Human Capital Mobility and Economic Performance: Microeconomic Evidence from Natural Experiments

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2016

vorgelegt von

Nadzeya Laurentsyeva

Referent:	Prof. Dr. Florian Englmaier
Korreferentin:	Prof. Dr. Monika Schnitzer
Promotionsabschlussberatung:	10. Mai 2017

Datum der mündlichen Prüfung: 3. Mai 2017

Namen der Berichterstatter: Florian Englmaier, Monika Schnitzer, Oliver Falck

Acknowledgements

This work would not have been possible without guidance and encouragement of my main advisor, Florian Englmaier. While giving, probably, the maximum research freedom, Florian supported me in developing ideas and, most importantly, in bringing them to 'life'. His encouragement was crucial for me to not get intimidated by the arising problems. On the contrary, I was motivated to search for viable solutions, with experiment and failure being part of the process. I am also very grateful to my second advisor, Monika Schnitzer, who was always available for advice when we approached her and whose critical insights stimulated us to move the work forward. I am thankful to Oliver Falck for giving his feedback and for serving as my third advisor.

I thank all senior and junior faculty members for their openness and readiness to help with research. I also greatly benefited from taking part in MGSE courses and workshops. I thank, in particular, Carsten Eckel - the coordinator of the GRK1928 - for his constant support and a positive attitude. Joachim Winter, Andreas Haufler, Martin Watzinger, Davide Suverato, Andreas Steinmayr, Anna Gumpert, and Lisandra Flach provided very valuable comments throughout the entire research work.

I appreciate the support I received from Silke Englmaier and Manuela Beckstein as well as from the 'GRK gurus', Julia Zimmermann and Florian Urbschat. I also gratefully acknowledge the financial support from the DFG Research Training Group 1928: Microeconomic Determinants of Labour Productivity.

I enjoyed working with great coauthors - Yvonne Giesing, Thomas Fackler, Michele Battisti, and Andreas Bastgen. In addition, I thank all fellow MGSE PhD colleagues, whose inputs and suggestions were mostly helpful. I am also thankful to my new colleagues from CEPS and MEDAM - for bringing 'policy-relevance' to my current work and for being a strong commitment to meet my submission deadlines.

Yet, three years in Munich brought me much more than the PhD. @Yvonne: once, I thought I was doing the PhD in a wrong way, because I realised I enjoyed it so much. Thanks for sharing our research and non-research quests, runs, ups and downs! @Bettina, Martina, Dana, and Franzi: while migration research often emphasises immigrants' integration challenges, this was hardly the case for me in Munich thanks to the best ever flatmates! A special thanks goes also to my Belarusian and 'IES' friends who accepted the 'alpine' challenge and visited me in Munich.

Finally, I am enormously grateful to my parents, Andrei and Halina, who have always believed in me and whose support I constantly feel, no matter how far away from each other we live.

Nadzeya Laurentsyeva, July 2017

Contents

Introduction

Human capital mobility, broadly defined as the free movement of knowledge, skills, and abilities embodied in people, is important for the efficient allocation of this scarce resource. Yet, numerous factors, such as political borders, administrative obstacles, physical and social distances limit mobility, thus exposing firms and individuals to certain constraints. These barriers have changed over time and have evolved in both directions. While some obstacles to human capital mobility disappeared with stronger economic integration and technology advancements, new barriers arose, for instance, as some countries restricted labour mobility or as political conflicts imposed new legal and psychological borders between people. The changes to human capital mobility modify constraints, under which people and firms optimise, and hence affect economic decisions and outcomes. Therefore, from a research perspective, these changes provide good opportunities to investigate the role of human capital and human interactions for economic performance. This knowledge, apart from being of academic interest, can help firms in designing optimal personnel policies and can guide decisions of policy-makers.

This thesis exploits two natural experiments to analyse the economic consequences of changes in human capital mobility. The first two Chapters use the introduction of free labour mobility in Europe to examine the impact of higher international labour mobility on firm productivity and innovation in source countries. The third Chapter uses a somewhat reverse setting: an unexpected political conflict introduced new barriers to international collaboration on an online open-source platform - an environment, where in theory no legal or physical barriers to human capital mobility exist. My work in Chapter 3 aims at uncovering the nature of these barriers and at estimating their role for performance of international teams. All three Chapters use credible exogenous variation to establish a causal link between changes in human capital mobility and economic outcomes. The conducted microeconomic analysis allows to identify the channels, related to firm or individual decision-making, which could explain the observed effects and establish the role of human capital and human interactions for firm productivity, innovation, and team performance. Below, I summarise the results and contributions of each of the three Chapters.

The first two Chapters focus on the effects of free labour mobility for countries that have become net generators of migration flows. The consequences of free labour mobility for source countries of migrants remain controversial. On the one hand, skilled emigration might lead to brain-drain and hence exacerbate the existing income differences among countries. On the other hand, removing legal barriers to human capital mobility may lead to better allocation of resources, stimulate knowledge flows and international collaboration. Policy decisions without proper understanding of costs and benefits of free labour mobility might lead to suboptimal outcomes.

When estimating the effects of emigration, an empirical challenge arises due to the endogeneity of migration flows. To overcome this problem, both Chapters exploit a novel instrumental variable strategy, which is based on changes in labour mobility legislations. In particular, we exploit differences in labour mobility regulations across European countries and industries throughout 2004-2014. These regulations were subject to the political decisions in destination countries and hence can be treated as exogenous to economic conditions in migrants' source countries, which are the focus of the empirical analysis.

Chapter 1 (based on joint work with Yvonne Giesing) examines the impact of emigration on firm total factor productivity (TFP). We conduct our empirical analysis using the panel firm-level data from Eastern and Central European countries, which following their accession to the European Union (EU) experienced unprecedentedly high emigration rates. Our identification builds on the transitional provisions applied by old EU member states, which could temporarily restrict the access to their labour markets for citizens from new member countries. While these transitional provisions were in force, emigration opportunities for citizens from new member countries varied, depending on their country of origin and the industry they were qualified to work in. According to our results, firms in industries that were exposed to higher labour mobility of their workers, experienced a drop in TFP. This result is statistically and economically significant and robust to various measures of TFP and firm profits. Yet, we also note the substantial heterogeneity among firms. More productive foreign-owned and patenting firms avoid losses in TFP, suggesting that these firms are better able to adjust to the changing environment. We further develop a simple conceptual framework that characterises the plausible channel behind the negative effect on TFP. Better emigration opportunities increase turnover of skilled workers, thus leading to the loss in the firm-specific human capital and lowering incentives for firms to invest in the intensive training of new employees. As a consequence, the firm-specific stock of knowledge, which represents part of the firm's TFP, decreases. The empirical evidence is consistent with this channel. We observe higher worker turnover in the affected industries, furthermore, firms' personnel and training costs increase.

This Chapter makes several contributions to the literature. First, we contribute to the literature on the economic impacts of migration (Clemens, 2013; Docquier and Rapoport, 2012; Freeman, 2006; Grossmann and Stadelmann, 2011, 2013) by providing the causal microeconomic evidence of migration effects for source countries. We combine the novel identification strategy with the detailed firm-level dataset, which allows to obtain credible estimations of the effect. Second, we complement the emerging migration literature, which focuses on the firm as the level of analysis (Dustmann and Glitz, 2015; Kerr and Kerr, 2013; Kerr et al., 2014; Paserman, 2013; Peri, 2012). Looking at the firm level allows to account for firm heterogeneity and to shed light on the possible mechanisms of the migration effect. Third, apart from migration literature, we make a contribution to the works on the determinants of firm total factor productivity (Bartelsman et al., 2013; Bloom et al., 2012, 2016; Fox and Smeets, 2011). While TFP is one of the best predictors of firm growth, it is often referred to as "the measure of our ignorance" (Syverson, 2011). We discuss theoretically and then provide empirical evidence for the channel that links firm total factor productivity to skilled emigration and, thus, emphasise the firm-specific human capital as an important determinant of TFP.

