relationship between labor endowments and innovation tends to be more positive for the case of labor recruitment. There are several potential reasons for this. First, labor recruitment might be associated with a greater planning security compared to labor supply shocks, possibly making firms more willing to expand their innovative activities. Second, the findings on labor recruitment concern a period of close-to-full employment and severe labor shortages. As a result, the increase in production flexibility through labor recruitment may be more effective in greasing the wheels of non-automation innovation. Third, given that the Guest Worker Program was in place around three decades earlier than the allocation of ethnic Germans, the relationship between labor endowments and innovation might vary with the state of the technological frontier. Finally, the difference in results may lie in the different research design: while Chapter 2 exploits *exogenous* variation in labor supply, Chapter 3 presents a descriptive analysis on *demand-driven* labor recruitment.

Exploiting the same placement procedure of the German Guest Worker Program from an *individual* perspective, Chapter 4 “Growing up in Ethnic Enclaves: Language Proficiency and Education of Immigrant Children” (joint with Alexander M. Danzer, Marc Piopiunik and Ludger Woessmann) studies the effect of regional ethnic concentration on immigrant childrens’ human capital acquisition. More precisely, we focus on the outcomes language proficiency and educational attainment. As the initial job location depended on the labor recruitment by German firms and regional preferences of guest workers were not considered, the policy rules out self-selection of individual guest workers into specific regions. The guest workers primarily originated from Italy, Greece, Spain, Turkey and Yugoslavia. We identify the effect of ethnic concentration on immigrant childrens’ human capital acquisition by observing several immigrant groups who are exposed to varying concentrations of co-ethnics within the same region. A beneficial feature of this set up is that it allows us to include a rich set of fixed effects. Region fixed effects ensure that any region-specific peculiarities are accounted for to the extent that they are common across guest-worker ethnicities. Ethnicity fixed effects ensure that any ethnicity-specific differentials in integration are accounted for to the extent that they are common across regions. For the empirical analysis, we construct a novel dataset on immigrant children with their parents. It contains rich information on individual characteristics and potential mediating factors such as parents’ language skills, inter-ethnic contacts and contacts to natives. It is based on survey data from the German Socio-Economic Panel, a large annual household survey that is representative of the resident population in Germany. We complement these individual data with ethnic concentration measures at different levels of regional aggregation based on social-security and census data.

Our analysis shows that ethnic concentration increases the likelihood of school dropout and lowers host-country language proficiency. We further show that the effect on immigrant childrens’ language proficiency is mediated by parents’ host-country language skills. For this mediation analysis, Chapter 3 addresses measurement error in the self-reported parental language skills by implementing an instrumental variable approach that exploits parents’ responses on the same survey item from consecutive years (Dustmann and Van Soest, 2002).

We contribute to a growing literature on ethnic enclaves and the human capital acquisition of immigrant children (see e.g., Grönqvist, 2006; Cortes, 2006) by exploiting a placement policy, by focusing on the exposure to ethnic enclaves with a predominant share of low educated co-ethnics and by introducing novel measures of speaking and writing abilities. With the exception of a study by Åslund *et al.* (2011), who use a refugee placement policy in Sweden, other prior research does not place major emphasis on addressing bias from self-selection into ethnic enclaves. In contrast to our study, Åslund *et al.* (2011) find that ethnic enclaves with significant shares of highly-skilled co-ethnics are education-enhancing. The findings of this chapter can be reconciled with those of Åslund *et al.* (2011) in a sense that the impact of ethnic concentration might depend on the average quality of the ethnic environment, the so-called ethnic capital introduced by (Borjas, 1992). Finally, given that guest workers were

employed upon arrival in Germany, our study can further advance the literature by neutralizing the potential channel related to ethnic networks and parents' labor market integration (see e.g., Edin *et al.*, 2003; Damm, 2009).

Human capital in the form of language skills and educational attainment plays a crucial role for immigrant childrens' economic and social integration (Dustmann and Glitz, 2011; Chiswick and Miller, 2015). Within the context of ongoing public debates about the integration of immigrant children, the last chapter's findings on ethnic concentration and human capital are highly policy relevant.

The following three chapters are self-contained and can be read independently. References for all three chapters are listed in a joint bibliography at the end of this thesis.

Chapter 2

Labor Supply and Automation Innovation

2.1 Introduction

Are man and machine substitutes? Economists have long assumed that technological progress does not make workers irrelevant as factor of production; however, recent empirical evidence suggests that the income share in national income has been falling since the 1980s and that one potential explanation may be the increased capital intensity of production, i.e., labor-saving technological progress (Salomons *et al.*, 2018; Karabarbounis and Neiman, 2014). This recent evidence also revives an older theoretical debate that labor supply can affect firms' investments into labor-saving innovation (Habakkuk, 1962; Hicks, 1932). In a more recent theoretical contribution, Acemoglu (2010) suggests that labor scarcity may induce technology progress if the new technology is *strongly labor saving*.³ The substitutability between labor and innovation should be particularly relevant in tasks where automation is technically feasible and efficiency enhancing. Despite the topic's economic as well as societal relevance, empirical studies on the adjustment of automation innovation to changes in labor supply have remained scarce – which is probably due to a lack of suitable exogenous variation in labor supply and the unavailability of systematic data to quantify automation innovation.

This paper provides first evidence of the causal relationship of regional labor supply on automation innovation by firms. For identification, we rely on plausibly exogenous labor supply shocks from immigration, which has been shown to affect regional wages and employment (e.g., Card, 1990; Borjas, 1994; Dustmann *et al.*, 2008, 2017b; Peri and Sparber, 2009). Our analysis takes advantage of the quasi-experimental placement of ethnic Germans across German regions during the 1990s and 2000s as a source of exogenous variation in regional labor supply.⁴ Following the collapse of the Soviet Union, approximately 2.49 million ethnic Germans entered Germany between 1990 and 2006, most of them from the low end

³*Strongly labor saving* implies that technological advances decrease the marginal product of labor.

⁴Glitz (2012) is the first study to exploit the placement of the ethnic Germans to investigate the labor market effects of the resulting labor supply shocks. Regarding the regionality of effects, note that production activity tends to cluster in the same location as innovative activity (e.g., Paci and Usai, 2000).

of the skill distribution or with poor prospects of receiving recognition for their outdated skills which had been acquired in a different economic system (Koller, 1993; Bundesverwaltungsamt, 2019). With their predominant work experience in manual occupations prior to migration, the migrants competed with the existing labor force for low-skilled manual jobs. To ensure a more even distribution of these immigrants across regions, most German states introduced an allocation policy, which became binding in the middle of our study period from 1992 to 2006. We exploit this allocation policy to disentangle the effect of positive labor supply shocks on automation innovation from potential bias that would occur if immigrants self-selected into specific regions. The dispersion of migrants across Germany thus provides a unique setting for causal inference.

We construct a panel dataset at the labor market-year level comprising novel measures of automation innovation and a rich set of regional characteristics. The automation innovation measures are based on patent data from the European Patent Office. Although patents are assigned to specific technology classes during examination, the available bibliographic information does not tell whether a patent covers an automation or non-automation invention. We therefore identify automation patents by relying on a text-based classification algorithm. We link automation as well as non-automation patent applications to labor market regions based on the inventors' addresses using the OECD REGPAT Database (Maraut *et al.*, 2008). These regional characteristics include GDP per capita, the unemployment rate, and precise measures of three skill groups and twelve occupation groups based on high-quality administrative data from the Institute for Employment Research (IAB).

We analyze the effect of labor supply on automation in a difference-in-difference framework. We hereby compare automation innovation in region-year pairs with differential labor supply shocks resulting from the quasi-experimental placement of ethnic Germans. We use a pure spatial approach that captures the total automation innovation effects of labor supply shocks at the regional level, taking skill downgrading and complementarities across skill groups into account.⁵ The main outcome variable is the regional share of automation patents. The key explanatory variable is operationalized as the lagged exogenous ethnic German inflow divided by the total workforce in the previous year. By including region fixed effects, we account for time-constant systematic differences across regions. By including time fixed effects, we control for nationwide time trends in automation innovation. Our identification rests on the common trends assumption for which we provide empirical support.

We find that the exogenous labor supply shock of ethnic Germans had a statistically significant negative effect on automation innovation activities in the respective region: an increase by one ethnic German per 1,000 employed workers led to a decline in the regional share of automation innovation by 0.9 percentage points and in the level of automation innovation by 4.3 percent. According to these estimates, the average annual inflow of 2.54 ethnic Germans per 1,000 employed workers corresponds to a decline in the number of automation

⁵Examples of prior studies with a similar spatial approach identifying the labor market effects of immigration include Boustan *et al.* (2010) and Dustmann *et al.* (2017b).

patents by about 3 patents. Given that the average annual number of automation patents per region during our period of observation is about 26.7 patents, this is an economically sizeable effect. In contrast, we do not find any significant effects of labor inflows on the level of non-automation innovation.

A distinguishing feature of our work is that we explore the mechanism of how labor supply affects automation innovation. Prior research shows that the inflows of ethnic Germans had no effect on wages (Glitz, 2012), a result which is compatible with Germany's heavy labor market regulations and strong unions. To shed light on search costs as a potential channel, we explore effect heterogeneity by pre-existing labor market tightness across a large number of labor market regions: the substitution between workers and automation innovation is much stronger in regions where labor is scarce compared to regions plagued by unemployment. These findings are in line with the established notion that labor inflows relax labor supply shortages in tight markets, but have little impact on labor abundant markets.

The impact on automation innovation is concentrated in the first and second year after the labor supply shock, in mechanical engineering and among corporate patent applicants. We illustrate that the effect originates from industries that employ large numbers of low- and un-skilled workers. A variety of robustness checks with respect to the estimation technique, to lag structures, to the weighting of the data, to the observation period, to regional subgroups, and to alternative measures of automation innovation corroborate our causal interpretation.

This study contributes to several strands of literature. First, our study provides causal evidence on the regional effects of labor supply on automation innovation across various industries in a contemporary setting. While several scholars highlight the role of relative factor supplies for the direction of technological change (e.g., Acemoglu, 2002, 2007; Hanlon, 2015; Kiley, 1999), empirical tests on labor supply are scarce and remain inconclusive: San (2019) finds increased invention activities in the US farming industry due to labor shortages following the exclusion of Mexican workers in the 1960s. In contrast, Doran and Yoon (2019) study the effect of US mass immigration in the early 20th century and find a *positive* relationship between low-skilled worker inflows and the overall rate of innovation. Using cross-country variation on wages and firm-level data on distinct patenting activities, the results from Dechezleprêtre *et al.* (2019) suggest that an increase in low-skilled wages increases automation innovation.

Second, our paper relates to the literature on the role of factor endowments for technology adoption (e.g., Zeira, 1998). Prior research suggests that firms absorb shifts in regional labor supply by switching to the most cost-efficient production technology available (e.g., Hanson and Slaughter, 2002; Dustmann and Glitz, 2015; Zator, 2019). In fact, Clemens *et al.* (2018), Lewis (2011), Imbert *et al.* (2019), and Monras (2019) find that low-skilled labor supply explains firm adoption of production technologies. We complement these findings and show that firms do not only *adopt* but also *develop* these technologies as evidenced by their increased automation patenting activities. Furthermore, our results indicate that small and young firms react more strongly than large and old firms. Finally, we find suggestive

evidence that not only internal but also external demands drive these R&D responses. In light of national, if not global, product markets, a *regional* demand-pull mechanism seems at first glance surprising, but is in line with the geographical bias of buyer-supplier relationships (Bernard *et al.*, 2019) and technology transfer (Almeida and Kogut, 1999; Audretsch and Feldman, 1996).

Lastly, we complement the recent literature concerned with the effect of immigration on the overall level of innovation (e.g., Hornung, 2014; Hunt and Gauthier-Loiselle, 2010; Kerr *et al.*, 2015; Moser *et al.*, 2014). Most previous studies analyse the effects of high-skilled immigration; more relevant from a labor replacement perspective, however, is the effect of low-skilled immigration.

The remainder of this paper is structured as follows: Section 2.2 describes the institutional background of the quasi-experimental placement of ethnic Germans in the 1990s and the 2000s. Section 2.3 presents the datasets used in the empirical part of the paper and the underlying patent classification algorithm. Section 2.4 describes the econometric model and provides empirical support for the identifying assumption. We present the results in Section 2.5 and robustness analyses in Section 2.6. Section 2.7 concludes.

2.2 Institutional Background: Germany’s Migration Placement Policy

After the fall of the Iron Curtain, Germany experienced a massive permanent resettlement of ethnic Germans from Eastern Europe (Klose, 1996). Approximately 2.49 million ethnic Germans – around 3.1 percent of Germany’s population and 6.7 percent of its workforce – immigrated between 1990 and 2006 (Bundesverwaltungsamt, 2019). The incoming ethnic Germans were descendants of German speaking emigrants to Eastern Europe in the 18th and 19th centuries (Bade, 1990).⁶ Prospective ethnic German immigrants had to apply for visa at the German embassy in their home country and provide proof of German ancestry. Successful applicants were granted entry in Germany subject to annual immigration quotas (from 1993: 225.000, from 1999: 100.000). Annual inflows of ethnic Germans to Germany amounted to about 200.000 per year until 1995, before they fell to around 100.000 per year thereafter (see Figure 2A-4 on inflows from the Former Soviet Union (FSU)).⁷

Upon arrival in Germany, these immigrants were sent to central admission centers and naturalized, implying that they could immediately take up work (Dietz, 2006; Ohliger, 2008). To prevent the emergence of residential enclaves, the government had enacted a regional allocation policy for ethnic Germans in 1989; however, the rule remained inoperative until 1996.

⁶Many ancestors had followed a resettlement offer by Russian Empress Catherine the Great in 1763 (granting land and religious freedoms).

⁷Between 1992 and 2006, 95.7 percent of ethnic Germans originated from the successor states of the USSR (Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russian Federation, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan). After the implementation of the placement policy in 1996, this share increased to over 98 percent.

During these early years many ethnic Germans self-selected into clusters of co-ethnics, so that newly arrived ethnic Germans comprised 20 percent or more of the population in some regions (Klose, 1996). In consequence, most West German federal states except Bavaria and Rhineland-Palatinate started adhering to the allocation rule bindingly from March 1996, with Lower Saxony following in April 1997 and Hesse in January 2002.⁸ The policy assigned immigrants to one of the federal states according to historical state quotas that were originally developed for budget rules (*Koenigsteiner Schluessel*⁹). Within the respective states, incoming ethnic Germans were further allocated to counties according to relative population size.¹⁰ Immigrants were unable to choose their destination in Germany and their allocation was not determined by labor market considerations, such as educational endowments (see Glitz, 2012, for more details). After their placement, immigrants were bound to stay in their allocated county for at least three years; non-compliance was heavily sanctioned with the loss of most welfare benefits. Therefore, compliance with the rules was very high, actual immigration matched the allocated quotas well (Dietz, 2006) and the policy was considered successful (Federal Constitutional Court (1 BvR 1266/00, Rn. 1-56)).

The majority of incoming ethnic Germans were of working age and with working experience in their countries of origin (see Figure 2A-5b for details). Most were low-skilled and had worked in manual occupations, such as farmers, laborers, transport workers, operatives and craft workers, according to the official statistics *Jahresstatistik fuer Aussiedler*, published annually by *Bundesverwaltungsamt* (see also Figure 2A-5a for details). The few formally high-skilled migrants faced considerable barriers to the recognition of their qualifications and experienced significant skill downgrading (e.g., Eckstein and Weiss, 2004; Danzer and Dietz, 2014). Despite their German ancestry, many immigrants had a limited account of the German language. Over the period 1992 to 2002 the share of immigrants with working experience (Figure 2A-5a) and their occupational distribution (Figure 2A-5b) were quite stable, suggesting a constant quality of immigration cohorts.

To summarize, the immigration of ethnic Germans provides a quasi-experimental setting, which helps overcoming the potential bias from the self-selection of immigrants into specific regions: In the absence of a placement policy immigrants might have either chosen regions with declining trends in (labor-replacing) automation innovation or booming regions with high levels of automation innovation, depending on their belief about labor market conditions. The resultant reverse causality would have biased the empirical estimates. In line with our considerations, earlier studies have documented the exogeneity of regional inflows of

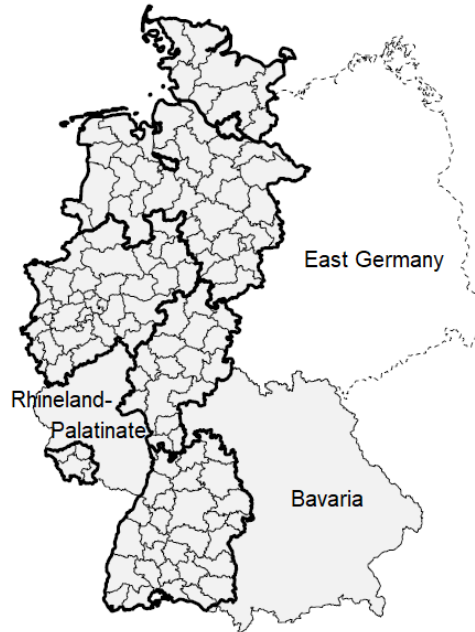
⁸Figure 2.1 shows the West Germany's states with the implementation of the Assigned Place of Residence Act. See Appendix Table 2A-9 for details regarding the analysis sample and the implementation of the assigned place of residence act.

⁹The state quotas are based on population size (with weight 1/3) and tax revenues (with weight 2/3). A comparable rule for the UK is the Barnett Formula.

¹⁰Meeting these quotas was of utmost priority, but anecdotal evidence suggests that the allocation was relaxed in particular cases: ethnic Germans could be exempted if they were able to provide proof of registered employment and sufficient housing in a different county and were willing to waive social benefits. Yet, only 11 percent of ethnic Germans did not receive any kind of social benefits during the first three years after arrival, according to Haug and Sauer (2007).

ethnic Germans with respect to regional conditions of the labor market (Glitz, 2012), crime (Piopiunik and Ruhose, 2017), or the capacity to innovate (Jahn and Steinhardt, 2016). In Section 2.4 we provide empirical support for the identifying assumption that immigrant inflows can be considered as exogenous to regional automation innovation.

Figure 2.1: West German states with allocation policy



Notes: West Germany's states with the implementation of the Assigned Place of Residence Act. The black lines denote state borders. With the exception of Bavaria and Rhineland-Palatinate all states in West Germany introduced the allocation policy. See Appendix Table 2A-9 for details on the analysis sample. Figure based on a shapefile of the Federal Republic of Germany from Eurostat and a reference file on counties and labor market regions from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

2.3 Data

Our analysis exploits a regional panel data set with yearly information spanning the time period 1992-2006; it covers the pre-allocation period from 1992 to 1995 and the period of legally enforced allocation thereafter.¹¹ The research is staged at the level of German labor market regions (*Arbeitsmarktregionen*), following Glitz (2012). These labor market regions comprise one or several counties and have been designed – based on commuter flows – to capture regional labor markets (Federal Office for Building and Regional Planning, 2019). Following prior studies, we restrict our analysis to labor markets in West Germany, since East Germany experienced severe adjustment processes following the German reunification,

¹¹We do not analyze the years before 1992, because the selection of ethnic Germans was different before the collapse of the Soviet Union. Also, regional data on migrant inflows are not systematically available before 1992. After 2006 the number of incoming ethnic Germans was negligible.

and exclude states without binding allocation policy (Bavaria and Rhineland-Palatinate).¹² Our analysis sample contains 127 labor market regions for 15 years.¹³ The research combines regionalized information on automation innovation, immigrant inflows and regional economic conditions from several data sources.

2.3.1 Automation Innovation

The automation innovation measure is based on all patent applications filed at the European Patent Office (EP patents) by inventors located in the West German allocation states with a priority date between 1992 and 2010. The focus on EP patents has advantages. First, as EP patents provide the option for Europe-wide instead of just domestic patent protection, the underlying inventions tend to be of higher economic value. Second, practically all EP patents feature English descriptions, regardless of their origin; this liberates our text-based method from language-specific adjustments.¹⁴ We rely on the text and bibliographic data as provided by the PATSTAT database (2017 Autumn Edition). In our main analysis, we focus on patent applications of corporate entities.¹⁵ In a subsequent falsification exercise, we focus specifically on those patents with at least one non-corporate applicant (e.g., a natural person or a university).

Classification of patents

We classify patent applications into automation and non-automation innovations by searching the invention descriptions for word stems related to automation. Following standard text mining procedures, we first pre-process the English abstracts to reduce the high-dimensionality of text features: this involves tokenizing, case-folding each alphabetic token into lowercase and removing punctuation, numbers and stop words. We map all words that refer to the same basic concept to their linguistic root using the English version of a commonly employed stemming algorithm (Porter *et al.*, 1980). For example, this algorithm transforms the words “automate” and “automation” to their common word stem “automat”. The idea is that the relevant information in words is stored in their linguistic root, not in their gram-

¹²No regional data on ethnic German inflows were recorded in the state of Bavaria (Glitz, 2012). With Baden-Wuerttemberg and North Rhine-Westphalia our sample includes the most innovative and most populous federal states.

¹³For details on the assignment of labor market regions to the analysis sample, see also Table 2A-9 in the Appendix. We exclude the labor market region of Ulm from our analysis because it is partly in the state of Bavaria, which did not implement the placement policy.

¹⁴If not directly available, we draw on English publications within the same DOCDB patent family available from other patent offices, where English is the official language (e.g., the USPTO and the UKIPO). For more than 99.86 percent of the patent applications the English abstract is available in PATSTAT. Patent applications with a missing English abstract are excluded for the construction of the regional measures of automation innovation.

¹⁵We identify an applicant’s entity with the sector categorization provided by PATSTAT. We consider applicants as a corporate entity if they are labeled as either “Company”, “Company Gov Non-Profit” or “Company Hospital”. In the very rare case of missing applicant entities, we assume that the applicant entity is a company. We consider applicants as non-corporate entities if they are labeled as either “hospital”, “individual”, “governmental non-profit university”, “governmental non-profit” or “university”.

mathematical form. We classify patent applications into automation patents or non-automation patents using a simple Boolean search: if a pre-processed English abstract contains one of the following stemmed substrings (“automat”, “execut”, “detect”, “input”, “system”, “display”, “output”, “inform”), we classify the patent application as an automation patent and as a non-automation patent if otherwise.¹⁶ Even though this classification method does not rely on any sophisticated semantic-based analysis, it performs well in manual checks. In later robustness checks, we show that our main findings are robust to variations in the set of automation keywords.

Figure 2A-6 shows the shares of automation innovation across 34 technological areas. These shares are consistent with common expectations what technologies most likely yield automation. The share of automation innovation is high in technological areas such as “IT Methods”, “Telecom” and “Transportation”, and low in technological areas such as “Organic Chemistry” and “Polymers”.

We further cross-check our classification method by comparing the intersection of the sample of EP patents with the classified US patents data from Mann and Püttmann (2018).¹⁷ Both automation indicators are highly correlated: our measure is equal to theirs in 77.3 percent of cases and the correlation coefficient corresponds to 0.45.¹⁸ Overall, our automation patent classification shows a lower recall rate, which renders our approach more conservative compared to theirs.

We assign each automation and non-automation patent application to one of five main technology areas. For this purpose, we map the IPC classes using the concordance table developed by the Fraunhofer ISI and the Observatoire des Sciences et des Technologies in cooperation with the French Patent Office (Schmoch, 2008).¹⁹ The five main technology areas are: “Chemistry”, “Electrical Engineering”, “Instruments”, “Mechanical Engineering” and “Other Technology Areas”. This allows us to explore the effects of labor supply shocks on automation innovation across different technology areas. For illustrative purposes, Figure 2A-7 presents the share of patents across technology areas separately by each automation keyword appearing in the patent abstracts. There is significant variation in the appearance of these keywords across technological areas: for example, while the keyword “detect” frequently appears in the technology area instruments, it less commonly appears in patents related to mechanical engineering. Table 2A-29 presents examples of automation patents from our sample with their full English abstract.

Figure 2A-8 shows the annual number of patent applications related to automation and

¹⁶The set of keywords is borrowed from Mann and Püttmann (2018), who created a manually labelled training dataset of 560 granted patents to eventually classify all USPTO patents as automation or non-automation inventions.

¹⁷We can link 41 percent of the EP patents in our dataset to their equivalent in the US dataset through a common DOCDB patent family number.

¹⁸For this comparison, it is important to point out that the out-of-sample error rate of the algorithm by Mann and Püttmann (2018) is equal to 22.6 percent, thereby reducing the potential positive correlation to our measure of automation innovation.

¹⁹The IPC codes are aggregated into 34 technology areas. These technology areas are associated with one of the five main technological areas.

non-automation between 1992 and 2006. Although there was a general rise in the total number of patents, the increase was particularly strong for patents covering automation inventions. The number of automation patents has increased from about 1,500 per year in 1992 to more than 4,500 in 2006. Figure 2A-9 in the Appendix presents the share of automation patents by main technology areas over time. The shares of automation patents differ considerably between classes but remain overall fairly stable during our observation period.

Regionalization of patents

We assign all patent applications to regions using the inventors' region of residence as inferred from the geocoded address information in the OECD REGPAT Database (March 2018 edition).²⁰ If a patent application lists multiple inventors, we divide the respective patent application equally among all the inventors' regions of residence using fractional counts.²¹ Then, we construct a labor market region-year panel dataset on the number of automation and non-automation patent applications for the years 1992 to 2010.²² For this purpose, we use the priority year of the patent applications to capture the time of inventive activity (e.g., OECD, 2009). Figure 2A-11 in the Appendix shows the distribution of the average share of automation patents across labor market regions between the years 1992 and 2006.

2.3.2 Ethnic German Inflows and Other Regional Data

To construct our key explanatory variable, we obtain data on the annual regional inflows of ethnic German immigrants from official registers provided by each state's federal admission centers.²³ The variable is defined as the ethnic German inflow in $t - 2$ divided by the regional workforce in $t - 3$ and scaled by a factor of 1,000. We allow for up to two lags in order to better reflect the delayed impact of labor supply shocks on innovation and patenting. Since immigrants were locked into their initially assigned region for the first three years, our set-up does not suffer from post-placement mobility of ethnic Germans.

We combine the ethnic German inflow data with a rich set of regional control variables:²⁴ the size of the labor force and the share of unemployed workers stem from the German Employment Office, population size, gross domestic product, and different measures of gross value added stem from the *Working Group Regional Accounts VGRdL*²⁵, and the regional

²⁰Note that we have merged the labor market regions "Osterode" and "Goettingen" into one region to make the patent data compatible with the regional data on education groups and skill groups.

²¹This also applies to patent applications with foreign co-inventors.

²²We assign NUTS3 regions to labor market regions using a reference file from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

²³While we use the inflow data made available by Glitz (2012) and Piopiunik and Ruhose (2017), the original data comes from the admission centers in each state for the years 1992 to 2001. The original data from 2002 to 2006 comes from the Bundesarbeitsgemeinschaft Evangelische Jugendsozialarbeit e.V., Jugendmigrationsdienste.

²⁴See Table 2A-10 for an overview of regional characteristics and their sources.

²⁵The statistical offices of all German states and the Federal Statistical Office are members of the *Working Group Regional Accounts VGRdL*.

shares of immigrants stem from INKAR online and from Glitz (2012).²⁶ We supplement our spatial panel dataset with three skill groups (medium skilled, high skilled, engineers and scientists), twelve occupation groups and one old age group (employees aged 55 or older) using employment data from the Institute for Employment Research (IAB) Establishment History Panel. With the inclusion of skill groups we proxy for regional innovative capacity and R&D personnel, and with occupation groups we account for the industrial structure of districts (in which the establishments are located). Since recent research has shown that population ageing might stimulate the adoption of automation technologies (Acemoglu and Restrepo, 2019; Zator, 2019), we also include a control variable for the share of old age workers. Since the IAB data use the *full population* of all establishments with at least one employee subject to security contributions in Germany,²⁷ the data provide precise measures of these regional groups.

2.3.3 Descriptive Statistics

Table 2.1 reports the summary statistics of our region-year panel dataset. The average number of total patent applications filed in a labor market region in a given year is 91.8. While the average number of automation patents corresponds to 24.0, the large standard deviation of 56.9 patent applications reflects the substantial variation in automation innovation capacity across German regions, with a number of high performing regions. We log-transform the patent count variables for the empirical analysis to account for this right-skewed distribution.²⁸ The average share of automation patents in total patents in a region equals 23.0 percent.²⁹

The average annual ethnic German inflow corresponds to 535.9 immigrants per region. While the mean of the interaction term between the binding placement dummy and the inflow rate is equal to 1.7, the average inflow during the binding allocation years (1996-2006) corresponds to 2.54 ethnic Germans per 1,000 workers in a region.

Table 2.1 also reports summary statistics on lagged regional control variables such as population size, the overall share of the non-native population (average 8.5 percent), GDP per capita (average 23,000 EUR), and the unemployment rate (average 10 percent). The table further reports fractions of the three different skills groups and twelve occupation groups.

²⁶The data from Glitz (2012) is originally from the German Statistical Office. We compared the data of the two different sources for the overlapping years finding a very high degree of consistency.

²⁷From 1999 onwards, the data set also covers marginally employed persons.

²⁸Since a small minority of regions contains zero filed patent applications in a given year, we add plus one to every region-year observation before transforming the variables into logs.

²⁹The share of automation patents is not available for two observations because there were zero patent applicants in two region-year pairs.

2.4 Empirical Framework

As our main empirical approach, we employ a difference-in-difference design using region-year pairs from the pre-allocation and from the period of the binding allocation rule.³⁰ The continuous treatment is the lagged region-year specific inflow of ethnic Germans. As advocated by Dustmann *et al.* (2016b), we employ a pure spatial approach relating automation innovation to immigrant inflows. Our approach covers the *total* innovation effect of immigration-induced labor supply shocks at the region level taking into account complementarities across regional skill groups. This approach is also immune to misclassification due to skill downgrading.³¹

We estimate the impact of ethnic German inflows on (i) the share of automation innovation and (ii) the level of automation innovation. Exploiting the quasi-exogenous placement of ethnic Germans over time, we make use of the panel structure of our dataset in which we observe the regional innovative activity and the inflows of ethnic German in consecutive years before and after the introduction of the binding placement policy. We estimate the following difference-in-difference model with region and time fixed effects at the region-year level:

$$\begin{aligned} \frac{A_{rt}}{N_{rt}} = & \beta_0 + \beta_1 I_{rt-2} + \beta_2 A_{rt-2} \\ & + \beta_3 A_{rt-2} \times I_{rt-2} + X'_{rt-3} \vartheta + \delta_r + \tau_t + \eta_{rt} + \epsilon_{rt}, \end{aligned} \quad (2.4.1)$$

where $\frac{A_{rt}}{N_{rt}}$ is the share of automation patents filed in period t in region r . I_{rt-2} corresponds to the lagged inflow of ethnic Germans allocated to region r in period $t-2$ divided by the pre-existing workforce in $t-3$.³² A_{rt-2} is a dummy variable that equals 1 if the state-wide allocation policy was binding in region r in $t-2$. The coefficient β_1 captures the effect of the potentially endogenous inflows before the introduction of the binding placement policy. The interaction term $A_{rt-2} \times I_{rt-2}$ corresponds to a continuous treatment. The sum of the coefficients β_1 and β_3 captures the total effect of the immigration-induced labor supply shocks during the binding allocation period. If the positive labor supply shocks indeed reduced automation innovation, we would expect a negative value as the sum of β_1 and β_3 . We report the total effects and the corresponding p-values in all tables. By including a vector of region-fixed effects δ_r , we control for time-invariant unobserved factors across regions such as the general capacity to innovate in automation technologies. We account for national time trend in automation innovation that affects all regions uniformly by including a vector of time-fixed effects τ_t . To test the robustness of results, we include a rich set regional

³⁰See Table 2A-9 in the Appendix for details on the analysis sample of region-year observations and the state-level implementation of the assigned place of residence act. While the allocation policy remained in effect until 2009, regional data on ethnic German inflows for the years after 2007 are not available.

³¹Note that the occurrence of skill downgrading is particularly pronounced in the first years after arrival of immigrants (Dustmann *et al.*, 2017b).

³²Note that we also investigate the robustness of results by employing alternative lag structures ranging from t to $t-3$. Using alternative empirical models with an overlapping data structure (Section 2.6.4) or poisson regressions (Section 2.6.3) yields similar results.

time-variant control variables X'_{rt-3} that might be related to automation innovation such as population size, the overall share of the non-native population, the unemployment rate, GDP per capita, different variables on gross value added, the share of employees older than 55, three skill groups and twelve occupation groups. In the most conservative model, we include year-by-state fixed effects η_{rt} to account for systematic changes over time in automation innovation that are common to all regions within the same state, such as state-level support infrastructure or subsidies. The error term is ϵ_{rt} . We cluster standard errors at the region level to account for correlations within regions over time. Our regression models identify the effect of labor supply shocks on automation innovation from the variation of ethnic German inflows over time within the same region. We weight the region-year observations with pre-determined regional population sizes as of 1991.³³

While we investigate intensity effects in regression equation 2.4.2, we analyze scale effects using a similar set up:

$$\begin{aligned} \ln(A_{rt} + 1) = & \beta_0 + \beta_1 I_{rt-2} + \beta_2 A_{rt-2} \\ & + \beta_3 A_{rt-2} \times I_{rt-2} + X'_{rt-3} \vartheta + \delta_r + \tau_t + \eta_{rt} + \epsilon_{rt}, \end{aligned} \quad (2.4.2)$$

where $\ln(A_{rt} + 1)$ is the log-number of automation patents filed in period t in region r .

The identification strategy hinges on the assumption that the inflows of ethnic Germans are exogenous to unobserved factors that may affect (automation) innovation. Prior research has elaborately shown that these inflows were exogenous to the regional labor market and to the overall regional innovative capacity (Glitz, 2012; Jahn and Steinhardt, 2016). Under the identifying assumption, the allocated inflows should be unrelated to current automation innovation since firm responses are unlikely to result in an immediate change in observed patenting activities. In Section 2.5.1 we show that the allocated inflows are indeed not associated with the share of automation innovation, the level of automation innovation and the level of non-automation in the year of placement. Moreover, the lack of pre-trends in our event study analyses in Section 2.5.1 solidifies the identifying assumption.

2.5 Results

In this section, we first present our main results (Section 2.5.1) and then examine the underlying mechanisms (Section 2.5.2). To this end, we explore potential effect heterogeneity with respect to labor market tightness, firm characteristics, and regionality.

2.5.1 Main Results

Table 2.2 shows the results of estimating equation (1): the effect of the ethnic German inflows on the regional share of automation innovation is negative and highly significantly

³³Alternatively, weighting observations with regional GDP in 1991 yields similar results.

different from zero, irrespective of specification. The total effect of the ethnic German inflows on automation innovation and its corresponding p-value is reported for each specification: increasing the inflow rate by one ethnic German per 1,000 employed workers leads to a significant decline in the share of automation innovation by around 0.75 percentage points (column 4, full specification). Notably, the magnitudes of the estimated coefficients increase only modestly when we include time-variant regional controls and state-by-year fixed effects. The results are also robust to controlling for the regional time-variant shares of twelve occupation groups and three skill groups (column 4). The fact that the coefficients barely change across specifications is consistent with prior research showing that the inflows are orthogonal to regional (labor market) conditions (Glitz, 2012). Overall, the results indicate that an exogenous increase in the low skilled workforce significantly shifts the direction of technological change away from automation innovation.

We repeat the analysis with the level of automation innovation as dependent variable: in line with the negative impact on the share of automation, we also find a negative effect of the exogenous inflows on the level of automation innovation (Table 2.3, column 1-4). The point estimates suggest that an increase in the inflow rate by one ethnic German per 1,000 workers is followed by a decline in the number of automation patents by 4.3 percent (column 4, full specification), which represents about 3 automation patent fewer for the average region (which has about 27 automation patents per year). Given that the average inflow rate is about 2.54 ethnic Germans per 1,000 workers during the years of the binding allocation policy, this constitutes an economically significant effect.

Both findings suggest that regional labor supply shocks play a critical role in technological progress related to automation. Taken together, our findings are consistent with theoretical predictions that labor scarcity will encourage labor-saving innovation.

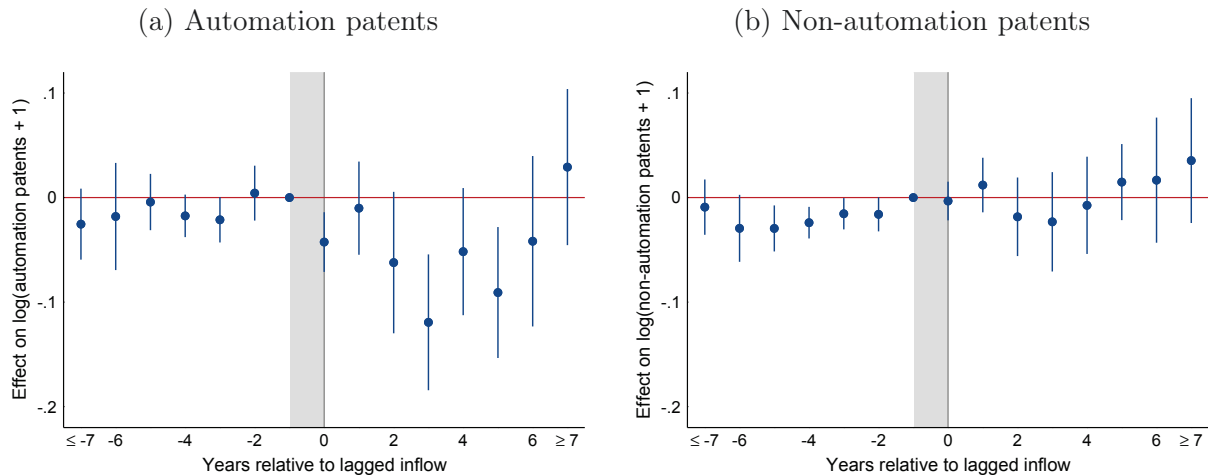
Finally, we run the empirical analysis with the level of non-automation innovation as outcome variable (Table 2.3, column 5-8). After controlling for time-variant variables, the estimates of the total automation innovation effect are very close to zero and statistically insignificant indicating that non-automation did not adjust to the regional labor supply shocks (column 7-8). The insignificant effects of the labor supply shock on non-automation innovation serve as empirical evidence against alternative explanations that systematic unobserved factors might drive the main results on the effects of the inflows on innovation in general.

Dynamics of the Effect

In order to visualize how automation innovation adjusts to labor supply over time, we perform an event study by regressing the level of innovation on the lagged inflow rates interacted with time dummies that capture the response relative to the year of the inflow. Figure 2.2a (Figure 2.2b) displays the estimated lead and lag effects on the level of automation patents (non-automation patents) by year relative to the lagged inflow rate. The estimated effect in $t = 0$ refers to the first labor supply shock resulting from the binding placement policy. As evident in Figure 2.2a, we find significant negative effects on automation innovation during

most years after the introduction of the binding placement policy. By contrast, the labor supply shocks do not have any from zero distinguishable effect on the level of non-automation innovation.

Figure 2.2: Event study: the effect of the ethnic inflow rate on the level of patents



Notes: The left-hand (right-hand) figure displays the estimated effects of the ethnic inflow rate on the level of automation patents (non-automation patents) by year relative to the lagged inflow rate. The outcomes are regressed on the ethnic inflow rates interacted with time dummies indicating the years relative to the introduction of the binding allocation policy. Full set of time-variant controls, fixed effects (region, time, year-by-state) and relative year fixed effects included. Standard errors clustered on the region level.

We further explore the effects across specific lag structures using analogous OLS models. For this purpose, we estimate the following model with region and time fixed effects at the region-year level:

$$\begin{aligned} \frac{A_{rt}}{N_{rt}} = & \beta_0 + \beta_1 I_{rt-L} + \beta_2 A_{rt-L} \\ & + \beta_3 A_{rt-L} \times I_{rt-L} + X'_{rt-L-1} \vartheta + \delta_r + \tau_t + \eta_{rt} + \epsilon_{rt}, \end{aligned} \quad (2.5.1)$$

where $\frac{A_{rt}}{N_{rt}}$ is the regional share of automation patents in region r filed in year t . I_{rt-L} corresponds to the inflow of ethnic Germans in year $t-L$ divided by the size of the workforce in the previous year. X'_{rt-L-1} corresponds to the full set of control variables. Since it is a priori unclear which lag structure $L \in \{0, 1, 2, 3\}$ between the labor supply shock and the response in innovative activity is empirically most pertinent, we evaluate the effect for each of the inflows in the years t , $t-1$, $t-2$, and $t-3$ in Table 2.4.

We find a significantly negative effect of the ethnic German inflows on the share and level of automation innovation in the years $t-1$ and $t-2$ (Table 2.4, column 2-3 and 6-7). In contrast, we find no significant effects for the year of treatment t and the more distant year $t-3$. These results suggest that the impact of the positive labor supply shock on automation innovation materializes in the first and second year and, hence, early after the exogenous shock.