Chapter 2 of this thesis (based on joint work with Yvonne Giesing and Thomas Fackler) complements Chapter 1 by focusing on potential benefits of emigration. We combine industry-level migration and patenting data from 32 European countries and show that emigration can foster innovation in source countries. We attribute this positive effect to the knowledge flows trigged by emigrants.

Similarly to Chapter 1, we exploit changes in the European labour mobility laws throughout 2004-2014 to construct an instrumental variable for migration and thus to overcome the endogeneity problem between emigration and innovation. We obtain two main empirical results. First, innovation, as measured by the number of patent applications, increases in the number of emigrants. This effect is strong enough to lower patenting asymmetries between more and less advanced countries. Second, we use patent citations as a proxy for knowledge flows and show that migration stimulates cross-border knowledge flows from destination back to origin countries of migrants. While skilled emigrants do not patent in their home country anymore, they can contribute to knowledge and technology diffusion, thus improving the patent production in origin countries and helping them to catch-up with the technology frontier. For all the three results, the effects are quantitatively more pronounced when we consider only migration of people with patenting potential.

This Chapter complements the research by Bosetti et al. (2015), Choudhury (2015), Kaiser et al. (2015), Kerr (2008), and Kerr and Lincoln (2010) on the effects of labour mobility for innovation. First, we establish causality between skilled emigration and higher patenting in the source countries. Second, because we possess comparable patenting data for source and destination countries of migrants, we can extend this result and show that emigration helps convergences between less and more advanced countries. The latter result is important, given the wide-spread concerns that skilled emigration can exacerbate inequalities between countries. By investigating the channel behind the positive effect of emigration, we contribute to the literature on the determinants of knowledge flows (Jaffe et al., 1993; Singh and Marx, 2013; Thompson and Fox-Kean, 2005). We provide evidence from the European context that labour mobility fosters cross-border knowledge diffusion and thus increases innovation efficiency in the source countries of migrants. The empirical setting allows separating this effect from other potential channels such as trade and foreign direct investment.

Bringing the results of the two Chapters together shows that international human capital mobility generates both new opportunities and new challenges. Firms and policy-makers can address the challenges and maximise benefits if they put the right policies in place. For firms, our results would call for active human resource strategies, for instance, to invest in better training technologies aimed at new workers. From the policy perspective, rather than trying to limit skilled emigration, origin countries can benefit more by reducing internal search frictions, facilitating inflows of foreign workers, and preserving ties with emigrants who work in more advanced economies.

While Chapters 1 and 2 investigate the economic consequences of removing legal barriers to the international labour mobility, Chapter 3 estimates the consequences of *imposing* barriers to international collaboration in the environment, where legal or physical obstacles to human capital mobility should not play a role. In Chapter 3, I study the consequences of an exogenous political conflict for performance of international teams on GitHub, the world's largest open-source platform. In particular, I exploit the unexpected crisis between Russia and Ukraine following the annexation of Crimea in March 2014 and analyse the impact of this conflict on the online collaboration between Ukrainian and Russian programmers. Using microdata from GitHub, I show that the conflict reduced the scope of the Ukrainian-Russian collaboration on the platform. The empirical approach comprises the difference-in-difference and triple-difference methods and allows to argue that this reduction was the direct consequence of the political conflict. In addition, I provide empirical evidence that the economic reasons, such as higher transaction costs or career expectations, cannot rationalise the observed changes. Rather, the drop in collaboration concords with the identity-based explanation (following Akerlof and Kranton (2000)). The political conflict, though external to the working environment on the open-source platform, modified prescriptions related to the national identities of programmers and thus shifted their preferences for teammates or projects. This shift in preferences distorted existing and future collaborations, profitable from an economic perspective. I further discuss and link to the empirical evidence two potential channels, through which identity could affect team performance: first, by increasing communication costs due to poorer interaction between team members and second, by lowering incentives to contribute to projects associated with a "hostile" social group.

Chapter 3 adds to the literature on performance of diverse teams (Bandiera et al., 2005, 2009; Lazear, 1999; Mas and Moretti, 2009) by providing empirical evidence for one of the risks that diverse teams face: external events can exacerbate social differences within teams and hence inhibit performance. My results also complement the findings of theoretical and experimental literature on social identity and group performance (Charness et al., 2007; Chen et al., 2014; Chen and Li, 2009) by emphasising the importance of common (non-conflicting) identity for efficient team collaboration. Lastly, Chapter 3 contributes to the literature on microeconomic effects of international or ethnic tensions (Fisman et al., 2013; Hjort, 2014; Ksoll et al., 2010; Marx et al., 2015; Rohner et al., 2013a,b) by providing causal evidence that even in the settings where national or ethnic borders should play no role, identity-based conflicts can impose social barriers on international collaboration and hence inhibit the efficient allocation of human capital.

Chapter 1

Firms Left Behind: Emigration and Firm Productivity*

1.1 Introduction

The emigration of high-skilled workers poses a challenge for many countries, not only in the developing world. As workers leave their firms to follow better opportunities abroad, policy-makers and managers complain about skill shortages and emphasise the negative effects of brain drain. However, whether there is a causal link from skilled emigration to firm productivity is not clear. Scarcity of firm-level data from emigrants' countries of origin and the endogeneity of migration flows inhibit from going beyond anecdotal evidence. The direction of causation could well go the other way with migrants leaving the least productive firms or a change in unobservable variables triggering both lower firm performance and higher emigration rates. Yet, identifying firm-level effects of emigration and thoroughly disentangling the mechanisms is indispensable for the design of appropriate policies in source countries.

Central and Eastern Europe is a region that has experienced particular high emigration rates in recent years. Following the EU accession of Central and Eastern European countries in 2004 and 2007, migration flows from new member states (NMS) to old

*This chapter is based on joint work with Yvonne Giesing.

EU member states have increased considerably. In 2003, the number of NMS migrants residing in other EU countries amounted to 846,000 people and by 2014 this number had reached 3.95 million. Although the skill level of emigrants varies across destination countries, on average, NMS migrants have been positively selected. As of 2014, 25% of the post-accession NMS emigrants had tertiary education. To compare, among NMS non-migrants, people with university degree accounted for 13.5%.[1] Despite important positive consequences of free labour mobility in terms of lower unemployment and a better skill match, there have been growing concerns that the emigration of skilled workers has created a severe challenge for source countries (Kahanec, 2012; OECD, 2013; Zaiceva, 2014).

This paper investigates the causal effects of skilled emigration on firm performance. As 'skilled', we denote individuals with either tertiary education or a professional qualification. To identify the effect of interest, we exploit changes in EU labour mobility legislation from 2004 to 2014. The transitional provisions applied by old EU member states created a quasi-experimental setting by allowing earlier or later free labour mobility for certain categories of NMS workers. While these transitional provisions were in place, emigration opportunities for NMS citizens varied, depending on their country of origin and the industry they were qualified to work in. Using firm-level data from NMS countries, we show that firms in industries that were exposed to higher outflows of skilled workers experienced a drop in total factor productivity (TFP). The estimates are qualitatively robust to various measures of TFP and firm profits.

Apart from analysing the reduced-form effects of legislation changes on firm productivity, we also perform 2SLS regressions to estimate the effect for firms which effectively experienced skill shortages due to higher emigration rates. Changes in EU labour mobility laws strongly predict skill shortages as reported by firms in NMS. This allows us to use the legislation changes as an instrument. We argue for the validity of this instrument: detailed sector- and country-specific legislation changes had not been anticipated and are uncorrelated with other integration-related events, such as the free movement of goods or capital. Using annual data from the European Commission Business Survey, we find

[1] Source: Eurostat LFS Data. Only migrants, who entered the old EU countries after the EU accession are taken into account.

that a one percentage point increase in instrumented skill shortages leads to a 1.6 percent drop in firm TFP.

To analyse more thoroughly how emigration reduces firm productivity, we develop a simple theoretical framework that illustrates one plausible channel behind this result. Better emigration opportunities induce more skilled workers to quit their jobs. This results in higher job turnover rates that reduce the existing firm-specific human capital and lower firms' incentives to invest in firm-specific training of new employees. As turnover increases and more workers have to be trained, intensive training programmes become costlier. Consequently, the stock of firm-specific skills and knowledge decreases. The effect is captured by TFP, as this form of human capital is not fully accounted for in wages. Our results are consistent with this mechanism. We find evidence for higher turnover of workers in sectors that are strongest hit by emigration and document an increase in firms' personnel and training costs. This mechanism fits well into the previous literature. Konings and Vanormelingen (2015) find that the productivity of workers increases by more than their wage after they have participated in training. Consequently, if trained workers are leaving, this is captured by labour productivity and residual TFP. Jäger (2016) shows that longer-tenured workers are harder to replace with outsiders. For more studies on the relationship between job turnover, firm-specific human capital, and firm productivity we refer to Brown and Medoff (1978), Shaw (2011), Strober (1990), and Yanadori and Kato (2007). The firm has several ways to adapt to higher quitting rates of skilled workers. It can substitute labour with capital (see Dustmann and Glitz (2015) for the case of immigration), substitute high-skilled with low-skilled workers or improve training technology for new hires.

Panel data allow us to account for firm heterogeneity and to explore the link between firms' characteristics and their sensitivity and adaptation to higher quitting rates of workers. We find that innovating and foreign-owned firms substantially increase their per-employee personnel costs. These firms are apparently able to (at least, partly) match wages offered abroad and provide more training, and therefore prevent the loss of firm-specific human capital.

This paper makes three key contributions to the literature. The first and main contribution is that we analyse the effects of emigration at the firm level. So far, the economic effects of emigration and brain drain have focused on the aggregate level (Clemens, 2013; Docquier and Rapoport, 2012; Freeman, 2006; Grossmann and Stadelmann, 2011, 2013). We expect that the migration literature can gain richer insights into the consequences of migration by investigating firm-level outcomes. Kerr et al. (2014), Kerr et al. (2013) and Kerr (2013), for instance, are encouraging the firm-level approach for the analysis of migration. Accounting for firm level outcomes, adaptation mechanisms and firm heterogeneity is important as it shapes the observable effect of migration on macro outcomes.

While there is an emerging migration literature that focuses on the firm as the unit of analysis, it has focused on immigration until now. Peri (2012), Kerr and Kerr (2013), Kerr et al. (2014), Paserman (2013), Mitaritonna et al. (2014) and Ottaviano et al. (2015) study the effects of immigration on firm productivity in the US, Israel, France and the UK respectively. They find that an increase in the supply of foreign-born workers positively affects firm productivity due to a faster growth of capital and the specialisation of natives in more complex tasks. Lewis (2013) furthermore finds that besides increased investment, firms also adapt new technology. Using firm-level German data, Dustmann and Glitz (2015) analyse how industries and firms respond to changes in the local labour supply. They find that immigration alters the local skill composition and investigate three adaptation mechanisms: a change in factor prices, a within-firm change in skill intensity, and an adjustment through the entry and exit of firms. Our research is complementary to this literature. While these authors look at the effects of immigration on firms, we focus on the consequences of emigration. Moreover, we propose a plausible mechanism that links the outflow of skilled workers to firms' total factor productivity. This mechanism emphasises the role of firm-specific human capital for firms' performance and allows drawing concrete policy recommendations for firms.

The second contribution is the creation of an instrument that circumvents the endogeneity of migration. To the best of our knowledge, this paper is the first to exploit industry-level variation in labour mobility laws to causally evaluate the effect of emigration on firm performance. Due to a lack of firm level data for source countries and the endogeneity of migration, the causal analysis is not trivial. To address these issues, we create an extensive dataset that merges migration and firm level data on the country, year and industry level to exogenous labour mobility legislation changes. We are thus able to show that emigration imposes binding skill shortages for firms and lowers TFP via a loss of firm specific human capital.

Third, we add to the literature on the consequences of EU enlargement. This is of very high relevance to policy makers in Brussels, in accession countries, and in candidate countries, for instance Serbia. In particular, we complement the research that investigates the consequences of the recent emigration wave from the NMS. Mayr and Peri (2009) develop a model to study the consequences of European free labour mobility on human capital in the sending countries and differentiate between brain drain and brain gain due to return migration and increased incentives to invest in education. Dustmann et al. (2015) and Elsner (2013) estimate the effects of emigration on wages in Poland and Lithuania and find that wages increase for the stayers. Our contribution is to illustrate that, while firms in general experience a drop in TFP, there are various adaptation mechanisms for firms. Moreover, we suggest policies that concerned governments can implement to mitigate the negative effects.

The paper is organised as follows. The next section outlines a theoretical framework to motivate and structure our empirical analysis. Section 3 provides background information on the EU opening and transitional provisions regarding free labour mobility, which helps to understand our identification strategy. Section 4 describes the data, followed by Section 5 that presents the empirical specification. Section 6 discusses the results including heterogeneous effects, while Section 7 provides robustness checks. Section 8 concludes.

1.2 Theoretical Framework

1.2.1 General Setting

Our theoretical framework illustrates the consequences of skilled emigration at the firm level in the source country. Using a partial-equilibrium framework, we generate predictions about changes in firms' factor demand, training provision, and TFP.

We assume that there are frictions in the labour market: job separations occur at an exogenous rate and in order to fill vacant positions firms post costly vacancies. One trigger for job separations, for instance, is an easier access to foreign labour markets, which induces higher emigration. If the job separation rate increases, firms in source countries experience higher skill shortages. In this setting, skill shortages are not a disequilibrium phenomenon, but correspond to some measure of search frictions (for example, the number of posted vacancies for skilled employees).

We allow firm-specific human capital to explicitly enter the production function. A higher labour turnover destroys part of the firm-specific human capital. Since the latter is not fully captured by wages, this loss translates to a drop in TFP. In this way, we characterise one possible micro channel, through which skilled emigration directly affects firm productivity.[2]

The economy consists of a representative firm that produces output according to the production function:

$$Y = Af(K, L_s, L_u) \tag{1.1}$$

$Af()$ is a general production function, where K is the capital input and L_s and L_u are the skilled and unskilled labour inputs. $f()$ increases in the production factors K, L_s, L_u; exhibits diminishing marginal returns to K, L_s, L_u and is twice-differentiable. Each period L_s and L_u workers are involved in the production process. At the end of the period, a

[2] On a macro level, this problem was examined by Grossmann and Stadelmann (2011). In their overlapping generations model, the drop in TFP is attributed to less firm entry and, consequently, to the reduction in human capital externalities of skilled employees.

proportion δ_s (δ_u) of skilled (unskilled) job matches are destroyed. The total turnover rate δ is defined as the number of separations over the total number of employees. To fill the positions with new workers, a firm posts vacancies. For simplicity, we assume that vacancies are matched with probability one. In equilibrium, the number of job separations must equal the number of matched vacancies:

$$V_i = \delta_i L_i, i = s, u. \tag{1.2}$$

Posting vacancies creates a search cost of c_s (c_u) per period.