By contrast, we do not find any significant effects of the positive labor supply shocks on

the level of non-automation across different lag structures (Table 2A-11 in the Appendix).

Results by Technology Area

Next, we assess the innovation response to the labor supply shock across across the five main technology areas (Chemistry, Electrical Engineering, Instruments, Mechanical Engineering, and Other Fields). We construct the dependent variables for each technology area. That is, the average “Share of automation patents (Mechanical Engineering)” corresponds to the regional number of automation patents in mechanical engineering divided by the total regional number of patents in mechanical engineering. We assume effect heterogeneity since automation innovation is probably more likely to economize on manual labor in some areas (such as mechanical engineering) than in others. Figure 2A-10 visualizes the number of low- and unskilled workers by specific technology areas in Germany over time.³⁴ In fact, the largest workforce of low- and unskilled workers is employed in industries related to mechanical engineering and chemistry. Moreover, close to 60 percent of the ethnic Germans (who used to work in their country of origin) were employed in fairly low-skilled and largely manual occupations such as farmers, laborers, transport workers, operatives and craft workers (see Figure 2A-5a).

The effect of the labor supply shock on automation innovation indeed varies considerably across technology areas (Table 2.5). To show this we define the outcome variables related to specific main technology areas, as share of automation innovation (panel A), level of automation innovation (panel B) and level of non-automation innovation (not reported). The effect is especially pronounced in the area of mechanical engineering: an increase in the inflow rate by one ethnic German per 1,000 workers is associated with a decline in this specific share of automation innovation by 0.54 percentage points (column 4, Panel A). In addition, we find some evidence of negative effects on automation innovation related to chemistry. In contrast, we find such systematic and significant negative effects neither for any other technology area, nor for the number of non-automation innovation in general (not reported).

2.5.2 Discussion of Mechanism

As a major contribution of our study, we can shed light on the mechanism through which labor supply affects innovation. We specifically investigate the tightness of the labor market, firm characteristics, and the regional demand for innovation.

Labor Market Tightness

We investigate in how far the size of the labor supply effect on automation innovation depends on pre-existing labor scarcity. The idea is that labor inflows to regions characterized by labor

³⁴To this end, we calculated the size of the workforce in industries related to the different main technology areas using data from the Institute for Employment Research (IAB) and concordance tables between industries and technologies by Dorner and Harhoff (2018).

shortages may have larger marginal effects on automation innovation because positive labor supply shocks might more effectively relax labor supply constraints and reduce search costs for workers. Prior research has shown that search costs of firms are higher in tight than in slack labor markets (Muehlemann and Leiser, 2018).³⁵ To the contrary, labor inflows may have smaller marginal effects on automation innovation in periods of labor abundance characterized by high unemployment. To shed light on effect heterogeneity by labor market conditions, we split our sample into tight and slack labor market regions.³⁶

We split the sample at the 75-percentile of the pre-determined unemployment rate in the year before the placement of ethnic Germans became binding, i.e., in the year 1995.³⁷ The results in Table 2.6 indicate that there are strong negative effects of labor supply shocks on the share and level of automation innovation (column 1 and column 5) for regions with tight labor markets. At the same time, the negative effects of labor supply shocks on the share and level of automation innovation (column 2 and column 6) are absent and insignificant in regions with high pre-determined unemployment, i.e., slack labor markets.

This interesting effect heterogeneity is robust to using an alternative definition of labor market tightness: we split the sample by the 75-percentile of the unemployment rate in $t - 3$, the year before the actual labor supply shock in $t - 2$. This definition captures more precisely the pre-existing labor scarcity. While we again find significant negative effects of labor supply shocks in tight labor markets on automation innovation (column 3 and 7), the total effects are even slightly positive and highly insignificant in slack labor markets. To test the robustness regarding different thresholds, we present analogous regressions using the 50 pctl. (90 pctl.) of the unemployment rates in Table 2A-25 (Table 2A-26). The results are robust to these modifications.

Next we repeat the analysis using the level of *non-automation* innovation as the outcome variable. Table 2A-27 reports the results. Strikingly, our findings show that there are *no heterogenous effects* of labor supply shocks on non-automation innovation by pre-existing labor market tightness. This suggests that the role of pre-existing labor scarcity is specific to *automation innovation*.

While labor supply shocks might influence automation innovation also through *wages*, adjustments are, in practice, hampered by wage rigidity (see for example Card *et al.* (1996)). In fact, the German labor market is notoriously inflexible owing to strict labor market regulation and powerful unions.³⁸ Accordingly, Glitz (2012) and D'Amuri *et al.* (2010) find little to no effect of immigration on wages in Germany in the 1990s. Hence, wage effects are very unlikely to mediate the effects of labor supply on automation innovation in our setting.

³⁵Also, job-finding rates are higher in tight labor markets (Shimer, 2005).

³⁶Our approach is loosely related to Buchheim *et al.* (2019) and Nakamura and Steinsson (2014) in classifying regional labor markets into a tight or slack category.

³⁷As an exception, it refers to the year 1996 for regions in Lower Saxony, and to the year 2001 for regions in Hesse.

³⁸For example, the collective bargaining coverage in Germany was larger than 80 percent in 1990 and equal to 68 percent in 2000 (OECD, 2004). For comparison, note that the degree of unionisation is substantially higher in Germany compared to the US where the collective bargaining coverage was only 14 percent in 2000.

In sum and consistent with search costs for workers playing a key role, labor supply shocks have larger marginal effects on automation innovation in tight labor markets.

Firm Characteristics

We examine what innovators appear to be most responsive to the labor supply shock (see Figure 2.3 for an overview of all results related to firm characteristics). We first distinguish between corporate entities or non-corporate entities, such as universities, governmental institutions, or natural persons. These entities clearly differ with respect to their objective function (in simplified terms: profit orientation vs. non-profit maximizing goals) and their exposure to competition. We therefore expect them to differ in their response to the labor supply shock. Since the labor-automation link should be most vivid for firms under cost pressure, we hypothesize that corporate entities respond substantially stronger than non-corporate entities to the immigration of ethnic Germans. Focusing on patents filed by non-corporate entities as our dependent variables, we find no statistically significant effect of the labor supply shock on the share and the absolute number of automation patents (Figure 2.3a and 2.3b). These results on non-corporate applicants are in stark contrast to the effects on innovation by firms. In line with our hypothesis, only corporations seem to have adjusted their automation innovation activities to the labor supply shock.

We next explore potential heterogeneity in innovation responses between firms of different size, age, and locality. We proxy firm size as the cumulative number of previously filed patents, firm age as the number of years since the first patent filed, and locality as the number of distinct inventor regions of all patents of a given firm.³⁹ We split the sample of patents at the respective medians of these three variables stratified by technology area and year. Obviously, these variables constitute crude approximations of the actual firm characteristics. Nonetheless, we do find indicative evidence for heterogeneous firm responses.

Figure 2.3 presents the point estimates of the combined effect of $Inflow\ rate_{t-2}$ and $Allocation_{t-2} \times Inflow\ rate_{t-2}$ on the share of automation innovation, and the level of (non-)automation innovation.⁴⁰ We find a significant negative effect on automation innovation (share and level) by small firms and young firms. In contrast, the same point estimates are smaller (in absolute size) and less precisely estimated for large firms and old firms. We find no substantial difference in the effect when distinguishing firms by whether their innovation activities are geographically concentrated or dispersed. This is somewhat surprising, as we would expect that local firms react more strongly to regional labor supply shocks. However, note that we do not observe whether innovation activities and manufacturing happen within the same region or in different ones. Distinguishing firms along this dimension would be more meaningful and may result reveal the expected differential effect.

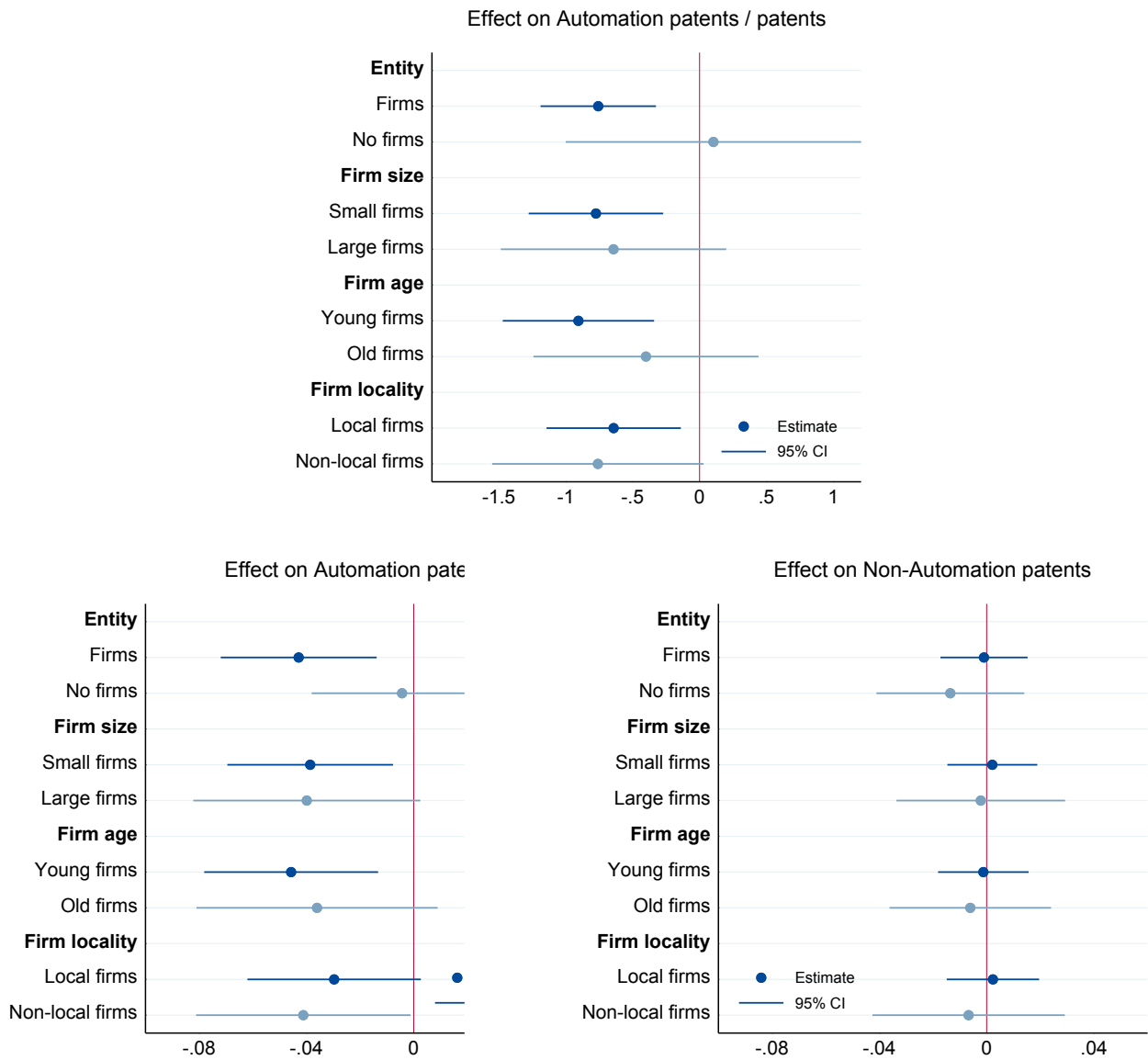
It bears mentioning that, again, we do not find any significant effects (or differences) on

³⁹We rely on the PATSTAT standardized name (PSN ID) for the patent applicants to construct firm patent portfolios.

⁴⁰The corresponding regression results can be found in Tables 2A-12, 2A-13 and 2A-14 in the Appendix.

the level of non-automation.

Figure 2.3: Effect on automation innovation by different originators



Notes: This figure presents the point estimates of the combined effect of $Inflow\ rate_{t-2}$ and $Allocation_{t-2} \times Inflow\ rate_{t-2}$ on innovation by different originators, i.e., patents filed by different patent applicants. Figure 2.3a displays the estimated effects on the share of automation patents. Figure 2.3b (Figure 2.3c) displays the estimated effects on the level of automation patents (non-automation patents). The corresponding regression results can be found in Tables 2A-12, 2A-13 and 2A-14 in the Appendix.

Demand for Automation Innovation

In this section, we explore whether the decrease in automation innovation activities is a response to lower *internal* demand (i.e., the innovating firm itself demands less automation technology) or to a lower *external* demand (i.e., other firms demand less automation technology).

With no information on whether automation technologies are used internally and/or

externally, we leverage a different classification of patented technologies: process vs. product innovation. In general, process innovations relate to how goods or services are created whereas product innovations refer to the outcomes of such procedures. As a result, we assume that process innovations are more likely linked to internal use, whereas product innovations are more likely linked to market activities. Differences in the effect for process vs. product automation innovation may therefore indicate whether the local labor supply shock is channeled through changes in internal or external demand for automation innovation.

We test this empirically and classify patents into process vs. non-process innovation in a similar manner to Ganglmair and Reimers (2019).⁴¹ There are on average 24.0 non-automation and only about 11.4 automation patents related to processes filed in a region in a given year. The average share of automation patents that is related to process innovation per region corresponds to 27.9 percent. Table 2.7 reports the effects of the ethnic German inflows on automation innovation that is related to processes and products. Product automation innovation (column 2, 4 and 6) seems to respond more strongly than automation innovation related to processes (column 1, 3 and 5). The latter estimates are, however, less precisely estimated due to the small share of all patents belong to the process and automation category.⁴² In summary, the results for process and product innovation indicate that firm innovation responses are largely driven by a lower external demand; i.e., other firms demand fewer automation technologies for adoption. This observation is in line with the findings of prior literature. Clemens *et al.* (2018), Lewis (2011), Imbert *et al.* (2019), and Monras (2019) all find that low-skilled labor supply determines firm decisions to adopt (new) production technologies.

In light of national, if not global, product markets, a *regional* demand-pull mechanism seems at first glance surprising. First, local firms demanding automation technologies may source them from other firms outside of the region. Likewise, firms developing automation technologies can supply these to other firms outside of the regions. Both aspects should weaken the relationship between local labor supply and local automation innovation. Given that we still find a significant effect, demand and supply for automation technologies seems to be fairly localized.⁴³ This inference is not too far-fetched, as previous studies argue for considerable geographical bias of buyer-supplier relationships (Bernard *et al.*, 2019) and technology transfer (Almeida and Kogut, 1999; Audretsch and Feldman, 1996). Moreover, the arguably higher specificity of automation technologies, and the need for their continuous maintenance, may further explain the spatial proximity between originator and user.

Unfortunately, market activities are hard to observe as product purchases and technology licensing deals most often remain undisclosed to the public. Nonetheless, we exemplify the geographical bias in the market for technology by looking at the geographical distance

⁴¹We searched for the keywords “method”, “process”, and “procedure” in the patent claim text. There was no patent claim text available for around 12 percent of patent applications.

⁴²Firms may refrain from patenting process innovations and instead decide to avoid misappropriation through secrecy (cf. Levin *et al.*, 1987; Ganglmair and Reimers, 2019). We cannot exclude that some selection into patenting leads to us to underestimate the true effect on automation process innovations.

⁴³As our spatial approach does not capture potential spillovers, we likely underestimate the overall effect.

between the location of the cited patent and the citing patent. Figure 2A-12 shows that indeed more than half of the citations to patents in our sample originate within a proximity of 20 km. Albeit an imperfect measure⁴⁴, we consider this as indicative for geographical bias in the technology market. To further explore whether close-by external demand drives the responses in automation innovation, we split our sample of regions into small and large labor markets. We assume that large labor market regions also show higher economic activity and therefore play a larger role for focal firms relative to other regions. If local external demand drives the effect on automation innovation, we would expect a stronger (weaker) effect of the local labor supply shock in large (small) labor market regions. Indeed, Table 2.8 shows that the effect on the share and the level of automation innovation is much larger in magnitude and more precisely estimated for the sample of large labor market regions compared to the sample of small labor market regions.

2.6 Robustness Analyses

In this section, we present additional results with weighted patent counts (Section 2.6.1) and alternative measures of automation innovation (Section 2.6.2). We also assess the labor supply-innovation nexus for the pre- and post-allocation period separately (Section 2.6.4). Moreover, we repeat our main analysis with an alternative estimation model (Section 2.6.3) and alternative samples (Section 2.6.5). All tests confirm the robustness of our previously reported findings.

2.6.1 Accounting for Patent Value

Some patent applications turn into highly successful patents while others remain close to irrelevance. To account for these differences with respect to future patent value, we resort to a weighting scheme in which patent applications receive different impact depending on their patent grant status⁴⁵ (Table 2A-15, col. 1), the size of their DOCDB patent family (col. 2), and the number of received US patent forward citations within the first 3 years (col. 3). The first weighting scheme allows to only consider patent applications with a successful examination, the second puts greater emphasis on innovative activities closely related to the particular invention, and the third accounts for the short- to medium run impact of each patent for future innovative activity. All estimated effects on the weighted share of automation patents (Panel A) and the weighted number of automation patents (Panel B) confirm that the positive labor supply shock leads to a reduction in automation innovation. To the contrary, we do not find any significant effects on the weighted measures of the level of non-automation innovation (not reported).

⁴⁴Technology adoption does not necessarily lead to patented follow-on inventions. Moreover, citations may also be the result of knowledge spillovers, which are localized as well.

⁴⁵While a granted patent application has a weight of one, a non-granted application has a weight of zero.

2.6.2 Alternative Measures of Automation Innovation

While the recent literature has successfully applied keyword search in patent texts (see e.g., Dechezleprêtre *et al.*, 2019), the choice of relevant keywords can be disputed. Therefore, we examine the sensitivity of our findings to using three alternative keywords-based measures of automation innovation (Table 2A-16). First, we construct one reduced keyword measure by searching the pre-processed English abstracts for the following substrings: “automat”, “execut”, “input”, “system”, “output” and “inform”. Second, the extended keyword measure utilizes the following ten instead of eight keywords: “automat”, “execut”, “detect”, “input”, “system”, “display”, “output”, “inform”, “signal” and “sensor”. Finally, we construct a very conservative classifier by searching only for the unambiguous keyword “automat”. We analogously classify the underlying patent application as an automation patent as soon as one of the corresponding keywords appears in the English abstract. Reassuringly, the reduced and the extended keyword classifications as well as the very conservative classification confirm our baseline results on the effect of labor supply shocks on the share of automation patents, the level of automation patents and the level of non-automation patents.

2.6.3 Alternative Estimation Model: Poisson Regression

As patents are categorized as count data, we test the robustness of our OLS level findings with analogous poisson quasi-maximum likelihood regressions:

$$A_{rt} = \exp \left(\gamma_0 + \gamma_1 I_{rt-2} + \gamma_2 A_{rt-2} + \gamma_3 A_{rt-2} \times I_{rt-2} + X'_{rt-3} \vartheta + \delta_r + \tau_t + \eta_{rt} \right) + \epsilon_{rt}, \quad (2.6.1)$$

where A_{rt} denotes the number of number of filed automation patents in region r and year t . I_{rt-2} corresponds to the number of ethnic German inflows in region r and year $t-2$ divided by the workforce in the previous year. X'_{rt-3} represents a vector of the full set of control variables in year $t-3$. Again, we control for region fixed effects δ_r , time fixed effects τ_t and year-by-state fixed effects η_{rt} . We obtain the estimates using a Poisson pseudo-likelihood regression with multiple levels of fixed effects, as described by Correia *et al.* (2019). We continue to cluster standard errors at the regional level.

The results in Table 2A-17 confirm our main findings that the level of automation patents (column 1) declines in response to regional labor supply shocks. Once again, the level of non-automation patents remains unaffected by these shocks (column 2).

2.6.4 Separate Analyses for Pre- vs. Post-Binding-Allocation Periods

While our difference-in-difference approach exploits observations from before the introduction of the binding allocation rule, we follow the approach by Glitz (2012) and focus on

those region-year pairs with a binding allocation rule (1996-2006) in this section.⁴⁶ For this purpose, we first estimate the following fixed effects model using only region-year pairs from the binding-allocation period starting in 1996:

$$\frac{A_{rt}}{N_{rt}} = \beta_0 + \beta_1 I_{rt-2} + X'_{rt-3} \vartheta + \delta_r + \tau_t + \eta_{rt} + \epsilon_{rt}, \quad (2.6.2)$$

where $\frac{A_{rt}}{N_{rt}}$ is the regional share of automation patents in region r filed in year t . I_{rt-2} corresponds to the inflow of ethnic Germans in year $t-2$ divided by the size of the workforce in the previous year. Throughout this section, we weight the region-year observations with pre-determined regional population sizes as of 1995.

Table 2A-18 quantifies the effects of the supply shock induced by the inflow of ethnic Germans on the share of automation patents (column 1), the level of automation patents (column 2) and the level of non-automation patents (column 3). Again, we find negative effects of similar size on automation innovation confirming our difference-in-difference estimates from the baseline model.

To the contrary, we do not find any significant effects of the ethnic German inflows on automation innovation or non-automation using only region-year pairs from the pre-binding allocation period (see Table 2A-19).

As an additional robustness test, we use an *overlapping observation model* to avoid the arbitrariness in choosing a specific lag structure.⁴⁷ As a further advantage, the overlapping data structure exploits the data most efficiently (see e.g., Harri and Brorsen, 2009). We estimate the following model with region and time fixed effects at the region-year level:

$$\frac{\sum_{z=0}^2 A_{rt+z}}{\sum_{z=0}^2 N_{rt+z}} = \beta_0 + \beta_1 \frac{\sum_{z=-2}^0 Inf_{rt+z}}{L_{rt-3}} + X'_{rt-3} \vartheta + \delta_r + \tau_t + \eta_{rt} + \epsilon_{rt}, \quad (2.6.3)$$

where $\frac{\sum_{z=0}^2 A_{rt+z}}{\sum_{z=0}^2 N_{rt+z}}$ is the regional share of automation patents filed over the period t to $t+2$ in region r . Our main explanatory variable $\frac{\sum_{z=-2}^0 Inf_{rt+z}}{L_{rt-3}}$ corresponds to the cumulative inflow of ethnic Germans allocated to region r over the period $t-2$ to t divided by the workforce in $t-3$. We include regional time-variant control variables X'_{rt-3} . We cluster standard errors at the region level to account for correlations between regions over time. Given the overlapping data structure, we additionally report p-values calculated using the wild cluster bootstrap-t method by Cameron *et al.* (2008) to account for within-group dependence in estimating standard errors with a limited number of clusters.

Table 2A-20 reports results based on this overlapping observation model: an increase of one ethnic German inflow per 1000 workers is followed by a decline in the share of automation patents (number of automation patents) by 0.29 percentage points (0.18 percent). Both

⁴⁶See Table 2A-9 in the Appendix for details regarding the region-year pairs from the binding allocation period.

⁴⁷See also Glitz and Meyersson (2020) who use a similar overlapping observations model in their primary specification.

conventional p-values and p-values calculated using the wild cluster bootstrap-t method indicate that these effects are significant. At first glance, the estimated effects on automation innovation are smaller compared to our difference-in-difference results. However, note that these estimates capture the effect of labor supply shocks on automation innovation not in a single year but over a three-year period from t to $t + 2$. Once again, the ethnic German inflows have no effect on the level of non-automation innovation (column 3).

2.6.5 Alternative Samples

Omitting Specific Regions

On a national level, we find large differences in innovative activities across regions: for instance, while in 1991 (i.e., prior to the migration allocation) only 13 patent applications were filed in the labor market region “Hameln”, we count 613 patent applications in the labor market region “Stuttgart”.

To investigate whether our baseline results on labor supply and automation innovation are driven by potentially influential observations with unusually high or low regional levels innovation, we rerun our regressions omitting those regions. More precisely, we exclude any region-year data points from our sample if the pre-determined regional number of filed patent applications in 1991 is below (above) the 10th (90th) percentile. Our results are robust to excluding regions with low (Table 2A-21, col. 1, 3 and 5) or high (col. 2, 4 and 6) levels of innovative activity. The estimated effects of the ethnic German inflow rate on the share of automation patents (column 1-2) and the level of automation patents (column 3-4) are sizeable and significant across all specifications. Similar to our full-sample analysis, we find negligible and insignificant effects on the level of non-automation innovation (column 5-6).

We also investigate whether our results are robust to excluding regions that signed the so-called Gifhorn declaration. In the first years after the breakdown of the Soviet Union, a small number of regions in Germany had received disproportionate large inflows of ethnic Germans. In 1995, these regions signed the Gifhorn declaration that asked for making a more equal regional distribution of ethnic Germans mandatory (Niedersaechsische Landeszentrale fuer Politische Bildung, 2002).⁴⁸ After the implementation of the binding allocation policy, fewer incoming ethnic Germans were allocated to these regions. Results excluding these seven regions mirror our key findings on the effects of labor supply on the share (Table 2A-22, col. 1) and level of automation patents (col. 2). At the same time, we again estimate a zero effect of these shocks on the level of non-automation innovation (col. 3), confirming that our results are not sensitive to the exclusion of specific regions from the estimation sample.

⁴⁸The following counties signed the Gifhorn declaration: Wolfsburg, Salzgitter, Gifhorn, Nienburg/Weser, Cloppenburg, Emsland and Osnabrueck.

Addressing Outliers

In this section, we check if the results are sensitive to outliers in the ethnic inflow rate. For this purpose, we winsorize the inflow rate by replacing low (high) values by the 5th (95th) percentile. Table 2A-23 reports the results: once again, we confirm a negative effect of the labor supply shocks on both outcomes of automation innovation (col. 1-2) and an insignificant zero effect on non-automation innovation (col. 3).

Excluding Transition Years

In this robustness check, we exclude region-year pairs in which the binding allocation policy was implemented. The labor supply shocks induced by the ethnic German inflows during these transition years might be partly endogenous: while the states Baden-Wuerttemberg, Bremen, Hamburg, North Rhine-Westphalia and Schleswig-Holstein introduced the policy on March 1, 1996, the Saarland did so on March 11, 1996. Lower Saxony introduced the policy on April 4, 1997. Table 2A-24 shows that the results are robust to the exclusion of these transition region-year pairs.

2.7 Discussion and Conclusion

Economic theory suggests substitutability between labor input and labor-saving investments in automation innovation. Exploiting the placement of ethnic German immigrants across German regions, we analyze substitution effects between labor supply and automation innovation. Our difference-in-difference estimates show that the exogenous labor supply shocks led to an absolute as well as relative decline in regional automation innovation. These effects are concentrated to automation innovation technologies from industries with a high share of low- and unskilled workers (mechanical engineering and chemistry). Moreover, the degree of substitution between labor supply and automation innovation is moderated by pre-determined labor market tightness: identical inflows have no or weaker effects in regions with high unemployment. Our paper highlights the link between a greater availability of workers and the pressure of firms to invent cost-saving labor-replacing techniques.

Our study complements current research on the consequences of the rising adoption of automation technologies. Our findings suggest potential *feedback effects* of automation technologies on employment: if the adoption of automation technologies in one industry or firm frees up labor, the resulting increase in labor supply may *counterbalance* future automation innovation in other industries or firms.

Our research question bears policy relevance because labor supply induced shifts in automation innovation can influence the demand for and the relative remuneration of labor. A reduction in automation innovation will shift the production technology of firms towards labor intensive production. Firms will hire more workers for automatable jobs. Ironically, positive low-skilled labor supply shocks can shield human labor from being replaced by new

machines. The increased labor demand might counterbalance some of the labor market competition comparable native workers experience with immigrant workers, dampening the effects of immigration on wages and employment in the medium and long run. As a consequence, labor supply shocks can have redistributive effects in the relative remuneration of labor vs. capital.

Our paper fills an important gap in the literature by providing first empirical evidence on labor abundance and an endogenous response in automation innovation. Note that the impact of labor supply on automation innovation likely depends on the time period, the institutional framework, the level of observation, and the automation technology in question. Further research should explore other settings to investigate the causal role of labor supply for automation innovation.

Table 2.1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max	N
<i>Innovation</i>					
Patents	91.77	172.13	0.00	1938.92	1849
Log Patents	3.81	1.19	0.00	7.57	1849
Automation patents	24.00	56.94	0.00	718.72	1849
Log Automation patents	2.40	1.22	0.00	6.58	1849
Automation patents / patents	23.01	11.14	0.00	88.89	1847
Non-automation patents	67.77	118.09	0.00	1251.35	1849
Log Non-automation patents	3.55	1.16	0.00	7.13	1849
<i>Ethnic German Inflows</i>					
Inflow _{t-2}	535.92	666.64	0.00	7342.00	1849
Inflow rate _{t-2}	3.89	3.57	0.00	39.31	1849
Allocation _{t-2} × Inflow rate _{t-2}	1.73	1.84	0.00	12.26	1849
<i>Population and Economic Indicators</i>					
Log Population _{t-3}	12.51	0.73	11.27	14.83	1849
Share of immigrants _{t-3}	8.49	3.47	1.88	25.64	1849
GDP per capita _{t-3}	22.95	4.84	12.07	44.78	1849
GVA total _{t-3}	8.61	0.84	6.94	11.46	1849
GVA production _{t-3}	7.52	0.82	5.89	10.32	1849
GVA tertiary _{t-3}	8.14	0.90	6.43	11.23	1849
<i>Labor Market</i>					
Log Labor Force _{t-3}	11.65	0.74	10.37	14.03	1849
Unemployment rate _{t-3}	0.10	0.03	0.03	0.18	1849
Share age > 55 _{t-3}	0.11	0.02	0.06	0.17	1849
<i>Occupation Groups</i>					
Share agricultural _{t-3}	0.02	0.01	0.01	0.09	1849
Share unskilled manual _{t-3}	0.16	0.05	0.05	0.35	1849
Share unskilled services _{t-3}	0.15	0.04	0.09	0.28	1849
Share unskilled commercial and admin. _{t-3}	0.09	0.01	0.06	0.16	1849
Share skilled manual _{t-3}	0.17	0.03	0.09	0.27	1849
Share skilled services _{t-3}	0.05	0.01	0.03	0.08	1849
Share skilled commercial and admin. _{t-3}	0.18	0.03	0.10	0.32	1849
Share technicians _{t-3}	0.05	0.01	0.02	0.13	1849
Share semiprofessions _{t-3}	0.06	0.01	0.03	0.10	1849
Share engineers _{t-3}	0.02	0.01	0.01	0.07	1849
Share professions _{t-3}	0.01	0.01	0.00	0.04	1849
Share managers _{t-3}	0.02	0.01	0.01	0.05	1849
<i>Skill Groups</i>					
Share medium skilled _{t-3}	0.73	0.03	0.61	0.81	1849
Share high skilled _{t-3}	0.07	0.03	0.03	0.18	1849
Share research and development _{t-3}	0.02	0.01	0.00	0.06	1849

Notes: Summary statistics of the region-year panel dataset computed for the period 1994 to 2008. Inflow_{t-2}: ethnic German inflows in $t - 2$. Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. *Data sources:* See Table 2A-10 in the Appendix.

Table 2.2: The effect of the ethnic German Inflows on the share of automation innovation

	(1)	(2)	(3)	(4)
Inflow rate _{t-2}	0.069 (0.112)	0.106 (0.105)	0.110 (0.132)	0.139 (0.132)
Allocation _{t-2} × Inflow rate _{t-2}	-0.766*** (0.193)	-0.907*** (0.206)	-0.884*** (0.208)	-0.896*** (0.211)
Log Population _{t-3}			-47.652* (26.891)	-65.640** (31.622)
Log Labor Force _{t-3}			22.653 (15.172)	18.834 (16.523)
Unemployment rate _{t-3}			-15.026 (36.732)	-13.729 (36.853)
Share of immigrants _{t-3}			0.085 (0.241)	-0.116 (0.231)
GDP per capita _{t-3}			-0.228 (0.490)	-0.957 (0.663)
GVA total _{t-3}			33.700 (24.359)	26.945 (24.431)
GVA production _{t-3}			-7.165 (6.476)	-1.229 (6.912)
GVA tertiary _{t-3}			-16.391 (11.700)	4.003 (13.373)
Share age > 55 _{t-3}			-81.685 (53.888)	-5.282 (60.548)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Year-by-State fixed effects	No	Yes	Yes	Yes
Occupation + Skill groups	No	No	No	Yes
Observations	1847	1847	1847	1847
R-squared	0.556	0.581	0.584	0.591
Within R-squared	0.007	0.009	0.015	0.033
Total effect	-0.697	-0.801	-0.774	-0.757
P-value	0.000	0.000	0.000	0.001

Notes: OLS regressions. Dependent variable: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.3: The effect of the ethnic German inflows on the level of (non-)automation innovation

Dep. Var.:	Automation patents				Non-Automation patents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inflow rate _{t-2}	-0.008 (0.006)	-0.007 (0.006)	-0.005 (0.007)	-0.007 (0.007)	-0.015*** (0.006)	-0.016*** (0.005)	-0.011* (0.006)	-0.015** (0.006)
Allocation _{t-2} × Inflow rate _{t-2}	-0.046*** (0.015)	-0.048*** (0.015)	-0.041*** (0.015)	-0.035** (0.015)	-0.005 (0.010)	0.004 (0.009)	0.008 (0.009)	0.014 (0.008)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Occupation + Skill groups	No	No	No	Yes	No	No	No	Yes
Observations	1849	1849	1849	1849	1849	1849	1849	1849
R-squared	0.953	0.955	0.956	0.958	0.972	0.975	0.975	0.977
Within R-squared	0.014	0.014	0.037	0.070	0.016	0.017	0.042	0.089
Total effect	-0.054	-0.055	-0.046	-0.043	-0.019	-0.012	-0.003	-0.001
P-value	0.000	0.000	0.003	0.004	0.059	0.208	0.721	0.904

Notes: OLS regressions. Dependent variable: Column 1-4: Automation patents: number of automation patents filed in year t (entered in logs). Column 5-8: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.4: The effect of the ethnic German inflows on automation innovation - alternative lag structure

Dep. Var.:	Automation patents / patents				Automation patents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inflow rate _t	0.131 (0.121)				-0.004 (0.008)			
Allocation _t × Inflow rate _t	-0.112 (0.244)				0.000 (0.015)			
Inflow rate _{t-1}		0.049 (0.138)				-0.009 (0.007)		
Allocation _{t-1} × Inflow rate _{t-1}		-0.743*** (0.223)				-0.025* (0.014)		
Inflow rate _{t-2}			0.139 (0.132)				-0.007 (0.007)	
Allocation _{t-2} × Inflow rate _{t-2}			-0.896*** (0.211)				-0.035** (0.015)	
Inflow rate _{t-3}				0.039 (0.113)				-0.009 (0.006)
Allocation _{t-3} × Inflow rate _{t-3}				-0.133 (0.217)				-0.001 (0.015)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1847	1847	1847	1847	1849	1849	1849	1849
R-squared	0.581	0.579	0.591	0.582	0.957	0.958	0.958	0.957
Within R-squared	0.034	0.035	0.033	0.021	0.083	0.077	0.070	0.047
Total effect	0.019	-0.694	-0.757	-0.094	-0.004	-0.034	-0.043	-0.010
P-value	0.939	0.002	0.001	0.657	0.792	0.015	0.004	0.501

Notes: OLS regressions. Dependent variable: Column 1-4: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 5-8: Automation patents: number of automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.5: Effect on automation innovation across technology areas

Main Technology Area:	Chemistry	Electrical Engineering	Instruments	Mechanical Engineering	Other Fields
<i>A: Automation patents / patents</i>	(1)	(2)	(3)	(4)	(5)
Inflow rate _{t-2}	-0.113 (0.162)	0.113 (0.418)	0.089 (0.565)	0.369* (0.215)	-0.100 (0.276)
Allocation _{t-2} × Inflow rate _{t-2}	-0.772** (0.332)	-0.359 (0.650)	-1.274 (0.994)	-0.916*** (0.299)	-0.705 (0.580)
Observations	1798	1690	1666	1837	1667
R-squared	0.236	0.350	0.293	0.391	0.215
Within R-squared	0.015	0.024	0.017	0.029	0.022
Total effect	-0.885	-0.246	-1.185	-0.547	-0.805
P-value	0.009	0.707	0.174	0.079	0.202
Dep var mean	12.067	40.059	41.915	20.287	15.882
<i>B: Automation patents</i>	(6)	(7)	(8)	(9)	(10)
Inflow rate _{t-2}	-0.003 (0.007)	0.005 (0.009)	0.005 (0.010)	-0.004 (0.008)	-0.007 (0.007)
Allocation _{t-2} × Inflow rate _{t-2}	-0.021 (0.015)	-0.039* (0.020)	-0.018 (0.023)	-0.035** (0.016)	-0.013 (0.021)
Observations	1849	1849	1849	1849	1849
R-squared	0.913	0.929	0.906	0.927	0.785
Within R-squared	0.073	0.086	0.058	0.052	0.036
Total effect	-0.024	-0.033	-0.013	-0.039	-0.020
P-value	0.107	0.076	0.520	0.017	0.369
Dep var mean	0.752	1.255	1.152	1.547	0.527
Region and year fixed effects	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Occupation + skill groups	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Dependent variables: Panel A: Automation patents / patents: number of automation patents which are in a specific technology area divided by the number of patents in that technology area. Panel B: Automation patents: number of automation patents which are in a specific technology area (entered in logs). Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 2.6: Heterogeneous effects on automation innovation by labor market tightness

Labor Market Tightness:	Tight	Slack	Tight	Slack
	$U_0 < 75$ pctl	$U_0 \geq 75$ pctl	$U_{t-3} < 75$ pctl	$U_{t-3} \geq 75$ pctl
<i>A: Automation Patents / Patents</i>	(1)	(2)	(3)	(4)
Inflow rate _{t-2}	0.038 (0.169)	0.130 (0.231)	-0.027 (0.158)	0.334 (0.296)
Allocation _{t-2} × Inflow rate _{t-2}	-0.981*** (0.214)	-0.600 (0.741)	-0.809*** (0.207)	-0.225 (0.895)
Observations	1373	459	1371	457
R-squared	0.666	0.504	0.680	0.502
Within R-squared	0.055	0.078	0.057	0.085
Total effect	-0.943	-0.470	-0.836	0.109
P-value	0.000	0.481	0.001	0.888
<i>B: Automation patents</i>	(5)	(6)	(7)	(8)
Inflow rate _{t-2}	-0.008 (0.007)	-0.010 (0.015)	-0.010 (0.007)	-0.003 (0.020)
Allocation _{t-2} × Inflow rate _{t-2}	-0.043*** (0.014)	-0.010 (0.046)	-0.042*** (0.014)	0.024 (0.046)
Observations	1374	460	1371	459
R-squared	0.968	0.922	0.966	0.934
Within R-squared	0.075	0.173	0.087	0.159
Total effect	-0.051	-0.020	-0.052	0.020
P-value	0.001	0.659	0.001	0.617
Region and year fixed effects	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Occupation + skill groups	Yes	Yes	Yes	Yes

Notes: OLS regressions. Dependent variables: Panel A: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Panel B: Automation patents: number of automation patents filed in year t (entered in logs). Sample splits by the 75-percentile of the unemployment rate in the year before the binding placement (column 1-2) and by the 75-percentile of the unemployment rate in $t - 3$ (column 3-4). Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.7: Effect of ethnic German inflows on automation innovation – process and non-process innovation

Dep. Var.:	Automation patents / patents		Automation patents		Non-Automation patents	
	Process	Product	Process	Product	Process	Product
	(1)	(2)	(3)	(4)	(5)	(6)
Inflow rate _{t-2}	0.365 (0.234)	-0.039 (0.146)	0.006 (0.008)	-0.020*** (0.008)	-0.017** (0.008)	-0.020*** (0.007)
Allocation _{t-2} × Inflow rate _{t-2}	-0.958** (0.402)	-1.015*** (0.294)	-0.021 (0.017)	-0.041** (0.017)	0.037*** (0.013)	0.011 (0.009)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1835	1843	1849	1849	1849	1849
R-squared	0.512	0.326	0.940	0.932	0.959	0.964
Within R-squared	0.033	0.028	0.079	0.062	0.070	0.074
Total effect	-0.593	-1.054	-0.015	-0.062	0.020	-0.009
P-value	0.121	0.000	0.346	0.001	0.145	0.265