We represent TFP as $A = t^\gamma$. In our setting, the firm TFP consists entirely of firm-specific knowledge t. This tacit knowledge makes all the input factors more productive. It could be, for instance, a collection of the firm's best practices, a code of conduct, or tricks of an internal IT system. In order to employ this knowledge in the production, the firm has to train all skilled workers in using it. We assume that there is no training needed for unskilled workers. If a skilled worker leaves and the firm hires a new worker as a replacement, it has to pay the training costs for the new worker, which are proportional to the amount of firm-specific knowledge to learn. Given a turnover rate δ_s, the total training costs per period would amount to $\delta_s L_s c_t t$, where $\delta_s L_s$ is the number of newly hired skilled workers. c_t denotes the costs of training, which we set equal to 1. The total training costs can also be interpreted as the loss of firm-specific human capital due to worker turnover. We treat the amount of training per worker t as adjustable when the firm hires new skilled workers. For instance, if it becomes too expensive to teach a particular firm practice to all the new hires, the firm can drop this practice, thus reducing its knowledge t. If there is no turnover, $\delta_s = 0$, the firm-specific knowledge stays constant.

1.2.2 The Firm's Optimisation Problem

The firm chooses inputs K, L_s, L_u to maximise profits Π. In addition, when hiring skilled workers, the firm decides on t - the amount of firm-specific knowledge to teach. The exogenous variables are the output price (P), wages (w_s, w_u), the interest rate (r), the job destruction rate (δ_s, δ_u), and the vacancy costs (c_s, c_u).

$$\Pi = PY - \sum_{i=s,u} (w_i L_i + c_i V_i) - rK - V_s t \tag{1.3}$$

s.t.

$$V_i = \delta_i L_i, i = s, u;$$

$$Y = t^{\gamma_\tau} f(K, L_s, L_u)$$

Using the constraint to substitute for V_i yields the total personnel costs of skilled workers: $L_s(w_s + c_s\delta_s + t\delta_s)$. These costs comprise wages, search costs, and training expenses. Similarly, the total personnel costs of unskilled workers are equal to $L_u(w_u + c_u\delta_u)$.

The emigration of skilled workers raises δ_s and results in a higher turnover ($\frac{\delta_s L_s}{L}$). The marginal hiring costs of a skilled worker ($\delta_s(c_s+t)$) increase.[3] Thus, emigration augments the marginal personnel costs of a skilled worker ($w_s+\delta_s(c_s+t)$) and affects the relative input demand of the firm. Further, the incentives for training change. The higher turnover rate makes training more expensive, which consequently reduces the optimal level of the firm-specific knowledge t. This result follows from the fact that the firm has to teach all its specific knowledge t to all newly hired skilled workers.[4] Therefore, when δ_s increases, it becomes more expensive for a firm to sustain its knowledge level due to higher training costs. We provide a proof of the comparative statics results for a general production function in the Appendix.

[3] The model is generalisable to the situation in which both skilled and unskilled workers emigrate. In this case turnover would increase for both groups but firm-specific human capital would only be lost for skilled workers.

[4] For instance, unless all of the firm's sales managers know how to use a Customer Relationship Management (CRM) system, there will be very poor coordination among them. This may lead to both the sales managers and the CRM system being unproductive.

1.2.3 Comparative Statics

We are interested in the effect of emigration on firm productivity. If workers obtain the possibility to emigrate to a country with higher wages, this results in a higher quitting probability. This can be triggered by exogenous political events such as the EU accession. In the model, the introduction of free labour mobility that resulted in higher emigration rates can thus be represented by higher job separation rates δ_s and δ_u. In the comparative statics, we focus on the effect of raising δ_s, because it has direct implications for firm TFP.

Proposition: *An increase in the job separation rate δ_s reduces the firm's TFP through the reduction in firm-specific knowledge t.*

1. *An increase in δ_s raises the marginal hiring costs of a skilled worker. This corresponds to an increase in the personnel costs $w_s + \delta_s(c_s + t)$. Depending on the elasticity of substitution between the inputs, firms might find it optimal to substitute high-skilled workers with low-skilled workers or with capital. The ratio $\frac{L_s}{L}$ decreases and/or the ratio $\frac{C}{L}$ increases.*
2. *An increase in δ leads to a lower provision of training (t) per hired skilled worker because higher turnover rates increase marginal training costs. This results in a negative effect on the firm's TFP. However, the total training costs $\delta_s L_s t$ might increase as, on the extensive margin, due to a higher δ_s, the firm has to train more workers.*

In our simple framework, we assume that wages are exogenously given, which is a realistic assumption if we consider an average small or medium-sized firm. Emigration lowers the available supply of skilled labour and should lead to a general increase of w_s. This will increase personnel costs $w_s + \delta_s(c_s + t)$ and thus lower the relative demand for skilled workers. Provided δ_s is now kept constant, the effect on the training provision t will be of a second order. Hence, if emigration leads only to the adjustment of wages, we would not observe a strong negative effect on firm TFP.

1.3 Transitional Provisions for the Free Movement of Workers from New Member States

Before testing the derived predictions with the data, we provide background information on the transitional provisions applied by old EU member states from 2004 to 2014. This section shows how the gradual opening of the EU labour markets created time, country, and industry-level variation in the emigration rates of NMS citizens.

In 2004, ten Eastern and Southern European Countries joined the EU: Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia. While free mobility of goods and capital was introduced either prior to or at the point of accession by all countries, free labour mobility was initially restricted. Some EU15 countries[5] feared an inflow of cheaper labour. The EU Commission thus allowed the old member states to unilaterally restrict their labour markets by national laws for a period of up to seven years. These transitional arrangements were applied to all new members in the same way, except Malta and Cyprus. We thus denote the remaining eight countries as NMS8. In 2007, Bulgaria and Romania (NMS2) joined the European Union, also facing the transitional agreement rules.

The option to unilaterally restrict labour markets generated different rules within the EU. While Ireland, Sweden, and the UK decided to open their labour markets immediately in 2004 without any restrictions, other countries delayed the access or applied special job schemes in certain industries. Denmark, Greece, Spain, and Portugal, for instance, removed restrictions only in 2009. France, Belgium, Netherlands, and Austria opened their labour markets gradually, allowing only workers in certain industries and introducing quotas. Germany kept the labour market almost completely closed until the expiration of the transitional agreements (2011 for NMS8; 2014 for NMS2). Other EFTA members, Iceland, Liechtenstein, Norway, and Switzerland, also applied transitional provisions and we thus include them in our analysis (EU15+4 denote all countries that applied tran-

[5] Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom (EU15).