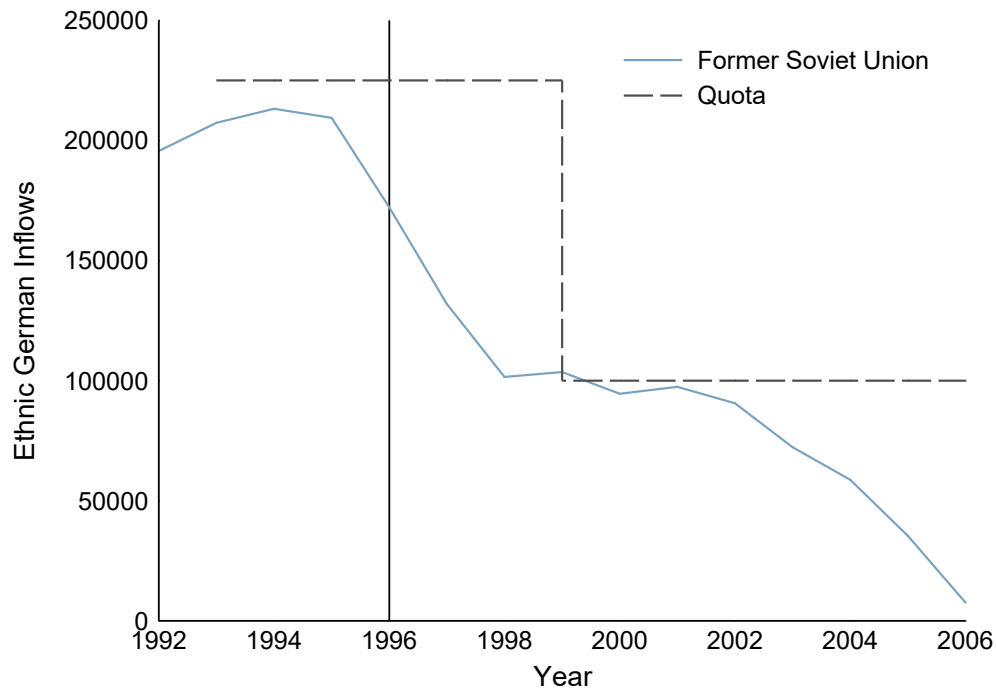
Notes: OLS regressions. Dependent variables refer to patent applications that are related to processes (column 1, 3 and 5) and non-processes (column 2, 4 and 6). Dependent variables: Column 1-2: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 3-4: Automation patents: number of automation patents filed in year t (entered in logs). Column 5-6: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.8: Effect of ethnic German inflows on automation innovation – by labor market size 1991

Dep. Var.:	Automation patents / patents		Automation patents		Non-Automation patents	
	Small	Large	Small	Large	Small	Large
	(1)	(2)	(3)	(4)	(5)	(6)
Inflow rate _{t-2}	0.085 (0.196)	0.112 (0.186)	0.001 (0.011)	-0.015* (0.008)	-0.003 (0.009)	-0.021** (0.010)
Allocation _{t-2} × Inflow rate _{t-2}	-0.588 (0.458)	-0.888*** (0.242)	-0.013 (0.028)	-0.039** (0.017)	0.018 (0.018)	0.010 (0.010)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes
Observations	931	901	933	901	933	901
R-squared	0.421	0.721	0.837	0.961	0.918	0.977
Within R-squared	0.027	0.074	0.060	0.117	0.093	0.129
Total effect	-0.503	-0.776	-0.011	-0.053	0.015	-0.010
P-value	0.224	0.007	0.650	0.004	0.305	0.339

Notes: OLS regressions. Dependent variables: Column 1-2: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 3-4: Automation patents: number of automation patents filed in year t (entered in logs). Column 5-6: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

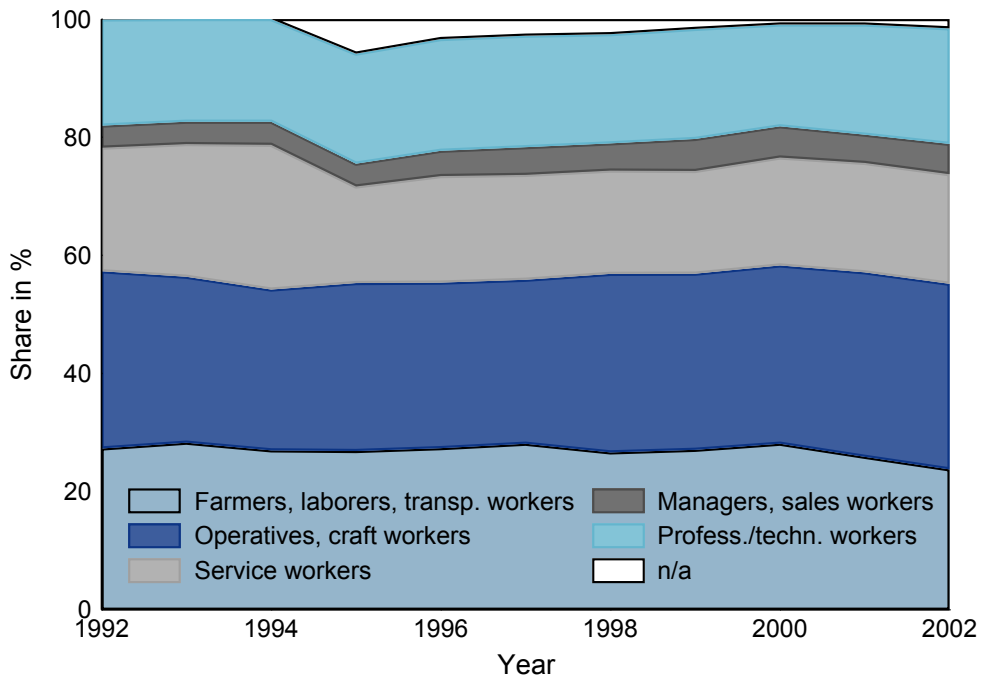
Figure 2A-4: Ethnic German inflows by arrival year



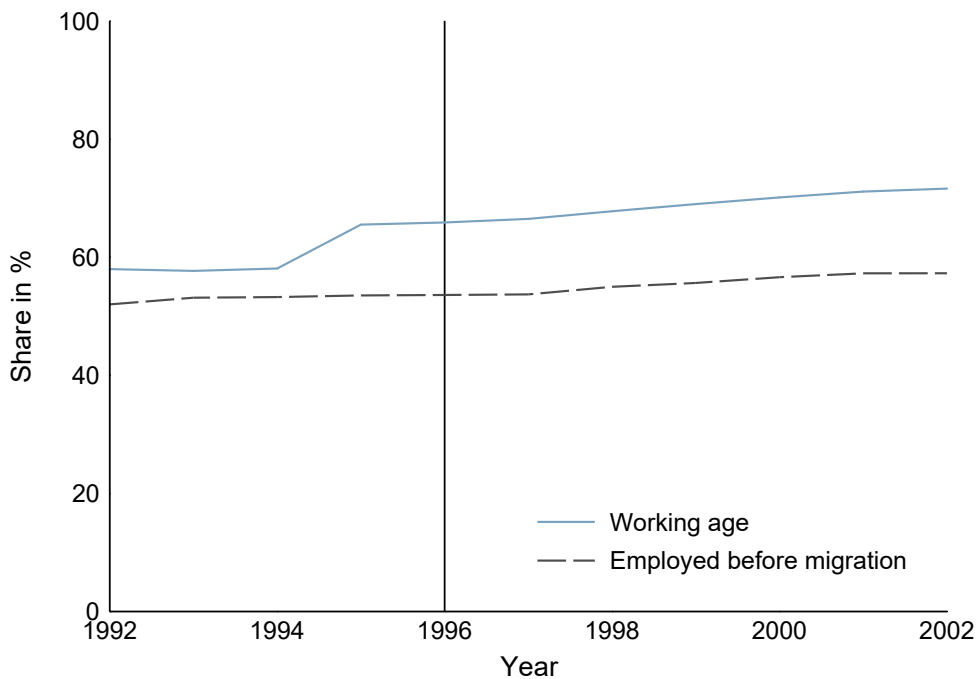
Notes: Annual ethnic German inflows from the former Soviet Union to Germany. More than 95.7 percent of the incoming ethnic Germans came from the former Soviet Union during the analysis period from 1992 to 2006. The following countries are considered to be part of the former Soviet Union: Armenia, Azerbaijan, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russian Federation, Tajikistan, Turkmenistan, Ukraine, Uzbekistan and Belarus. The German authorities introduced a yearly quota of 225,00 ethnic Germans per year in 1993. This quota was further reduced in 1999 to 100,000 arrivals per year. All federal states in West Germany except Bavaria and Rhineland-Palatinate made the allocation policy binding after 1996. Most states adhered to the allocation policy from March 1996 with Lower Saxony following in April 1997 and Hesse in January 2002. Figure based on data from the German Federal Office of Administration (*Bundesverwaltungsamt*).

Figure 2A-5: Occupations and demographics of incoming ethnic Germans

(a) Last occupation in country of origin by arrival year

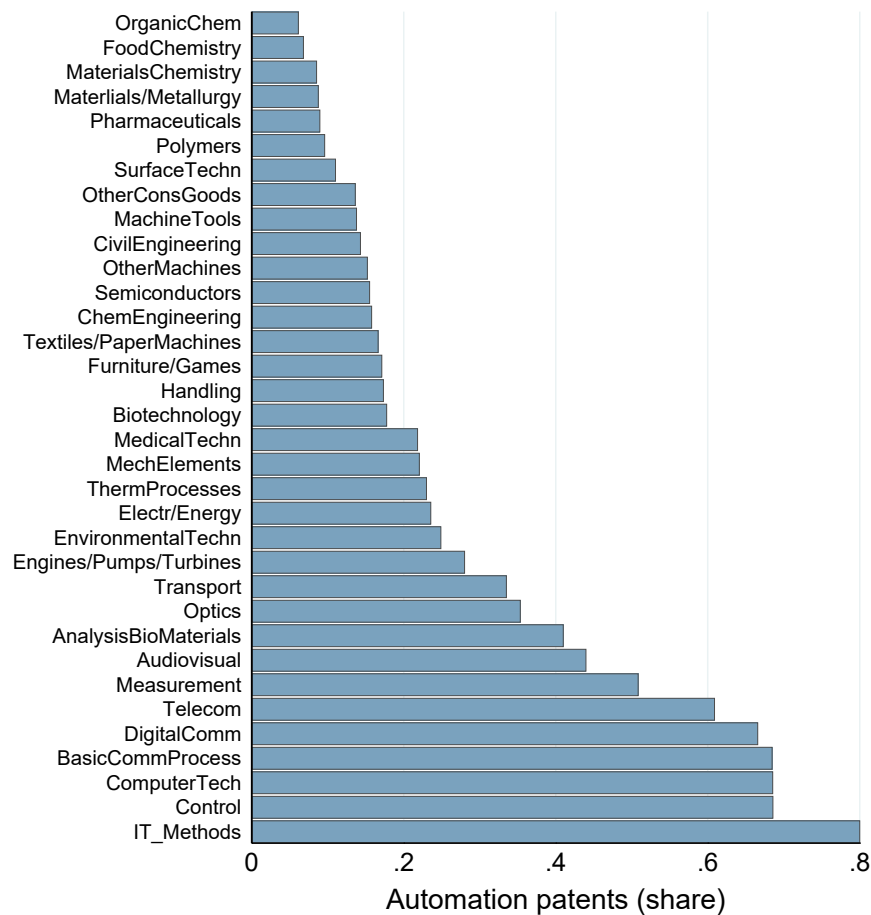


(b) Working before migration and working age by arrival year



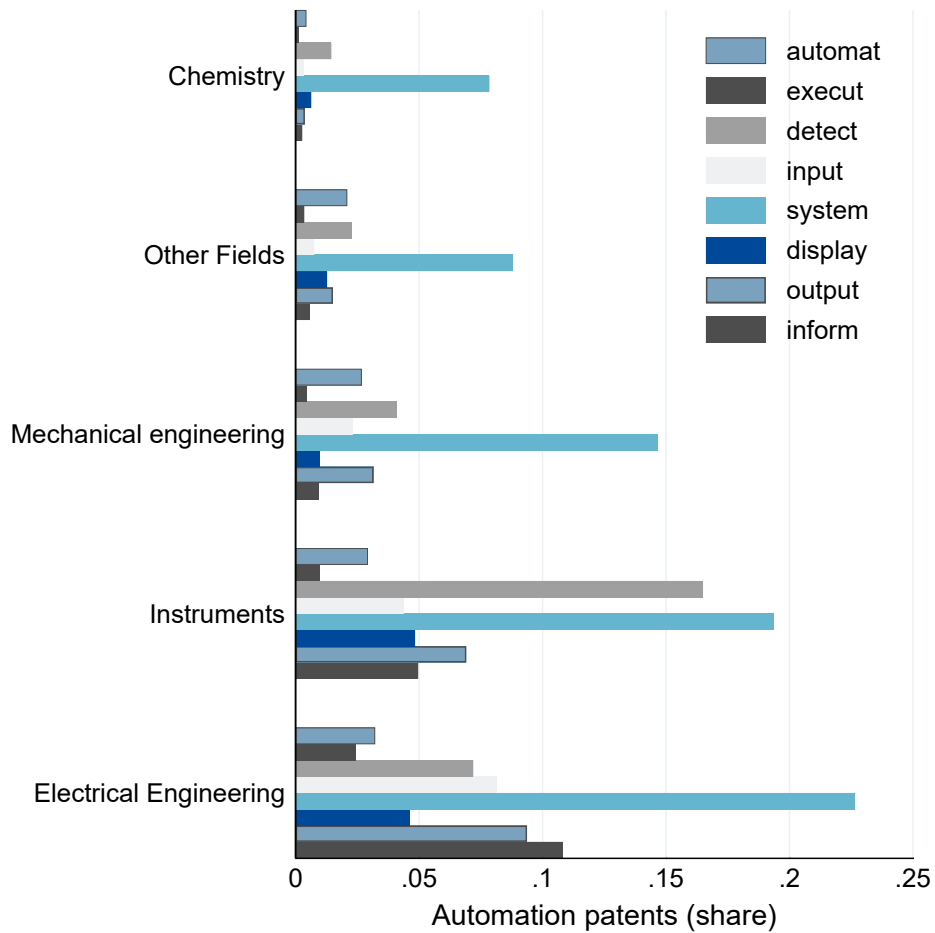
Notes: Top picture: Occupation shares of the last occupation in the country of origin of incoming ethnic Germans by arrival year. Bottom picture: Shares with respect to working before migration and working age (age between 18-64 (1992-1994) and 15-64 (1995-2002)) by arrival year. Own calculations based on data from Glitz (2012). Original data from the *Jahresstatistik fuer Aussiedler*, published annually by *Bundesverwaltungsamt*.

Figure 2A-6: Automation innovation across technology areas



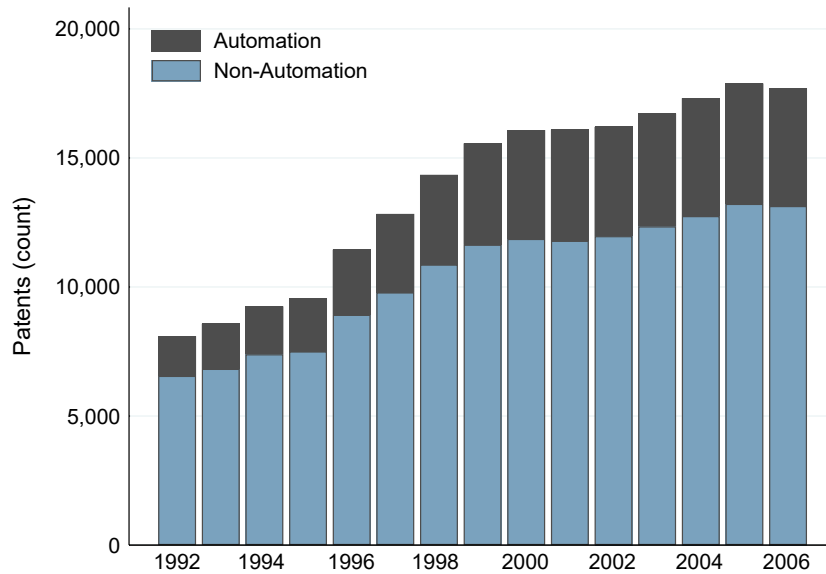
Notes: The share of automation patents across technology areas. Based on all patent applications filed at the European Patent Office by at least one inventor located in one of the allocation states with a priority date between 1992 and 2006. Source: PATSTAT. Own calculations of the shares of automation patents for 34 technology areas. We classify patents into automation patents or non-automation patents by searching keywords related to automation in the pre-processed English abstracts. See text for more details on the classification and the technology areas which are constructed using IPC classes and a concordance table developed by the Fraunhofer ISI and the Observatoire des Sciences et des Technologies in cooperation with the French Patent Office (Schmoch, 2008).

Figure 2A-7: Automation keywords across technology areas



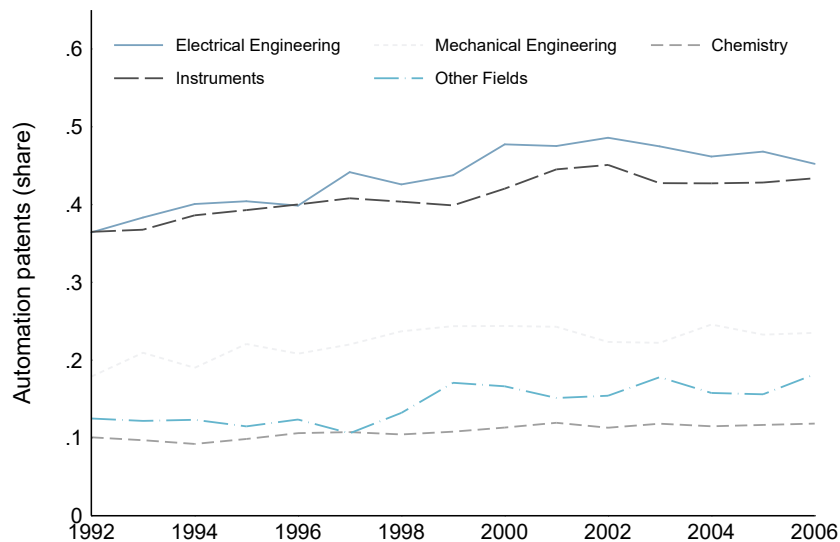
Notes: The share of patents across technology areas by each automation keyword appearing in the patent abstracts. Based on all patent applications filed at the European Patent Office by at least one inventor located in one of the allocation states with a priority date between 1992 and 2006. Source: PATSTAT. Own calculations of the shares for five technology areas. See text for more details on technology areas which are constructed using IPC classes and a concordance table developed by the Fraunhofer ISI and the Observatoire des Sciences et des Technologies in cooperation with the French Patent Office (Schmoch, 2008).

Figure 2A-8: The level of automation patents and non-automation patents



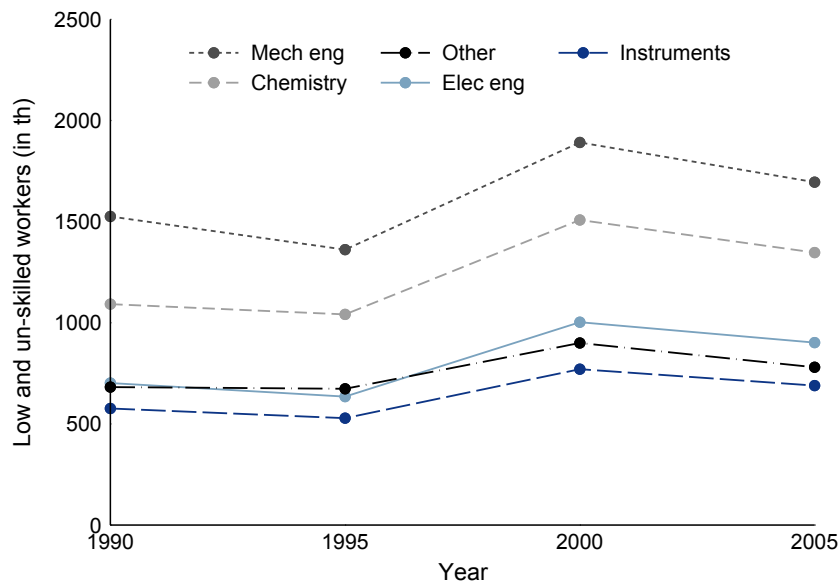
Notes: Figure shows the annual number of patent applications filed at the European Patent Office by at least one inventor located in one of the allocation states with a priority date between 1992 and 2006. We classify patents into automation patents or non-automation patents by searching keywords related to automation in the pre-processed English abstracts. See text for more details on the classification. Source: PATSTAT, own calculations.

Figure 2A-9: The share of automation patents by technology area



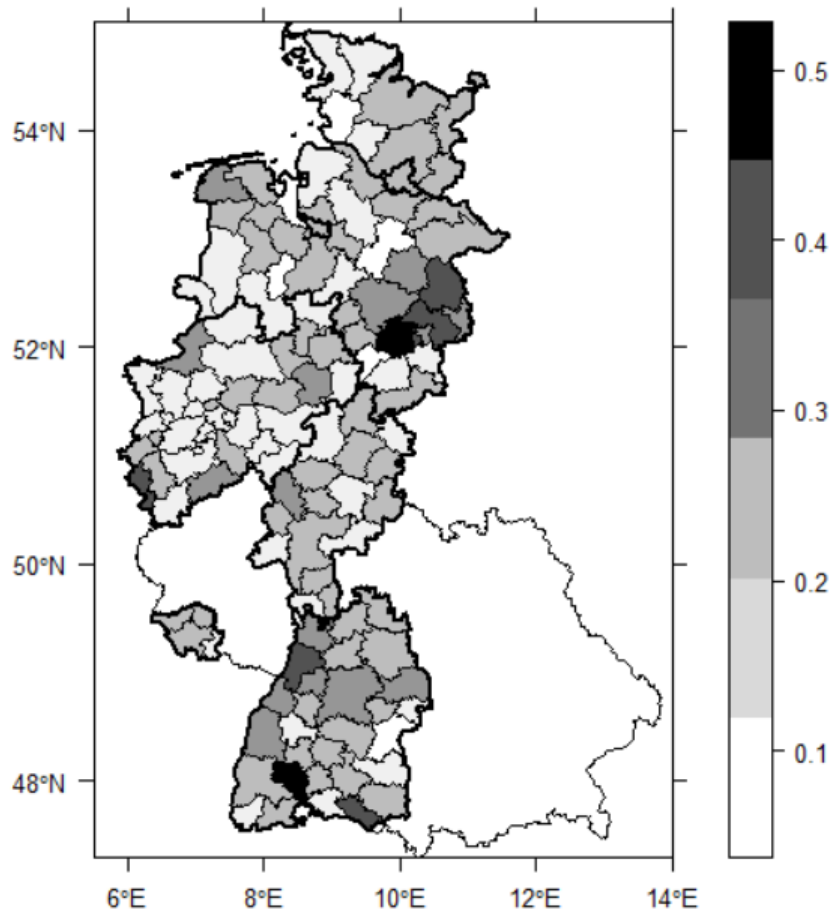
Notes: Figure shows the annual share of patent applications related to automation by technology area. Based on all patent applications filed at the European Patent Office by at least one inventor located in one of the allocation states with a priority date between 1992 and 2006. We classify patents into automation patents or non-automation patents by searching keywords related to automation in the pre-processed English abstracts. See text for more details on the classification. Source: PATSTAT, own calculations.

Figure 2A-10: The number of low-skilled workers by technology area



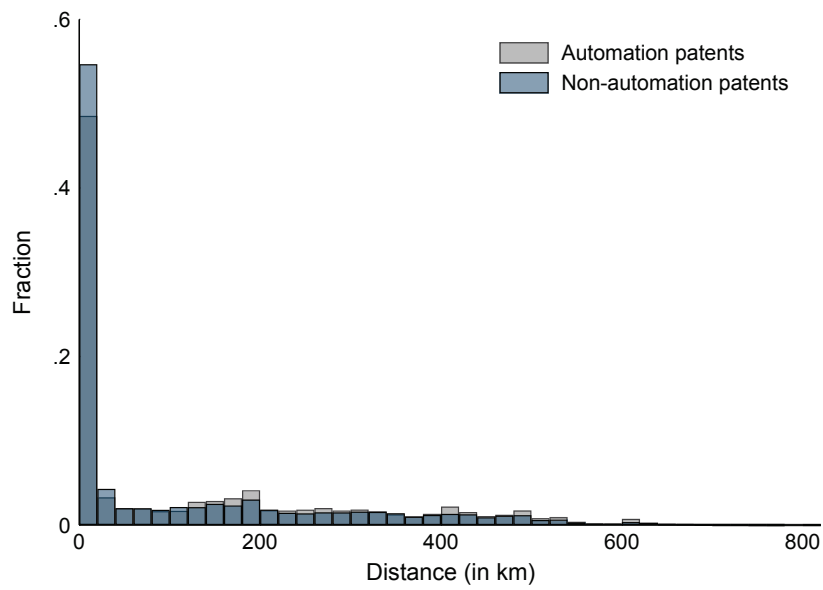
Notes: Figure shows the annual employment of unskilled and low skilled workers in industries related to specific technology areas. Own calculations based on concordance tables between industries and technologies by Dorner and Harhoff (2018) and employment data from the Institute for Employment Research.

Figure 2A-11: Share of automation patents across German regions



Notes: The share of automation patents across labor market regions in West German allocation states. Own calculations of the share of automation patents for 127 labor market regions. Based on all patent applications filed at the European Patent Office by at least one inventor located in one of the allocation states with a priority date between 1992 and 2006. The black lines denote state borders. Figure based on a shapefile of the Federal Republic of Germany from Eurostat and a reference file on counties and labor market regions from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

Figure 2A-12: Distance between focal patent and citing patent



Notes: Histogram of citing distance for automation and non-automation patents in our sample. Bin width corresponds to 20 kilometers. Citing patents with foreign inventor excluded. Own calculations.

Table 2A-9: Analysis sample

Federal State of Germany	Labor Market Regions	Analysis Sample	Pre Binding Allocation	Binding Allocation			
		Region-Year Observations	Region-Year Observations	Region-Year Observations	Start Year	End Year	Implementation Date
Baden-Wuerttemberg	28	420	112	308	1996	2006	01.03.1996
Bremen [*]	0	0	0	0	1997	2006	01.03.1996
Hamburg ^{**}	0	0	0	0	1996	2006	01.03.1996
North Rhine-Westphalia	36	540	144	396	1996	2006	01.03.1996
Schleswig-Holstein	8	120	32	88	1996	2006	01.03.1996
Lower Saxony	34	505	170	335	1997	2006	07.04.1997
Saarland	4	60	16	44	1996	2006	11.03.1996
Hesse	17	204	119	85	2002	2006	01.01.2002
Total	127	1849	593	1256			

Notes: Analysis sample and the implementation of the assigned place of residence act. Regional level: labor market region. We have merged the regions "Osterode" and "Goettingen" into one region to make the data compatible with the regional employment data. We conservatively exclude the region "Ulm" because it is partly located in the state of Bavaria which did not implement the placement policy. The regions "Bremen" and "Bremerhaven" are partly in the state of Bremen and contain counties that are in Lower Saxony^{*}. We conservatively include these regions only from 1997 onward when Lower Saxony implemented the allocation policy. The region "Hamburg" contains the state of Hamburg, three counties in Schleswig-Holstein and one county in Lower Saxony^{**}. Since the counties in Schleswig-Holstein and the city of Hamburg are dominant, we follow Glitz (2012) and include this region from 1996 onward when Hamburg and Schleswig-Holstein adopted the placement policy. The region "Mannheim" contains also one county in Hesse. We conservatively include this region only from 2002 onward when Hesse implemented the allocation policy. For the construction of the state fixed effects and year-by-state fixed effects, we assign the region "Hamburg" to the state of Schleswig-Holstein, the regions "Bremen" and "Bremerhaven" to the state of Lower Saxony and the region "Mannheim" to the state of Hesse. Data on ethnic German inflows are not available for the region "Hannover" (Lower Saxony) for the years 2002-2006 and for regions within the state of Hesse for the years 1992-1994. To merge the inflow data with the patent data, we account for territorial reforms in the regions using historical files of changes between the various NUTS versions from Eurostat. There were only two territorial reforms in the labor market regions of the allocation states: in 2001, "Hannover, Kreisfreie Stadt" and "Hannover, Landkreis" were united into "Region Hannover." Likewise in 2009, "Aachen, Kreisfreie Stadt" and "Aachen, Kreis" were united into "Staedteregion Aachen."

Table 2A-10: Regional characteristics

Variable	Source and description
Ethnic inflows	Glitz (2012) and Piopiunik and Ruhose (2017), original data from 1992 to 2001 from the admission centers in each state and from 2002 to 2006 from Bundesarbeitsgemeinschaft Evangelische Jugendsozialarbeit e.V., Jugendmigrationsdienste.
Population	1992 and 1994-2008: Working Group Regional Accounts VGRdL. 1991 and 1993: Glitz (2012), original data from the German Statistical Office.
Share of immigrants	1995-2008: INKAR online. 1991-1994: Glitz (2012), original data from the German Statistical Office.
GDP, GVA, GVA Production, GVA Tertiary	Working Group Regional Accounts VGRdL. We impute the 1993 values using the 1992 values and the 1993 national growth rate. We impute the 1991 values using the 1992 values and the 1992 national growth rate.
Labor force	Federal Employment Agency. Own calculation of the dependent labor force based on total unemployment and the unemployment rate.
Unemployment rate	Federal Employment Agency. Refers to the dependent labor force.
Skill groups	Establishment History Panel from the Institute for Employment (customized analysis, based on the population of establishments with at least one employee subject to social security in Germany). The shares of high skilled employees (with university degree or applied university degree), medium skilled employees (with school degree and vocational education but no higher degree) and the share of engineers and scientists (employees with a degree from a university or a university of applied sciences and with specific occupation classifications).
Occupation groups	Establishment History Panel from the Institute for Employment (customized analysis, based on the population of establishments with at least one employee subject to social security in Germany). The shares of 12 different occupation groups (agricultural, unskilled manual, unskilled services, unskilled commercial and admin., skilled manual, skilled services, skilled commercial and admin., technicians, semiprofessions, engineers, professions, managers).
Age > 55	Establishment History Panel from the Institute for Employment (customized analysis, based on the population of establishments with at least one employee subject to social security in Germany). The employment share of workers older than 55.

Table 2A-11: The effect of the ethnic German inflows on the level of non-automation innovation - alternative lag structure

	(1)	(2)	(3)	(4)
Inflow rate _t	-0.013*			
	(0.007)			
Allocation _t × Inflow rate _t	0.008			
	(0.009)			
Inflow rate _{t-1}		-0.014**		
		(0.007)		
Allocation _{t-1} × Inflow rate _{t-1}		0.022**		
		(0.009)		
Inflow rate _{t-2}			-0.015**	
			(0.006)	
Allocation _{t-2} × Inflow rate _{t-2}			0.014	
			(0.008)	
Inflow rate _{t-3}				-0.016***
				(0.005)
Allocation _{t-3} × Inflow rate _{t-3}				0.008
				(0.010)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes
Observations	1849	1849	1849	1849
R-squared	0.976	0.976	0.977	0.976
Within R-squared	0.117	0.096	0.089	0.056
Total effect	-0.005	0.008	-0.001	-0.007
P-value	0.638	0.461	0.904	0.423

Notes: OLS regressions. Dependent variable: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_t: ethnic German inflow in year t scaled by the workforce in the previous year. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-12: Effect of ethnic German inflows on automation innovation – by firm size

Dep. Var.:	Automation patents / patents			Automation patents			Non-Automation patents		
	Small firm	Large firm	No firm	Small firm	Large firm	No firm	Small firm	Large firm	No firm
Innovator:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Inflow rate _{t-2}	-0.069 (0.132)	0.210 (0.348)	-0.198 (0.243)	-0.012 (0.007)	0.002 (0.011)	0.007 (0.007)	-0.013** (0.006)	-0.008 (0.010)	0.005 (0.006)
Allocation _{t-2} × Inflow rate _{t-2}	-0.705*** (0.243)	-0.854** (0.372)	0.301 (0.555)	-0.027 (0.016)	-0.042* (0.022)	-0.011 (0.017)	0.015* (0.009)	0.006 (0.016)	-0.019 (0.016)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1845	1814	1771	1849	1849	1849	1849	1849	1849
R-squared	0.399	0.484	0.155	0.933	0.943	0.840	0.966	0.955	0.910
Within R-squared	0.022	0.023	0.009	0.033	0.090	0.015	0.063	0.078	0.024
Total effect	-0.774	-0.643	0.103	-0.039	-0.040	-0.004	0.002	-0.002	-0.014
P-value	0.003	0.133	0.853	0.015	0.065	0.799	0.803	0.890	0.332
Dep var mean	21.100	26.025	21.267	1.904	1.656	0.819	3.100	2.512	1.691

Notes: OLS regressions. Dependent variables constructed using patent applications from small firms (Column 1, 4, 7) or medium-sized firms (Column 2, 5, 8) or large firms (Column 3, 6, 9): Column 1-3: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 4-6: Automation patents: number of automation patents filed in year t (entered in logs). Column 7-9: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-13: Effect of ethnic German inflows on automation innovation – by firm age

Dep. Var.:	Automation patents / patents		Automation patents		Non-Automation patents	
	Young firm	Old firm	Young firm	Old firm	Young firm	Old firm
Innovator:	(1)	(2)	(3)	(4)	(5)	(6)
Inflow rate _{t-2}	-0.207 (0.161)	0.582** (0.248)	-0.009 (0.007)	-0.006 (0.010)	-0.005 (0.006)	-0.025** (0.010)
Allocation _{t-2} × Inflow rate _{t-2}	-0.699** (0.288)	-0.983** (0.377)	-0.036** (0.017)	-0.030 (0.022)	0.004 (0.009)	0.019 (0.015)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1839	1822	1849	1849	1849	1849
R-squared	0.380	0.470	0.933	0.928	0.962	0.951
Within R-squared	0.022	0.043	0.034	0.087	0.057	0.083
Total effect	-0.905	-0.401	-0.046	-0.036	-0.001	-0.006
P-value	0.002	0.347	0.006	0.115	0.886	0.689
Dep var mean	21.744	24.145	1.802	1.604	2.942	2.553

Notes: OLS regressions. Dependent variables constructed using patent applications from young firms (Column 1, 3, 5) or old firms (Column 2, 4, 6): Column 1-2: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 3-4: Automation patents: number of automation patents filed in year t (entered in logs). Column 5-6: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-14: Effect of ethnic German inflows on automation innovation – by firm regionality

Dep. Var.:	Automation patents / patents		Automation patents		Non-Automation patents	
	Local firm	Non-local firm	Local firm	Non-local firm	Local firm	Non-local firm
Innovator:	(1)	(2)	(3)	(4)	(5)	(6)
Inflow rate _{t-2}	-0.096 (0.138)	0.200 (0.320)	-0.015** (0.008)	0.002 (0.011)	-0.013** (0.006)	-0.010 (0.009)
Allocation _{t-2} × Inflow rate _{t-2}	-0.546** (0.258)	-0.960** (0.369)	-0.014 (0.017)	-0.043** (0.020)	0.016* (0.009)	0.003 (0.019)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1843	1809	1849	1849	1849	1849
R-squared	0.383	0.432	0.927	0.936	0.962	0.955
Within R-squared	0.022	0.029	0.041	0.076	0.079	0.069
Total effect	-0.642	-0.760	-0.030	-0.041	0.002	-0.007
P-value	0.013	0.059	0.072	0.044	0.789	0.711
Dep var mean	20.982	26.437	1.827	1.565	3.019	2.385

Notes: OLS regressions. Dependent variables constructed using patent applications from local firms (Column 1, 3, 5) or non-local firms (Column 2, 4, 6): Column 1-2: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 3-4: Automation patents: number of automation patents filed in year t (entered in logs). Column 5-6: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-15: Effect on automation innovation (quality weighted)

Patent value weights:	Granted patents only	Weighted by family size	Weighted by fwd cites
<i>A: Automation patents / patents</i>	(1)	(2)	(3)
Inflow rate _{t-2}	0.076 (0.151)	0.124 (0.151)	0.558* (0.294)
Allocation _{t-2} × Inflow rate _{t-2}	-0.937*** (0.232)	-1.117*** (0.230)	-2.644*** (0.498)
Observations	1846	1847	1823
R-squared	0.515	0.538	0.407
Within R-squared	0.028	0.027	0.034
Total effect	-0.860	-0.994	-2.086
P-value	0.000	0.000	0.000
<i>B: Automation patents</i>	(4)	(5)	(6)
Inflow rate _{t-2}	-0.013* (0.007)	-0.013 (0.009)	0.007 (0.014)
Allocation _{t-2} × Inflow rate _{t-2}	-0.037*** (0.014)	-0.046** (0.019)	-0.115*** (0.027)
Observations	1849	1849	1849
R-squared	0.949	0.927	0.897
Within R-squared	0.067	0.045	0.049
Total effect	-0.050	-0.059	-0.108
P-value	0.000	0.002	0.000
Region fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes

Notes: OLS regressions. Dependent variables: Panel A: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Panel B: Automation patents: number of automation patents filed in year t (entered in logs). Each patent application is weighted with patent grant status (column 1), the number of patents within the same DOCDB family (column 2) or US patent citations within the first 3 years (column 3). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-16: Effect of ethnic German inflows on automation innovation – alternative sets of automation keywords

Dep. Var.:	Automation patents / patents			Automation patents			Non-Automation patents		
	Reduced	Extended	automat	Reduced	Extended	automat	Reduced	Extended	automat
Set of Automation Keywords:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Inflow rate _{t-2}	0.112 (0.121)	0.176 (0.126)	0.068* (0.035)	-0.007 (0.007)	-0.006 (0.006)	0.003 (0.009)	-0.014** (0.006)	-0.015** (0.006)	-0.013** (0.006)
Allocation _{t-2} × Inflow rate _{t-2}	-0.882*** (0.201)	-0.958*** (0.212)	-0.302*** (0.063)	-0.042*** (0.015)	-0.034** (0.014)	-0.085*** (0.015)	0.013 (0.009)	0.015* (0.008)	0.005 (0.010)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1847	1847	1847	1849	1849	1849	1849	1849	1849
R-squared	0.536	0.659	0.342	0.954	0.961	0.835	0.977	0.976	0.978
Within R-squared	0.033	0.036	0.040	0.069	0.072	0.073	0.092	0.088	0.098
Total effect	-0.770	-0.781	-0.234	-0.049	-0.040	-0.082	-0.001	0.000	-0.008
P-value	0.000	0.000	0.000	0.001	0.004	0.000	0.893	0.982	0.376
Dep var mean	19.582	26.710	2.248	2.259	2.531	0.731	3.594	3.500	3.787

Notes: OLS regressions. Dependent variables constructed using an reduced (Column 1, 4, 7) or extended set (Column 2, 5, 8) of automation keywords or the keyword "automat" (Column 3, 6, 9): Column 1-3: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 4-6: Automation patents: number of automation patents filed in year t (entered in logs). Column 7-9: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-17: Effect of ethnic German inflows on automation innovation – poisson regressions

Dep. Var.:	Automation patents	Non-Automation patents
	(1)	(2)
Inflow rate _{t-2}	-0.012 (0.008)	-0.014** (0.007)
Allocation _{t-2} × Inflow rate _{t-2}	-0.046*** (0.016)	0.008 (0.008)
Region and year fixed effects	Yes	Yes
Year-by-State fixed effects	Yes	Yes
Controls	Yes	Yes
Occupation + Skill groups	Yes	Yes
Observations	1849	1849
Log-likelihood	-5147.551	-6758.882

Notes: Poisson pseudo-likelihood regression with multiple levels of fixed effects, as described by Correia *et al.* (2019). Dependent variables: Column 1: Automation patents: number of automation patents filed in year t . Column 2: Non-automation patents: number of non-automation patents filed in year t . Inflow rate_{t-2}: ethnic German inflow in year $t - 2$ scaled by the workforce in the previous year. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-18: Effect of ethnic German inflows on automation innovation – allocation period only

Dep. Var.:	Automation patents / patents	Automation patents	Non-Automation patents
	(1)	(2)	(3)
Inflow rate _{t-2}	-0.698*** (0.246)	-0.044*** (0.016)	-0.009 (0.008)
Region fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes
Observations	1256	1256	1256
R-squared	0.663	0.964	0.982
Within R-squared	0.044	0.083	0.094

Notes: OLS regressions. Throughout all regressions, we only include region-year pairs from the binding allocation period. Dependent variables: Column 1: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 2: Automation patents: number of automation patents filed in year t (entered in logs). Column 3: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-19: Effect of ethnic German inflows on automation innovation – non-binding allocation period only

Dep. Var.:	Automation patents / patents	Automation patents	Non-Automation patents
	(1)	(2)	(3)
Inflow rate _{t-2}	-0.003 (0.232)	-0.005 (0.011)	0.008 (0.009)
Region fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes
Observations	591	593	593
R-squared	0.601	0.968	0.984
Within R-squared	0.037	0.099	0.180

Notes: OLS regressions. Throughout all regressions, we only include region-year pairs from the non-binding allocation period. Dependent variables: Column 1: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 2: Automation patents: number of automation patents filed in year t (entered in logs). Column 3: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-20: Effect of ethnic German inflows on automation innovation – overlapping observations

Dep. Var.:	Automation patents / patents	Automation patents	Non-Automation patents
	(1)	(2)	(3)
OL: Inflow rate _{t-2}	-0.295** (0.117)	-0.018** (0.009)	-0.001 (0.005)
Region and year fixed effects	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes
Observations	1002	1002	1002
R^2 within	0.16	0.27	0.32
p-value WCB	0.024	0.052	0.863

Notes: OLS regressions. Dependent variables: Column 1: Automation patents / patents: cumulated number of automation patents over the three-year period t to $t + 2$ divided by the cumulated number of patents. Column 2: Automation patents: cumulated number of automation patents over the three-year period t to $t + 2$ (entered in logs). Column 3: Non-automation patents: cumulated number of non-automation patents over the three-year period t to $t + 2$ (entered in logs). Inflow rate_{t-2}: cumulative ethnic German inflows over the three-year period $t - 2$ to t scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-21: Effect of ethnic German inflows on automation innovation – exclusion of regions with low or high innovative capacity

Dep. Var.:	Automation patents / patents		Automation patents		Non-Automation patents	
	Low capacity	High capacity	Low capacity	High capacity	Low capacity	High capacity
	(1)	(2)	(3)	(4)	(5)	(6)
Inflow rate _{t-2}	0.084 (0.144)	0.050 (0.143)	-0.012* (0.007)	-0.008 (0.007)	-0.019*** (0.007)	-0.008 (0.006)
Allocation _{t-2} × Inflow rate _{t-2}	-0.855*** (0.212)	-0.808*** (0.241)	-0.035** (0.015)	-0.033* (0.017)	0.014 (0.009)	0.009 (0.009)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1669	1684	1669	1686	1669	1686
R-squared	0.643	0.535	0.957	0.903	0.976	0.944
Within R-squared	0.040	0.034	0.081	0.073	0.100	0.093
Total effect	-0.772	-0.758	-0.047	-0.041	-0.005	0.000
P-value	0.001	0.003	0.003	0.015	0.579	0.980

Notes: OLS regressions. We exclude regions in column 1, 3 and 5 (2, 4 and 6), if the pre-existing regional number of patent applications in 1991 is below (above) the 10 (90) percentile. Dependent variables: Column 1-2: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 3-4: Automation patents: number of automation patents filed in year t (entered in logs). Column 5-6: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-22: Effect of ethnic German inflows on automation innovation – exclusion of Gifhorn regions

Dep. Var.:	Automation patents / patents	Automation patents	Non-Automation patents
	(1)	(2)	(3)
Inflow rate _{t-2}	0.034 (0.161)	-0.010 (0.008)	-0.015** (0.007)
Allocation _{t-2} × Inflow rate _{t-2}	-1.011*** (0.217)	-0.043*** (0.015)	0.014* (0.008)
Region fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes
Observations	1757	1759	1759
R-squared	0.598	0.959	0.978
Within R-squared	0.038	0.064	0.067
Total effect	-0.977	-0.053	-0.001
P-value	0.000	0.001	0.865

Notes: OLS regressions. Throughout all regressions, we exclude regions that signed the Gifhorn declaration. Dependent variables: Column 1: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 2: Automation patents: number of automation patents filed in year t (entered in logs). Column 3: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-23: Effect of ethnic German inflows on automation innovation – winsorized inflow rate

Dep. Var.:	Automation patents / patents	Automation patents	Non-Automation patents
	(1)	(2)	(3)
Inflow rate _{t-2}	0.189 (0.187)	-0.006 (0.011)	-0.017* (0.009)
Allocation _{t-2} × Inflow rate _{t-2}	-0.949*** (0.233)	-0.036** (0.016)	0.016* (0.010)
Region fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes
Observations	1847	1849	1849
R-squared	0.591	0.958	0.977
Within R-squared	0.033	0.070	0.086
Total effect	-0.760	-0.043	-0.001
P-value	0.001	0.004	0.910

Notes: OLS regressions. Ethnic inflow rates are winsorized at the 5th percentile and the 95th percentile. Dependent variables: Column 1-2: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 3-4: Automation patents: number of automation patents filed in year t (entered in logs). Column 5-6: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-24: Effect of ethnic German inflows on automation innovation – exclusion of transition years

Dep. Var.:	Automation patents / patents	Automation patents	Non-Automation patents
	(1)	(2)	(3)
Inflow rate _{t-2}	0.113 (0.133)	-0.008 (0.007)	-0.013** (0.006)
Allocation _{t-2} × Inflow rate _{t-2}	-1.003*** (0.284)	-0.040** (0.019)	0.014 (0.012)
Region fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes
Observations	1737	1739	1739
R-squared	0.583	0.958	0.977
Within R-squared	0.032	0.066	0.086
Total effect	-0.891	-0.048	0.000
P-value	0.001	0.010	0.991

Notes: OLS regressions. Throughout all regressions, we exclude region-year pairs from transition years. Dependent variables: Column 1: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Column 2: Automation patents: number of automation patents filed in year t (entered in logs). Column 3: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Inflow rate_{t-2}: ethnic German inflows in $t - 2$ scaled by the workforce in $t - 3$. Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-25: Heterogeneous effects on automation innovation by labor market tightness (50 pctl)

Labor Market Tightness:	Tight		Slack	
	$U_0 < 50$ pctl	$U_0 \geq 50$ pctl	$U_{t-3} < 50$ pctl	$U_{t-3} \geq 50$ pctl
A: <i>Automation Patents / Patents</i>	(1)	(2)	(3)	(4)
Inflow rate _{t-2}	-0.228 (0.205)	0.176 (0.162)	-0.248 (0.228)	0.232 (0.149)
Allocation _{t-2} × Inflow rate _{t-2}	-0.482* (0.269)	-0.701** (0.302)	-0.615** (0.254)	-0.723** (0.312)
Observations	885	947	891	926
R-squared	0.708	0.548	0.724	0.546
Within R-squared	0.075	0.045	0.075	0.046
Total effect	-0.709	-0.525	-0.862	-0.491
P-value	0.035	0.085	0.003	0.108
B: <i>Automation patents</i>	(5)	(6)	(7)	(8)
Inflow rate _{t-2}	-0.019* (0.011)	-0.006 (0.009)	-0.018 (0.012)	-0.005 (0.009)
Allocation _{t-2} × Inflow rate _{t-2}	-0.039* (0.019)	-0.023 (0.020)	-0.045** (0.017)	-0.023 (0.020)
Observations	885	949	891	928
R-squared	0.972	0.948	0.974	0.944
Within R-squared	0.115	0.101	0.094	0.103
Total effect	-0.058	-0.028	-0.062	-0.028
P-value	0.008	0.161	0.001	0.177
Region and year fixed effects	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Occupation + skill groups	Yes	Yes	Yes	Yes

Notes: OLS regressions. Dependent variables: Panel A: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Panel B: Automation patents: number of automation patents filed in year t (entered in logs). Sample splits by the 50-percentile of the unemployment rate in the year before the binding placement (column 1-2) and by the 50-percentile of the unemployment rate in $t-3$ (column 3-4). Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-26: Heterogeneous effects on automation innovation by labor market tightness (90 pctl)

Labor Market Tightness:	Tight	Slack	Tight	Slack
	$U_0 < 90$ pctl	$U_0 \geq 90$ pctl	$U_{t-3} < 90$ pctl	$U_{t-3} \geq 90$ pctl
<i>A: Automation Patents / Patents</i>	(1)	(2)	(3)	(4)
Inflow rate _{t-2}	0.080 (0.143)	-0.085 (0.288)	0.106 (0.139)	0.085 (1.197)
Allocation _{t-2} × Inflow rate _{t-2}	-0.945*** (0.212)	-0.339 (1.811)	-0.990*** (0.226)	1.426 (2.347)
Observations	1653	194	1651	176
R-squared	0.650	0.326	0.643	0.466
Within R-squared	0.044	0.094	0.045	0.148
Total effect	-0.865	-0.424	-0.884	1.511
P-value	0.000	0.811	0.000	0.390
<i>B: Automation patents</i>	(5)	(6)	(7)	(8)
Inflow rate _{t-2}	-0.005 (0.006)	-0.013 (0.011)	-0.003 (0.006)	-0.103* (0.051)
Allocation _{t-2} × Inflow rate _{t-2}	-0.040*** (0.015)	-0.003 (0.074)	-0.045*** (0.015)	0.137* (0.069)
Observations	1654	195	1652	177
R-squared	0.960	0.871	0.962	0.943
Within R-squared	0.067	0.311	0.085	0.335
Total effect	-0.045	-0.017	-0.047	0.034
P-value	0.003	0.819	0.002	0.483
Region and year fixed effects	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Occupation + skill groups	Yes	Yes	Yes	Yes

Notes: OLS regressions. Dependent variables: Panel A: Automation patents / patents: number of automation patents filed in year t divided by the number of patents. Panel B: Automation patents: number of automation patents filed in year t (entered in logs). Sample splits by the 90-percentile of the unemployment rate in the year before the binding placement (column 1-2) and by the 90-percentile of the unemployment rate in $t-3$ (column 3-4). Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-27: Heterogeneous effects on non-automation innovation by labor market tightness

Labor Market Tightness:	Tight	Slack	Tight	Slack
	$U_0 < 75$ pctl	$U_0 \geq 75$ pctl	$U_{t-3} < 75$ pctl	$U_{t-3} \geq 75$ pctl
	(1)	(2)	(3)	(4)
Inflow rate $_{t-2}$	-0.011 (0.007)	-0.015 (0.016)	-0.011* (0.006)	-0.012 (0.017)
Allocation $_{t-2} \times$ Inflow rate $_{t-2}$	0.012 (0.009)	0.003 (0.032)	0.006 (0.010)	0.025 (0.027)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Year-by-State fixed effects	Yes	Yes	Yes	Yes
Occupation + Skill groups	Yes	Yes	Yes	Yes
Observations	1374	460	1371	459
R-squared	0.983	0.955	0.983	0.964
Within R-squared	0.083	0.222	0.115	0.165
Total effect	0.001	-0.011	-0.005	0.013
P-value	0.874	0.690	0.572	0.484

Notes: OLS regressions. Dependent variables: Non-automation patents: number of non-automation patents filed in year t (entered in logs). Sample splits by the 75-percentile of the unemployment rate in the year before the binding placement (column 1-2) and by the 75-percentile of the unemployment rate in $t - 3$ (column 3-4). Controls: log population, log labor force, unemployment rate, share of immigrants, GDP per capita, log GVA total, log GVA production, log GVA tertiary and Share age > 55. Occupation + skill groups: employment shares of 12 occupation groups and 3 skill groups. Regressions estimated at the region-year level, weighted by regional population in 1991. *Data sources:* See Table 2A-10 in the Appendix. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2A-28: Patent-level summary statistics

Variable	Mean	Std. Dev.	Min	Max	N
<i>Main</i>					
Automation patents	0.25	0.43	0.00	1.00	264061
Automation patents (Extended)	0.29	0.45	0.00	1.00	264061
Automation patents (Reduced)	0.21	0.41	0.00	1.00	264061
Automation patents (automat)	0.02	0.14	0.00	1.00	264061
<i>Technology Areas</i>					
Electrical	0.16	0.37	0.00	1.00	263858
Instruments	0.12	0.33	0.00	1.00	263858
Chemistry	0.28	0.45	0.00	1.00	263858
Mechanical	0.36	0.48	0.00	1.00	263858
Other	0.08	0.27	0.00	1.00	263858
Automation (Electrical)	0.07	0.26	0.00	1.00	263858
Automation (Instruments)	0.05	0.22	0.00	1.00	263858
Automation (Chemistry)	0.03	0.17	0.00	1.00	263858
Automation (Mechanical)	0.09	0.28	0.00	1.00	263858
Automation (Other)	0.01	0.11	0.00	1.00	263858
<i>Quality Weights</i>					
Granted	0.62	0.49	0.00	1.00	263915
Automation (Granted)	0.15	0.36	0.00	1.00	263915
DOCDB Family Size (weighted)	6.29	5.80	1.00	160.00	263915
Automation: DOCDB Family Size (weighted)	1.38	3.30	0.00	160.00	263915
DOCDB citations (weighted)	2.37	8.56	0.00	1388.00	263915
Automation: DOCDB citations (weighted)	0.72	5.83	0.00	1388.00	263915
<i>Processes</i>					
Process	0.46	0.50	0.00	1.00	231855
Non-Process	0.54	0.50	0.00	1.00	231855
Automation (Process)	0.13	0.34	0.00	1.00	231855
Automation (Non-Process)	0.12	0.32	0.00	1.00	231855
Automation (Process+Mechanical)	0.03	0.18	0.00	1.00	231855
Non-Automation (Process+Mechanical)	0.08	0.28	0.00	1.00	231855
Automation (Non-Process+Mechanical)	0.05	0.22	0.00	1.00	231855
Non-Automation (Non-Process+Mechanical)	0.20	0.40	0.00	1.00	231855

Notes: Summary statistics of patent applications with a priority year between 1990 and 2010 filed by inventors located in the allocation states. *Data source:* PATSTAT. The measures of automation innovation by one of five main technology areas are based on mapped IPC classes and the concordance table developed by the Fraunhofer ISI and the Observatoire des Sciences et des Technologies in cooperation with the French patent office (Schmoch, 2008). See text for more details regarding the classification of patents into automation and non-automation patents.

Table 2A-29: Examples of automation patents

EP application number	EP19940115956
Title	Method for controlling dryers in brick factories
Assignee	INNOVATHERM Prof Dr Leisenberg GmbH
Abstract	According to a method for controlling dryers in the ceramic industry, which are operated with an external and/or internal heating and are operated in conjunction with a tunnel furnace, and in which, by means of an empirical or mathematical model of the drying process, the state of drying-out of the chambers is approximately determined, the variation of the air condition values and/or of the convection capacity of the individual chambers is automatically selected or calculated as a function of the products, the available drying time and the available furnace waste heat, from the point of view of a minimum total energy or energy cost expenditure, and is started as a drying programme. This mode of operation achieves an automatic optimisation of the heat consumption in a targeted manner, using the prescribed boundary conditions, and, as a result, the costs of energy and personnel are reduced in a concern without large investment in terms of installation being necessary.
EP application number	EP20010969514
Title	Method and device for analysing chemical or biological samples
Assignee	BASF LYNX BIOSCIENCE AG
Abstract	The invention relates to a method and related device for analysing chemical or biological samples. Chemical or biological samples and/or targets (probes) are applied to an outer cylindrical lateral area of a carrier in the form of individual defined spots, or are loaded into bore holes in the form of liquid drops, said bore holes being recessed in the lateral area of the carrier. The carrier is introduced into a recess in the holder, said recess being essentially complementary to the cylindrical lateral area, the samples and/or targets are influenced by means of physical and/or chemical interactions, and the accordingly modified spots are then analysed. The invention also relates to the use of a novel carrier system for examining chemical or biological samples, which contrary to conventional planar biochips is characterised by a cylindrical geometry, whereby substances can be applied, immobilised for example, on the functionalised lateral area of the cylinder or in the radial bore holes recessed in the cylinder casing. An analysis system having clearly defined reaction volumes is implemented by co-operating with a complementary holder, said analysis system being easily standardised and highly automated .
EP application number	EP19990942756
Title	Method for automatically controlling and selecting the bodies of slaughtered poultry
Assignee	CSB SYST SOFTWARE ENTWICKLUNG, Csb-System Software-Entwicklung & Unternehmensberatung AG
Abstract	The invention describes a method for automatically controlling and selecting the bodies of slaughtered poultry. The invention aims at providing a very easy and cost-effective method for automatically controlling and selecting the bodies of slaughtered poultry. According to the invention, said aim is achieved in that the body of the slaughtered poultry to be controlled is selected by conducting color analysis of the light reflected from the visible surface parts, said light being detected as a diffuse color mixing light eliminating spatial contours using a measuring technique, wherein the measured value is used for selection as integrating value for the totality of visible surface parts.

Notes: Source: PATSTAT.

Chapter 3

Labor Recruitment and (Non-)Automation Innovation

3.1 Introduction

Prior research has focused on the innovation impact of high-skilled labor recruitment (see e.g., Kerr and Lincoln, 2010; Doran *et al.*, 2014; Dimmock *et al.*, 2019). Yet there is little research on the relationship between low-skilled labor recruitment and innovation. This is an important omission given that economic theory has been suggesting for a long time that labor abundance may, first, encourage *labor-complementary* innovation (see e.g., Kremer, 1993; Acemoglu, 2010) and, second, discourage *labor-saving* innovation (see e.g., Hicks, 1932; Habakkuk, 1962). There are two likely reasons for the lack of research: first, there has been a lack of systematic data on labor-complementary and labor-saving innovation as the bibliographic information of patents does not contain these economic characteristics. Second, data limitations on recruited workers and worker self-selection into labor markets make it difficult to estimate the association between labor recruitment and innovation.

This paper provides the first evidence for the relationship between low-skilled labor recruitment and (non-)automation innovation at the regional level.⁴⁹ More precisely, it investigates 1) the *complementarity* of labor recruitment and *non-automation* innovation and 2) the *substitutability* of labor recruitment and *automation* innovation. On the one hand, labor recruitment might allow firms to grow, to extend the scope of production and to enter new markets: the increased flexibility of production might induce firms to undertake exploratory R&D investments leading to novel products or processes. On the other hand, labor recruitment might reduce the need for innovation activities aimed at automating the production.

This paper utilizes the massive demand-driven allocation of labor migrants during the German Guest Worker Program. Lasting from 1955 to 1973, this program provides a unique

⁴⁹Since innovative activity tends to cluster in the same location as production (see e.g., Paci and Usai, 2000), changes in regional (non-)automation innovation might be an important adjustment mechanism to regional labor recruitment.

opportunity to investigate the relationship between labor recruitment and innovation. Aiming at the reduction of labor shortages in the booming post-war Germany, the recruitment program was among the largest ones ever undertaken: 2.6 million foreign workers were employed in 1973 when the program was terminated. By that time, the nationwide share of migrants in the labor force stood at 11.9 % (Federal Employment Agency, 1974). Within the context of bilateral guest worker treaties, the pro-active labor recruitment by German firms induced regional inflows of migrants to fill the numerous open positions in the economy. Most importantly for this study, the demand-driven recruitment process arguably circumvents the immigrant self-selection into regions as the initial job location of incoming guest workers depended on firms' labor recruitment and regional preferences of guest workers were not considered. The guest workers constituted a rather homogeneous group of young and low-skilled labor migrants, often performing manual labor in German firms. They primarily originated from Italy, Greece, Spain, Turkey and Yugoslavia. Finally, the long program duration allows to study the dynamics of the relationship between labor recruitment and innovation. To the best of my knowledge, this paper presents the first analysis on the German Guest Worker Program at a regional level.

For this purpose, this paper introduces a new source of data on the universe of foreign workers at the regional level. I collect rich data over the period 1964-1973 from archival documents published by the Federal Employment Agency. These administrative data are of longitudinal nature and of high-quality, which I confirm in several consistency checks. These detailed data on the foreign workforce by country-of-origin allow me to distinguish between workers from countries with and without guest worker treaties. The absence of territorial reforms allows to systematically track the regional stock of guest workers over the most relevant years of the German Guest Worker Program. Using these novel data, I exploit the substantial regional variation in the increase in the stock of guest workers across regions and years to examine the relationship between labor recruitment and innovation.

A further distinguishing feature of my work is the novel regional data on the level of non-automation patents and automation patents. I construct these innovation outcomes using historic patent data from the German Patent and Trademark Office.⁵⁰ To define the set of automation patents and non-automation patents, I employ a keyword classification using the patent full texts.⁵¹ To assign non-automation and automation patents to regions, I extract the inventor addresses from the inventor field and from the patent full text. Further, patent filings provide useful information on the exact timing of inventive activities. This is the first spatial panel data on innovation during the period of the German Guest Worker Program.

Finally, this paper utilizes regional information from the 1961 full population census on the size of the manufacturing sector. Using modern methods of geographical science, I harmonize the county-level data from the Census to reflect the level of the regions from my analysis.

⁵⁰The title of the historic patent database is *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986*.

⁵¹In this regard, this paper relates to recent research on the classification of patents (see e.g., Webb, 2019).

For the empirical analysis, I relate regional (non-)automation innovation to the lagged stock of guest workers using poisson regressions. An advantage of this spatial approach is that it captures the association between labor recruitment and innovation at the regional level, while taking skill-complementarities across regional skill groups into account (Dustmann *et al.*, 2016b). Utilizing variation in the stock of guest workers across region-year pairs from 1964 to 1973, the longitudinal nature of my data allows to employ a rich set of fixed effects to control for unobserved factors: region fixed effects control for systematic time-invariant differences across regions such as the general innovative capacity or time-constant sectoral specialisation. Time fixed effects account for year-specific shocks that affect all regions in the same way such as general trends in innovation. To proxy for general economic conditions, I include total regional employment.

Overall, the results of this paper point towards labor recruitment having different associations with non-automation innovation and automation innovation. First, this paper finds a significant and positive relationship between the regional stock of guest workers and non-automation innovation. Models with different lag structures suggest that the association materializes in the second and third year after the labor recruitment. The estimated elasticity of the three-years lagged workforce of guest workers with respect to non-automation innovation is 0.17. Results are heterogeneous with regard to technological areas. More specifically, the positive association between labor recruitment and non-automation innovation tends to be concentrated in textiles, paper, metallurgy, transportation and performing operations. Overall, these findings are consistent with labor recruitment being *complementary* to non-automation innovation. The positive association is confined to the guest workers. The rather small stock of foreign workers from non-guest-worker-countries is not associated with non-automation innovation. Note that the relationship between labor recruitment and the overall level of innovation is also positive and significant.

Second, my results indicate an insignificant association between labor recruitment and automation innovation. This average association masks considerable heterogeneity across regions: while there are no significant associations in regions with small labor markets, there is a significant degree of substitution between labor recruitment and automation innovation in regions with large labor markets. This heterogeneity might be explained by larger labor markets capturing a greater part of the total innovation impact of labor recruitment. Prior research (see e.g., Lewis, 2011; Dustmann and Glitz, 2015; Monras, 2019) shows that the adoption of production technologies is a key mechanism to absorb regional labor supply shocks. Given that buyer-supplier relationships tend to be regionally clustered (Bernard *et al.*, 2019), larger regions might capture stronger demand effects for products related to automation innovation within regional borders. Within this context, a recent study by Danzer *et al.* (2020) provides suggestive evidence that external demand partly drives the effects of labor supply shocks on regional automation innovation.

Additional tests show that the main results are robust to the exclusion of regions with low or high levels of patenting activities. Also, the findings are robust to using an alternative

empirical model with overlapping observations.

This exploratory paper cannot attribute causality to the estimated associations because the demand-driven labor recruitment might be endogenous to (non-)automation innovation. In this context, there has been little prior research on the endogeneity between regional labor recruitment and (non-)automation innovation. While my analysis on the lag structures suggests that reverse causality is an unlikely explanation for the found patterns, it limits endogeneity concerns only to a certain extent. Time-variant unobserved factors that influence both innovation and labor recruitment would bias my estimates: for example, it could be that regions that are booming and have increasing levels of innovation might recruit more guest workers. Related to the latter, a sub-sample analysis shows that the association between labor recruitment and innovation tends to be less positive after taking into account regional unemployment rates. Finally, there could be also an alternative channel if guest workers in a region increase the demand for new products. Notwithstanding the limitations of the descriptive analysis, this paper might be considered as a first step towards estimating the impact of labor recruitment on (non-)automation innovation.

This paper relates to several strands of research. First, there is a nascent literature investigating the relationship between labor supply and labor-saving innovation. Exploiting the introduction of immigration barriers in the US, San (2019) shows that negative shocks to agricultural labor supply increase agricultural innovation. Investigating immigration restrictions in the US, Doran and Yoon (2019) find that labor supply shocks influence the direction of innovation. Exploiting the placement of ethnic Germans in the 1990s and 2000s, Danzer *et al.* (2020) find negative regional effects of exogenous labor supply shocks on automation innovation. While all these studies exploit variation in labor inputs resulting from the *supply side* of the labor market, this paper expands on existing work by analysing variation largely resulting from the *labor recruitment* by firms. Furthermore, my paper complements this research by providing evidence that the regional degree of substitution between labor recruitment and automation innovation depends on the size of the labor market.

Furthermore, this paper relates to research on factor supplies and technology (e.g., Acemoglu, 2002, 2007; Hanlon, 2015; Kiley, 1999).⁵² It expands on existing work by examining the relationship between labor endowments induced by labor recruitment and novel measures of (non-)automation innovation.

This paper also adds to the literature on immigration and innovation (e.g., Hornung, 2014; Hunt and Gauthier-Loiselle, 2010). While most studies focus on the relationship between high-skilled innovation and the level of innovation, the combination of an informative low-skilled labor recruitment program, the high-quality data on guest workers and the novel measures of (non-)automation innovation introduced by this paper advances the existing literature.

Furthermore, there is related research on labor migration and labor market outcomes

⁵²Within this context, there is a related recent study by Dechezleprêtre *et al.* (2019) on firm exposure to cross-country wages and automation innovation.

(see e.g., Mithas and Lucas Jr, 2010; Peri *et al.*, 2015; Bratu, 2018). While prior studies suggest that adjustments in wages, employment and production technology are common, my descriptive findings indicate that adjustments in (non-)automation innovation might be also relevant.

Finally, this paper contributes to the literature on immigration and ethnic concentration within the context of the German Guest Worker Program. Exploiting the demand-driven placement procedure from an *individual* perspective, prior research examines the effects of ethnic concentration on immigrants' outcomes, such as their host cost-country language skills, contacts to natives, educational achievement and ethnic identity (see e.g., Danzer and Yaman, 2016; Danzer *et al.*, 2018; Danzer and Yaman, 2013; Constant *et al.*, 2013). While the analysis of these studies is based on data from *after* the recruitment stop, the analysis of this paper refers to the *period during* the German Guest Worker Program. Most importantly, this paper expands on existing work by exploiting the same placement procedure to isolate the regional labor recruitment by firms and by constructing a novel dataset on the guest workers. As the research community was not aware of the underlying archival documents from the Federal Employment Agency, my dataset might stimulate future research on this unique setting of the German Guest Worker Program.

My research question is highly relevant for policy-makers because labor immigration can be an important tool for countries to reduce labor shortages. Understanding the relationship between labor recruitment and (non-)automation innovation might help for the optimal design of labor immigration policies.

The remainder of the paper proceeds as follows: Section 3.2 outlines the institutional background of the German Guest Worker Program and the demand-driven allocation of guest workers to regions. Section 3.3 describes the spatial longitudinal dataset on the universe of foreign workers and the measures of (non-)automation innovation. It also provides some summary statistics. Section 3.4 lays out the econometric framework. Section 3.5 presents the results. Section 3.6 presents robustness checks. Section 3.7 concludes.

3.2 The German Guest Worker Program

Lasting from 1955 to 1973, the German Guest Worker Program offers a rich yet unexploited opportunity to investigate the relationship between labor recruitment and innovation. In the late 1950s and the 1960s, there were strong labor shortages in post-war Germany. Aggregate unemployment was very low leading to severe difficulties for many firms to hire native workers. As a consequence, West Germany signed bilateral guest worker treaties with Italy in 1955, Greece and Spain in 1960, Turkey in 1961, South Korea and Morocco in 1963, Portugal in 1964, Tunisia in 1965, and Yugoslavia in 1968. Expanding labor recruitment by German firms triggered massive inflows of migrants from the sending countries to fill the many open positions in the economy.

Figure 3.1 depicts the annual stock of foreign workers in Germany between 1955 and 1973.

The number of foreign workers increased substantially in most years when the program was in place.⁵³ During the 10-year analysis period from 1964 to 1973, the stock of foreign workers had risen from around 900,000 to 2.35 million. By the end of the recruitment program in 1973, the share of foreign employed workers stood at 11.6 % (Federal Employment Agency, 1974).

To isolate the regional labor recruitment by German firms, I exploit a useful particularity of this immigration wave. Firms pro-actively recruited foreign workers from the guest worker countries by sending recruitment requests to local branches of the Federal Employment Agency. These recruitment requests were then forwarded to branches in the sending countries. Potential applicants were screened in the country of origin and sent to Germany in mass transports (Federal Employment Agency, 1962). Due to this demand-driven placement procedure, the initial region of work of the guest workers was predominantly determined by German firms' current labor recruitment activities. Most importantly for this study, this demand-driven placement circumvents immigrant self-selection into regional labor markets.⁵⁴ Consistent with this placement policy, earlier studies provide support that the initial job location was exogenous from the perspective of individual guest workers (Danzer and Yaman, 2013, 2016; Constant *et al.*, 2013; Danzer *et al.*, 2018). Further, the regional mobility of guest workers was restricted as they had to stay with the initial employer for at least two years and in the same occupation for five years (Dahmen and Kozłowicz, 1963). Also, any incentives to move to other regions were low because guest workers were employed immediately upon arrival. As a result, the spatial distribution of guest workers over time was largely determined by firms' initial labor recruitment.

In the absence of this placement policy, I could not isolate the labor recruitment by firms because immigrants might self-select into labor markets potentially taking into account ethnic networks, living costs, employment opportunities, wages or climate (see e.g., Edin *et al.*, 2003; Albert and Monras, 2017; Hunt, 1992).

The majority of the incoming guest workers was low educated, male, young and did not have sufficient German language skills. Given these characteristics, the guest workers were prone for employment in labor-intensive manual jobs. Guest workers were recruited in the 1960s mainly into booming sectors such as manufacturing or vehicle construction. A significant part of the activities in these industries were manual tasks and could be carried out by unskilled workers because production processes were highly mechanized and semi-automated (Scholten, 1968). They were often employed in routine tasks such as assembly line work (Striso, 1968). Figure 3.2 shows the foreign workforce by nine occupations classes in 1972: while there were only few foreign workers employed in agriculture, the vast majority

⁵³As an exception, the recruitment of guest workers briefly decreased during the 1966-1967 economic downturn.

⁵⁴There was also an alternative demand-driven recruitment channel: guest workers already in Germany could recommend relatives or friends in the country of origin as potential workers to their employer. Interested employers could then request these recommended workers. This type of recruitment was negligible in the early years. In practice, a large share of requested workers couldn't be hired due to various reasons (Federal Employment Agency, 1972a).

of migrants were employed in occupations related to manufacturing, construction and the iron and metal industry.

3.3 Data

The empirical analysis utilizes a spatial panel dataset spanning the 10-year period from 1964 to 1973. It is based on a combination of data from several sources: Section 3.3.1 presents the guest worker data. Section 3.3.2 describes the construction of the regional measures of (non-)automation innovation. Section 3.3.3 describes the data on the regional manufacturing sector from the 1961 full population census. Section 3.3.4 introduces the data on regional unemployment. Section 3.3.5 provides summary statistics of the final dataset.

3.3.1 Guest Worker Data

The spatial panel data on the universe of foreign workers in Germany is based on administrative data from the Federal Employment Agency. The Federal Employment Agency published relevant spatial information in their annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer*. Of particular importance for this study is that the underlying publications contain detailed information on regional employment of foreign workers by country of origin: the dataset covers information on the stock of foreign workers from almost all sending countries. These sending countries are Italy, Greece, Spain, Turkey, Yugoslavia, Portugal, Tunisia, and Morocco.⁵⁵ The detailed information allows to distinguish between *guest workers* and *other foreign workers*.⁵⁶ To approximate the size of the workforce associated with the German Guest Worker Program, I calculate the regional number of guest workers by summing up all foreign workers from countries with active bilateral recruitment agreements. The data also contains regional information on total regional employment allowing to control for economic conditions.⁵⁷ The guest worker data refers to the 10-year period 1964-1973 covering the most relevant years of the German Guest Worker Program.⁵⁸ The stock of foreign workers is always measured every September.⁵⁹

The annual information provided is at the regional level of the so-called *labor office*

⁵⁵As an exception, while regional data on workers from South Korea were not recorded, the nationwide number of South Korean workers was negligible: the annual nationwide number of workers from South Korea was below 8000 throughout all the years of the German Guest Worker Program. Note that information on the regional number of foreign workers is sometimes missing if the nation-wide number of workers from that country-of-origin was low. For example, while regional data on workers from Morocco is missing for the early year of 1965, the nationwide number of workers from Morocco was negligible in this year. In such cases, I replace these missing values with zeros to calculate the regional number of guest workers.

⁵⁶Note that my data does not allow a further regional disaggregation by occupation.

⁵⁷Regarding missing values on total regional employment in 1964 and 1966, I impute the 1964 (1966) values using the 1963 (1965) values and the 1964 (1966) national growth rate in employment.

⁵⁸My analysis does not cover the beginning of the 1960s due to numerous territorial reforms during these years.

⁵⁹As an exception, the data from the year 1973 refers to January.

districts. In the remainder of the paper, I will refer to the labor office districts as regions. These regions tend to be larger than counties and consist of around one or several counties. On average, the total workforce of one region corresponds to approximately 152.000 workers in 1970. Throughout most years under consideration, there were 141 labor office districts in West Germany. A beneficial feature of the regional level is that there were only two territorial reforms during the analysis period allowing to systematically track the stock of guest workers over time.⁶⁰ I conservatively exclude the two region-year cells which were affected by territorial reforms using information from Federal Employment Agency (1964).⁶¹ Figure 3.3 maps the 141 labor office districts in West Germany.

Several quality checks confirm the high quality of my data. I compare the aggregated regional data on foreign workers with data at the federal district level, finding a very high degree of consistency. With few exceptions, the numbers match exactly.

3.3.2 (Non-)Automation Innovation

This paper measures innovation using patents. The underlying region-year panel dataset on (non-)automation innovation is based on patents filed at the German Patent and Trademark Office. It includes patents by inventors located in the West German regions with a filing date between 1964 and 1975. It is based on bibliographic data and complete invention descriptions from the historic patent database *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986*. For the construction of my dataset, I focus on the second publication of the patent applications.⁶²

Following recent research (see e.g., Webb, 2019), I employ a keyword classification to split the patents of the sample into two mutually exclusive groups: non-automation patents and automation patents. I define the set of automation patents as those that use the substring “*automat*” in their complete invention description. The word stem “*automat*” has been shown to be highly indicative of automation innovation in English invention descriptions (Mann and Püttmann, 2018).⁶³ Furthermore, it is a key word stem used for the classification of patents in recent studies on automation innovation (Danzner *et al.*, 2020; Dechezleprêtre *et al.*, 2019). Since “*automat*” is a largely unambiguous keyword, the resulting classifier can be considered as rather conservative. Although simple, this classification performs well in manual checks. Figure 3.7 shows the annual number of (non-)automation patents filed by at least one inventor

⁶⁰For comparison, there were substantial, ongoing territorial county reforms in the 1960s and 1970s.

⁶¹More precisely, I exclude the regions Bremen and Bremerhafen in the year 1964.

⁶²The second publication corresponds to granted patents if the patent application was published before 1968. Since there was a reform in the German patent system in 1968, the second publication corresponds to the so-called *Auslegeschriften* for the vast majority of patent applications published after 1968. Note that granted patents and the *Auslegeschriften* are similar: *Auslegeschriften* have also been examined by the patent office before publication. If no opposition was filed within the opposition period of three months after the publication of the *Auslegeschrift*, the underlying patent application was also granted. I identify all relevant patent documents using the publication number and the so-called *Schriftart*.

⁶³While the invention descriptions used in this study are in German, the substring “*automat*” captures also German words related to automation.

located in West Germany between 1964 and 1975.⁶⁴ On average, the share of automation patents is about 12 percent.

For a deeper analysis, I assign each patent to one of the eight main technology areas of the International Patent Classification (IPC).⁶⁵ Since the degree of substitution or complementarity between labor recruitment and innovation might vary across technology areas, these more detailed measures allow to shed light on heterogeneous associations. Figure 3.4 shows the share of automation patents across the eight technology areas of the IPC. While the share of automation patents is, as expected, high in technology areas related to performing operations or physics, it is low in technology areas such as chemistry.

To capture the time of inventive activity, I use the filing year of the patent. I approximate regional innovative activity with the inventor location (e.g., OECD, 2009). For this purpose, I extract the inventor location from the inventor field or from the patent full text.⁶⁶ For the majority of the patents in my sample, information on inventors' city of residence is available. In a first step, I exclude patent-inventor pairs if an inventor is located in a foreign country. Foreign inventors are identified by searching for foreign country names, country codes and major foreign cities in the inventor information. After extracting and pre-processing the inventors' city of residence, I assign cities to labor office districts using a list of 700 West German cities with their corresponding labor office districts. The list of West German cities is taken from Federal Employment Agency (1971). In manual checks, I find that the quality of the assignment of patents to labor office districts is high.⁶⁷

Finally, I aggregate the patent data to the region-year level by counting the number of non-automation patents and automation patents filed in a given region and year. If a patent has multiple inventors,⁶⁸ I assign equally weighted fractions to the inventor's region of residence. Figure 3.5 (Figure 3.6) shows the regional distribution of the level of non-automation patents (automation patents).

⁶⁴Note that Figure 3.7 is based on all patents of my final sample. Only patents which I could assign to at least one region in West Germany are included. The reduction in the number of filings between 1966 and 1968 coincides with the 1966-1967 recession in West Germany.

⁶⁵These technology areas are: A Human Necessities; B Performing Operations, Transporting; C Chemistry, Metallurgy; D Textiles, Paper; E Fixed Constructions; F Mechanical Engineering, Lighting, Heating, Weapons; G Physics and H Electricity. The information on the IPC class is available for around 72 percent of the patents in my sample.

⁶⁶For granted patents published before 1968, I extract the available inventor locations from the patent full text. Usually, the inventors' city of residence appears after the inventors' name in the patent text.

⁶⁷I could assign about 40 percent of the patents with potentially domestic inventors to labor office districts. Note that patents which were not assigned to a labor office district tend to originate from foreign inventors with missing city information. It could also be that the inventors are located in Berlin which is not included in my analysis sample or in rather small cities which were not contained in the list of West German cities. Finally, it could also be that the inventor information for some domestic inventors is missing, both in the patent full text and in the inventor field.

⁶⁸If an inventor field falsely contains more than one ZIP Code and more than one inventor, I adjust the patent data accordingly.

3.3.3 Regional Data on the Size of the Manufacturing Sector

For further analysing the regionality of these associations, I add regional data on the size of the manufacturing sector and the total labor market size from the 1961 full population census. The corresponding data are taken from Schmitt *et al.* (1994). Since the data are provided at the territorial status of the 328 German counties in 1987, I harmonize these county-level data to reflect the geography of the labor office districts from my analysis. For this purpose, I harmonized the projections of 1) a digitized map of labor office districts from Federal Employment Agency (1972b) and 2) a map of the counties in 1987 from Max Planck Institute for Demographic Research and CCG (2011) and Bundesamt für Kartographie und Geodäsie (2011). Then, I calculate the geographical overlap between the two regional units. When counties are split and assigned to neighbouring labor office districts, I assign the corresponding values fractionally based on area size.

Figure 3.8 plots the geographical overlap between labor office districts and counties. The regional boundaries of both units are often very similar. A typical labor office district contains around one or more counties. Finally, a comparison of total employment from the census data with total employment from the Federal Employment Agency shows a high degree of consistency supporting my approach of harmonizing these data.⁶⁹

3.3.4 Unemployment

I add regional unemployment rates to the panel data using information from the annual series *Arbeitsamtsstatistik* published by the Federal Employment Agency. The unemployment rates are available for the years 1967-1972.⁷⁰

3.3.5 Summary Statistics

Table 3.1 shows the summary statistics for the main variables of interest at the region-year level. The main sample includes 1,408 region-year pairs for the period 1964 to 1973. The mean number of non-automation patents corresponds to 22.49. On average, there are 3.07 automation patents in a given year and region. Note that the corresponding standard deviations are roughly twice as large as the means suggesting significant heterogeneity in the regional level of patenting activities. The average share of automation patents corresponds to about 12.2 percent (not shown). While there were on average 8,736 guest workers in a region-year pair, the corresponding number of other foreign workers is about 2,317. In the average region, total employment is about 148,000. Consistent with the severe labor shortages during the German Guest Worker Program, the average unemployment rate corresponds to only about 0.9 percent during the period 1967-1972. The average number of regional workers in manufacturing is about 68,000 in 1961 (not shown).

⁶⁹The correlation coefficient between the two measures of total employment in 1970 is about 0.97.

⁷⁰Note that regional unemployment rates are not available before 1967.