Table 1.1: Overview of the Gradual Opening of the EU15+4 Labour Markets

Country	NMS8 (2004 entry)	NMS2 (2007 entry)	Sectoral Exceptions
Austria	2011	2014	NMS8 (2007-2010), NMS2 (2007-2013): Construction, Manufacturing of Electronics and Metals, Food and beverage services (restaurant business), other sectors with labour shortages
Belgium	2009	2014	-
Denmark	2009	2009	-
Finland	2006	2007	-
France	2008	2014	NMS8 (2005-2007), NMS2 (2007-2013): Agriculture, Construction, Accommodation and food services (tourism and catering), other sectors with labour shortages
Germany	2011	2014	NMS8 (2004-2010), NMS2 (2007-2013): sectors with labour shortages
Greece	2006	2009	-
Iceland	2006	2012	-
Ireland	2004	2012	-
Italy	2006	2012	NMS8 (2004-2005): sectors with labour shortages; NMS2 (2007-2011): Agriculture, Construction, Engineering, Accommodation and food services (tourism and catering), Domestic work and care services, other sectors with labour shortages; Occupations: Managerial and professional occupations
Lichtenstein	2011	2016	-
Luxembourg	2008	2014	NMS2 (2007 - 2013): Agriculture, Viticulture, Accommodation and food services (tourism and catering)
Netherlands	2007	2014	NMS8 (2004-2006), NMS2 (2007-2013): International transport, Inland shipping, Health, Slaugther-house/meet-packaging, other sectors with labour shortages
Norway	2009	2012	NMS8 (2004-2008), NMS2 (2007-2011): sectors with labour shortages
Portugal	2006	2009	-
Spain	2006	2009	Reintroduction of restrictions for Romanians: 11/08/2011 - 31/12/2013
Sweden	2004	2007	-
Switzerland	2011	2014	-
United Kingdom	2004	2014	NMS2 (2007-2013): Agriculture, Food manufacturing

Notes: Column 2 shows the year of the labour market opening of the respective country for NMS8 countries, column 3 shows the year of the labour market opening of the respective country for the NMS2 countries. Column 4 shows, which sectors were exempt from restrictions.
Source: European Commission.

sitional provisions). Table 1.1 provides an overview of the precise opening dates and industry details per country.

This sequential opening by country, year and industry had a significant effect on migration rates. Constant (2011) and Kahanec (2012) provide descriptive evidence of EU migration flows following the enlargement. They show that the transitional agreements influenced the movement of migrants. The UK and Ireland, for example, have become the main EU destination country for Polish, Slovakian and Latvian workers. Kahanec et al. (2014) apply a difference-in-differences analysis and confirm that outward migration from the NMS increased with the EU entry, but its full potential was hampered by the presence of transitional arrangements.

One might argue that the restriction of a country's labour market is endogenous and related to local labour market conditions. Germany, for instance, experienced high unemployment rates during the mid-2000s and this was one of the reasons for its labour market restrictions. However, while the transitional arrangements are endogenous to labour market conditions and firm productivity in the *receiving* country, they are exogenous to firm outcomes in the *source* countries.

There are additional worries that concern the validity of the instrument. One possible identification problem could arise if the decisions to open a particular industry by EU15 countries were to some extent endogenous to conditions in the new member countries. For example, mobility restrictions might have been directed at the NMS citizens working in countries and industries with high volumes of EU15 FDIs. This is not the case for the following reasons. First, EU15 countries could not differentiate transitional provisions across countries in NMS8 and NMS2 groups. Second, this proposition is hard to reconcile with significant time-variation in the removal of provisions. One might further suggest that the industry-specific timing of labour market openings coincided with trade liberalization. Yet, all new EU member countries had signed and enforced Free Trade Agreements with the EU prior to their accession. It is plausible to conclude that the application of transitional provisions by the EU15 was driven mainly by their own economic conditions and is thus exogenous to firm outcomes in the NMS.

The transitional agreements have not only affected the employed people in the new member states, but have also given new opportunities to the unemployed. One might assume that the unemployed had the highest incentives to leave their countries and to look for work abroad. This would bias our estimated coefficient towards zero as the emigration of unemployed workers would not lead to the loss of firm-specific human capital and would leave firm productivity unaffected. Another concern is that people might change industries as they migrate. This will again bias our estimate towards zero. It is plausible to assume though that people have the smallest emigration costs if the industry they work in opens for immigration. If they eventually work in another industry after migration, this does not affect our results as we are only interested in the fact that they left and it does not matter in which industry they actually work in their destination country.

1.4 Data Description

For our analysis we use firm-level financial and survey data, aggregate industry- and country-level indicators, detailed migration data, and information from EU labour legislation.

We obtain firm-level data from Bureau Van Dijk's AMADEUS database that provides standardised annual balance-sheet and profit information for European public and private companies. We work with an unbalanced panel of about 110,000 firms located in NMS. The period covered ranges from 2000 to 2013, and there are five annual observations for each firm on average. The sample includes companies in manufacturing, construction, retail trade and services. Apart from financial reports, the dataset provides information on firms' patenting activities, ownership structures, export markets, and exit status (such as bankruptcy or liquidation).

We include firms with at least two years of available financial data to calculate the TFP index. As a note of caution, we might not capture companies at the lower tail of the productivity distribution if they are less likely to be included in the sample. Based on observables, though, firms in the regression samples are not statistically different from the full sample (see table A.1). We used the largest possible number of firms with non-missing observations. The number of firms across regression results slightly varies due to differences in the availability of variables.

To obtain data on the training of employees, we complement this data with firm-level information from the Business Environment and Enterprise Performance Surveys (BEEPS) administered by the European Bank for Reconstruction and Development (EBRD) in all NMS. The survey was conducted in 2002, 2005, 2009 and 2012 and contains an extensive questionnaire on firms' self-reported financial performance, workforce composition, management practices, innovation, and perceptions of the business environment (including the availability and quality of human capital). The survey data provides a representative sample of manufacturing, construction, service, and retail trade firms. In total, there are 13,972 firm-year observations, of which 2,556 (with 1,293 unique firms) make up an unbalanced panel.

Disaggregated emigration data by country and industry does not exist.[6] Therefore, we cannot perform a meaningful OLS or first-stage regression using migration data. For our baseline estimations, we thus conduct reduced-form regressions where the main explanatory variable Free Movement (FM) is constructed directly using the legislation information. To construct the FM variable, we use the Labour Reforms database (section on labour mobility) of the EU Commission and complement it with information from the national legislation of EU15+4.

To shed more light on one potential channel, we measure if the opening of labour markets predicts firms' labour shortages. The measure of skill shortages is taken from the EU Commission Business Survey, which is conducted quarterly in all EU member countries by the Directorate General for Economic and Financial Affairs (DG ECFIN). The survey addresses representatives of the manufacturing, service, retail trade, and construction sectors and asks for firms' assessment and expectations of the business development. Among other questions, the survey's participants are asked to evaluate factors limiting their production (such as labour, access to finance, demand, and equipment). The EU commission publishes information on a two-digit NACE industry level, thus the obtained measure is equal to the share of firms in a given industry reporting to be constrained by labour. To match the data to other datasets, we aggregated quarterly indicators to annual levels. As an alternative measure, we consider firms' replies from the BEEPS survey, which asks respondents to evaluate the importance of 'inadequately educated labour' as an obstacle for businesses. To make it more comparable with the EU Commission Survey, we aggregate individual firm responses on a two-digit industry level.

As additional covariates, we use aggregated (two- and four-digit NACE) industry level data, which is available for all EU member states and is harmonised by Eurostat. The structural business statistics database contains annual information on industries' performance, including output, investment, employment, and personnel costs. Macroeconomic

[6]The Eurostat Labour Force Survey provides information on the industry, education, and occupation of immigrants, but aggregates the country-of-origin information. While observing immigrants in EU15+4, we can only see if they come from NMS8 (2004 entry) or NMS2 (2007 entry). Even if the detailed origin information were available, though, it would likely be noisy and the labour force sample would have small numbers in the specific country-industry-year cell.

controls (GDP, FDI, unemployment, interest rates) come from the Worldbank statistical database.

1.5 Econometric Specification

The aim of the empirical analysis is to test the predictions of our model. We thus want to establish how the exogenous increase in the emigration influences TFP, personnel costs, training, and the capital/labour ratio of firms. For identification, we exploit legislation changes, which generate exogenous variation in emigration. In this section, we first present the baseline reduced-form regressions of firm outcomes on legislation changes. The latter are summarised by the Free Movement variable described in Section 1.5.3 below. We then present the specification for the 2SLS regression. As there is no disaggregated country- and industry-specific emigration data for Eastern Europe, we do not explicitly use the Free Movement variable as the instrument for emigration. Instead, to estimate the treatment effects (LATE), we conduct 2SLS regressions with a measure of skill shortages, which are reported on the industry level. The 2SLS coefficients thus capture the effect in industries, where emigration created *binding* skill shortages.

1.5.1 Baseline Model: Reduced Form

The reduced-form empirical specification is described below:

$$Y_{fict} = \beta_1 FM_{ict-l} + \beta_2 X_{fict} + \beta_3 I_{ict} + \beta_4 J_{it} + \beta_5 C_{ct} + \tau_t + \nu_f + \epsilon_{fict} \tag{1.4}$$

where Y_{fict} are different performance measures of a firm (f) in industry (i), country (c) and year (t). FM_{ict-l} indicates the Free Movement variable. We include it in equation (1.4) with a lag of length l. β_1 is the reduced-form effect of the legislation change on a firm-level outcome. X_{fict} is a set of time-varying firm controls, such as age and capacity utilization. I_{ict} includes country-specific industry controls such as total investment, average mark-up (ratio of revenues to costs), and inward FDI. These variables account for variation due to other shifters of labour demand within an industry of a particular country, namely, technical change or higher competition. J_{it} are controls that are common to all countries such as industry-specific total sales and skill shortages. C_{ct} is a vector of macroeconomic covariates, accounting for country-wide changes: the GDP growth rate and FDI inflows. All monetary variables are in natural logarithms. τ_t are time dummies. ν_f represent firm fixed effects, and ϵ_{fict} is the error term. In the baseline empirical model, we consider only within-firm variation. Such a specification allows us to take care of firm unobserved time-invariant heterogeneity (as initial management ability or quality of business ideas) and other constant characteristics of a firm's location or industry-specific production technologies.

The focus of this project is to estimate the effect of emigration on firm total factor productivity. We compute firm productivity in several ways: using a TFP-index and a semi-parametric approach as in Levinsohn and Petrin (2003). The latter method allows us to overcome the simultaneity bias between firms' inputs and unobserved productivity shocks. For details regarding the TFP calculation, we refer to the Appendix A.3. As alternative measures of productivity, we consider firm profits: $\frac{EBIT}{Assets}$ calculated as the ratio of earnings before interest and tax over assets. Using a number of different productivity measures ensures that the effects we find are not driven by measurement issues.

To understand if our additional model predictions for firms' adjustment hold, we look at several other outcome variables and use the same regression equation. In particular, we are interested in the effects on the capital/labour ratio, the personnel costs, and training.

1.5.2 Two Stage Least Squares Model with Skill Shortages

Due to a lack of disaggregated migration data, we cannot directly test the relevance of the Free Movement variable for actual emigration rates from NMS. Instead we can go one step further in the causality chain and check if the EU15+4 labour mobility laws explain the increase in skill shortages as reported by NMS firms. The first-stage regression takes the following form:

$$SH_{ict} = \gamma_1 FM_{ict-l} + \gamma_2 I_{ict} + \gamma_3 J_{it} + \gamma_4 C_{ct} + \tau_t + \kappa_{ic} + u_{ict} \tag{1.5}$$

SH_{ict} is the industry-country-year measure of skill shortages. γ_1 is the coefficient of interest and reflects the marginal contribution of the Free Movement variable, given industry- and country-specific time-varying covariates $(I_{ict}, \gamma_3 J_{it}, C_{ct})$, and time dummies (τ_t). By including industry-country fixed effects κ_{ic}, we identify the Free Movement effect only from within-industry variation in the propensity to emigrate.

We run a second-stage regression, similar to (1.4), but instead of the Free Movement variable, use the instrumented measure of skill shortages. The coefficient $\hat{\beta}_1$ thus captures the productivity effect of skill shortages caused by the transitional provisions. It is identified only for industries where the legislation changes created binding skill constraints for firms.

1.5.3 Construction of the Free Movement Variable

In our model, we analyse an exogenous increase in the turnover rate due to emigration. In the data, this corresponds to the opening of the EU15+4 labour markets for NMS, which induced emigration and therefore increased the turnover rate. These openings are captured by the Free Movement (FM) variable. We construct it directly by aggregating information about EU15+4 labour mobility laws. We use it as the main explanatory variable in our baseline empirical specification and as the instrument for skill shortages in the 2SLS regression.

A country-industry-year cell makes up one observation. Industries are represented at the NACE two-digit level. The main period under consideration is from 2000 to 2014 (from the accession of NMS8 countries to the termination of all transitional provisions applied to NMS2). First, for each observation we construct a set of 15 dummies D_{cc_jit}, with each dummy corresponding to one of the EU15+4 countries, c_j. A dummy takes the value of 1 if according to the legislation of an old EU member, its corresponding industry i is open to labour migrants from a given new member state c. For example, the UK completely opened up its labour market for the NMS8 group in 2004. Therefore UK dummies for all industries for all NMS8 countries equal 1 starting from 2004. In contrast, France held the transitional provisions for the 2004-entrants until 2008. Prior to 2008, the French government applied a special job scheme, which allowed for free labour market access only in construction, tourism, and catering. France dummies for NMS8 industries take a value of 0 until 2008, except for the three mentioned sectors. Figure A.1 in the Appendix shows how the legislation dummies enter our dataset.

One of the limitations of the legislation dummies is low industry-level variation. Austria, Germany, France, Italy, and the Netherlands, for instance, did not explicitly specify which industries are open to labour migrants from new member states, but rather allowed for special job schemes in sectors that experienced skill shortages. The dummies also do not capture different capacities of EU15+4 markets to absorb migrants. To account for this, we multiply the legislation dummies D_{cc_jit} by a measure of skill shortages in a given industry of a j_{th} EU15+4 country. For this, we use the share of firms (in destination industries) reporting to be constrained by the labour factor. These data are available from the EU Commission Business Survey. This modification controls for implicit legislation changes and for differences in labour market conditions across and within industries in old EU members.[7] Easiness to find a job, which increases in sectors experiencing skill shortages, can be another important criteria for worker mobility. A possible concern with such a modification is that skill shortages in the old EU member states might not be fully exogenous to firm productivity in NMS countries, due, for example, to common technology shocks. We can control for this by including industry-specific time dummies or an average measure of skill shortages in a given industry for all EU members. Another concern is that labour demand could increase in EU15+4 industries, which after the EU enlargement had become more competitive relative to their rivals from new member states. In this case, however, one would expect to see negative tendencies in NMS firm performance already prior to the outflow of workers. We can also control for higher product-market competition by including a mark-up measure.

[7]This allows to capture, for example, a decrease in demand for foreign labour force during and after the economic crisis in 2008-2009. At this time, many labour markets were already open for NMS citizens, but effective job possibilities were limited. De-jure, only Spain reacted to the worsening of economic conditions by reintroducing restrictions for Romanian citizens in 2011.