3.4 Empirical Model

In this section, I introduce the main empirical model to investigate the association between labor recruitment and (non-)automation innovation. The estimation equation relates the regional number of patents to the lagged stock of guest workers. This spatial approach captures the association between labor recruitment and innovation at the regional level taking complementarities across regional skill groups into account (Dustmann *et al.*, 2016b). To account for the skewed distribution in the regional level of patents and to alleviate the issue of zeros in the innovation outcomes, I employ poisson quasi-maximum likelihood regressions as described by Correia *et al.* (2019). Furthermore, I drop separated observations because they might lead to incorrect estimates.⁷¹ Exploiting the longitudinal nature of my data, I estimate the following regressions with a rich set of fixed effects:

$$Y_{rt} = \exp \left(\beta_0 + \beta_1 G_{rt-2} + \beta_2 O_{rt-3} + X'_{rt-3} \vartheta + \delta_r + \tau_t \right) + \epsilon_{rt}, \quad (3.4.1)$$

where Y_{rt} denotes either the level of non-automation patents or the level of automation patents in region r and year t . I compare patents between regions with differential regional stocks of guest workers. The coefficient of interest is β_1 , which corresponds to the association between the two-years lagged regional stock of guest workers G_{rt-2} and the outcome Y_{rt} . While the main specification assumes a two-years lag between labor recruitment and innovative activities, Section 3.5.2 explores different lag structures. Region fixed effects δ_r absorb time invariant factors that are specific to regions, such as area size, the general innovative capacity or time-invariant sectoral specialization. Any nationwide time-varying technological trends that affect all regions in the same way is controlled by the inclusion of year fixed effects τ_t . The main association between labor recruitment and innovation can be estimated because of the variation in the stock of guest workers over time within regions. I also include the lagged stock of other foreign workers O_{rt-3} . Finally, I include lagged total regional employment X'_{rt-3} to proxy for labor market conditions and general economic conditions. The control variables refer to the year prior to stock of guest workers.⁷² In Section 3.6.1, I will test the robustness of the results by including regional unemployment rates. All standard errors are clustered at the region level.

My estimates of the coefficients cannot claim causality, but are of descriptive nature. The estimation may be affected by an endogeneity problem if regions choose to recruit guest workers with particular trends in (non-)automation innovation. The estimate of β_1 might be biased if, for instance, some regions have reached the technological boundary in automation innovation and recruit more guest workers. On the other hand, regions that recruit more workers might be booming regions with increasing levels of innovation. While the fixed

⁷¹As a result, the number of observations may slightly vary across different regression specifications.

⁷²The inclusion of control variables from the same year of the stock of guest workers might partly capture the effect of labor recruitment.

effects account for many unobserved factors, I unfortunately cannot include a large set of time-variant variables due to the lack of employment data prior to 1975. This would be the first year for which more extensive labor market data is available from the Institute for Employment Research.

3.5 The Association between Labor Recruitment and (Non-)Automation Innovation

This section presents the results. Section 3.5.1 presents the main results on the association between labor recruitment and (non-)automation innovation. Section 3.5.2 explores the dynamics of the associations. Section 3.5.3 sheds light on the association heterogeneity by pre-existing regional labor market size. Section 3.5.4 investigates heterogeneous associations across technology areas.

3.5.1 Main Results

Table 3.2 displays the main results on the regional associations between labor recruitment and (non-)automation innovation. The results from different specifications (column 1 - 3) indicate that an increase in the stock of guest workers is followed by a significant increase in the level of *non-automation innovation*: an increase in the stock of guest workers by 1 percent is associated with an increase in the level of non-automation innovation by 0.17 percent (column 3). The magnitudes of the coefficients do not substantially change when the time-variant controls on other foreign workers (column 2) and total employment (column 3) are included. This pattern is consistent with non-automation innovation being complementary to labor recruitment. An explanation for this could be that labor recruitment might allow to extend the scope of production as well as to enter new markets. The increased flexibility of the production might induce firms to undertake exploration R&D investments to develop new products or processes.

I repeat the analysis with the level of *automation innovation* as the dependent variable. In contrast to the significant positive association between the stock of guest workers and non-automation innovation, I find an insignificant negative association between the stock of guest workers and automation innovation (column 4-6).

Given the low level of automation innovation in the 1960s, the association between labor recruitment and the overall level of innovation is still positive and significant (not reported). Interestingly, there is no significant association between the stock of other foreign workers and (non-)automation innovation.

3.5.2 Dynamics of the Associations

Since there is some arbitrariness in choosing a specific lag structure between the labor recruitment and the response in innovative activity, I explore the associations across different lag structures $L \in \{1, 2, 3, 4\}$ using analogous poisson quasi-maximum likelihood regressions:

$$Y_{rt} = \exp \left(\beta_0 + \beta_1 G_{rt-L} + \beta_2 O_{rt-L-1} + X'_{rt-L-1} \vartheta + \delta_r + \tau_t \right) + \epsilon_{rt}, \quad (3.5.1)$$

Table 3.3 shows that the positive associations between labor recruitment and non-automation innovation are concentrated in the second and third year after the increase in the stock of guest workers. This reflects that increases in labor recruitment might need time to be reflected in the output measures of innovative activities. Figure 3.9 visualizes these associations across different lag structures.

Next, I run the empirical analysis with the level of automation innovation as the outcome variable (Table 3.4): while the association becomes negative in the second year after the recruitment, all associations are not indistinguishable from zero. Figure 3.10 provides a visual representation of the associations across lag structures.

There appear to be no measurable associations between labor recruitment and innovation in the same year (column 1 of Table 3.3 and Table 3.4), respectively. The causal effect of labor recruitment on current innovation should also be negligible because innovative activities are unlikely to respond in the same year. Hence, the insignificant association between labor recruitment and current innovation lends some credibility to my descriptive approach.

3.5.3 Associations by Regional Labor Market Size

This section sheds light on the extent to which the associations between labor recruitment and automation innovation are heterogeneous across regions. In this context, it is important to note that my spatial approach captures *regional* associations. If the response in innovation does not take place within the same region, my approach will not capture it. In other words, my results do not capture spillovers across regions. My estimates thus represent lower bound estimates for the overall effect.

I hypothesize that the degree of regional substitution between labor recruitment and automation innovation might positively depend on the regional size of the labor market. Large regional labor markets may capture a larger part of the overall effect of labor recruitment on automation innovation. Prior research shows that the adoption of production technologies is a key mechanism how regional labor supply shocks are absorbed (see e.g., Lewis, 2011; Dustmann and Glitz, 2015; Monras, 2019). Given that buyer-supplier relationships tend to be regionally clustered (Bernard *et al.*, 2019), larger regions might capture stronger demand effects for products related to automation innovation within regional borders. Within this

context, a recent study by Danzer *et al.* (2020) provides suggestive evidence that external demand partly drives the effects of labor supply shocks on regional automation innovation.

To investigate heterogeneity by regional labor market size, I split the sample by the 50-percentile of the pre-existing regional size of the manufacturing labor market in 1961.⁷³ The results in Table 3.5 suggest a significant negative association between the stock of guest workers and automation innovation in regions with a large manufacturing labor market (column 2). At the same time, there is an insignificant positive association between labor recruitment and automation innovation in regions with a small manufacturing labor market (column 1). This heterogeneity is robust to using total regional employment as an alternative measure of labor market size (Table 3.5, column 3 and column 4).

The significant substitutability between labor recruitment and automation innovation in regions with large labor markets is consistent with economic theory on labor abundance and labor-saving innovation (see e.g., Hicks, 1932; Habakkuk, 1962). These findings are also consistent with anecdotal evidence. A common concern was that the employment of guest workers has reduced firms' rationalisation measures as well as investments into technological progress because companies might have been more likely to do expansion investments (Striso, 1968). Finally, the significant substitutability of regional labor recruitment and automation innovation and is consistent with recent research on the regional effects of labor supply on automation innovation (Danzer *et al.*, 2020).

In contrast, I do not find such regional heterogeneities regarding labor recruitment and non-automation innovation (Table 3.6). The size of the regional labor market does not seem to play a significant role for non-automation innovation.

3.5.4 Results by Technology Area

In this section, I explore heterogeneous associations between labor recruitment and non-automation innovation through dividing the regional non-automation patents into the eight technology areas of the International Patent Classification. The same regression is run on the different groups separately. Figure 3.11 presents the corresponding results. The associations between labor recruitment and innovation vary with the technology area: while there are no measurable associations related to human necessities, I find some evidence of positive associations between labor recruitment and non-automation innovation related to performing operations, transporting, fixed constructions, textiles and paper. The lack of statistical significance might be explained by the relatively low number of non-automation patents per technology area.

This paper does not investigate heterogeneous associations between labor recruitment and automation innovation across technology areas because there are, on average, only about three automation patents in each region. Running separate regressions for each technology areas would introduce substantial measurement error because there would be many region-

⁷³A substantial number of guest workers were employed in the manufacturing sector.

year pairs with zero automation patents.

3.6 Robustness Checks

This section investigates the robustness of the results by controlling for pre-existing unemployment rates (Section 3.6.1), by excluding regions with low and high levels of patenting activities (Section 3.6.2) and by estimating an alternative empirical model (Section 3.6.3).

3.6.1 Accounting for Pre-Existing Unemployment Rates

In this section, I explore the sensitivity of my estimates to the inclusion of pre-existing unemployment rates. Note that my descriptive results might potentially suffer from endogeneity. For example, if a low regional unemployment rate leads to a stronger labor recruitment of guest workers and if a low unemployment rate is associated with higher levels of regional innovation, the estimates of β_1 are upwards biased.

The analysis in this section is restricted to the period beginning in 1968 as this is the first year where regional information on the lagged unemployment rate is available. Within the context of the demand-driven labor recruitment, the variation in the pre-existing labor market tightness may be a key determinant of guest worker labor recruitment. Hence, including regional unemployment rates eliminates both a substantial number of observations (i.e. statistical power) and variation in the stock of guest workers resulting from differences in unemployment.

Table 3.7 reports results with pre-existing unemployment rates as control variables. For comparison, the table also reports specifications using only region-year pairs where the lagged unemployment rate is available. In this sub-sample, the positive association between the two-years lagged stock of guest workers and non-automation is insignificant (column 1). While the association between the two-years lagged stock of guest workers and non-automation innovation becomes much smaller after controlling for the unemployment rate (column 2), the coefficient of the three-years lagged stock of guest workers is still sizeable and significant (column 4). At the same time, the negative coefficient regarding automation innovation increases in magnitude but remains statistically insignificant after controlling for the regional unemployment rate (column 6 and 8). Overall, the findings based on this sub-sample point towards the associations between labor recruitment and non-automation innovation (automation innovation) being less positive (more negative) when taking into account pre-existing labor market conditions.

3.6.2 Excluding Specific Regions

This section provides a robustness analysis by selectively excluding regions to account for the large regional differences in the level of pre-existing patenting activities. The distribution of region-specific patenting is highly skewed. For example, while there is only one patent in the

region “Saarlouis” in 1963, there are more than 600 patents in the region “Muenchen”. Table 3.8 illustrates that the results are robust to excluding regions with very low or high levels of pre-existing innovative activity. More precisely, I exclude all regions in column 1 and 3 (2 and 4), if the regional number of patents in 1963 is below (above) the 10 (90) percentile. The association between the stock of guest workers and the level of innovation does not substantially change after excluding these regions. Furthermore, the positive association between labor recruitment and non-automation innovation is still significant.

3.6.3 Overlapping Observations Model

To test for sensitivity with respect to the empirical model, I employ an *overlapping observation model*. It has two key advantages. First, it allows for exploiting the data more efficiently and reduces classical measurement errors by using cumulative numbers of regional patents over several years (see e.g., Harri and Brorsen, 2009). Second, it eases the challenge of selecting the appropriate lag structure.⁷⁴ I estimate the following model with the same set of fixed effects:

$$\ln\left(\sum_{z=0}^2 Y_{rt+z}\right) = \beta_0 + \beta_1 \sum_{z=-2}^0 G_{rt+z} + O_{rt-3} + X'_{rt-3} \vartheta + \delta_r + \tau_t + \eta_{rt} + \epsilon_{rt}, \quad (3.6.1)$$

where $\sum_{z=0}^2 Y_{rt+z}$ is the level of (non-)automation patents filed over the period t to $t+2$ in region r . The key explanatory variable $\sum_{z=-2}^0 G_{rt+z}$ corresponds to the cumulative stock of guest workers in r over the period $t-2$ to t . I include the three-years lagged stock of foreign workers and total employment as control variables. Conventional standard errors are clustered at the region level. Since the overlapping data structure introduces additional correlation and the estimation uses a limited number of clusters, I account for within-group dependence in estimating standard errors by reporting p-values calculated with the wild cluster bootstrap-t method by Cameron *et al.* (2008).

Results (see Table 3.9) suggest that an increase in the stock of guest workers by 1 percent is associated with an increase in the level of non-automation patents by 0.21 percent (column 1). This association is significant according to both conventional p-values and those based on the wild cluster bootstrap-t method. Again, I find a negative though statistically insignificant association between labor recruitment and automation innovation (column 2). To sum up, this robustness check confirms the main findings.

3.7 Conclusion

The German Guest Worker Program was one of the largest labor recruitment programs to reduce labor shortages. This paper represents the first analysis on the German Guest Worker

⁷⁴Glitz and Meyersson (2020) employ a similar overlapping observations model.

program at the region level. For this purpose, this paper introduces novel spatial panel datasets on guest workers and new measures of regional automation innovation. Exploiting the demand-driven allocation of guest workers to regions, this paper analyses the degree of substitutability (complementarity) of labor recruitment and (non-)automation innovation.

My findings point into the direction that regional labor recruitment has different associations with non-automation innovation and automation innovation. First, this paper finds a positive association between labor recruitment and non-automation innovation, respectively. This pattern is consistent with economic theory suggesting that labor abundance is complementary to certain types of innovation (see e.g., Acemoglu, 2010). Second, while there is an insignificant average association between labor recruitment and automation innovation, there is a significant degree of substitution between labor recruitment and automation innovation in regions with large labor markets. The latter finding is consistent with economic theory suggesting substitutability between labor abundance and labor-saving innovation (see e.g., Hicks, 1932).

Note that this exploratory paper cannot attribute causality to the estimated associations because it does not exploit exogenous variation in low-skilled labor recruitment. Still, the results can be regarded as important first steps towards estimating the impact of labor recruitment on (non-)automation innovation.

Overall, these descriptive findings imply that regional labor recruitment may change existing patterns of comparative advantage across regions. Understanding how labor recruitment and innovation interact helps policy makers to design targeted labor immigration policies. Further research should explore the causality nexus between labor recruitment and (non-)automation innovation.

Table 3.1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max	N
<i>Innovation</i>					
Non-automation patents	22.49	45.14	0.00	530.51	1408
Automation patents	3.07	6.93	0.00	93.27	1408
Patents	25.56	51.70	0.00	616.53	1408
<i>Foreign Employment</i>					
Guest workers	8736.01	12 986.75	21.00	118 292.00	1408
Log Guest workers	8.36	1.27	3.09	11.68	1408
Other foreign workers	2317.33	3613.31	123.00	39 774.00	1408
Log Other foreign workers	7.11	1.08	4.82	10.59	1408
<i>Employment by country-of-origin</i>					
Italians	2542.78	3896.92	10.00	33 091.00	1408
Greeks	1438.88	2481.07	2.00	23 735.00	1408
Spaniards	1136.78	1673.03	0.00	14 754.00	1408
Turks	1839.17	2710.68	1.00	25 537.00	1408
Portuguese	240.87	466.06	0.00	6010.00	1408
Yugoslavs	1519.30	3837.44	0.00	46 056.00	1408
Moroccans	10.70	110.46	0.00	3408.00	1408
Tunisians	7.52	44.54	0.00	768.00	1408
<i>Labor Market</i>					
Total employment	148 061.50	112 897.70	30 360.53	870 072.25	1408
Log Total employment	11.72	0.58	10.32	13.68	1408
% Unemployed	0.88	0.74	0.10	5.20	846
Year	1968.51	2.87	1964.00	1973.00	1408

Notes: Summary statistics of the region-year panel dataset computed for the period 1964 to 1973. *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Regional unemployment rates are taken from from the annual series *Arbeitsamtsstatistik* published by the Federal Employment Agency.

Table 3.2: Association between labor recruitment and (non-)automation innovation

Dep. Var.:	Non-automation patents			Automation patents		
	(1)	(2)	(3)	(4)	(5)	(6)
Log Guest workers _{t-2}	0.201** (0.084)	0.172** (0.084)	0.174** (0.082)	-0.146 (0.145)	-0.139 (0.149)	-0.136 (0.149)
Log Other foreign workers _{t-3}		0.065 (0.063)	0.059 (0.064)		-0.016 (0.088)	-0.019 (0.091)
Log Total employment _{t-3}			0.105 (0.181)			0.076 (0.231)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1406	1406	1406	1358	1358	1358
Log-likelihood	-3697.021	-3695.371	-3694.469	-2028.661	-2028.647	-2028.585

Notes: Poisson pseudo-likelihood regressions. Dependent variable: Column 1-3: Non-automation patents: number of non-automation patents filed in year t . Column 4-6: Automation patents: number of automation patents filed in year t . *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.3: Association between labor recruitment and non-automation innovation - alternative lag structure

	(1)	(2)	(3)	(4)
Log Guest workers _t	0.046 (0.089)			
Log Guest workers _{t-1}		0.127 (0.087)		
Log Guest workers _{t-2}			0.174** (0.082)	
Log Guest workers _{t-3}				0.234*** (0.090)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Other foreign workers	Yes	Yes	Yes	Yes
Total employment	Yes	Yes	Yes	Yes
Observations	1406	1406	1406	1265
Log-likelihood	-3888.313	-3758.417	-3694.469	-3338.125

Notes: Poisson pseudo-likelihood regressions. Dependent variable: Non-automation patents: number of non-automation patents filed in year t . *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.4: Association between labor recruitment and automation innovation - alternative lag structure

	(1)	(2)	(3)	(4)
Log Guest workers _t	0.049 (0.145)			
Log Guest workers _{t-1}		-0.146 (0.145)		
Log Guest workers _{t-2}			-0.136 (0.149)	
Log Guest workers _{t-3}				-0.005 (0.153)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Other foreign workers	Yes	Yes	Yes	Yes
Total employment	Yes	Yes	Yes	Yes
Observations	1358	1358	1358	1222
Log-likelihood	-2070.103	-2029.707	-2028.585	-1840.675

Notes: Poisson pseudo-likelihood regressions. Dependent variable: Automation patents: number of automation patents filed in year t . *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: Association between labor recruitment and automation innovation by regional labor market size

Regions with:	Manufacturing Workforce		Total Workforce	
	Small	Large	Small	Large
	(1)	(2)	(3)	(4)
Log Guest workers _{t-2}	0.092 (0.196)	-0.372* (0.198)	0.174 (0.235)	-0.354* (0.195)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Other foreign workers	Yes	Yes	Yes	Yes
Total employment	Yes	Yes	Yes	Yes
Observations	650	708	650	708
Log-likelihood	-689.186	-1329.474	-706.759	-1313.482

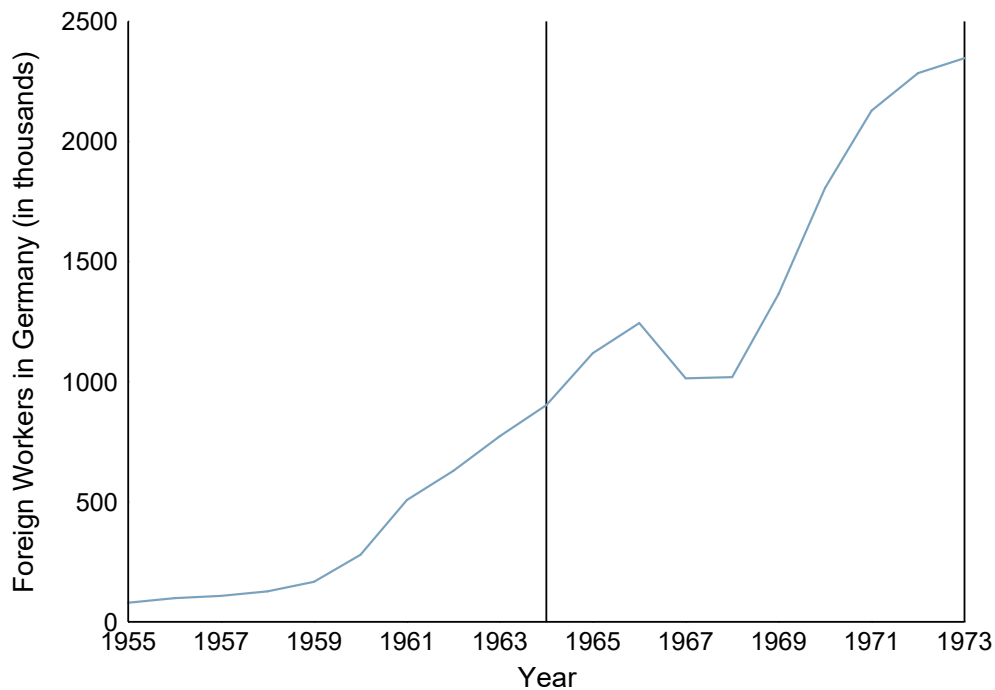
Notes: Poisson pseudo-likelihood regressions. Dependent variable: Automation patents: number of automation patents filed in year t . Sample splits by the 50-percentile of the manufacturing workforce in the year 1961 (column 1-2) and by the 50-percentile of the total workforce (column 3-4). *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Data on the manufacturing workforce and the total workforce of the year 1961 from Schmitt *et al.* (1994). Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.6: Association between labor recruitment and non-automation innovation by regional labor market size

Regions with:	Manufacturing Workforce		Total Workforce	
	Small	Large	Small	Large
	(1)	(2)	(3)	(4)
Log Guest workers _{t-2}	0.064 (0.118)	0.115 (0.104)	0.183 (0.152)	0.144 (0.102)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Other foreign workers	Yes	Yes	Yes	Yes
Total employment	Yes	Yes	Yes	Yes
Observations	698	708	698	708
Log-likelihood	-1440.352	-2238.302	-1474.273	-2210.013

Notes: Poisson pseudo-likelihood regressions. Dependent variable: Non-automation patents: number of non-automation patents filed in year t . Sample splits by the 50-percentile of the manufacturing workforce in the year 1961 (column 1-2) and by the 50-percentile of the total workforce (column 3-4). *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbereicht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Data on the manufacturing workforce and the total workforce of the year 1961 from Schmitt *et al.* (1994). Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.1: Foreign workers in West Germany



Notes: Annual stock of foreign workers in West Germany. Analysis period covers the 10-year period 1964-1973. Figure based on data from the Federal Employment Agency.

Table 3.7: Association between labor recruitment and (non-)automation innovation - controlling for unemployment rates

Dep. Var.:	Non-automation patents				Automation patents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Guest workers _{t-2}	0.102 (0.101)	-0.010 (0.111)			-0.033 (0.236)	-0.238 (0.244)		
% Unemployed _{t-3}		-0.093*** (0.033)				-0.187*** (0.060)		
Log Guest workers _{t-3}			0.303*** (0.106)	0.227** (0.113)			0.059 (0.246)	-0.065 (0.257)
% Unemployed _{t-4}				-0.055 (0.035)				-0.095 (0.077)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other foreign workers	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Total employment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	846	846	705	705	810	810	670	670
Log-likelihood	-2224.044-2218.880-1805.522-1803.972-1259.644-1256.759-1032.256-1031.614							

Notes: Poisson pseudo-likelihood regressions. Dependent variable: Column 1-4: Non-automation patents: number of non-automation patents filed in year t . Column 5-8: Automation patents: number of automation patents filed in year t . *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Regional unemployment rates are taken from from the annual series *Arbeitsamtsstatistik* published by the Federal Employment Agency. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.8: Association between labor recruitment and (non-)automation innovation – exclusion of regions with low or high innovative capacity

Dep. Var.:	Non-Automation patents		Automation patents	
	Low capacity	High capacity	Low capacity	High capacity
Excluding regions with:	(1)	(2)	(3)	(4)
Log Guest workers _{t-2}	0.177** (0.084)	0.130* (0.077)	-0.122 (0.152)	-0.110 (0.169)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Other foreign workers	Yes	Yes	Yes	Yes
Total employment	Yes	Yes	Yes	Yes
Observations	1260	1260	1250	1220
Log-likelihood	-3473.842	-3129.057	-1942.020	-1696.900

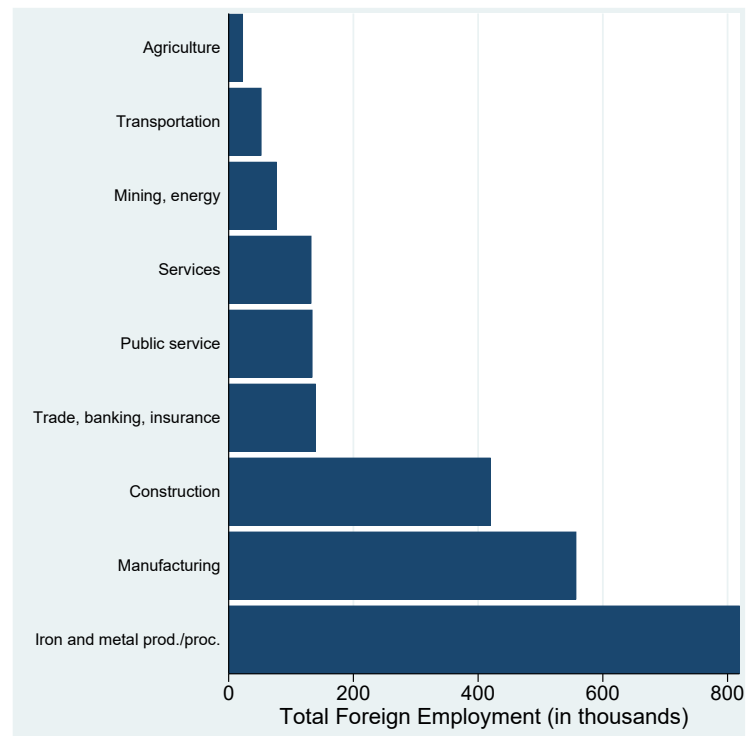
Notes: Poisson pseudo-likelihood regressions. I exclude regions in column 1 and 3 (2 and 4), if the pre-existing regional number of patents in 1963 is below (above) the 10 (90) percentile. Dependent variables: Column 1-2: Non-automation patents: number of non-automation patents filed in year t . Column 3-4: Automation patents: number of automation patents filed in year t . *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.9: Association between labor recruitment and (non-)automation innovation – overlapping observations

Dep. Var.:	Non-Automation patents	Automation patents
	(1)	(2)
OL: Log Guest worker _{t-2}	0.214* (0.118)	-0.168 (0.143)
Region and year fixed effects	Yes	Yes
Other foreign workers	Yes	Yes
Total employment	Yes	Yes
Observations	1112	1112
R^2 within	0.46	0.28
p-value WCB	0.098	0.280

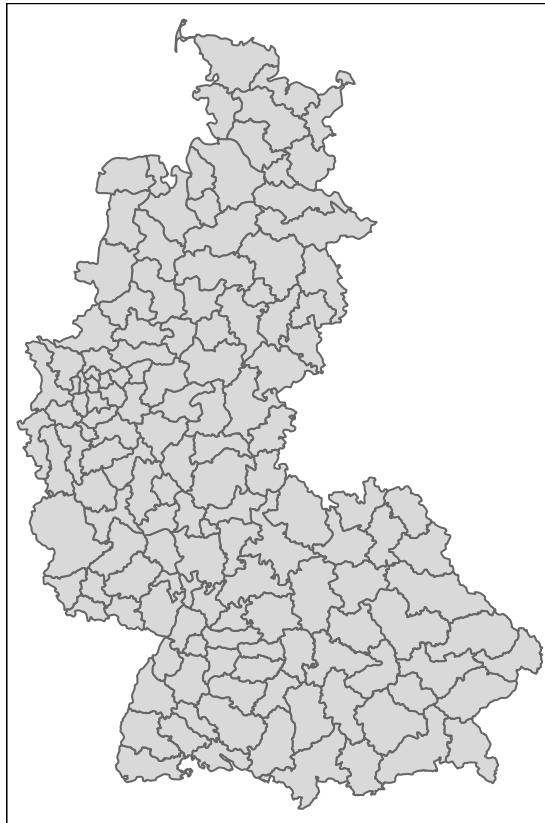
Notes: OLS regressions. Dependent variables: Column 1: Non-automation patents: the cumulative number of non-automation patents over the period t to $t + 2$ (entered in logs). Column 2: Automation patents: the cumulative number of automation patents over the period t to $t + 2$ (entered in logs). Log Guest worker_{t-2}: cumulative stock of guest workers over the three-year period $t - 2$ to t (entered in logs). *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency. The p-value WCB is calculated with the wild cluster bootstrap-t method by Cameron *et al.* (2008). Standard errors clustered on the region level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.2: Foreign workers in West Germany by occupation



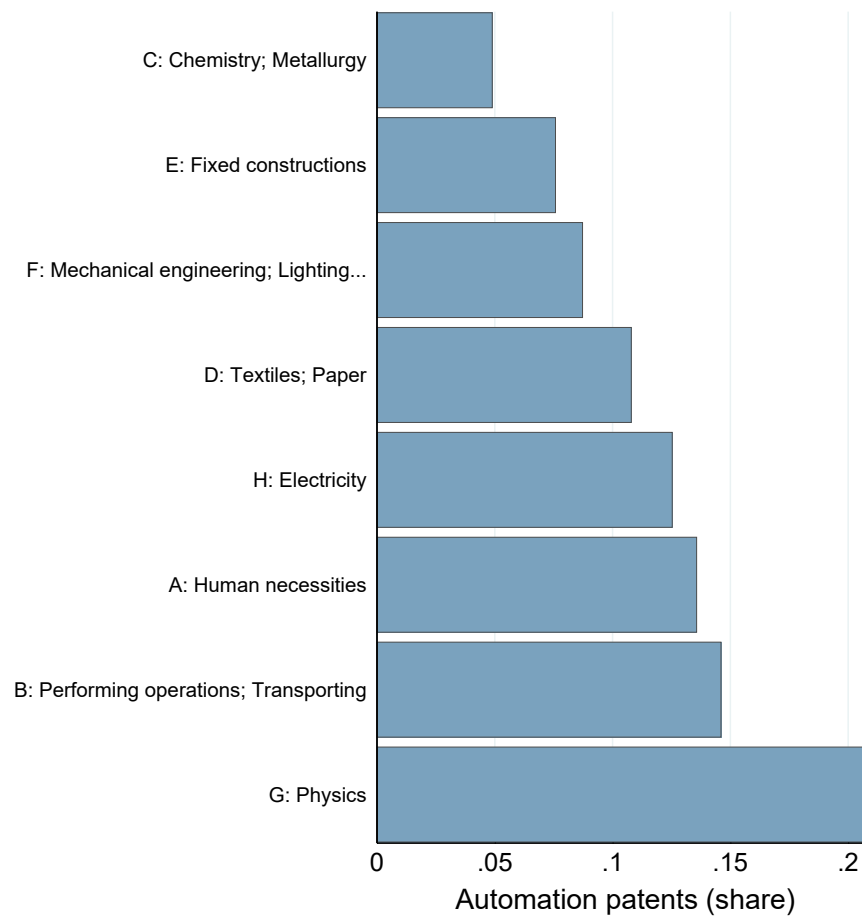
Notes: Foreign workers in West Germany by occupation classes in 1972. Figure based on data from Federal Employment Agency (1974).

Figure 3.3: West German labor office districts



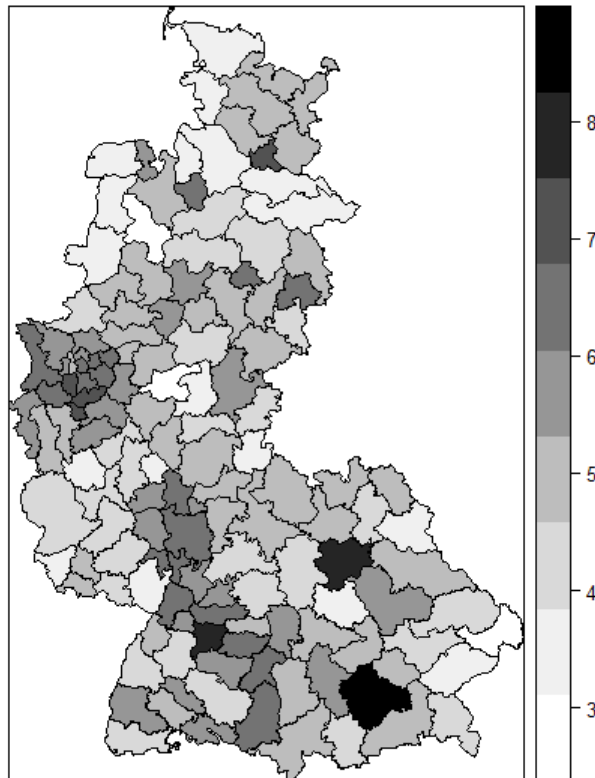
Notes: West Germany's 141 labor office districts. Figure based on a digitized map of labor office districts in West Germany from Federal Employment Agency (1972b).

Figure 3.4: Automation innovation across technology areas



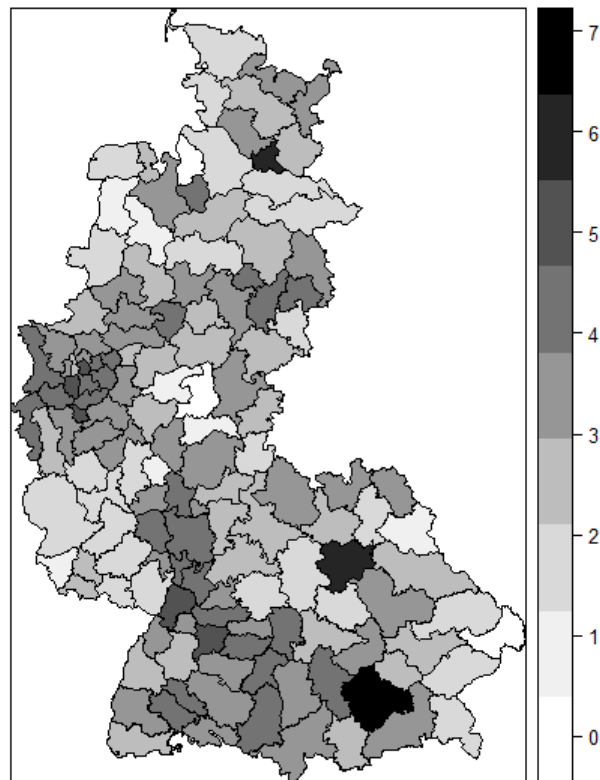
Notes: The share of automation patents across the eight main technology areas of the IPC. Figure based on the patents of my analysis sample. See text for more details on the classification of patents into automation patents and non-automation patents. Own calculations. Source: historic DPMA patent database *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986*.

Figure 3.5: Level of non-automation patents across regions



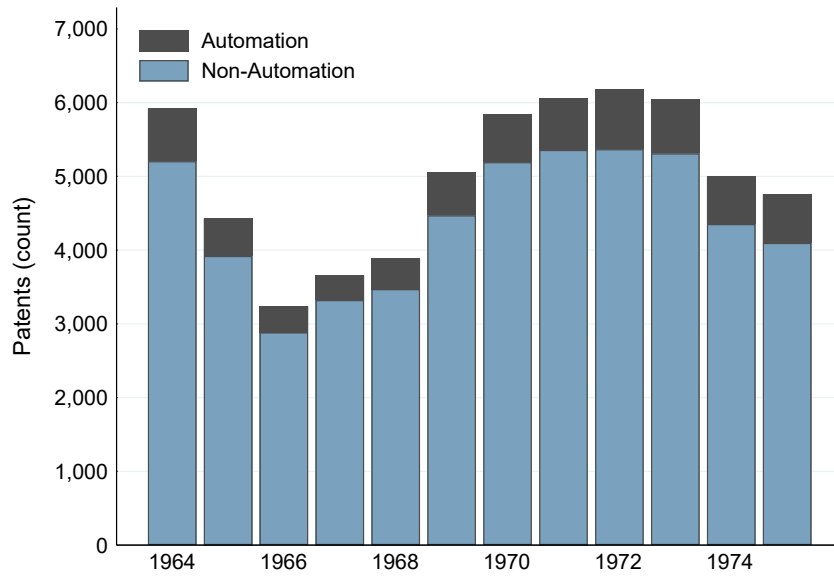
Notes: The level of non-automation patents across labor office districts in West Germany (entered in logs). Own calculations of the level of non-automation patents for 141 labor office districts. Based on the patents of my sample filed between 1964 and 1973. Figure based on a digitized map of labor office districts in West Germany from Federal Employment Agency (1972b).

Figure 3.6: Level of automation patents across regions



Notes: The level of automation patents across labor office districts in West Germany (entered in logs). Own calculations of the level of automation patents for 141 labor office districts. Based on the patents of my sample filed between 1964 and 1973. Figure based on a digitized map of labor office districts in West Germany from Federal Employment Agency (1972b)

Figure 3.7: The level of (non-)automation patents



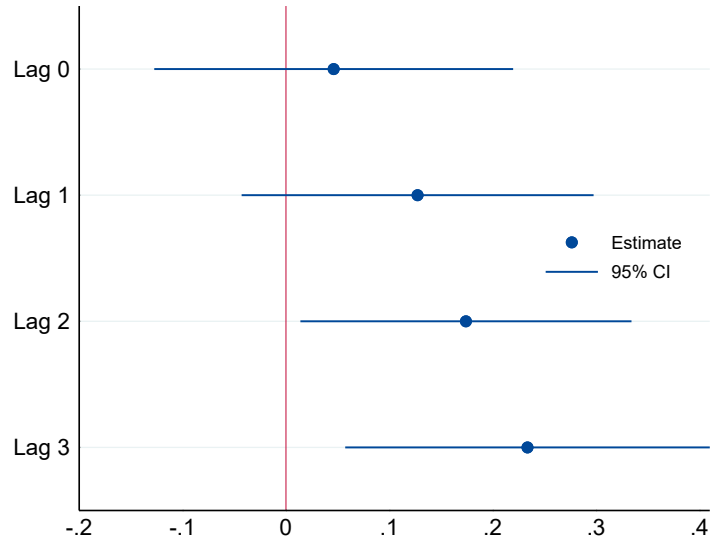
Notes: Figure shows the annual number of patents filed at least one inventor located in West Germany between 1964 and 1975. See text for more details on the classification of patents into automation patents and non-automation patents. Source: historic DPMA patent database *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986*.

Figure 3.8: Geographical overlap between labor office districts and counties



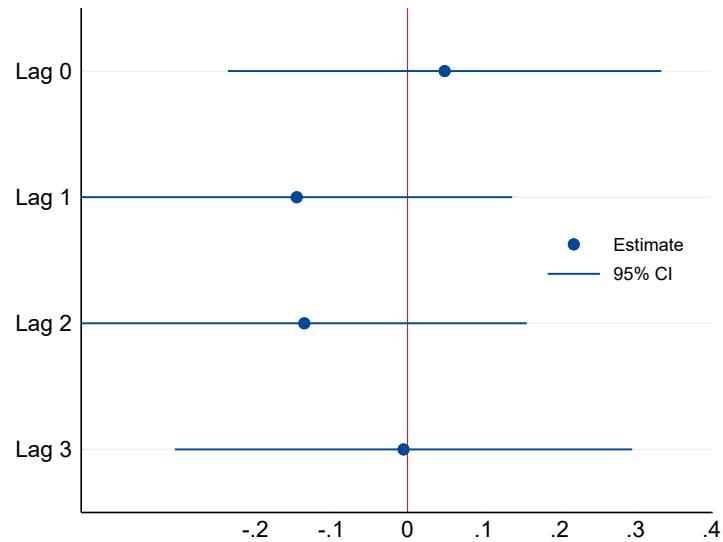
Notes: Figure shows the geographical overlap between West Germany's 141 labor office districts and the 327 West German counties. The thick black lines denote borders of labor office districts. A labor office district usually contains around one or more counties. Figure based on shapefiles of 1) labor office districts from Federal Employment Agency (1972b) and 2) counties from Max Planck Institute for Demographic Research and CCG (2011) and Bundesamt für Kartographie und Geodäsie (2011).

Figure 3.9: Association between labor recruitment and non-automation innovation - alternative lag structure



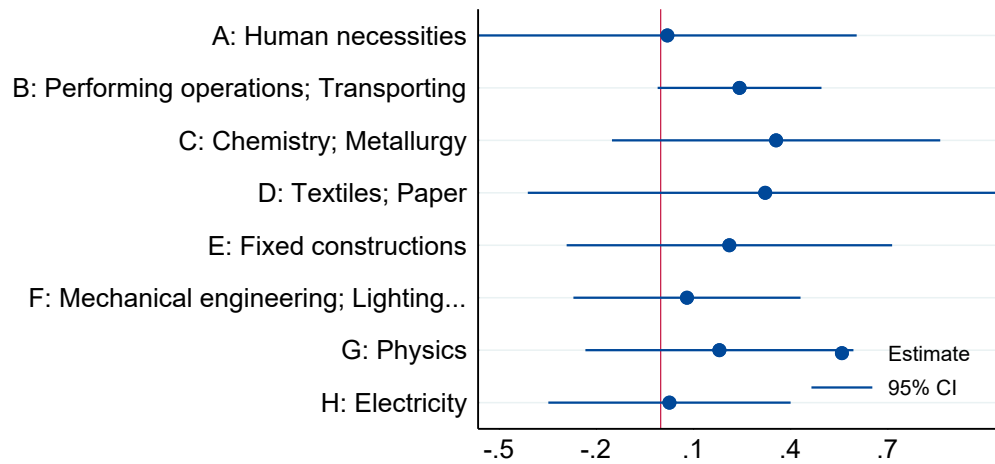
Notes: Poisson pseudo-likelihood regressions. Dependent variable: Non-automation patents: number of non-automation patents filed in year t . Figure shows the effects of the stock of guest workers across alternative lag structures. *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency.