To summarise the set of 19 dummies in a single measure, we apply special weights that reflect how strongly the opening of a particular EU15+4 labour market affects the citizens of a given new member state. It is reasonable to assume that labour migrants, for example, from Estonia were more sensitive to the opening of the Finnish labour market than the Portuguese one. One approach is to use bilateral distances between the two largest cities of each source and destination country as a measure of proximity: the shorter the distance, the larger is the weight for a corresponding EU15+4 labour market.

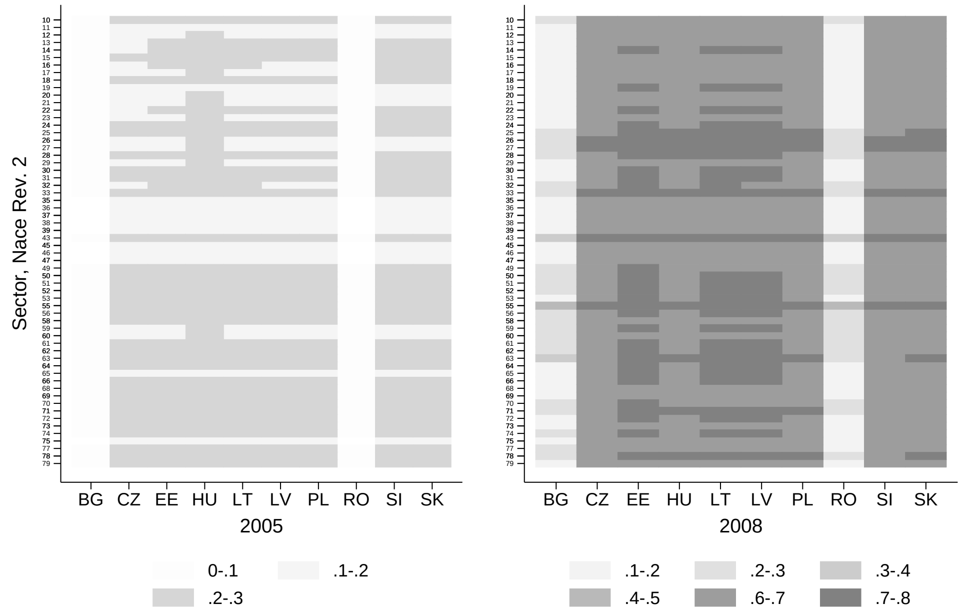
The legislation information is summarised in one variable:

$$FM_{cit} = \sum_{j=1}^{19} w_{c,c_j} \cdot D_{cc_jit} \tag{1.6}$$

FM_{cit} is the value for one observation (source country-industry-year). D_{cc_jit} denotes the legislation dummy for openness of the labour market in a j_{th} old EU member's corresponding industry for the citizens of a given source country in a given year and w_{c,c_j} denote the weights. To ensure the comparability of different versions of Free Movement variables, we standardise them to be in the range [0;1]. Figure 1.1 illustrates the variation in the Free Movement variable across industries of NMS.

To investigate the plausibility of our identifying assumption, we check if firms' outcomes prior to 2004 predict changes in the legislation over 2004-2014. We also run several placebo tests. We report the results in Section 1.7.

Figure 1.1: Variation in the Free Movement Variable



Notes: This graph shows the variation in the instrument. We compare different industries (y-Axis) in different countries (x-Axis) in 2005 and 2008. The darker the shading, the stronger these industries in these countries have been exposed to emigration.

1.6 Empirical Results

This section presents and discusses the empirical results and compares them with the model's predictions. All regressions include firm fixed effects and thus capture within-firm variation in performance as a response to changes in an industry's exposure to emigration.

1.6.1 Reduced Form Regressions

Table 1.2 presents the reduced form estimations: we regress firm outcomes directly on the Free Movement (FM) variable. We use a one-period lag for the Free Movement variable to account for some inertia between the legislation change and the migration decisions. All dependent variables are in natural logarithms, and the Free Movement variable is in the range from 0 to 1. The coefficients may be interpreted as the log point ($\simeq$ percent) change in dependent variables when the FM increases from 0 (no free labour mobility

Table 1.2: Free Movement Effect on Firm Performance, Reduced Form, Amadeus Data

	(1) TFP index	(2) TFP LP	(3) ROA	(4) Pers. costs	(5) C/L
$L.FM_{ict}$	-0.273***	-0.234***	-0.0344**	0.270***	0.172
	(0.0696)	(0.0619)	(0.0141)	(0.0628)	(0.106)
$Mark-up_{ict}$	0.212***	0.186***	0.0906***	-0.133***	-0.0824**
	(0.0526)	(0.0350)	(0.0165)	(0.0279)	(0.0404)
$L.log\_investment_{ict}$	0.00178	-0.00707	-0.00556***	0.00521	0.0243**
	(0.00680)	(0.00500)	(0.00203)	(0.00792)	(0.0106)
$L.log\_FDI\_inward_{ict}$	-0.00125	2.82e-05	-0.000435	0.00242	0.00502**
	(0.00143)	(0.00114)	(0.000509)	(0.00148)	(0.00219)
$Log\_total\_sales_{it}$	0.00350	-0.00910	0.000439	0.0516***	0.00880
	(0.0100)	(0.00913)	(0.00262)	(0.00851)	(0.0126)
$Mean\ skill\ sh._{it}$	0.0763	0.160	0.175***	0.0702	-0.0866
	(0.158)	(0.117)	(0.0384)	(0.115)	(0.169)
$L.log\_FDI_{ct}$	0.0123***	0.0108***	0.00231***	0.00819***	0.0113***
	(0.00248)	(0.00226)	(0.000617)	(0.00175)	(0.00229)
$D.log\_GDP_{ct}$	1.520***	1.301***	0.179***	0.397***	0.130
	(0.163)	(0.142)	(0.0443)	(0.0996)	(0.135)
Observations	546,661	322,938	542,500	529,567	529,567
Number of firms	108,413	71,652	107,585	105,572	105,572
R^2	0.074	0.040	0.053	0.105	0.122
Dummies	f y	f y	f y	f y	f y
Clusters	2660	2521	2630	2618	2618

Notes: The table presents reduced-form estimates of the free movement. All specifications are estimated with firm fixed effects and time dummies. Dependent variables: TFP index, TFP LP - TFP estimated with the Levinsohn and Petrin (2003) procedure, ROA - return on assets, Pers. costs - personnel costs per employee, C/L - capital-labour ratio. L.FM - Free Movement variable (distance-weighted, interacted with skill shortages in destination industries), 1 year lag. Standard errors (in parentheses) are clustered on country-industry (NACE 4-digit) level.
*** p<0.01, ** p<0.05, * p<0.1

within EU for workers qualified to work in a particular industry) to 1 (maximum exposure to free labour mobility in our sample).

For the main sample of firms, the effect of free movement on productivity is negative, which confirms the prediction of our model. The result is robust to different measures of productivity, to the exclusion of outliers (firms with sales below the 1st and above the 99th percentiles), and to the exclusion of firms that entered the market after 2002.

The maximum annual increase in the value of the FM variable in our sample is equal to 0.52 (for certain industries in Romania in 2007), while on average NMS industries experienced a maximum annual increase of 0.25. We can use this information to give a quantitative interpretation of our result. One year following the maximum increase in labour mobility, a firm's TFP drops by $0.25 \cdot 0.234 = 0.059$ or 5.9 log points. Given an average TFP of 29,500 EUR (estimated with the Levinsohn and Petrin (2003) method), this translates to annual losses of about 1,700 EUR per firm.