Figure 3.10: Association between labor recruitment and automation innovation - alternative lag structure



Notes: Poisson pseudo-likelihood regressions. Dependent variable: Automation patents: number of automation patents filed in year t . Figure shows the effects of the stock of guest workers across alternative lag structures. *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency.

Figure 3.11: Association between labor recruitment and non-automation innovation across technology areas



Notes: Poisson pseudo-likelihood regressions. Dependent variable: number of automation patents which are in one of the eight main technology areas of the IPC. *Data sources:* Innovation outcomes based on the historic patent dataset *Volltexte (Bibliografische Daten, Beschreibung, Ansprüche) 1877-1986* from the German Patent and Trade Mark Office. Guest worker data and total employment from the annual series *Anwerbung Vermittlung Beschäftigung Ausländischer Arbeitnehmer Erfahrungsbericht* and *Ausländische Arbeitnehmer* published by the Federal Employment Agency.

Chapter 4

Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children

4.1 Introduction

With the recent arrival of large numbers of refugees in Europe, many societies wonder about the best policies to integrate immigrants. One central issue is the regional allocation of immigrants. To prevent ethnic ghettoization, many European countries adopted dispersal policies that assign refugees across regions (Dustmann *et al.*, 2017a). Existing evidence tends to suggest, though, that enclaves may in fact facilitate the labor-market integration of immigrants (Schüller, 2016), presumably through positive network effects within ethnic groups (Dustmann *et al.*, 2016a). However, for the successful integration of immigrants into host-country societies in the long run, the intergenerational effects of ethnic concentration on the immigrants' children seem even more important. To that extent, immigrant children's proficiency in the host-country language and their educational attainment play a particular role for long-term employment opportunities and for cultural and social integration (Dustmann and Glitz, 2011; Chiswick and Miller, 2015). On the one hand, children's language acquisition and educational integration may benefit from ethnic enclaves that provide useful information, reduced discrimination, and positive role models. On the other hand, immigrant children may also be hindered by limited exposure to native children, reduced options for language acquisition, lower socio-economic opportunities of families, and negative role models. In this paper, we study the effect of regional ethnic concentration on the language proficiency and educational attainment of immigrant children.

Our analysis exploits the placement policy of the German guest-worker program. Between 1955 and 1973, the German government actively recruited (mainly low-skilled) foreign workers to fill labor shortages. The guest workers were enlisted in various countries of origin and then quasi-exogenously placed across West German firms. The German Socio-Economic Panel (SOEP) allows us to extract a sample of roughly 1,000 children whose

parents immigrated into Germany from five different countries of origin during the period of the guest-worker program. In contrast to administrative datasets, the SOEP household panel provides information on these children's host-country language proficiency, as well as their educational attainment. In addition, the SOEP contains rich information on parents' speaking and writing abilities, friendships with Germans, and indicators for parents' social and labor-market integration that allows us to analyze factors that may mediate the effect of ethnic concentration on child outcomes. We merge the SOEP data on individual immigrant children with administrative data on the regional concentration of different ethnicities.

The initial regional assignment of guest workers provides us with plausibly exogenous variation in ethnic concentration across regions, circumventing bias from endogenous sorting of immigrants into enclaves of co-ethnics. We show that demographics of guest-worker parents and their children are balanced across regions with low and high ethnic concentration. To account for any type of region-specific or ethnicity-specific differences, our models additionally include region and ethnicity fixed effects. Region fixed effects ensure that any region-specific peculiarities are accounted for to the extent that they are common across guest-worker ethnicities. Ethnicity (country-of-origin) fixed effects ensure that any ethnicity-specific differentials in integration are accounted for to the extent that they are common across regions. Thus, we identify the effect of ethnic concentration on immigrant children's host-country language proficiency and educational attainment by observing different (exogenously placed) immigrant groups who are exposed to differential concentrations of co-ethnics within the same region, thereby circumventing bias from endogenous location choices of immigrants and from unobserved factors such as differing baseline willingness or disposition to integrate of different ethnic groups.

Our results indicate that growing up in ethnic enclaves significantly reduces immigrant children's proficiency in the host-country language and their educational attainment. In particular, a one log-point increase in the size of the own ethnic group in the region - equivalent, e.g., to increasing an ethnicity's share in the regional population from 1.0 percent to 2.8 percent - leads to a reduction in the German speaking proficiency of the children of the guest-worker generation by 19 percent of a standard deviation and a reduction in the German writing proficiency by 17 percent of a standard deviation. In addition, a one log-point increase in exposure to own-ethnic concentration increases the likelihood that the immigrant child drops out of school without any degree by 5.6 percentage points (compared to an average of 7.1 percent). Although less robust, there is some indication that ethnic enclaves also reduce the probability of obtaining an intermediate or higher school degree. Concerning effect heterogeneities, we find that effects tend to be larger for those immigrant children who were born abroad, whereas there are no significant gender differences.

Importantly, the rich background information on children and parents contained in the SOEP allows us to analyze several mediating factors. Potential mechanisms underlying the negative effect of growing up in ethnic enclaves include parents' lower host-country language proficiency, reduced interactions with natives, and lower wages and employment opportuni-

ties of immigrant parents. We find that differences in parents' ability to speak the German language - which is strongly related to their children's German language proficiency - can in fact account for much of the effect of growing up in ethnic enclaves. In particular, once parental German speaking abilities are controlled for, the estimated effect of ethnic concentration on children's language proficiency is reduced to close to zero. For this analysis, it proves essential to address measurement error in the self-reported parental language measure by implementing an instrumental variable (IV) approach that uses parents' responses on the same survey item from consecutive years (leads and lags) as instruments (Dustmann and Van Soest, 2002). While measures of parental writing abilities, friendships with German children, visits from Germans at home, parental unemployment, and household income are also significantly related to immigrant children's language proficiency, they do not account for the negative effect of ethnic concentration. Furthermore, none of the investigated mechanisms can explain the negative enclave effect on school dropout.

Our results are robust to a number of sensitivity analyses. In particular, we use alternative functional forms for the measure of ethnic concentration, instrument ethnic concentration at the time of observation by the ethnic concentration observed a decade earlier, use social-security and census data to construct the ethnic concentration measure, measure ethnic concentration at different levels of regional aggregation, and account for interview mode (which may influence self-assessed German language proficiency). We also show that neither return migration nor regional migration within Germany were selective with respect to ethnic concentration and that ethnic concentration did not affect family size.

Our paper contributes to several strands of literature. A growing literature in sociology and economics addresses the link between ethnic enclaves and the human capital acquisition of immigrant children. Grönqvist (2006) finds lower university attainments among refugees in enclaves in Sweden, Cortes (2006) finds no test score disadvantages for enclave schools in the US, and Jensen and Rasmussen (2011) establish lower test scores among immigrant children in enclaves in Denmark. While these papers do not place major emphasis on addressing bias from self-selection into ethnic enclaves, Åslund *et al.* (2011) use a refugee placement policy in Sweden and find that the concentration of highly educated co-ethnics positively affects the achievement of immigrant students in school.

Sharing their emphasis on identification, our main contribution beyond the Swedish study lies in the target population and results. Guest workers in Germany arrived with a signed labor contract and were quasi-exogenously distributed across regions, whereas Sweden assigned refugees according to municipal integration capacities (that is, with respect to labor-market and educational opportunities). Since all guest workers in Germany were employed upon arrival, our study can switch off one potential channel of how enclaves affect human capital outcomes of children. Furthermore, while Åslund *et al.* (2011) focus on a heterogeneous group of partly well-educated refugees, the focus of our study is on children of a rather homogeneous group of low-skilled labor migrants. More than one third of guest workers had left school without any degree, and only 7 percent had completed over twelve years of education

(compared to 27 percent of other immigrants in Germany at the time). The corresponding share in the Swedish study is more than five times as large (39 percent). Crucially, our findings stand in sharp contrast to the Swedish study, which concludes that ethnic enclaves are education-enhancing. By contrast, ethnic enclaves inhibit language take-up and school graduation in the German case. Any effect of ethnic concentration may thus strongly depend on the skill and employment levels of co-ethnics in the enclave. Our analysis also differs from Åslund *et al.* (2011) in additional ways. Their study focuses on GPA scores and on-time school graduation, whereas we use long-run panel data to assess final school graduation as economically important outcome. Importantly, our survey data allow us to investigate the underlying mechanisms in detail, yielding important insights. Finally, unlike the Swedish refugee settlement, the German guest-worker placement implied a guaranteed immigration permit, immediate placement, and an immediate job start. Location preferences of guest workers were not considered, and they were not free to change residence during a lock-in period. In this sense, the experimental setup is even stricter and cleaner in our setting.

A vast literature studies the effects of ethnic enclaves on the economic integration of adult immigrants (for an overview see Schüller (2016)). Using dispersal policies in Sweden and Denmark, respectively, Edin *et al.* (2003) and Damm (2009) find positive network effects of ethnic concentration on immigrants' labor-market outcomes. By contrast, studying the same setting as in our paper, Danzer and Yaman (2016) and Constant *et al.* (2013) find negative effects of ethnic concentration on adult immigrants' proficiency in the host-country language and their cultural integration, respectively. In a different German setting, Battisti *et al.* (2016) find positive short-term but negative long-term effects of ethnic concentration on labor-market outcomes, with the negative effect being related to lower human capital investments and larger job mismatch.

Beyond immigrant integration, another large literature studies the effect of spatial segregation and concentration on the economic success of racial minorities, usually finding negative effects (e.g., Cutler and Glaeser, 1997; Fryer, 2011). More generally, a growing literature studies the effect of exposure to different quality neighborhoods during childhood on children's outcomes in the short and long run (e.g., Chetty, Hendren, and Katz, 2016; Chetty and Hendren, 2018; Gibbons, Silva, and Weinhardt, 2013, 2017).

We contribute to this literature by estimating well-identified effects of growing up in low-skilled ethnic enclaves on the language proficiency and educational attainment of immigrant children and by providing a rich analysis of mediating factors. Our findings indicate that parents' limited proficiency in speaking the host-country language is a key mediating factor of the negative impact of ethnic enclaves on immigrant children's language proficiency. By contrast, limited interaction with natives and parental economic conditions do not seem to be leading mechanisms. Overall, the opportunity to benefit from large social networks of co-ethnics may be particularly relevant for newly arriving immigrants, but less so for the long-term integration of the children of settled immigrants. More generally, most of the arguments in favor of ethnic enclaves tend to relate to the labor-market integration of adult

immigrants but bear less relevance for integration beyond the labor market. Regarding the cultural and educational integration of the second generation of immigrants, our results suggest that the fear of ghettoization that underlies the dispersal policies of several European countries may not be totally misplaced.

In what follows, Section 4.2 provides institutional background on the German guest-worker program. Section 4.3 describes the SOEP household data and the administrative data used to compute ethnic concentrations. Section 4.4 introduces our empirical model and shows balancing of demographic characteristics across regions with low and high ethnic concentration. Section 4.5.1 presents our main results on the effect of ethnic concentration on immigrant children's outcomes. Section 4.6 investigates the relevance of several potential mediating factors. Section 4.7 provides a number of robustness analyses. Section 4.8 concludes.

4.2 Institutional Background on the German Guest-Worker Program

The German guest-worker program was one of the largest guest-worker programs worldwide. West Germany (hereafter, Germany) signed bilateral guest-worker treaties with Italy in 1955, Greece and Spain in 1960, Turkey in 1961, and Yugoslavia in 1968. During a period of rapid economic growth in the 1960s and early 1970s, increasing demand for low-skilled workers induced a massive inflow of labor migrants to fill the numerous open positions in the economy. Given that all treaties were designed to attract low-skilled and mainly young workers, the guest workers constitute a rather homogeneous immigrant population that is, on average, less educated than the German workers. Due to the severe economic recession triggered by the oil crisis, Germany stopped the recruitment of guest workers in 1973. By that time, 2.6 million foreign workers were employed in Germany, implying that 12 percent of the labor force were foreigners (Federal Employment Agency, 1974).

To take up employment, guest workers were required to hold a valid work permit (*Arbeitsserlaubnisbescheinigung*). The formal process of obtaining this permit was initiated at the foreign branches of the German Federal Employment Agency in the guest-worker countries, which was similar for all source countries.⁷⁵ Potential workers were screened for basic literacy and underwent medical check-ups.⁷⁶ Then, guest workers were matched with German

⁷⁵The foreign branches of the German employment agency were called *Deutsche Kommission* in Greece, Italy, and Spain, *Deutsche Verbindungsstelle* in Turkey, and *Deutsche Delegation* in Yugoslavia. Italians could later enter Germany more freely within the European Economic Community (EEC) framework, but were placed by an internal recruitment branch within Germany (*Zentralstelle für Arbeitsvermittlung*). The German embassy in Yugoslavia opened a second track for guest-worker applications in 1970 to account for the high number of applicants. For more details, see Dohse (1985) and Federal Employment Agency (1962).

⁷⁶At this occasion, applicants also received information on the working and living conditions in Germany. Guest workers were predominantly low-skilled due to the nature of labor demand in the construction, mining, metal, and ferrous industries at that time and because the governments of the sending countries preferred emigration from underdeveloped and disaster-ridden areas Penninx and Van Renselaar (1976).

employers. The employers could submit recruitment requests together with blank work contracts to their local labor offices, which forwarded them to the foreign branches after initial approval.⁷⁷

German firms received almost no information about their requested workers before arrival and in practice generally could not select workers based on job skills or country of origin (Feuser, 1961; Fassbender, 1966; Voelker, 1976). Successful applicants got a work contract from a specific German company and a one-year work permit that was only valid for employment at the specific firm (Feuser, 1961). Recruited workers were then transferred to Germany in groups.⁷⁸ After having stayed with their initial employer for at least two years and in the same occupation (and, in practice, in the same region for most guest workers) for at least five years, guest workers could receive an upgrade of their work permit (*Erweiterte Arbeitserlaubnisbescheinigung*) that included free job choice (Dahnen and Kozłowicz, 1963).⁷⁹

Given that the initial location in Germany depended on current labor demand, the initial location was exogenous from the perspective of an individual guest worker. Most importantly, the guest-worker recruitment process generated exogenous variation in ethnic concentrations that allows us to estimate the causal effect of ethnic concentration on immigrant children's outcomes. The regional variation in ethnic concentration differed across ethnicities for at least three reasons. First, regional labor demand fluctuated over time, and guest-worker recruitment started in the different countries of origin at different points in time. For instance, the relative size of the guest-worker population in North Rhine-Westphalia declined between 1962 and 1973, whereas the share of guest workers in Lower Saxony and Bremen increased. Second, labor supply in the countries of origin varied over time. More guest workers were recruited from countries that had temporarily abundant labor supply, like Turkey in 1964 (Federal Employment Agency, 1965). Third, while guest workers from Spain and Portugal arrived by train in Cologne in the West, the port of entry for the remaining guest-worker groups (Greek, Italian, and Turkish) was Munich in the South, generating additional variation in the ethnic concentration across nationalities.

In 1973, the guest-worker recruitment was officially stopped. However, immigration of family members within the family reunification framework ensured high levels of inflows from guest-worker countries also afterwards. Those family members immigrated on the basis of the Aliens Act of 1965 and were granted a residence permit when joining a guest-worker family member.

⁷⁷The local labor office checked whether German workers were available for the open positions, whether housing was available for foreign workers, and whether the request fulfilled all conditions of the bilateral treaty.

⁷⁸Travel costs were covered by recruiting firms by paying a small flat fee for each recruited worker.

⁷⁹As an alternative recruitment process, employers were allowed to request guest workers by name if there was a personal relationship to that person, for example, through recommendations by relatives or friends who were already employed at that firm. Recruitment by name became more important as guest workers recommended their spouses. However, for various reasons, a large fraction of individuals who were requested by name were eventually not hired (Federal Employment Agency, 1972a).

4.3 Data

Our analysis uses individual-level information on guest workers and their children from the German Socio-Economic Panel (Section 4.3.1). We construct our main measure of ethnic concentration from a large employee sample of the Research Institute of the Federal Employment Agency (Section 4.3.2).

4.3.1 Survey Data on Guest Workers and their Children

We use information on guest workers and their children from the German Socio-Economic Panel (SOEP, version 30), a large annual household survey that is representative of the resident population in Germany. The first SOEP wave in 1984 strongly oversampled guest workers (by a factor of four). As a consequence, 1,393 of the 5,921 SOEP households originated from the five guest-worker countries, which comprised the largest foreigner populations in Germany at the time (Sample B). For each ethnicity, an independent random sample was drawn to allow for stand-alone analyses (Haisken-DeNew and Frick, 2005). The SOEP contains detailed information on individual characteristics, including educational attainment and, for foreigners, self-reported German speaking and writing proficiency.⁸⁰ The 1985 survey is the first wave that provides sufficient geographic information on the region of residence at the county level. Hence, we identify guest workers and their region of residence based on information in the 1985 wave. Using information from mothers' birth biography and pointers to their partners in 1985, we link parents to their children.⁸¹ While the SOEP does not contain a direct indicator of guest workers, we identify guest workers by their country of origin, year of immigration, and age at migration. The guest-worker immigration differed from previous immigration experiences in Germany in that guest workers were predominantly male, young, and low educated.⁸²

Our analysis sample consists of 1,065 guest-worker children with Greek, Italian, Spanish, Turkish, or Yugoslav background. To be included in the sample, children must have at least one parent who was aged 18 or older at immigration and who arrived in Germany during the period when the guest-worker program with her/his home country was in place. We restrict the sample to children aged 13 or younger at migration since the focus of our study is to investigate the impact of the region where children grow up.⁸³ We keep only children with

⁸⁰All questionnaires, in German and partly in English, are available at https://www.diw.de/en/diw_02.c.222729.en/questionnaires.html.

⁸¹We use only children for whom both mother and father could be identified.

⁸²Appendix Table 4A-1 compares demographic characteristics of guest workers and immigrants from other countries in 1985, based on all first-generation immigrants in the SOEP. Guest workers are, on average, five years younger and more likely to be employed in 1985, consistent with their demand-driven recruitment. Guest workers have around two years less education. Most strikingly, while 36 percent of guest workers have left school without any type of school degree, this share is only 3 percent among other immigrants. Only 7 percent of guest workers have more than 12 years of schooling, compared to 27 percent of immigrants from other countries.

⁸³We present heterogeneity results below for guest-worker children born in Germany vs. children born abroad.

at least one observation for self-reported German language proficiency or one observation for educational attainment.⁸⁴

We measure children's German language proficiency by two distinct outcomes: speaking proficiency and writing proficiency. Both language outcomes are self-reported and based on the following question: "In your opinion, how well do you speak and write German?" Answers are provided on a five-point scale: very well, well, fairly, poorly, and not at all. Children report their German language proficiency for the first time at the age of 17 or 18, i.e., when they are personally interviewed in the SOEP for the first time. An advantage of the panel data is that we observe multiple observations of self-reported language proficiency for each child (five observations per child on average), resulting in a large sample of language proficiency observations. An additional advantage of the panel data is that we can address measurement error in *parents'* language proficiency by instrumenting the self-reported language proficiency in a given year with their self-assessments in previous or succeeding years (see Section 4.6.1). In our sample of language proficiency, each observation is at the child-year level. This sample is based on the SOEP waves 1984-1987, 1989, 1991, 1993, and every two years from 1997 to 2005, including about 4,900 child-year observations.⁸⁵ We standardize each outcome of children's language proficiency to have mean 0 and standard deviation 1.

Children's educational attainment is also measured by two variables. The binary indicator "any school degree" equals 1 if the child obtained any type of school degree and 0 if the child dropped out of school without any degree. The binary indicator "at least intermediate school degree" equals 1 if the child obtained an intermediate school degree (*Realschulabschluss*) or a higher secondary school degree and 0 otherwise.⁸⁶ Children's educational attainment is based on the most recent available information in the SOEP.⁸⁷

Table 4.1 reports descriptive statistics of children's outcomes and demographic characteristics of children and their parents, separately for regions with low and high ethnic concentration (split at the ethnicity-specific median of the share of ethnic concentration in 1985). Immigrant children living in regions with a high co-ethnic concentration report lower German speaking proficiency (statistical significance at 12 percent) and lower writing proficiency (significant at the 10 percent level) than immigrant children living in low co-ethnic

⁸⁴The main reason for missing values on language proficiency and educational attainment is that households stopped participating in the SOEP survey before the children turned 17 years old and would be personally interviewed for the first time. The share of children with missing values on the outcomes does not differ between regions with low and regions with high co-ethnic concentration (see bottom of Table 4.1).

⁸⁵Our panel data set for children's language proficiency is unbalanced for two reasons. First, some children were younger than age 17 in 1985 and therefore did not participate in the personal interviews during the first years of our panel data. Second, some children (usually the entire household) left the SOEP survey before 2005.

⁸⁶In Germany, there are three types of secondary school degrees: basic (*Hauptschulabschluss*), intermediate (*Realschulabschluss*), and advanced (*Abitur*). A small share of children in our sample (2.9 percent) reported to have obtained another type of school-leaving certificate. While we assume that this other type of school-leaving certificate is equivalent to an intermediate school degree, the results do not depend on this assumption.

⁸⁷If the most recent available information indicates dropout or no school degree (yet), we checked for school-leaving degrees reported in previous waves. For only nine children, we adjusted the educational attainment variables based on previously reported school-leaving degrees.

concentration regions. Consistent with this finding, immigrant children in regions with high co-ethnic concentration are significantly less likely to obtain a school degree and slightly (and statistically insignificantly) less likely to obtain at least an intermediate school degree.

In terms of ethnicities, 37 percent of immigrant children in our sample are Turkish, 19 percent each are Italian and Yugoslav, 15 percent are Greek, and 10 percent are Spanish. We identify the ethnicity of the immigrant children primarily based on their first citizenship (94.2 percent of the children in our sample). In the case of a German citizenship or missing citizenship information, ethnicity is based on the children's country of birth or their parents' nationality (see Appendix Table 4A-2 for definitions of all individual-level variables).⁸⁸ A slight majority of immigrant children in the sample (57.1 percent) were born in Germany. The average year of birth is 1971, and the average age at migration is 2.8 years.

The SOEP also contains a rich set of additional individual characteristics, including the immigration history, educational attainment, and labor-market outcomes of adults.⁸⁹ This wealth of information allows us to investigate several potential mediating factors that may drive the effects of ethnic concentration. As potential mediating factors, we investigate parents' speaking and writing proficiency in German, parents' employment status, household income, visits from Germans at home, and whether the child's first friend is German. Parents' mediating factors are based on the average of mothers' and fathers' information.

4.3.2 Ethnic Concentration

We compute measures of the concentration of co-ethnics in the region separately for the five guest-worker nationalities (Greek, Italian, Spanish, Turkish, and Yugoslav) at the regional level of the so-called *Anpassungsschichten*. Typically comprising several counties, these regions constitute a regional labor market. In West Germany, there were 103 *Anpassungsschichten* in 1985 with an average population of about half a million people. While smaller geographic units may better reflect small-scale ethnic neighborhoods, a higher level of regional aggregation produces more conservative estimates and circumvents potential bias from the typical sorting of immigrants into close-by cities or across city districts. Since compared to the United States, the degree of ethnic and social segregation is low in Germany, *Anpassungsschichten* reflect the most suitable level of analysis (see Danzer and Yaman (2016), for a discussion of the trade-off between small and large units). In any case, our findings are fully confirmed in robustness analysis at the level of counties.

⁸⁸In the very few instances in which children have a German citizenship or information on citizenship is missing and the nationality of mother and father differs, we use mother's nationality or mother's country of birth.

⁸⁹As is typical for surveys, our data on guest workers and their children contain missing values for some variables. Since our set of control variables is large, dropping all children with any missing value would substantially reduce the sample size. We therefore impute missing values by using the mean of each control variable. For binary indicators, imputed means are rounded to the closest integer. To ensure that results are not driven by imputed values, all our estimations include imputation dummies for each variable. The share of imputed values is small for all imputed variables on demographic characteristics (at most 3.6 percent). The main results are similar when excluding all observations with imputed values.

For the measurement of ethnic concentration, we use the Sample of Integrated Labor Market Biographies of the Research Institute of the Federal Employment Agency. The SIAB is a 2 percent random sample of all individuals in Germany who are employed subject to social security, job seeking, or benefit recipients as contained in the Integrated Employment Biographies of the German social security system (Dorner *et al.*, 2011). We use data from 1985, the year when guest workers' region of residence is observed for the first time in the SOEP data. Ethnic concentration, our key explanatory variable, is measured by the logarithm of the size of the ethnic community in the region of residence in 1985 (see Appendix Table 4A-3 for definitions of regional variables). In our regression analyses, region fixed effects control for the size of the overall population in a region. While it is common to measure ethnic concentration as the log size of the own ethnicity (e.g., Edin *et al.*, 2003; Damm, 2009; Åslund *et al.*, 2011) below we also report the robustness of our results to using the share of the own ethnicity in the total regional population as an alternative measure (e.g., Chiswick and Miller, 2007; Danzer and Yaman, 2013, 2016). We match our measures of ethnic concentration to the individual-level SOEP data at the level of regions (*Anpassungsschichten*) and ethnicities.

The extensive demand-driven recruitment of guest workers generated substantial variation in ethnic concentrations across regions. Figure 4.1 shows the distribution of ethnic concentrations separately for each of the five ethnicities across the 103 West German regions (*Anpassungsschichten*) in 1985 (see Appendix Table 4A-4 for descriptive statistics). There are clear differences in the settlement structures between the guest-worker ethnicities. For example, while Spanish guest workers tend to be concentrated in central Germany, Italians and Yugoslavs are more concentrated in the southern regions. We exploit the differential concentrations of ethnicities across regions in our analyses by using only differences in ethnic concentrations within the same region.

For robustness analyses, we also use the 1987 German Census to compute alternative measures of ethnic concentration. Being based on a 2 percent employee random sample, the SIAB measure of ethnic concentration may contain classical measurement error, biasing our estimates toward zero. In addition, if the regional share of co-ethnics in the employee sample does not reflect the ethnic concentration in the overall population - for example, because of differential labor-market participation rates - there may be non-classical measurement error. In robustness analyses, we therefore also use an alternative measure of ethnic concentration based on data from the 1987 Census. An advantage of this alternative measure is that the 1987 Census includes the entire population in Germany. The depth of the Census data also allows us to perform robustness analyses that define ethnic enclaves at the level of the 328 West German counties. A major disadvantage of the 1987 Census is that it does not allow to compute ethnic concentrations for Spanish guest workers, which reduces the sample size and excludes one of the five guest-worker ethnicities.⁹⁰ In addition, the ethnicity measure

⁹⁰Individuals with Spanish citizenship are included in the category "other citizenship". In the SOEP data, Spanish guest-worker children make up about 10 percent of the analysis sample.

in the Census is based on citizenship information (as country of birth is not observed in the Census), and the 1987 Census measures ethnic concentrations two years later than the 1985 SIAB data. Appendix Figures 4A-1 and 4A-2 depict the distribution of the Census-based measures of ethnic concentration separately for the four ethnicities at the level of *Anpassungsschichten* and counties, respectively.

4.4 Empirical Model

In this section, we discuss the basic setup of our empirical model (Section 4.4.1) and show the balancing of demographic characteristics of guest workers and their children across regions with low and high concentrations of co-ethnics (Section 4.4.2).

4.4.1 Model Setup with Region and Ethnicity Fixed Effects

We aim to estimate the effect of ethnic enclaves on the language proficiency and educational attainment of immigrant children. Exploiting the quasi-exogenous placement of guest workers, our basic model setup expresses immigrant children's outcomes as a function of the concentration of their ethnicity in their region. Conditioning on fixed effects for ethnicities and regions, the model is identified from the concentration of an ethnicity in a particular region compared to the concentration of other guest-worker ethnicities in the same region.

When estimating the effect of ethnic enclaves on immigrant children's host-country language proficiency, we make use of the panel structure of the SOEP where immigrant children report their German language proficiency in multiple consecutive years. This allows estimating the following random effects model.⁹¹

$$lang_{icrt} = \sigma_c + \delta_r + \tau_t + \beta_1 EC_{cr} + C'_{icr} \beta_2 + P'_{icr} \beta_3 + \vartheta_{icrt} + \epsilon_{icrt}, \quad (4.4.1)$$

where $lang_{icrt}$ is the German speaking and writing proficiency, respectively, of child i of ethnicity (country-of-origin) c living in region r in year t . The key explanatory variable is the concentration of child i 's ethnicity in her region, EC_{cr} .⁹² C'_{icr} is a vector of child characteristics, including gender, year of birth, and age at migration. P'_{icr} is a vector of parent characteristics, including year of birth, year of arrival in Germany, education in country of origin, years of schooling, a migration indicator (which equals 0 for a few spouses

⁹¹Just like a pooled OLS model, the random effects model effectively uses the average language proficiency across all observed time periods. Pooled OLS regressions and cross-sectional regressions with average language skills as dependent variable yield very similar results (results available upon request).

⁹²As described in Section 4.3.2, ethnic concentration, EC_{cr} , is measured as the (log) size of child i 's ethnic community in her region of residence in 1985, the first year in which the SOEP provides sufficient geographical information on guest workers. Including the quadratic of EC_{cr} does not indicate any non-linearities in the effects of ethnic concentration on language proficiency or school dropout, speaking against the existence of crucial thresholds or tipping points.

who have no migration background),⁹³ and the number of children for mothers. All models include fixed effects for ethnicities, σ_c , fixed effects for regions, δ_r , and fixed effects for the year when the child reported her language proficiency, τ_t . The individual-specific effects, ϑ_{icrt} , are assumed to be i.i.d. random variables, and ϵ_{icrt} is an idiosyncratic error term. Throughout, we cluster standard errors at the region-by-ethnicity level, the level at which our measure of ethnic concentration varies.⁹⁴

To estimate the effect of ethnic concentration on immigrant children's educational attainment, we estimate the following OLS model using a cross-section of children:

$$educ_{icr} = \sigma_c + \delta_r + \theta_1 EC_{cr} + C'_{icr} \theta_2 + P'_{icr} \theta_3 + \epsilon_{icrt}, \quad (4.4.2)$$

where $educ_{icr}$ is the educational attainment of child i , measured either by a binary indicator for having obtained any school degree or by a binary indicator for having obtained at least an intermediate school degree. As before, we include controls for child and parent characteristics as well as ethnicity and region fixed effects. By including ethnicity fixed effects, we account for any differences between ethnicities, such as linguistic distance to the German language, cultural distance, school quality in the country of origin, and general willingness or disposition to integrate into the host country. By including region fixed effects, we exploit only variation in ethnic concentrations within the same region, but do not use systematic differences in ethnic concentrations across regions. Thus, we control for any differences across regions, such as unemployment rates, wage levels, overall share of migrants, school quality, and attitudes of the native population. Our model therefore identifies the effect of ethnic concentration on immigrant children's outcomes from the presence of several immigrant groups with differing community sizes within the same region.

4.4.2 Balancing Test by Degree of Ethnic Concentration

As argued above, the placement policy of the German guest-worker program led to quasi-exogenous variation in the regional placement of guest workers. We can test this assumption by comparing observable characteristics of the immigrant children and their parents between regions with low and high ethnic concentration of the respective ethnicity. To do so, we split the sample at the ethnicity-specific median of the share of ethnic concentration in the child's region of residence in 1985. As Table 4.1 shows, none of the demographic characteristics of immigrant children differs significantly (individually or jointly) across regions with low and high co-ethnic concentration. The same is true for the demographic characteristics of

⁹³Among the parents in our sample, 2.9 percent of mothers and 0.8 percent of fathers are of German nationality without migration background.

⁹⁴Our results are fully preserved when clustering standard errors at the regional level which would account for regional shocks that may affect all ethnicities similarly (not shown). Note that it is unlikely that labor-market shocks affect ethnicities within the same region differently since guest workers from different origin countries had similar qualification levels and worked in similar occupations (Appendix Table 4A-5).

mothers and fathers. Similarly, using the specification of our outcome model, there is no significant relationship between ethnic concentration and background characteristics when regressing the background characteristics on the share of ethnic concentration as well as ethnicity and region fixed effects (Appendix Table 4A-6). These balancing tests support our assumption that there was no systematic self-selection of guest workers into regions of differing ethnic concentration.

Beyond demographic backgrounds, the only exceptions where we find a significant difference between regions with low and high ethnic concentration are fathers' unemployment rates and household income. Interestingly, guest workers are better off in terms of employment and income in regions with high shares of co-ethnic concentration. If anything, this difference should work against finding any negative effect of ethnic concentration on children's outcomes. The unemployment difference observed for guest-worker fathers in the SOEP sample is qualitatively in line with the overall unemployment rates in 1985 from the Federal Employment Agency (see bottom of Table 4.1). Thus, the unemployment difference likely reflects the fact that guest workers were particularly demanded in regions with booming industries, which were still characterized by lower unemployment levels in 1985. Of course, the region fixed effects in our regression models account for any general difference across regions, exploiting only within-regional variation across different ethnicities. Furthermore, as we show below, differences in unemployment and household income do not explain the effect of ethnic concentration on children's outcomes.

The balancing of guest workers' demographic characteristics across regions with low and high ethnic concentration is particularly reassuring as we observe the location of guest workers in 1985 for the first time. As we do not observe the initial location to which guest workers had been assigned, we have to assume that any movement of guest workers across regions between their arrival in the 1960s/1970s and 1985 is orthogonal to our relationship of interest. Thus, the estimated coefficient on ethnic concentration would be biased downward (upward) if parents with adverse (advantageous) characteristics related to their child's outcomes moved to regions with high ethnic concentrations. The balancing results support our identifying assumption that guest workers in Germany did *not* systematically self-select into regions between their arrival and 1985.⁹⁵ Further supporting this assumption, we show in Section 4.7.7 below that the movement of guest workers across regions in the ten years after 1985, when we can observe them, was in fact orthogonal to ethnic concentration and to our outcome measures.

This is in line with existing work investigating the German guest-worker program. Previous studies also did not find any evidence of significant differences in demographic characteristics between guest workers living in regions with high concentrations of co-ethnics and

⁹⁵Alternatively regressing the continuous measure of the share of ethnic concentration on each demographic characteristic (including region and ethnicity fixed effects) yields only statistically insignificant coefficients, except for a significant positive coefficient on the variable indicating that parents had no schooling. In contrast, the coefficients on all four outcome variables are negative and statistically significant at about the one percent level (results available upon request).

those living in regions with low concentrations (Constant *et al.*, 2013; Danzer and Yaman, 2013, 2016). In contrast to the settings studied in some other papers (such as refugees in Sweden in Åslund *et al.* (2011)), the evidence against endogenous sorting of immigrants into ethnic enclaves in our setting is perfectly consistent with two specific features of the German guest-worker program.

First, as discussed above, guest workers were restricted in their residential choice as their work permit required them to stay in the initially assigned region for several years (Dahmen and Kozłowicz, 1963). Thus, the formal rules of the guest-worker program made it hardly possible for guest workers to move across regions during the initial years after their arrival.

Second, guest workers in Germany were well integrated into the labor market immediately upon arrival as they had been recruited specifically for the purpose to fill open positions in the German economy. As a result, the unemployment rate of foreigners in Germany was less than 1.5 percent in every year between 1968 and 1973 and was even lower than that of natives (Federal Employment Agency, 1974). Since guest workers - who migrated to Germany with the aim to work - had been employed immediately upon arrival, the incentive to move to other regions was very low. Accordingly, the current settlement structures of immigrants in Germany have been shown to still reflect the demand for labor in the 1960s and 1970s (Schönwälder and Söhn, 2009). Quite generally, ethnic segregation has been reasonably stable across workplaces and residential locations over the entire period from 1975 to 2008 (Glitz, 2014). In sum, the demographic characteristics of guest workers and their children are very similar across regions with low and high ethnic concentration. This finding supports our identification strategy of exploiting the quasi-exogenous placement of guest workers across West German regions to estimate the effect of ethnic enclaves on immigrant children's outcomes.

4.5 The Effect of Ethnic Concentration on Immigrant Children's Language Proficiency and Educational Attainment

This section presents our main results (Section 4.5.1) and subgroup analyses (Section 4.5.2). In the subsequent sections, we provide investigations of mediating factors and robustness analyses.

4.5.1 Main Results

Table 4.2 shows our main results on the effect of ethnic concentration on the host-country language proficiency of immigrant children. The results indicate that an increase in co-ethnic concentration significantly reduces immigrant children's speaking and writing proficiency in German. An increase in the size of the own ethnicity by one log-point is related to a decline

in speaking skills by 19 percent and in writing skills by 17 percent of a standard deviation. The magnitudes of the estimated coefficients barely change when we include controls for children's and parents' characteristics.

To facilitate interpretation of magnitudes, ethnic concentration would increase by one log-point, for example, if a Turkish child moved from the city of Bonn (with a share of Turks of about 1 percent) to the city of Munich (with a share of about 2.8 percent).⁹⁶ This change in the region of residence would, *ceteris paribus*, reduce the child's German speaking proficiency by 19 percent and her writing proficiency by 17 percent of a standard deviation, respectively. This is a modest effect, given that the difference between "poor" and "fair" German language proficiency is 1.39 standard deviations for speaking and 1.12 standard deviations for writing.

In line with the negative impact on host-country language proficiency, we also find a negative effect of ethnic concentration on immigrant children's educational attainment (Table 4.3). Living in an ethnic enclave substantially increases the likelihood of the child to drop out of school without any degree (columns 1 and 2). A one log-point increase in co-ethnic concentration increases the probability of dropping out of school by 5.6 percentage points. Given that the overall drop-out rate among immigrant children in our sample is only 7.1 percent, this is a huge effect. While results also point toward a negative impact on the probability of obtaining at least an intermediate school degree, the coefficient is much less precisely estimated and becomes zero when controlling for child and parent characteristics (columns 3 and 4).⁹⁷

Both findings - the negative effect on host-country language proficiency and the negative effect on obtaining any school degree - suggest that immigrant children who grew up in regions with high shares of (low-educated) co-ethnics suffer long-term disadvantages in human capital acquisition.

4.5.2 Subgroup Analysis

Next, we investigate effect heterogeneity by country of birth, gender, and ethnicity.⁹⁸ We start by investigating whether the negative effects of ethnic concentration on children's outcomes differ between children born abroad and children born in Germany. About 42 percent of the immigrant children in our sample were born abroad, entering Germany through a family reunification scheme. The first two columns of Table 4.4 suggest that the negative enclave effects on German speaking and writing proficiency are roughly 30 percent smaller for children who were born in Germany rather than abroad. As children born in Germany

⁹⁶An increase in the size of the ethnic community by one log-point corresponds to an increase by 172 percent. The difference in average ethnic concentration between low ethnic concentration and high ethnic concentration regions is 1.19 log-points.

⁹⁷Similarly, there is no evidence for a significant effect of ethnic concentration on obtaining an advanced school degree (*Abitur*) (not shown).

⁹⁸Analysis of effect heterogeneity by parents' educational background does not produce significant differences.

start learning the German language already in kindergarten and school, co-ethnic concentration may be less important for them compared to children born abroad who typically start learning the German language at an older age. Still, the ethnic-concentration impact is also significant for guest-worker children who were born in Germany. Furthermore, the smaller negative impact on the host-country language proficiency of children born in Germany does not translate into a smaller disadvantage in terms of dropping out of school (column 3).

The right panel of Table 4.4 investigates effect heterogeneity by child gender. Results indicate that the impact of ethnic concentration on children's language proficiency and educational attainment does not differ significantly between boys and girls, although the negative effect on school dropout may be slightly smaller (in absolute terms) for girls.

Subgroup analyses by ethnicity indicate little heterogeneity (Appendix Table 4A-7). Results suggest that the effect of ethnic concentration on German speaking and writing proficiency and on school dropout does not differ significantly for Greek, Italian, Spanish, Turkish, or Yugoslav guest-worker children. There is some indication, however, that ethnic concentration may have a more negative effect on the probability of obtaining at least an intermediate school degree for Italian and Turkish children, and a more positive one for Greek and Spanish children.

4.6 Mediating Factors

The effect of ethnic enclaves on immigrant children's outcomes may be mediated through numerous different channels, including parents' language skills, inter-ethnic contacts with natives, and economic conditions. Existing studies that rely on administrative data are usually restricted to looking at the enclave effect as a black box. By contrast, the rich SOEP survey data allow us to investigate several potential mediating factors at the child and parent level.

4.6.1 Parental Proficiency in the Host-Country Language

A first candidate for a mediating factor is parents' host-country language skills, as children's human capital accumulation may critically depend on the language proficiency of their parents. In fact, Danzer and Yaman (2016) find a strong negative effect of ethnic enclaves on the language skills of first-generation guest workers in Germany. In the SOEP, adult guest workers (i.e., the parents of our children) report their own German language proficiency in speaking and writing. Using the same random effects specification (without child controls) and the same definitions for language proficiency and ethnic concentration as in our main model, we find an effect of ethnic enclaves on the speaking proficiency of parents of -0.351 (standard error 0.081), but no significant effect on parents' writing proficiency (-0.072 , standard error 0.091).

In a standard descriptive analysis of potential mechanisms, we add different potential

mediating factors as control variables to our main models.⁹⁹ As indicated in column 2 of Table 4.5, parents' German speaking proficiency is significantly positively related to their children's German speaking proficiency.¹⁰⁰ Controlling for parents' German speaking proficiency reduces the effect of ethnic concentration and renders it statistically insignificant, although the negative point estimate remains quite sizeable. However, self-assessed language proficiency is likely measured with error. To circumvent downward bias in the estimated effect of parents' language proficiency, we follow the approach of Dustmann and Van Soest (2002) and exploit the panel dimension of the SOEP to instrument parents' speaking proficiency reported in a given year with their speaking proficiency reported in preceding (lag) and subsequent (lead) years.¹⁰¹

After accounting for random measurement error by instrumenting parents' speaking proficiency with their reported proficiency in the preceding and subsequent years, parents' German speaking proficiency can fully account for the effect of ethnic concentration on children's speaking proficiency. The IV estimate on parents' speaking proficiency (column 3) is three times as large as the OLS estimate, indicating that the latter suffers from substantial attenuation bias.¹⁰² Intriguingly, once the independent-over-time measurement error is accounted for, the point estimate of the effect of ethnic concentration on guest-worker children's German speaking proficiency is reduced to close to zero. This suggests that poor parental host-country language skills in ethnic enclaves are a main driver of the enclave effect on children's host-country language proficiency.

Columns 4 and 5 present equivalent analyses for parents' writing proficiency in German. While parents' German writing skills are also significantly related to their children's German speaking proficiency, controlling for them does not reduce the estimated effect of ethnic concentration by much.

Table 4.6 shows the same analyses for children's writing rather than speaking proficiency.

⁹⁹Causal interpretation of the mechanisms rests on two key assumptions of sequential ignorability (Imai *et al.*, 2010, 2011). The first one is the standard exogeneity assumption: the treatment (ethnic concentration) is assumed to be ignorable given the pre-treatment covariates, that is, it is statistically independent of potential outcomes and potential mediators. The second assumption is that the mediator variable is ignorable given ethnic concentration and pre-treatment covariates. This implies that conditional on ethnic concentration and pre-treatment covariates, the mediator is statistically independent of potential outcomes. This assumption would be violated, for example, if parents' underlying attitude toward education affected both their own language proficiency and their children's language proficiency. If we implement the non-parametric version of the causal mediating analysis proposed by (Imai *et al.*, 2010) using the Stata module "mediation", we obtain the same qualitative results. For example, parents' speaking proficiency accounts for about 80 percent of the impact of ethnic concentration on children's language proficiency, whereas all other mediating factors explain less than 15 percent (not shown).

¹⁰⁰Missing data on the self-reported language proficiency of parents reduce the sample size by 16 percent, but this does not qualitatively affect the estimate of our main effect (see column 1).

¹⁰¹If one of the two instruments is missing, the missing value is imputed with the other instrument. We add an imputation dummy taking on the value of one for observations with imputed values, zero otherwise. The same applies to parents' writing proficiency. Excluding observations with imputed leads and lags of parents' language proficiency yields qualitatively similar results. Note that the IV approach solves the issue of idiosyncratic (i.e., year-specific) measurement error but does not address the issue that immigrants may systematically over- or underrate their host-country language proficiency (Dustmann and Van Soest, 2002).

¹⁰²The F-statistic of the excluded instruments in the first stage is 40.7, and it is at least 10 in all following IV specifications (Appendix Table 4A-8).

We find similar associations of parents' German language proficiency with their children's writing proficiency as we found for children's speaking proficiency. Intriguingly, it is again only parents' *speaking proficiency* (column 3), rather than their writing proficiency (column 5), that reduces the estimated enclave effect on children's *writing proficiency* to close to zero. Thus, it appears that reduced speaking proficiency in the host-country language (and therefore likely reduced speaking of the host-country language at home), rather than limited writing proficiency in the host-country language, is a leading mechanism by which ethnic enclaves inhibit the language proficiency of immigrant children. One potential reason for the greater importance of parents' speaking proficiency is the oral communication of guest workers and their children at home. Furthermore, linguists point out that while speaking requires inter-personal communication, writing is based on structural learning (Blanche and Merino, 1989). The lack of language courses for guest workers may thus have suppressed the development of writing skills more than of speaking skills. As guest-worker parents not only have a lower average level of writing ability, but also lack the competency to self-assess this ability, the self-assessed measure of writing skills may suffer from greater measurement error (Danzer and Yaman, 2016).

4.6.2 Inter-Ethnic Contacts with Natives and Economic Conditions

Limited contacts to German natives may constitute a further mediating factor of the negative effect of co-ethnic concentration on children's host-country language proficiency. Prior research shows that guest workers in Germany who were placed in ethnic enclaves tend to interact less with natives (Danzer and Yaman, 2013), and reduced contact with natives may in turn affect the human capital acquisition of their children. As columns 6 and 7 of Tables 4.5 and 4.6 show, having personal contacts with natives - either measured by whether the child's first friend is German or whether parents regularly receive visits from Germans - is indeed significantly positively associated with the child's German speaking and writing proficiency.¹⁰³ Yet, controlling for the reduced contacts with natives does not significantly change the negative estimate of ethnic enclaves on children's host-country language skills.

Furthermore, differences in economic conditions such as parental unemployment or household income might explain the negative effect of ethnic enclaves on immigrant children's language proficiency. As column 8 of Tables 4.5 and 4.6 shows, parents' unemployment status is significantly associated with their children's host-country language proficiency in the expected way, but controlling for parental unemployment and household income does not affect the estimated effect of ethnic concentration on children's language proficiency at all.

Similar analyses indicate that none of the mediating factors analyzed here can account for the effect of ethnic enclaves on children's schooling outcomes. As shown in Table 4.7, parents' speaking ability is the only analyzed factor that is significantly associated with their

¹⁰³The two respective SOEP questions read as follows: "What is the nationality of the first person befriended?" [German national, other national] (answered by the children) and "Have you received German visitors in your home in the last 12 months?" [yes, no] (answered by the parents).

children's probability to obtain a school degree. Still, controlling for parents' speaking ability does not reduce the estimated effect of ethnic concentration on whether children obtain a school degree.¹⁰⁴

In sum, the negative effect of ethnic enclaves on immigrant children's host-country language proficiency can be fully accounted for by parents' lower host-country speaking proficiency. Parents' writing proficiency explains the negative enclave effect only to a small extent. By contrast, limited contacts to natives and economic factors do not appear to be relevant mediating factors of the negative enclave effects. None of the investigated mediating factors - parents' language skills, inter-ethnic contact, and economic conditions - can account for the detrimental effect of ethnic enclaves on the schooling success of immigrant children.¹⁰⁵

4.7 Robustness

In this section, we show that our results are robust to measuring ethnic concentration by ethnic shares (Section 4.7.1), instrumenting ethnic concentration in 1985 by ethnic concentration in 1975 (Section 4.7.2), measuring ethnic concentration with Census data (Section 4.7.3), measuring ethnic concentration at the county level (Section 4.7.4), and accounting for interview mode (Section 4.7.5). We also investigate return migration (Section 4.7.6), regional migration within Germany (Section 4.7.7), and family size (Section 4.7.8).

4.7.1 Measuring Ethnic Concentration by Ethnic Shares

There is no strong *a priori* argument for any specific functional form of the ethnic concentration measure. At least two different specific measures of ethnic concentration have been used in the literature. In our analyses so far, we followed Edin *et al.* (2003), Damm (2009), and Åslund *et al.* (2011) in using the logarithm of the size of the own ethnicity. In contrast, Chiswick (2009) and Danzer and Yaman (2013, 2016) measure ethnic concentration as the share of the own ethnicity in the total regional population. When using the share of the own ethnicity in the regional population as an alternative measure of ethnic concentration, results on guest-worker children's German speaking and writing proficiency and on school dropout are qualitatively similar to our main models (Table 4.8). Interestingly, the alternative concentration measure also produces significant results on the probability that guest-worker children obtain at least an intermediate school degree. Specifically, the point estimate suggests that a one percentage-point increase in the share of own-ethnics in the regional population reduces the likelihood of obtaining at least an intermediate school degree by 5.1 percent.

¹⁰⁴Similar analyses for obtaining at least an intermediate school degree as the child outcome do not indicate any significant enclave effects; only instrumented parental writing abilities and having a German as the first friend are significantly associated with this outcome (not shown).

¹⁰⁵This result is robust to including all mediating factors in the regression model simultaneously (not shown).

4.7.2 Instrumenting Ethnic Concentration in 1985 by Ethnic Concentration in 1975

As discussed in Section 4.4.2, we do not observe guest workers and their region of residence before 1985. While the balancing tests indicated no evidence of self-selection of guest workers across regions with different ethnic concentrations, the extent of ethnic concentration may have changed between the end of the German guest-worker program in 1973 and the observed ethnic concentration in 1985. To account for potential endogeneity of our main explanatory variable, we can instrument a region's ethnic concentration in 1985 by the region's ethnic concentration in 1975, i.e., towards the end of the German guest-worker recruitment program (Danzer and Yaman, 2013). 1975 is the first year of the SIAB data. This IV model rules out any bias from changes in ethnic concentrations in a given region during the decade before we first observe guest workers' region of residence. For instance, if economic conditions improved between immigration and 1985 in the initial placement region of a guest worker relative to other regions, we may expect an increase in ethnic concentration in this region owing to economically motivated in-migration. In this case, the ethnic concentration observed in 1985 differs from the one at quasi-exogenous placement.

Ethnic concentration in 1975 is a very strong instrument for ethnic concentration in 1985. The F statistic on the excluded instrument in the first stage is 236 in the regressions for language outcomes and 321 in the regressions for schooling outcomes.¹⁰⁶ In line with Schönwälder and Söhn (2009), this suggests that there is strong persistence in the settlement structures of guest workers between the end of the guest-worker program and 1985.

Table 4.9 presents the results of the IV model that uses only that part of the variation in ethnic concentration in 1985 that can be traced back to variation in ethnic concentration that already existed in 1975. For both speaking and writing proficiency, the enclave effect is somewhat stronger when instrumenting 1985 with 1975 ethnic concentration compared to the baseline model. The effect on school dropout does not change and the coefficient for obtaining at least an intermediate school degree remains insignificant. Similarly, all results on mediating factors are very similar in the IV model compared to the baseline model (not shown).

In sum, our baseline estimates are not biased by any change in ethnic concentration that occurred between 1975 and 1985. If anything, restricting the analysis to variation in ethnic concentration that already existed in 1975 leads to slightly larger estimates of the detrimental effect of ethnic enclaves on immigrant children's outcomes.

4.7.3 Measuring Ethnic Concentration with Census Data

Measuring the size of the immigrant population based on a 2 percent random sample of employees like the SIAB can lead to attenuation bias in estimating effects of immigration

¹⁰⁶The first-stage coefficient on the size of the ethnic community in 1975 is 0.85 (p=0.000) in the language sample and 0.84 (p=0.000) in the schooling sample.

measures (Aydemir and Borjas, 2011). To address potential measurement error in our preferred measure of ethnic concentration, we use data from the 1987 German Census, which includes the entire population in Germany. As the 1987 Census data do not allow identifying Spanish citizens, the Census analysis is restricted to the other four ethnicities. For each ethnicity, the correlation coefficient between our preferred 1985 SIAB measure and the 1987 Census measures of the (log) size of the ethnic community exceeds 0.96.

As the odd-numbered columns of Appendix Table 4A-9 indicate, replacing the 1985 SIAB measure of ethnic concentration with the 1987 Census measure yields very similar results to our main specifications. Furthermore, the even-numbered columns show IV models that instrument the 1987 Census measure of ethnic concentration with the concentration of guest workers in the mid-1970s using the SIAB 1975 data. These IV estimates, which simultaneously account for measurement error and changes in regional ethnic concentration after the end of the guest-worker program, are also quite similar to the baseline results. Again, the IV estimates are somewhat larger than the non-instrumented estimates. The results on mediating factors are also unaffected when using the 1987 Census data to compute measures of ethnic concentration, both in the non-instrumented and in the instrumented model (not shown). In sum, we do not find evidence that measurement error in our ethnic concentration measure has a substantial effect on our results.

4.7.4 Measuring Ethnic Concentration at the County Level

Our preferred regional level for measuring ethnic concentration are the *Anpassungsschichten*, as they comprise sufficiently large regions in order to circumvent bias from commuting within regional labor markets. While the much smaller regional entity of counties may more precisely measure immigrant children's exposure to co-ethnics, they also increase concerns of bias due to commuting and moving across county borders. Still, using the 1987 Census, which includes the entire population, we can test for robustness of our results to measuring ethnic concentration at the level of 328 counties rather than 103 *Anpassungsschichten*. However, the guest-worker children observed in the SOEP data live in only 114 different counties, reducing the variation used in the analysis.

When measuring ethnic concentration at the county level, the effects of ethnic concentration on children's speaking and writing proficiency are very similar to the estimates when measuring ethnic concentration at the *Anpassungsschicht* level (Appendix Table 4A-10). By contrast, the effect on obtaining any school degree becomes smaller and loses statistical significance. Besides the fact that Spanish guest-worker children are missing in the analysis, statistical power in the county-level analysis may be impaired by the fact that enclave effects are identified from fewer guest-worker children observed within the same region in the SOEP data. This likely affects in particular the analysis of school dropout, which on average is already rather low (7.1 percent). In fact, incidents of school dropout by guest-worker children are observed in only 42 of the 114 counties with guest-worker children in the SOEP. This suggests that models with county fixed effects exploit only very limited variation in school

dropout.

4.7.5 Accounting for Interview Mode

We also show that immigrants' self-reported language proficiency is not affected by the specific interview mode used in the SOEP, such as oral face-to-face interview or written interview by mail. Therefore, the first two columns of Appendix Table 4A-11 control for the interview mode used when guest-worker children report their levels of German language proficiency. Adding this control does not affect the estimated enclave effects on children's proficiency in speaking or writing German.

4.7.6 Investigating Return Migration

Acquiring host-country language skills and education is an investment decision that may depend on whether immigrants intend to stay in the host country or return to their home country (Dustmann and Glitz, 2011). To account for this possibility, columns 3-6 of Appendix Table 4A-12 include a binary indicator that equals 1 if guest-worker parents see their future in Germany (0 otherwise).¹⁰⁷ Adding this control variable does not affect our baseline estimates. Parents' intention to stay in Germany is positively associated with the children's outcomes, albeit statistically significantly only in the case of obtaining a school degree. We also investigate to what extent immigrant children in our sample actually returned to their home country during the first ten years of our analysis, and in particular, whether return migration was related to ethnic concentration. Following Dustmann (2003), we use the information that individuals provide when leaving the SOEP survey. We construct a binary indicator, $return_{icr}$, which equals 1 if an immigrant child left the SOEP between 1985 and 1995 reporting a move abroad, and 0 otherwise.¹⁰⁸ To predict return migration, we use the same explanatory variables as in our main model, in particular ethnic concentration in 1985:

$$return_{icr} = \sigma_c + \delta_r + \gamma_1 EC_{cr} + C'_{icr} + \epsilon_{icrt}, \quad (4.7.1)$$

During the observation period 1985-1995, 94 immigrant children (8.8 percent of our sample) left the SOEP and moved abroad. However, as indicated in Appendix Table 4A-12, moving abroad is unrelated to ethnic concentration. The result that there is no selective return migration strengthens our identifying assumptions.

¹⁰⁷The respective SOEP question reads as follows: "How long do you want to remain in Germany?" [up to 12 months, a few years, stay in Germany].

¹⁰⁸It is possible that children report other reasons for leaving the SOEP survey and later on return to their home country, which we would not observe.

4.7.7 Investigating Regional Migration within Germany

Our analysis requires the validity of the assumption that there was no systematic sorting of guest workers across regions between their initial placement in Germany and 1985, when we observe them for the first time in the SOEP data. While lack of data prevents us from investigating cross-regional mobility before 1985, we can analyze moving patterns after 1985, a period when overall regional mobility was higher. We find that only 6.9 percent of the 749 immigrant children who remained in the SOEP sample moved across regions between 1985 and 1995. Similarly, the moving rate was modest during the 20-year period between 1985 and 2005 (15.6 percent). This rather low mobility is consistent with Glitz (2014), who shows that regional ethnic concentrations were stable in Germany between 1975 and 2008. Most importantly, ethnic concentration in 1985 does not predict whether guest-worker families moved across regions over the next ten years (Appendix Table 4A-13). In sum, the low mobility and the unsystematic moving patterns with respect to ethnic concentration support our identifying assumption that guest workers did not systematically sort across regions before 1985.

4.7.8 Investigating Family Size

If ethnic concentration affected the number of children in a household, our analysis sample could be subject to selection. For example, ethnic concentration might affect fathers' labor-market success, which in turn might affect their decision to have children or to bring their existing families from their home countries to Germany. However, results in Appendix Table 4A-14 show that this is not the case. Ethnic concentration is neither related to the probability of having children in the household nor to the number of children living in the household. These results are also consistent with the fact that mothers do not have significantly different numbers of children in 1985 in regions with low and high ethnic concentration (Table 4.1). Therefore, our analysis seems to be unaffected by endogenous fertility and family reunification.

4.8 Conclusion

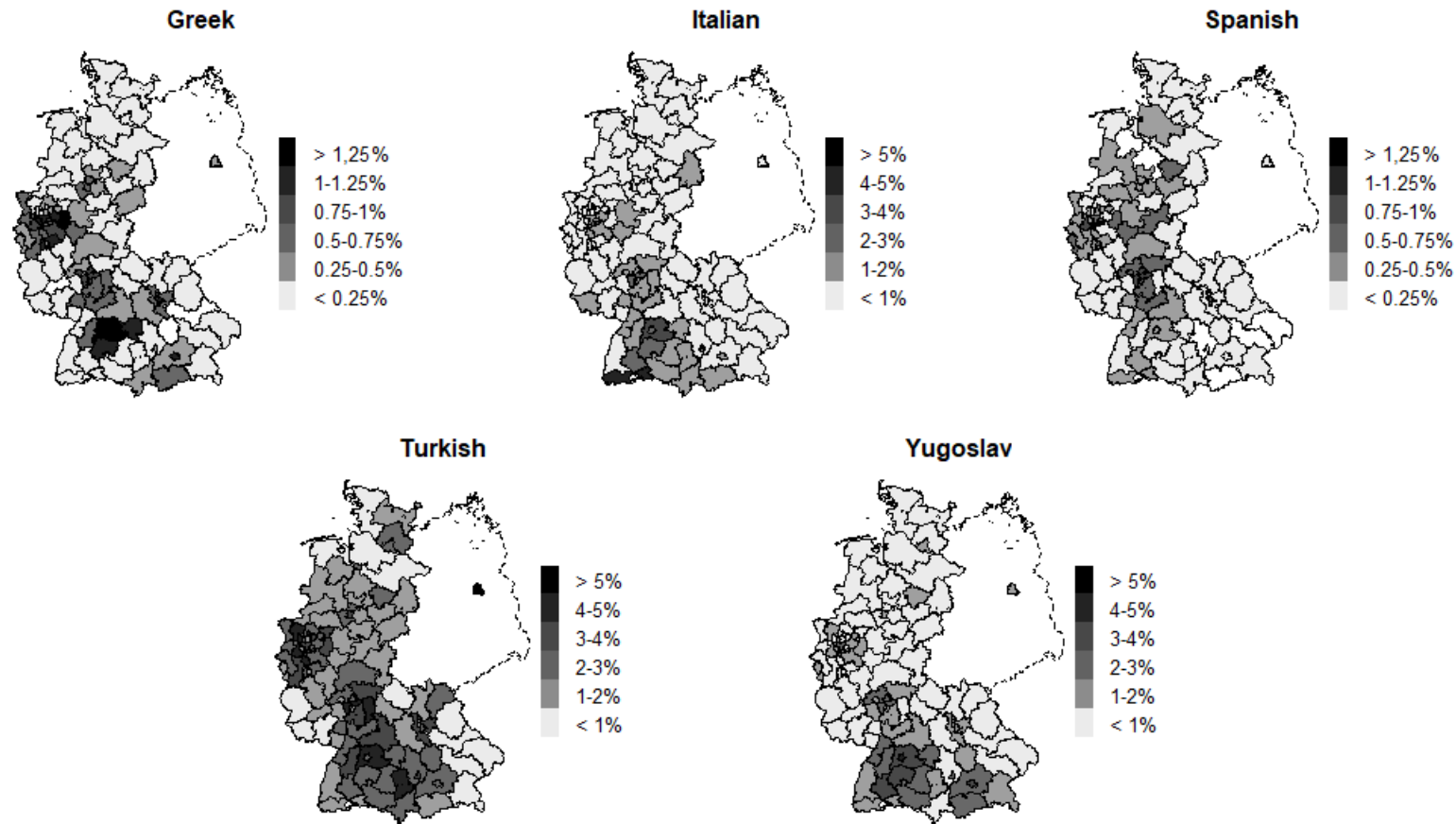
We exploit the quasi-exogenous placement of guest workers across Germany during the 1960s and 1970s to estimate the effect of growing up in ethnic enclaves on the language proficiency and educational outcomes of immigrant children. We find that growing up in regions with higher own-ethnic concentration significantly reduces immigrant children's proficiency in the host-country language and their educational attainment. For schooling outcomes, the effect is concentrated at the lower end of the educational distribution, although there is some indication that more academic school degrees may be affected as well. The enclave effects tend to be larger for immigrant children who were born abroad.

The rich information contained in the German Socio-Economic Panel, most importantly

on parents' host-country language proficiency, allows investigating several factors that might mediate the effect of ethnic concentration on child outcomes. We find that parents' German speaking proficiency completely explains the negative effect of ethnic enclaves on their children's German language proficiency. Parents' writing abilities explain only little, and contacts to natives and parents' economic conditions cannot account for the negative effect of ethnic enclaves on immigrant children's outcomes at all.

These findings imply that even children of immigrants who are well integrated into the labor market may suffer from worse human capital outcomes - host-country language proficiency and educational attainment - when growing up in regions with many, mainly low-educated, immigrants of their own ethnicity. Since the enclave effect on children's language proficiency is completely explained by parents' lower host-country language skills, our findings suggest that host-country language training for adult immigrants might have important positive spillover effects on their children. Language training for adult immigrants would complement current policies in Germany that emphasize language training for immigrant children themselves, which includes compulsory German language tests before starting school. More generally, our results indicate that the long-run cultural and social integration of immigrants, including the next generation, may be more successful when immigrants do not live in ethnic enclaves.

Figure 4.1: Ethnic concentrations across West Germany, 1985



Notes: Share of ethnicity in the total population of the region (*Anpassungsschicht*) of residence, 1985. Source: Institut für Arbeitsmarkt- und Berufsforschung (IAB). Own calculations of ethnic concentrations for 103 *Anpassungsschichten*. Figures based on a historical GIS datafile of the Federal Republic of Germany from the Max Planck Institute for Demographic Research and the Chair for Geodesy and Geoinformatics, University of Rostock (2011) and Bundesamt für Kartographie und Geodäsie (2011).

Table 4.1: Descriptive statistics by degree of ethnic concentration

Variable	Low EC	High EC	Diff.	P-Value	Obs.
<i>Outcomes (Children)</i>					
Speaking proficiency	0.08	-0.07	0.16	0.12	996
Writing proficiency	0.09	-0.08	0.17	0.07	996
Any school degree	0.95	0.91	0.05	0.01	1005
At least intermediate school degree	0.43	0.41	0.02	0.64	1005
<i>Children</i>					
First year of language assessment	1989.29	1988.99	0.30	0.53	996
Male	0.54	0.57	-0.02	0.44	1065
Year of birth	1971.28	1971.03	0.25	0.67	1065
Age at migration	2.55	2.95	-0.40	0.28	1065
Born in Germany	0.58	0.56	0.02	0.55	1065
Greek	0.15	0.15	0.00	0.98	1065
Italian	0.19	0.19	-0.00	0.99	1065
Spanish	0.10	0.09	0.01	0.85	1065
Turkish	0.37	0.38	-0.01	0.92	1065
Yugoslav	0.19	0.18	0.00	0.98	1065
<i>Mothers</i>					
Year of birth	1944.36	1943.90	0.46	0.57	1065
Year of immigration (for the foreign born)	1970.54	1970.29	0.26	0.70	1022
Age at migration (for the foreign born)	26.20	26.49	-0.29	0.70	1022
Born in Germany	0.04	0.04	0.01	0.74	1065
Migrant	0.97	0.98	-0.01	0.59	1065
Education in country of origin					
No schooling	0.23	0.23	-0.00	1.00	1065
Incomplete compulsory schooling	0.41	0.36	0.05	0.38	1065
At least compulsory schooling	0.36	0.41	-0.05	0.40	1065
Years of education	8.29	8.29	-0.00	1.00	1065
Never moved flat since arrival in Germany	0.13	0.13	0.01	0.84	1065
Children	3.67	3.70	-0.02	0.93	1065
Not employed (1984-1986)	0.57	0.51	0.05	0.36	1065
Unemployed (1984-1986)	0.06	0.04	0.02	0.27	1065
<i>Fathers</i>					
Year of birth	1939.78	1940.29	-0.51	0.48	1065
Year of immigration (for the foreign born)	1967.56	1967.71	-0.15	0.76	1056
Age at migration (for the foreign born)	27.73	27.46	0.28	0.64	1056
Born in Germany	0.01	0.01	0.00	0.70	1065
Migrant	0.99	0.99	-0.00	0.70	1065
Education in country of origin					
No schooling	0.09	0.12	-0.03	0.42	1065
Incomplete compulsory schooling	0.27	0.28	-0.01	0.79	1065
At least compulsory schooling	0.63	0.59	0.04	0.46	1065
Years of education	9.15	9.06	0.09	0.66	1065
Never moved flat since arrival in Germany	0.04	0.04	0.00	0.87	1065
Not employed (1984-1986)	0.09	0.08	0.01	0.66	1065
Unemployed (1984-1986)	0.10	0.04	0.06	0.01	1065
Household income (1984-1986)	1700.37	1821.21	-120.84	0.09	1065
<i>For Comparison</i>					
Official unemployment rate (1985)	0.11	0.08	0.03	0.00	1065
Information on language proficiency available	0.70	0.70	-0.00	1.00	1429
Information on school degree available	0.71	0.70	0.02	0.66	1429
Children	500	565			1065

Notes: Variable means by degree of ethnic concentration. Low vs. high EC split at the ethnicity-specific median of the share of ethnic concentration in 1985. P-values refer to two-sided tests with standard errors clustered at region-ethnicity level. Speaking/writing proficiency: first reported self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Household income, not employed, and unemployed refer to three-year means over 1984-1986. Information on language proficiency/school degree available: 1 if information on corresponding outcome is available in the SOEP data in at least one survey year, 0 otherwise. The F-statistic of joint significance of a regression of a high-concentration dummy on individual characteristics is 0.22 for children (p-value 0.992), 0.51 for mothers (0.865), and 3.53 for fathers (0.0005) (which includes household income). Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB), Federal Employment Agency (2017).

Table 4.2: Effect of ethnic concentration on host-country language proficiency

	Speaking proficiency		Writing proficiency	
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.189** (0.083)	-0.185** (0.081)	-0.167** (0.075)	-0.173** (0.069)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes
Child characteristics	No	Yes	No	Yes
Parent characteristics	No	Yes	No	Yes
Observations	4932	4932	4922	4922
R^2 overall	0.180	0.270	0.188	0.293

Notes: Random Effects Model. Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4.3: Effect of ethnic concentration on educational attainment

	Any school degree		At least intermediate school degree	
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.072*** (0.019)	-0.056*** (0.021)	-0.059 (0.052)	0.002 (0.049)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Child characteristics	No	Yes	No	Yes
Parent characteristics	No	Yes	No	Yes
Observations	1005	1005	1005	1005
Adjusted R^2	0.033	0.051	0.086	0.211

Notes: OLS regressions. Dependent variables: Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4.4: Subgroup analysis

	Country of birth				Child gender			
	Speaking proficiency (1)	Writing proficiency (2)	Any school degree (3)	At least intermediate degree (4)	Speaking proficiency (5)	Writing proficiency (6)	Any school degree (7)	At least intermediate degree (8)
Size of ethnic group in 1985	-0.211** (0.082)	-0.200*** (0.070)	-0.060** (0.024)	-0.022 (0.051)	-0.174** (0.082)	-0.154** (0.074)	-0.064*** (0.021)	0.009 (0.051)
Size of ethnic group * born in Germany	0.062** (0.028)	0.063** (0.026)	0.009 (0.018)	0.049** (0.021)				
Size of ethnic group * female					-0.027 (0.031)	-0.049 (0.038)	0.019* (0.010)	-0.016 (0.027)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	n.a.	n.a.	Yes	Yes	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4932	4922	1005	1005	4932	4922	1005	1005
R^2 overall	0.271	0.295			0.271	0.295		
Adjusted R^2			0.050	0.215			0.052	0.210

Notes: Columns 1-2 and 5-6: Random Effects Model. Columns 3-4 and 7-8: OLS regressions. Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4.5: Mediating factors - effect of ethnic concentration on host-country speaking proficiency

	Baseline		Mediating Factors					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1985	-0.178** (0.085)	-0.123 (0.084)	-0.007 (0.092)	-0.173** (0.086)	-0.135 (0.111)	-0.181** (0.083)	-0.169** (0.084)	-0.182** (0.085)
Speaking abilities, parents		0.165*** (0.019)						
Speaking abilities, parents, IV lead + lag			0.531*** (0.106)					
Writing abilities, parents				0.073*** (0.024)				
Writing abilities, parents, IV lead + lag					0.635*** (0.244)			
First friend German						0.226*** (0.059)		
Visits from Germans, parents							0.089** (0.045)	
Unemployed (1984-1986), parents								-0.685** (0.279)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household income	No	No	No	No	No	No	No	Yes
Observations	4125	4125	4125	4125	4125	4125	4125	4125
R^2 overall	0.270	0.291	0.271	0.275	0.209	0.281	0.273	0.276

Notes: Random Effects Model. Columns 3+5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Dependent variable: Speaking proficiency: self-assessed speaking ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4.6: Mediating factors - effect of ethnic concentration on host-country writing proficiency

	Baseline		Mediating Factors					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1985	-0.134*	-0.095	0.033	-0.127	-0.090	-0.138*	-0.126	-0.138*
	(0.079)	(0.078)	(0.095)	(0.080)	(0.105)	(0.077)	(0.079)	(0.078)
Speaking abilities, parents		0.121***						
		(0.021)						
Speaking abilities, parents, IV lead + lag			0.574***					
			(0.120)					
Writing abilities, parents				0.103***				
				(0.024)				
Writing abilities, parents, IV lead + lag					0.680**			
					(0.289)			
First friend German						0.252***		
						(0.054)		
Visits from Germans, parents							0.082*	
							(0.046)	
Unemployed (1984-1986), parents								-0.545*
								(0.281)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household income	No	No	No	No	No	No	No	Yes
Observations	4120	4120	4120	4120	4120	4120	4120	4120
R^2 overall	0.296	0.310	0.269	0.303	0.226	0.308	0.299	0.299

Notes: Random Effects Model. Columns 3+5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Dependent variable: Writing proficiency: self-assessed writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4.7: Mediating factors - effect of ethnic concentration on obtaining any school degree

	Baseline		Mediating Factors					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1985	-0.062*** (0.022)	-0.062*** (0.022)	-0.065*** (0.020)	-0.062*** (0.022)	-0.064*** (0.020)	-0.056** (0.022)	-0.062*** (0.022)	-0.065*** (0.022)
Speaking abilities, parents		0.002 (0.010)						
Speaking abilities, parents, IV lead + lag			0.042** (0.019)					
Writing abilities, parents				0.002 (0.012)				
Writing abilities, parents, IV lead + lag					0.007 (0.022)			
First friend German						0.027 (0.019)		
Visits from Germans, parents							0.007 (0.032)	
Unemployed (1984-1986), parents								-0.072 (0.095)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household income	No	No	No	No	No	No	No	Yes
Observations	943	943	943	943	943	943	943	943
Adjusted R^2	0.057	0.056	0.045	0.056	0.056	0.066	0.055	0.059

Notes: OLS regressions. Columns 3+5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Dependent variable: Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4.8: Measuring ethnic concentration by share of own ethnicity in regional population

	Speaking proficiency (1)	Writing proficiency (2)	Any school degree (3)	Intermediate school degree (4)
Share of own ethnicity in 1985	-0.080** (0.034)	-0.080*** (0.029)	-0.025** (0.011)	-0.051** (0.020)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes
Observations	4932	4922	1005	1005
R^2 overall	0.269	0.292		
Adjusted R^2			0.051	0.216

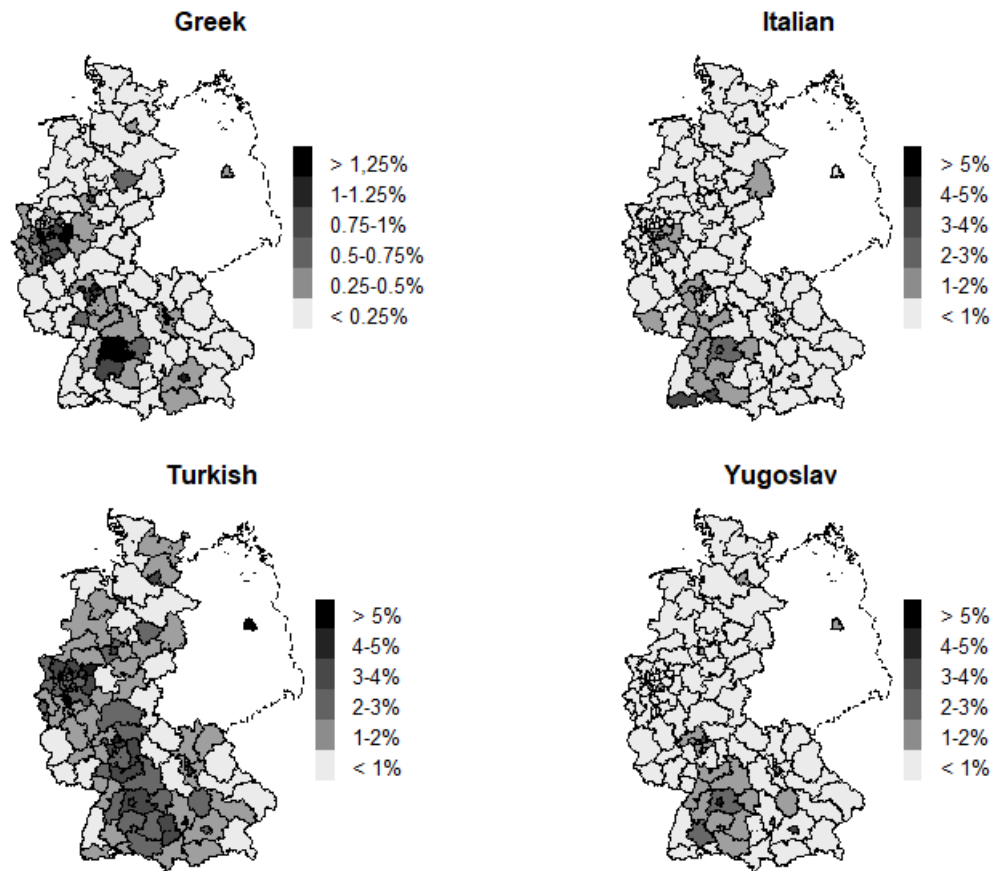
Notes: Columns 1-2: Random Effects Model. Columns 3-4: OLS regressions. Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Share of own ethnicity in 1985: share of own ethnicity in the population of the region (*Anpassungsschicht*) of residence, 1985. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4.9: Instrumental-variable estimates using ethnic concentration in 1975

	Speaking proficiency (1)	Writing proficiency (2)	Any school degree (3)	Intermediate school degree (4)
Size of ethnic group in 1985	-0.234** (0.103)	-0.183** (0.075)	-0.056*** (0.019)	0.032 (0.049)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes
Observations	4932	4922	1005	1005
R^2 overall	0.269	0.293		
Adjusted R^2			0.051	0.210
First-stage F-statistic	236.63	236.29	321.309	321.309

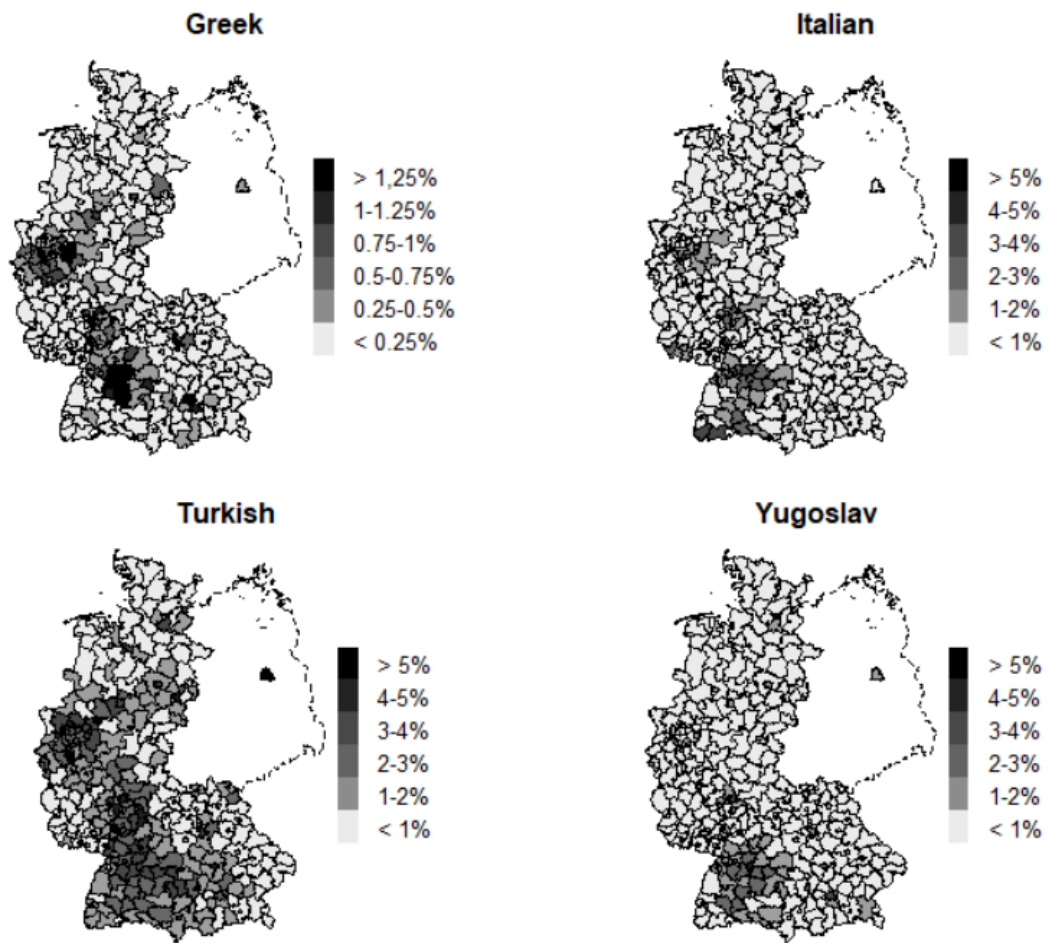
Notes: Columns 1-2: Random Effects Model. Columns 3-4: OLS regressions. Size of ethnic group in 1985 is instrumented by size of ethnic group in 1975 (both variables in logs). Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Figure 4A-1: Ethnic concentrations across West Germany: census 1987



Notes: Share of ethnicity in the total population of the region (*Anpassungsschicht*) of residence, 1987. Source: German Census 1987. Own calculations of ethnic concentrations for 103 Anpassungsschichten. Figures based on a historical GIS datafile of the Federal Republic of Germany from the Max Planck Institute for Demographic Research and the Chair for Geodesy and Geoinformatics, University of Rostock (2011) and Bundesamt für Kartographie und Geodäsie (2011).

Figure 4A-2: County-level ethnic concentrations across West Germany: census 1987



Notes: Share of ethnicity in the total population of the county of residence, 1987. Source: German Census 1987. Own calculations of ethnic concentrations for 328 counties. Figures based on a historical GIS datafile of the Federal Republic of Germany from the Max Planck Institute for Demographic Research and the Chair for Geodesy and Geoinformatics, University of Rostock (2011) and Bundesamt für Kartographie und Geodäsie (2011).

Table 4A-1: Immigrants from guest-worker and other countries

Variable	Guest-worker countries	Other countries	Diff.	P-Value	Obs.
Age	37.97	42.78	-4.80	0.00	2733
Male	0.55	0.46	0.08	0.01	2733
Year of immigration	1970.37	1967.78	2.59	0.00	2615
Age at migration	23.33	25.13	-1.80	0.04	2615
Years of education	9.07	11.27	-2.20	0.00	2709
Any school degree	0.64	0.97	-0.33	0.00	2684
More than 12 years of education	0.07	0.27	-0.20	0.00	2709
Married	0.81	0.75	0.06	0.02	2608
Children (female individuals only)	2.45	2.03	0.43	0.00	1264
Unemployed	0.08	0.06	0.03	0.09	2657
Not employed	0.31	0.44	-0.13	0.00	2733
Household income	1632.56	1649.50	-16.94	0.74	2673
Obs.	2466	267			2733

Notes: First-generation immigrants in survey year 1985. P-values are estimated using robust standard errors. Data source: German Socio-Economic Panel (SOEP).

Table 4A-2: Individual-level variables

Variable	Description
<i>Outcomes (Children)</i>	
Speaking proficiency	Generated from self-assessed speaking ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5), normalized to have mean 0 and standard deviation 1, <i>Random Effects Model</i> : each observation is a child-year observation based on self-reported language proficiency in the years 1984-1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005.
Writing proficiency	Generated from self-assessed writing ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5), normalized to have mean 0 and standard deviation 1, <i>Random Effects Model</i> : each observation is a child-year observation based on self-reported language proficiency in the years 1984-1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005.
Any school degree	Binary indicator that equals 1 if individual obtained any type of school degree and 0 otherwise. Based on most recent available educational level. If the most recent available information is <i>dropout, no school degree</i> or <i>no school degree yet</i> , we checked for school-leaving degrees reported in previous years. In nine cases, we adjusted the educational attainment variables based on previously reported school-leaving degrees.
At least intermediate school degree	Binary indicator that equals 1 if individual obtained at least an intermediate school degree and 0 otherwise. Based on the most recent available educational level. If the most recent available information is <i>dropout, no school degree</i> or <i>no school degree yet</i> , we checked for school-leaving degrees reported in previous years. In nine cases, we adjusted the educational attainment variables based on previously reported school-leaving degrees.
<i>Demographics of Children</i>	
Ethnicity dummies (Greek, Italian, Spanish, Turkish, Yugoslav)	Binary indicators primarily based on children's first citizenship (94.2 %). In case of German citizenship or no available citizenship information, these indicators are based on parents' joint nationality (1.0 %) or on children's country of origin (0.3 %). If children's ethnicity is not yet available and one parent is intermarried to a German or to a foreigner with different or missing nationality, we use the citizenship of the parent with the guest-worker background as a proxy for children's ethnicity (4.4 %). Regarding rare cases, if children's ethnicity is not available and both parents are migrants but their country of origin differs, we use the mother's nationality or country of origin. In the few cases of remaining missing children's ethnicities, we base children's ethnicity on father's country of origin or nationality. For more than 98.5 % of the children in our analysis sample, children's ethnicity corresponds to the father's country of origin. We could assign an ethnicity to all children in our sample.
Age at migration	Age at migration (in years). If a child is born in Germany, the variable is coded as zero.
<i>Demographics of Parents</i>	

Continued on next page

Table 4A-2 (*Continued*)

Variable	Description
Migrant	Binary indicator that equals 1 if individual has a migrant background and 0 otherwise. Based on variable "migback" of the SOEP Person-related meta-dataset.
Variables on education in country of origin	Three binary indicators for <i>No schooling</i> , <i>Incomplete compulsory schooling</i> , and <i>At least compulsory schooling</i> , based on survey question "Obtained School Degree Outside Germany" in survey year 1985.
Years of education	Amount of education or training (in years), generated variable by SOEP. Based on survey year 1985.
Never moved flat since arrival in Germany	Binary indicator equal to 1 if the individual's year of immigration is either equal to the year in which the household moved into the dwelling or is later than the year in which the household moved; 0 otherwise. Based on survey year 1985.
Children	Number of mother's children. Based on variable "sumkids" from the SOEP Birth Biography of Female Respondents.
Household income (1984-1986)	Mean of parents' adjusted household income over three years. Based on survey years 1984-1986.
Not employed (1984-1986)	Mean of an indicator for being not employed during the survey years 1984-1986.
Unemployed (1984-1986)	Mean of an unemployment dummy during the survey years 1984-1986.
<i>Mediating Factors</i>	
Speaking abilities, parents	Parents' speaking ability, generated from self-assessed speaking ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported speaking ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available speaking proficiency, largely based on the second survey year of the SOEP in 1985 (99 %).
Speaking abilities, parents, IV lead + lag	Parents' speaking ability is instrumented with the corresponding lead and lag to reduce measurement error (see Dustmann and van Soest, 2002), missing leads (lags) of parents' current language proficiency are imputed with available lags (leads), all regressions include imputation dummies for imputed leads or lags of parents' language proficiency, generated from self-assessed speaking ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported speaking ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available speaking proficiency, largely based on the second survey year of the SOEP in 1985 (99 %).

Continued on next page

Table 4A-2 (*Continued*)

Variable	Description
Writing abilities, parents	Parents' writing ability, generated from self-assessed writing ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported writing ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available writing proficiency, largely based on the second survey year of the SOEP in 1985 (99 %).
Writing abilities, parents, IV lead + lag	Parents' writing ability is instrumented with the corresponding lead and lag to reduce measurement error (see Dustmann and van Soest, 2001), missing leads (lags) of parents' current language proficiency are imputed with available lags (leads), all regressions include imputation dummies for imputed leads or lags of parents' language proficiency, generated from self-assessed writing ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported writing ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available writing proficiency, largely based on the second survey year of the SOEP in 1985 (99 %).
First friend German	Binary indicator equal to 1 if a child's first friend is German. Based on the first available variable on the nationality of the first-named friend. Based on survey years 1988, 1990, 1992, 1994, 1996, 2001, 2006, and 2011.
Visits from Germans, parents	Average of the following variable of mother and father: a binary indicator equal to 1 if mother or father received visits from Germans at home during the previous 12 months. For language proficiency estimations: refers to year of children's reported language proficiency. For educational attainment estimations: refers to average of the years 1985 and 1986.
Unemployed (1984-1986), parents	Average of the following variable of mother and father: Mean of an unemployment dummy over three years. Based on survey years 1984-1986.
Household income (1984-1986), parents	Mean of parents' household income over three years (in logs). Based on survey years 1984-1986.
<i>Robustness Checks</i>	
Stay in Germany, parents	Average of the following variable of mother and father: A binary dummy indicating the intent to stay in Germany. Based on the following answer categories: "I intend to stay in Germany forever" (= 1), "I intend to stay in Germany for several years" (= 0), "I intend to leave Germany within 12 months" (= 0). Based on survey year 1985.

Continued on next page

Table 4A-2 (*Continued*)

Variable	Description
Interview mode	Dummies based on a variable indicating the interview mode in the corresponding survey years of self-reported language proficiency (years 1984-1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005). Based on the following answer categories: "Oral Interview", "Written Questionnaire Interviewer", "Mixed Type", "Written Questionnaire No Interviewer", "Oral And Written", "Proxy", "Third Person Present", "No Third Person Present", "Computer Assisted Personal Interviewing", "Telephone Assistance", "Written, By Mail", and "Telephone Interview".

Notes: Source (for all variables): German Socio-Economic Panel (SOEP).

Table 4A-3: Regional variables

Variable	Description
Size of ethnic group in 1985	Log size of ethnic community (individuals of same ethnicity) in region (<i>Anpassungsschicht</i>) of residence, 1985. Log of 1 used in rare case of zero co-ethnics in the region; all regressions include a corresponding imputation dummy. See variable <i>Region of residence</i> for details on the assignment of children to 1985 regions. Based on a two percent sample of the German employee population (incl. recipients of social transfers) from the Institut für Arbeitsmarkt- und Berufsforschung (IAB).
Size of ethnic group in 1987	Log size of ethnic community (individuals of same ethnicity) in region of residence, based on German Census 1987, regional level: <i>Anpassungsschicht</i> (Table A5) or county (Table A6). Measure not available for immigrant children with Spanish ethnicity.
Official unemployment rate 1985	Unemployment rate in the year 1985, regional level: <i>Anpassungsschicht</i> , based on county-level data from Federal Employment Agency (2017).
Region of residence (used to construct ethnic concentration measures)	The region of residence (<i>Anpassungsschicht</i> or county) is primarily based on children's 1985 region of residence (94.7 %). If children's household IDs for the year 1985 are not available, the ethnic concentration measures are based on parents' 1985 region of residence for the following scenarios: children were born after 1985 (2.1 %), children had the same household ID as their parents in 1984 (1.5 %), children migrated to Germany after 1985 (0.2 %), or children joined the SOEP in a later wave than 1985 for other reasons (1.6 %). All children in our sample could be assigned to a 1985 region.

Table 4A-4: Ethnic concentration by ethnicity

	Mean	SD	Min	Max
Greek	0.95	0.63	0.06	2.25
Italian	1.51	0.95	0.16	4.03
Spanish	0.39	0.24	0.00	0.97
Turkish	3.05	1.20	0.29	6.19
Yugoslav	2.06	1.21	0.48	4.14
Total	2.00	1.39	0.00	6.19

Notes: Share of ethnicity in the total population of the region (*Anpassungsschicht*) of residence, 1985 (based on full sample of guest-worker children in SOEP). Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4A-5: Occupation classes by ethnicity

	Greeks	Italians	Spaniards	Turks	Yugoslavs
Agricultural	0.00	0.03	0.02	0.01	0.00
Manufacturing	0.68	0.62	0.58	0.64	0.64
Mining and Minerals	0.00	0.01	0.00	0.05	0.02
Services	0.17	0.16	0.21	0.09	0.18
Technical Occupation	0.00	0.01	0.03	0.01	0.03
Others	0.00	0.01	0.00	0.00	0.00
n.a.	0.15	0.16	0.17	0.19	0.13
Sum	1	1	1	1	1

Notes: Shares of different occupation classes across ethnicities. Information on the type of job is based on the *StaBuA 1992 Job Classification* by the German Federal Statistical Office (StaBuA) and was constructed using the raw occupation titles provided by the survey respondents. We assigned each minor occupation class into one of the main six classes using the official guidelines. The table is based on the immigrant fathers from our sample (n=1065) and refers to the year 1985. Data sources: German Socio-Economic Panel (SOEP).

Table 4A-6: Balancing test using continuous share of ethnic concentration

Variable	Coefficient	Standard Error	P-Value	Obs.
<i>Outcomes (Children)</i>				
Speaking proficiency	-0.15	0.04	0.00	996
Writing proficiency	-0.11	0.04	0.00	996
Any school degree	-0.03	0.01	0.00	1005
At least intermediate school degree	-0.06	0.02	0.00	1005
<i>Children</i>				
First year of language assessment	-0.02	0.24	0.92	996
Male	0.02	0.02	0.26	1065
Year of birth	-0.08	0.33	0.80	1065
Age at migration	0.23	0.24	0.34	1065
Born in Germany	-0.01	0.03	0.76	1065
<i>Mothers</i>				
Year of birth	-0.24	0.41	0.56	1065
Year of immigration for the foreign born	-0.33	0.30	0.27	1022
Age at migration	-0.19	0.48	0.70	1022
Born in Germany	-0.02	0.01	0.20	1065
Migrant	0.01	0.01	0.18	1065
No schooling	-0.02	0.03	0.59	1065
Incomplete compulsory schooling	0.01	0.03	0.86	1065
At least compulsory schooling	0.01	0.04	0.80	1065
Years of education	0.06	0.09	0.48	1065
Never moved flat since arrival in Germany	-0.01	0.02	0.39	1065
Children	0.05	0.13	0.72	1065
Not employed (1984-1986)	0.00	0.04	0.91	1065
Unemployed (1984-1986)	-0.00	0.01	0.75	1065
<i>Fathers</i>				
Year of birth	0.37	0.48	0.45	1065
Year of immigration for the foreign born	0.11	0.22	0.63	1056
Age at migration	-0.27	0.36	0.46	1056
Born in Germany	0.00	0.00	0.53	1065
Migrant	-0.00	0.00	0.53	1065
No schooling	0.05	0.01	0.00	1065
Incomplete compulsory schooling	0.00	0.04	0.92	1065
At least compulsory schooling	-0.05	0.04	0.15	1065
Years of education	-0.16	0.14	0.23	1065
Never moved flat since arrival in Germany	-0.00	0.01	0.76	1065
Not employed (1984-1986)	0.02	0.01	0.10	1065
Unemployed (1984-1986)	-0.01	0.01	0.36	1065
Household income (1984-1986)	-15.04	36.89	0.68	1065
<i>For Comparison</i>				
Information on language proficiency available	0.00	0.01	0.86	1065
Information on school degree available	-0.00	0.01	0.71	1065

Notes: Each row reports the results of a separate regression where the respective variable is regressed on the continuous share of ethnic concentration. All regressions include region and ethnicity fixed effects. Speaking/writing proficiency: first reported self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Household income, not employed, and unemployed refer to three-year means over 1984-1986. Information on language proficiency/school degree available: 1 if information on corresponding outcome is available in the SOEP data in at least one survey year, 0 otherwise. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB), Federal Employment Agency (2017).

Table 4A-7: Subgroup analysis by ethnicity

Interacted Ethnicity:	Greek (1)	Italian (2)	Spanish (3)	Turkish (4)	Yugoslav (5)
<i>Panel A: Speaking proficiency</i>					
Size of ethnic group in 1985	-0.176** (0.079)	-0.184** (0.081)	-0.200** (0.082)	-0.204** (0.085)	-0.167** (0.077)
Size of ethnic group * ethnicity	-0.032 (0.063)	-0.005 (0.077)	0.122 (0.113)	0.109 (0.110)	-0.074 (0.081)
<i>Panel B: Writing proficiency</i>					
Size of ethnic group in 1985	-0.177** (0.070)	-0.173** (0.073)	-0.187*** (0.072)	-0.187** (0.075)	-0.151** (0.063)
Size of ethnic group * ethnicity	0.015 (0.063)	-0.002 (0.084)	0.108 (0.114)	0.076 (0.088)	-0.094 (0.083)
<i>Panel C: Any school degree</i>					
Size of ethnic group in 1985	-0.051** (0.022)	-0.055*** (0.021)	-0.059*** (0.021)	-0.055** (0.022)	-0.060*** (0.021)
Size of ethnic group * ethnicity	-0.018 (0.024)	-0.007 (0.025)	0.030 (0.026)	-0.002 (0.020)	0.016 (0.026)
<i>Panel D: At least intermediate school degree</i>					
Size of ethnic group in 1985	-0.048 (0.044)	0.021 (0.048)	-0.017 (0.050)	0.023 (0.046)	0.011 (0.054)
Size of ethnic group * ethnicity	0.177*** (0.031)	-0.115*** (0.042)	0.157** (0.076)	-0.125*** (0.042)	-0.035 (0.041)
Region fixed effects	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variables: Panel A: Speaking proficiency: self-assessed speaking ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Panel B: Writing proficiency: self-assessed writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Panel C: Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. Panel D: At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Panels A and B additionally include dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4A-8: First-stage results using parents' leads and lags of language proficiency as instruments

Second-stage outcome: Reference estimation:	Speaking proficiency		Writing proficiency		Any School Degree	
	Table 5, Col 3 (1)	Table 5, Col 5 (2)	Table 6, Col 3 (3)	Table 6, Col 5 (4)	Table 7, Col 3 (5)	Table 7, Col 5 (6)
Speaking abilities, parents, lead	0.174*** (0.019)		0.153*** (0.019)		0.156*** (0.061)	
Speaking abilities, parents, lag	0.138*** (0.026)		0.116*** (0.027)		0.494*** (0.065)	
Writing abilities, parents, lead		0.099*** (0.020)		0.087*** (0.021)		0.173*** (0.053)
Writing abilities, parents, lag		0.049* (0.028)		0.035 (0.029)		0.455*** (0.069)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Size of ethnic group in 1985	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4125	4125	4120	4120	943	943
First-stage F-statistic	40.74	12.20	29.42	9.98	51.66	43.53

Notes: First-stage results using lead and lag of parents' speaking/writing proficiency as instruments. Columns 1-4: Random Effects Model. Columns 5-6: OLS regressions. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4A-9: Ethnic concentration measured in 1987 census

	Speaking proficiency		Writing proficiency		Any school degree		Intermediate school degree	
	Baseline (1)	IV (2)	Baseline (3)	IV (4)	Baseline (5)	IV (6)	Baseline (7)	IV (8)
Size of ethnic group in 1987	-0.238** (0.094)	-0.265** (0.111)	-0.167** (0.083)	-0.193** (0.084)	-0.063** (0.024)	-0.071*** (0.022)	0.000 (0.049)	0.004 (0.049)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	n.a.	n.a.	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4523	4523	4514	4514	907	907	907	907
R^2 overall	0.272	0.271	0.301	0.300				
Adjusted R^2					0.038	0.038	0.229	0.229
First-stage F-statistic		278.58		278.58		384.913		384.913

Notes: Columns 1-4: Random Effects Model. Columns 5-8: OLS regressions. Columns 2, 4, 6, and 8: Size of ethnic group in 1987 is instrumented by size of ethnic group in 1975 (both variables in logs). Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1987: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, German Census 1987. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors, clustered at the region-ethnicity level, in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Data sources: German Socio-Economic Panel (SOEP), German Census 1987.

Table 4A-10: Ethnic concentration measured at county level (1987 census)

	Speaking proficiency (1)	Writing proficiency (2)	Any school degree (3)	Intermediate school degree (4)
Size of ethnic group in 1987	-0.244** (0.097)	-0.209** (0.087)	-0.021 (0.035)	-0.003 (0.054)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes
Observations	4523	4514	907	907
R^2 overall	0.305	0.333		
Adjusted R^2			0.033	0.234

Notes: Columns 1-2: Random Effects Model. Columns 3-4: OLS regressions. Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1987: log size of ethnic community (individuals of same ethnicity) in county of residence, German Census 1987. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors, clustered at the county-ethnicity level, in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), German Census 1987.

Table 4A-11: Controlling for interview mode and return intention

	Speaking proficiency	Writing proficiency	Speaking proficiency	Writing proficiency	Any school degree	At least intermediate degree
	(1)	(2)	(3)	(4)	(5)	(6)
Size of ethnic group in 1985	-0.171** (0.083)	-0.161** (0.068)	-0.173** (0.080)	-0.162** (0.069)	-0.054** (0.021)	0.006 (0.049)
Stay in Germany, parents			0.105 (0.080)	0.085 (0.078)	0.036* (0.020)	0.016 (0.056)
Interview mode	Yes	Yes	No	No	No	No
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4932	4922	4932	4922	1005	1005
R^2 overall	0.305	0.319	0.272	0.295		
Adjusted R^2					0.052	0.212

Notes: Columns 1-4: Random Effects Model. Columns 5-6: OLS regressions. Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Stay in Germany, parents: a binary dummy indicating the intent to stay in Germany (average of the variable of mother and father). Interview mode: dummies for different types of interview method such as "Oral Interview" and "Written, By Mail". Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4A-12: Predicting return migration by immigrant children between 1985 and 1995

	(1)	(2)
Size of ethnic group in 1985	0.005 (0.035)	0.003 (0.035)
Male		-0.054*** (0.018)
Region fixed effects	Yes	Yes
Ethnicity fixed effects	Yes	Yes
Child characteristics	No	Yes
Observations	1065	1065
Adjusted R^2	0.023	0.082

Notes: OLS regressions. Dependent variable: Moved abroad: 1 if individual left the SOEP survey between 1985 and 1995 and the corresponding exit reason is "moved abroad", 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Child characteristics: dummies for birth cohort (2-year intervals) and age at migration. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4A-13: Predicting regional migration within Germany by immigrant children between 1985 and 1995

	(1)	(2)
Size of ethnic group in 1985	-0.046 (0.037)	-0.038 (0.038)
Male		-0.013 (0.021)
Region fixed effects	Yes	Yes
Ethnicity fixed effects	Yes	Yes
Child Characteristics	No	Yes
Observations	749	749
Adjusted R^2	0.098	0.104

Notes: OLS regressions. Dependent variable: Moved region: 1 if individual moved across regions (*Anpassungsschichten*) between 1985 and 1995, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Child characteristics: dummies for birth cohort (2-year intervals) and age at migration. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table 4A-14: Predicting the presence of children in the household

	Children in household		Number of children	
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.013 (0.059)	-0.016 (0.052)	0.034 (0.142)	0.038 (0.126)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Individual characteristics	No	Yes	No	Yes
Observations	780	780	780	780
Adjusted R^2	0.037	0.209	0.074	0.227

Notes: OLS regressions. Dependent variable: Children in household: 1 if children in the household in 1985, 0 otherwise. Number of children: number of children living in the household in 1985. Sample: male first-generation immigrations who migrated to Germany during the German guest-worker program. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Individual characteristics: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), and years of education in 1985. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Bibliography

- Acemoglu, D. (2002). Directed Technical Change. *The Review of Economic Studies*, 69 (4), 781–809.
- (2007). Equilibrium Bias of Technology. *Econometrica*, 75 (5), 1371–1409.
- (2010). When does Labor Scarcity Encourage Innovation? *Journal of Political Economy*, 118 (6), 1037–1078.
- , LeLarge, C. and Restrepo, P. (2020). Competing with Robots: Firm-Level Evidence from France. *American Economic Association Papers and Proceedings* (forthcoming).
- and Restrepo, P. (2019). Demographics and automation.
- Albert, C. and Monras, J. (2017). Immigrants’ Residential Choices and their Consequences.
- Almeida, P. and Kogut, B. (1999). Localization of Knowledge and the Mobility of Engineers in regional Networks. *Management Science*, 45 (7), 905–917.
- Åslund, O., Edin, P.-A., Fredriksson, P. and Grönqvist, H. (2011). Peers, Neighborhoods, and Immigrant Student Achievement: Evidence from a Placement Policy. *American Economic Journal: Applied Economics*, 3 (2), 67–95.
- Audretsch, D. B. and Feldman, M. P. (1996). R&D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, 86 (3), 630–640.
- Aydemir, A. and Borjas, G. J. (2011). Attenuation Bias in Measuring the Wage Impact of Immigration. *Journal of Labor Economics*, 29 (1), 69–112.
- Bade, K. J. (1990). Aussiedler-Rückwanderer über Generationen hinweg. In K. J. Bade (ed.), *Neue Heimat im Westen. Vertriebene. Flüchtlinge. Aussiedler*, Münster: Westfaelischer Heimatbund.
- Battisti, M., Peri, G. and Romiti, A. (2016). Dynamic Effects of Co-Ethnic Networks on Immigrants’ Economic Success, National Bureau of Economic Research No. 22389.
- Bernard, A. B., Moxnes, A. and Saito, Y. U. (2019). Production Networks, Geography, and Firm Performance. *Journal of Political Economy*, 127 (2), 639–688.
- Blanche, P. and Merino, B. J. (1989). Self-assessment of Foreign-Language Skills: Implications for Teachers and Researchers. *Language learning*, 39 (3), 313–338.
- Borjas, G. J. (1992). Ethnic capital and intergenerational mobility. *The Quarterly journal of economics*, 107 (1), 123–150.
- (1994). The Economics of Immigration. *Journal of economic literature*, 32 (4), 1667–1717.
- Boustan, L. P., Fishback, P. V. and Kantor, S. (2010). The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression. *Journal of Labor Economics*, 28 (4), 719–746.
- Bratu, C. (2018). Firm-and Individual-Level Responses to Labor Immigration.
- Buchheim, L., Watzinger, M. and Wilhelm, M. (2019). Job Creation in Tight and Slack Labor Markets. *Journal of Monetary Economics*.

- Bundesamt für Kartographie und Geodäsie (2011). VG 2500 Verwaltungsgebiete (Ebenen) 1:2.500.000. Stand 01.01.2009.
- Bundesverwaltungsamt (2019). (Spaet-)Aussiedler und ihre Angehörigen Zeitreihe 1950 - 2019.
- Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90 (3), 414–427.
- Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *ILR Review*, 43 (2), 245–257.
- , Kramarz, F. and Lemieux, T. (1996). Changes in the relative Structure of Wages and Employment: A Comparison of the United States, Canada, and France, National Bureau of Economic Research No. w5487.
- Chiswick, B. (2009). The Economics of Language for Immigrants: An Introduction and Overview. *The Education of Language Minority Immigrants in the United States*, 72–91.
- Chiswick, B. R. and Miller, P. W. (2007). *The Economics of Language: International Analyses*. Routledge.
- and — (2015). International Migration and the Economics of Language. In *Handbook of the economics of international migration*, Vol. 1, Elsevier, 211–269.
- Clemens, M. A., Lewis, E. G. and Postel, H. M. (2018). Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion. *American Economic Review*, 108 (6), 1468–87.
- Constant, A. F., Schüller, S. and Zimmermann, K. F. (2013). Ethnic Spatial Dispersion and Immigrant Identity, IZA Working Paper No. 7868.
- Correia, S., Guimaraes, P. and Zylkin, T. (2019). ppmlhdf: Fast Poisson Estimation with High-Dimensional Fixed Effects.
- Cortes, K. E. (2006). The Effects of Age at Arrival and Enclave Schools on the Academic Performance of Immigrant Children. *Economics of Education review*, 25 (2), 121–132.
- Dahmen, J. and Kozłowicz, W. (1963). *Ausländische Arbeitnehmer in der Bundesrepublik. Sozialpolitik in Deutschland*. Stuttgart: W. Kohlhammer Verlag.
- Damm, A. P. (2009). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labor Economics*, 27 (2), 281–314.
- D’Amuri, F., Ottaviano, G. I. and Peri, G. (2010). The Labor Market Impact of Immigration in Western Germany in the 1990s. *European Economic Review*, 54 (4), 550–570.
- Danzer, A. M. and Dietz, B. (2014). Labour Migration from Eastern Europe and the EU’s Quest for Talents. *JCMS: Journal of Common Market Studies*, 52 (2), 183–199.
- , Feuerbaum, C. and Gaessler, F. (2020). Labor Supply and Automation Innovation.
- , —, Piopiunik, M. and Woessmann, L. (2018). Growing Up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children. CESifo Working Paper No. 7097.
- and Yaman, F. (2013). Do Ethnic Enclaves impede Immigrants’ Integration? Evidence from a Quasi-Experimental Social-Interaction Approach. *Review of International Economics*, 21 (2), 311–325.
- and — (2016). Ethnic Concentration and Language Fluency of Immigrants: Evidence from the Guest-Worker Placement in Germany. *Journal of Economic Behavior & Organization*, 131, 151–165.
- Dechezleprêtre, A., Hémous, D., Olsen, M. and Zanella, C. (2019). Automating Labor: Evidence from Firm-Level Patent Data, Available at SSRN 3508783.
- Dietz, B. (2006). Aussiedler in Germany: From Smooth Adaptation to Tough Integration. In L. Lucassen, D. Feldmann and J. Oltmer (eds.), *Paths of Integration. Migrants in Western Europe (1880-2004)*, Amsterdam: Amsterdam University Press, 116–136.

- Dimmock, S. G., Huang, J. and Weisbenner, S. J. (2019). Give Me Your Tired, Your Poor, Your High-Skilled Labor: H-1B Lottery Outcomes and Entrepreneurial Success.
- Dohse, K. (1985). *Ausländische Arbeiter und bürgerlicher Staat: Genese und Funktion von staatlicher Ausländerpolitik und Ausländerrecht: vom Kaiserreich bis zur Bundesrepublik Deutschland*. Königstein: Hain.
- Doran, K., Gelber, A. and Isen, A. (2014). The Effects of High-Skilled Immigration Policy on Firms: Evidence from H-1B Visa Lotteries.
- and Yoon, C. (2019). Immigration and Invention: Evidence from the Quota Acts.
- Dorner, M. and Harhoff, D. (2018). A Novel Technology-industry Concordance Table Based on Linked Inventor-establishment Data. *Research Policy*, 47 (4), 768–781.
- , König, M., Seth, S. et al. (2011). *Sample of Integrated Labour Market Biographies. Regional File 1975-2008 (SIAB-R 7508)*. Tech. rep., Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for E.
- Dustmann, C. (2003). Return Migration, Wage Differentials, and the optimal Migration Duration. *European Economic Review*, 47 (2), 353–369.
- , Fasani, F., Frattini, T., Minale, L. and Schönberg, U. (2017a). On the Economics and Politics of Refugee Migration. *Economic Policy*, 32 (91), 497–550.
- and Glitz, A. (2011). Migration and Education. In *Handbook of the Economics of Education*, Vol. 4, Elsevier, 327–439.
- and — (2015). How Do Industries and Firms Respond to Changes in Local Labor Supply? *Journal of Labor Economics*, 33 (3), 711–750.
- , — and Frattini, T. (2008). The Labour Market Impact of Immigration. *Oxford Review of Economic Policy*, 24 (3), 477–494.
- , —, Schönberg, U. and Brücker, H. (2016a). Referral-based Job Search Networks. *The Review of Economic Studies*, 83 (2), 514–546.
- , Schönberg, U. and Stuhler, J. (2016b). The Impact of Immigration: Why do Studies reach such Different Results? *Journal of Economic Perspectives*, 30 (4), 31–56.
- , — and Stuhler, J. (2017b). Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment. *The Quarterly Journal of Economics*, 132 (1), 435–483.
- and Van Soest, A. (2002). Language and the Earnings of Immigrants. *ILR Review*, 55 (3), 473–492.
- Eckstein, Z. and Weiss, Y. (2004). On the Wage Growth of Immigrants: Israel, 1990–2000. *Journal of the European Economic Association*, 2 (4), 665–695.
- Edin, P.-A., Fredriksson, P. and Åslund, O. (2003). Ethnic Enclaves and the Economic Success of Immigrants-Evidence from a Natural Experiment. *The Quarterly Journal of Economics*, 118 (1), 329–357.
- Fassbender, S. (1966). Unternehmerische und betriebliche Probleme. *Probleme der ausländischen Arbeitskräfte in der Bundesrepublik, Beihefte der Konjunkturpolitik*, 13, 50–55.
- Federal Employment Agency (1962). *Anwerbung und Vermittlung ausländischer Arbeitnehmer - Erfahrungsbericht 1961*. Nürnberg.
- Federal Employment Agency (1964). *Die bezirkliche Gliederung der Bundesanstalt für Arbeitsvermittlung und Arbeitslosenversicherung*. Nürnberg.
- Federal Employment Agency (1965). *Ausländische Arbeitnehmer: Beschäftigung, Anwerbung, Vermittlung - Erfahrungsbericht 1964*. Nürnberg.

- Federal Employment Agency (1971). *Arbeitstatistik 1970 - Jahreszahlen*. Nürnberg.
- Federal Employment Agency (1972a). *Ausländische Arbeitnehmer: Erfahrungsbericht 1971*. Nürnberg.
- Federal Employment Agency (1972b). *Strukturdaten für die Dienststellen der Bundesanstalt für Arbeit aus den Ergebnissen der Arbeitsstättenzählung 1970*. Nürnberg.
- Federal Employment Agency (1974). *Ausländische Arbeitnehmer: Erfahrungsbericht 1972/73*. Nürnberg.
- Feuser, G. (1961). *Ausländische Mitarbeiter im Betrieb*. München: Verlag Moderne Industrie.
- Ganglmair, B. and Reimers, I. (2019). Visibility of Technology and Cumulative Innovation: Evidence from Trade Secrets Laws, ZEW-Centre for European Economic Research Discussion Paper 19-035.
- Glitz, A. (2012). The Labor Market Impact of Immigration: A Quasi-experiment Exploiting Immigrant Location Rules in Germany. *Journal of Labor Economics*, 30 (1), 175–213.
- (2014). Ethnic Segregation in Germany. *Labour Economics*, 29, 28–40.
- and Meyersson, E. (2020). Industrial Espionage and Productivity. *American Economic Review*.
- Grönqvist, H. (2006). Ethnic Enclaves and the Attainments of Immigrant Children. *European Sociological Review*, 22 (4), 369–382.
- Habakkuk, H. J. (1962). *American and British Technology in the Nineteenth Century: The Search for Labour Saving Inventions*. Cambridge University Press.
- Haisken-DeNew, J. P. and Frick, J. R. (2005). Desktop Companion to the German Socio-Economic Panel Study (GSOEP).
- Hanlon, W. W. (2015). Necessity is the Mother of Invention: Input Supplies and Directed Technical Change. *Econometrica*, 83 (1), 67–100.
- Hanson, G. H. and Slaughter, M. J. (2002). Labor-market Adjustment in Open Economies: Evidence from US States. *Journal of International Economics*, 57 (1), 3–29.
- Harri, A. and Brorsen, B. W. (2009). The Overlapping Data Problem. Available at SSRN 76460.
- Haug, S. and Sauer, L. (2007). Zuwanderung und Integration von (Spät-) Aussiedlern—Ermittlung und Bewertung der Auswirkungen des Wohnortzuweisungsgesetzes (Forschungsbericht des Bundesamtes für Migration und Flüchtlinge 3). Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Hicks, J. (1932). *The Theory of Wages*. London: Macmillan.
- Hornung, E. (2014). Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia. *American Economic Review*, 104 (1), 84–122.
- Hunt, J. (1992). The Impact of the 1962 Repatriates from Algeria on the French Labor Market. *ILR Review*, 45 (3), 556–572.
- and Gauthier-Loiselle, M. (2010). How Much Does Immigration Boost Innovation? *American Economic Journal: Macroeconomics*, 2 (2), 31–56.
- Imai, K., Keele, L., Tingley, D. and Yamamoto, T. (2011). Unpacking the black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review*, 105 (4), 765–789.
- , — and Yamamoto, T. (2010). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*, 51–71.
- Imbert, C., Seror, M., Zhang, Y. and Zylberberg, Y. (2019). Migrants and Firms: Evidence from China, CESifo Working Paper No. 7440.

- Jahn, V. and Steinhardt, M. F. (2016). Innovation and Immigration – Insights from a Placement Policy. *Economics Letters*, 146, 116–119.
- Jensen, P. and Rasmussen, A. W. (2011). The Effect of Immigrant Concentration in Schools on Native and Immigrant Children’s Reading and Math Skills. *Economics of Education Review*, 30 (6), 1503–1515.
- Karabarbounis, L. and Neiman, B. (2014). The Global Decline of the Labor Share. *The Quarterly journal of economics*, 129 (1), 61–103.
- Kerr, S. P., Kerr, W. R. and Lincoln, W. F. (2015). Skilled Immigration and the Employment Structures of US Firms. *Journal of Labor Economics*, 33 (S1), S147–S186.
- Kerr, W. R. and Lincoln, W. F. (2010). The Supply Side of Innovation: H-1B Visa Reforms and US ethnic Invention. *Journal of Labor Economics*, 28 (3), 473–508.
- Kiley, M. T. (1999). The Supply of Skilled Labour and Skill-Biased Technological Progress. *The Economic Journal*, 109 (458), 708–724.
- Klose, H.-U. (1996). Bevölkerungsentwicklung und Einwanderungspolitik. In *Integration und Konflikt: Kommunale Handlungsfelder der Zuwanderungspolitik*, Bonn: Friedrich-Ebert-Stiftung.
- Koller, B. (1993). Aussiedler nach dem Deutschkurs: Welche Gruppen kommen rasch in Arbeit. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 26 (2), 207–221.
- Kremer, M. (1993). Population Growth and Technological Change: One Million BC to 1990. *The Quarterly Journal of Economics*, 108 (3), 681–716.
- Levin, R. C., Klevorick, A. K., Nelson, R. R., Winter, S. G., Gilbert, R. and Griliches, Z. (1987). Appropriating the Returns from Industrial Research and Development. *Brookings papers on economic activity*, 1987 (3), 783–831.
- Lewis, E. (2011). Immigration, Skill Mix, and Capital Skill Complementarity. *The Quarterly Journal of Economics*, 126 (2), 1029–1069.
- Mann, K. and Püttmann, L. (2018). Benign Effects of Automation: New Evidence from Patent Texts, Available at SSRN 2959584.
- Maraut, S., Dernis, H., Webb, C., Spiezia, V. and Guellec, D. (2008). The OECD REGPAT Database: A Presentation, OECD Science, Technology and Industry Working Papers No. 2008/02.
- Max Planck Institute for Demographic Research and CCG (2011). MPIDR Population History GIS Collection (partly based on Bundesamt für Kartographie und Geodäsie (2011)).
- Mincer, J. (1984). Human capital and economic growth. *Economics of education review*, 3 (3), 195–205.
- Mithas, S. and Lucas Jr, H. C. (2010). Are foreign IT Workers cheaper? US Visa Policies and Compensation of Information Technology Professionals. *Management Science*, 56 (5), 745–765.
- Monras, J. (2019). Immigration, Internal Migration, and Technology Adoption, CEPR Discussion Paper No. DP13998.
- Moser, P., Voena, A. and Waldinger, F. (2014). German Jewish Emigrés and US invention. *American Economic Review*, 104 (10), 3222–55.
- Muehlemann, S. and Leiser, M. S. (2018). Hiring Costs and Labor Market Tightness. *Labour Economics*, 52, 122–131.
- Nakamura, E. and Steinsson, J. (2014). Fiscal Stimulus in a Monetary Union: Evidence from US Regions. *American Economic Review*, 104 (3), 753–92.
- OECD (2004). *OECD Employment Outlook: 2004*. Organisation for Economic Co-operation and Development.

- (2009). *OECD Patent Statistics Manual 2009*. Organisation for Economic Co-operation and Development.
- Ohliger, R. (2008). Country Report on Ethnic Relations: Germany. *EDUMIGROM Background Papers*.
- Paci, R. and Usai, S. (2000). Technological Enclaves and Industrial Districts: An Analysis of the Regional Distribution of Innovative Activity in Europe. *Regional studies*, 34 (2), 97–114.
- Penninx, R. and Van Renselaar, H. (1976). Evolution of Turkish Migration before and during the current European Recession. *Migration and Development: A Study of the Effects of International Labor Migration on Bogazliyan District*.
- Peri, G., Shih, K. and Sparber, C. (2015). Foreign and Native Skilled Workers: What Can We Learn from H-1B Lotteries?, National Bureau of Economic Research No. w21175.
- and Sparber, C. (2009). Task Specialization, Immigration, and Wages. *American Economic Journal: Applied Economics*, 1 (3), 135–69.
- Piopunuk, M. and Ruhose, J. (2017). Immigration, Regional Conditions, and Crime: Evidence from an Allocation Policy in Germany. *European Economic Review*, 92, 258–282.
- Porter, M. F. *et al.* (1980). An Algorithm for Suffix Stripping. *Program*, 14 (3), 130–137.
- Powell, W. W. and Snellman, K. (2004). The Knowledge Economy. *Annu. Rev. Sociol.*, 30, 199–220.
- Salomons, A. *et al.* (2018). Is Automation Labor-displacing? Productivity Growth, Employment, and the Labor share, NBER Working Paper No. w24871.
- San, S. (2019). Labor Supply and Directed Technical Change: Evidence from the Abrogation of the Bracero Program in 1964, Available at SSRN 3268969.
- Schmitt, K., Rattinger, H. and Oberndörfer, D. (1994). *Kreisdaten (Volkszählungen 1950-1987)*. Köln: GESIS Datenarchiv. Datenfile Version 1.0.0.
- Schmoch, U. (2008). Concept of a Technology Classification for Country Comparisons, Final Report to the World Intellectual Property Organisation (WIPO).
- Scholten, W. (1968). *Die Beschäftigungsstruktur der ausländischen Arbeitnehmer in der BRD*. Ruhr-Universität Bochum.
- Schönwälder, K. and Söhn, J. (2009). Immigrant Settlement Structures in Germany: General Patterns and urban Levels of Concentration of major Groups. *Urban Studies*, 46 (7), 1439–1460.
- Schüller, S. (2016). Ethnic Enclaves and Immigrant Economic Integration. *IZA World of Labor*.
- Seeni, A. and Brown, T. (2015). Measuring innovation performance of countries using patents as innovation indicators.
- Shimer, R. (2005). The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *American economic review*, 95 (1), 25–49.
- Striso, W. (1968). *Zur betriebswirtschaftlichen Integration der ausländischen Arbeitnehmer*. Köln: Kleikamp.
- Voelker, G. E. (1976). More foreign workers-Germany's labour problem No. 1. *Turkish Workers in Europe 1960-1975*.
- Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market.
- WIPO (2009). World Intellectual Property Indicators - 2009.
- Zator, M. (2019). Digitization and Automation: Firm Investment and Labor Outcomes, Available at SSRN 3444966.
- Zeira, J. (1998). Workers, Machines, and Economic growth. *The Quarterly Journal of Economics*, 113 (4), 1091–1117.