We can also see that firms adjust to emigration by increasing personnel costs. In our dataset, personnel costs include wages and other employee-related costs. We are thus not able to compare this aggregate data directly with our model predictions. However, the observed increase in personnel costs is consistent with more hiring and training expenses due to worker turnover. The annual increase in the Free Movement value of 0.25 would lead to $0.25 \cdot 0.27 = 0.0675$ or 6.75 log point increase in personnel costs per employee. With average annual employee costs of 7,840 EUR, this leads to additional 550 EUR per worker. The change in the capital/labour ratio is positive, but imprecisely estimated.

To confirm our results and to analyse additional variables, we perform the same regression using firm-level data from the BEEPS survey. Table 1.3 presents the reduced form estimates. BEEPS contains only a limited number of firms with available panel data. Therefore, in the reported specification we pooled firm observations together, adding firm-level covariates: lagged sales, capital, quadratic terms for firm age and lagged number of employees, share of foreign capital, share of export in sales. All regressions are estimated with country·year (c·y), country·industry(c·i), and industry·year(i·y) fixed effects. The remaining variation in dependent variables should come from country-industry-year changes in the value of the Free Movement variable. As with the Amadeus data, we find a negative effect of the EU labour market opening on firm TFP. Furthermore, we report significant increases in the share of trained employees by firms in industries, which have potentially experienced higher labour emigration. Combining this with our model predictions, it suggests that firms train more people as they increase their hiring due to turnover.

One assumption we are making to bring the model to the data is that the Free Movement variable affected average turnover in NMS industries. Using Eurostat LSF data, the results in Table 1.4 are in line with our hypothesis: industries exposed to higher labour mobility experience a decrease in average tenure (which corresponds to higher turnover).[8] The estimates are robust to the inclusion of country-specific time trends. To check for

[8] $Tenure$ can be expressed as $1/Turnover$

Table 1.3: Free Movement Effect on Firm Performance, Reduced Form, BEEPS Data

	(1) TFP index	(2) Wage	(3) Train	(4) New product
$L.FM_{ict}$	-0.541***	0.772	1.706***	-3.589
	(0.083)	(0.620)	(0.577)	(2.772)
$log\_lag\_l_{fict}$	-0.0501	-0.211***	0.0630***	0.0201
	(0.118)	(0.0347)	(0.0170)	(0.0159)
$log\_lag\_sales_{fict}$	0.215**	0.225***	0.0218**	-0.0144*
	(0.107)	(0.0221)	(0.00945)	(0.00831)
$\%\ foreign_{fict}$	-0.187	0.462***	0.179***	0.115
	(0.207)	(0.118)	(0.0637)	(0.0717)
$export\_share_{fict}$	-0.0135	0.0620	-0.105*	-0.0681
	(0.198)	(0.100)	(0.0589)	(0.0707)
Observations	1,344	5,432	5,078	2,179
R^2	0.971	0.227	0.243	0.247
Dummies	cy ci iy	cy ci iy	cy ci iy	cy ci iy
Robust	yes	yes	yes	yes
Clusters	296	591	574	290

Notes: The table presents reduced-form estimates of free movement on firm performance using BEEPS data. All specifications are estimated with country·year (c·y), country·industry(c·i), and industry·year(i·y) fixed effects. The variable $Train$ represents the share of trained workers in the total workforce. Additional firm-level covariates include lagged sales, capital, quadratic terms for firm age and number of employees, share of foreign capital, share of export in sales. $L.FM_{ict}$ represents the sum of legislation dummies, weighted by distance to a given old EU member-country and interacted with skill shortages in destination industries. Standard errors (in parentheses) are clustered on country-industry level.
*** p<0.01, ** p<0.05, * p<0.1

the presence of pre-trends, we add a one-period forward of the Free Movement variable (column 3), which turns out to be insignificant, as expected.

Table 1.4: Free Movement Effect on Tenure, Reduced Form, Eurostat Data

	(1) Mean tenure	(2) Mean tenure	(3) Mean tenure
$F.FM_{ict}$			-0.268
			(0.320)
$L.FM_{ict}$	-0.858***	-1.052***	-1.842***
	(0.144)	(0.219)	(0.272)
$L.log\_investment_{ict}$	-0.110***	-0.0899**	-0.107**
	(0.0389)	(0.0390)	(0.0422)
$log\_total\_sales_{ict}$	0.0261	0.00509	-0.00460
	(0.0907)	(0.0883)	(0.0943)
$L.log\_FDI_{ct}$	-0.172***	-0.148***	-0.166***
	(0.0336)	(0.0320)	(0.0394)
$D.log\_GDP_{ct}$	-1.200	-0.141	-3.564**
	(0.904)	(0.870)	(1.424)
Observations	1,873	1,873	1,564
Number of idc	314	314	312
R^2	0.142	0.136	0.208
Dummies	ic y	ic y	ic y
Clusters	314	314	312

Notes: The table presents reduced-form estimates of free movement on average tenure. All specifications are estimated with industry·country fixed effects and time dummies. L.FM - Free Movement variable, 1 year lag. In specification 1, we use only distance-weighted FM dummies. In specifications 2 and 3, FM dummies are interacted with skill shortages in destination industries. In specification 3, we add a forward lag of the FM variable to check for the absence of pre-trends. Standard errors (in parentheses) are clustered on country-industry level.
*** p<0.01, ** p<0.05, * p<0.1

1.6.2 Heterogeneity

In the main specification, we analyse the effect of free movement for the full sample of firms. To check for heterogeneous effects, we estimate specification 1.4 for different sub-samples of firms.

Tables 1.5 and 1.6 show the results for foreign-owned and innovating firms. The estimated effect of free movement on firm TFP is smaller compared to the full sample and loses its statistical significance. At the same time, the estimated coefficients for personnel costs and capital/labour ratios suggest that these firms adjust much stronger to the increased emigration opportunities of their workforce. Foreign-owned firms increase their personnel costs significantly more. They might be able to offer wage increases to retain workers and training to newcomers to teach firm-specific human capital. Patenting firms seem to adapt in particular through increasing the capital/labour ratio. These firms might also be able to provide an interesting work environment and have retention initiatives to keep their essential research staff. There is also evidence that innovating firms benefit from

Table 1.5: Free Movement Effect on Firm Performance, Reduced Form, Foreign-Owned Companies

	(1) TFP index	(2) TFP LP	(3) ROA	(4) Pers. costs	(5) C/L
$L.FM_{ict}$	-0.0571	-0.124	0.0191	0.396***	0.395***
	(0.0796)	(0.0805)	(0.0269)	(0.0642)	(0.110)
$Mark-up_{ict}$	0.105***	0.143***	0.0480***	-0.0509	-0.0546
	(0.0339)	(0.0357)	(0.0136)	(0.0399)	(0.0446)
$L.log\_investment_{ict}$	0.0127*	-0.0108	-0.00692*	-0.0115	-0.0274**
	(0.00768)	(0.00798)	(0.00375)	(0.0110)	(0.0140)
$L.log\_FDI\_inward_{ict}$	4.88e-06	0.000242	-0.000519	0.00119	0.00414
	(0.00170)	(0.00182)	(0.000606)	(0.00139)	(0.00260)
$Log\_total\_sales_{it}$	-0.0114	-0.0245*	-0.00649	0.0509***	0.0302*
	(0.0119)	(0.0130)	(0.00538)	(0.0121)	(0.0171)
$Mean\ skill\ sh._{it}$	-0.103	0.178	0.0425	0.0712	0.0603
	(0.139)	(0.154)	(0.0647)	(0.149)	(0.204)
$L.log\_FDI_{ct}$	0.00938***	0.00639*	0.000382	-0.00195	-0.00690*
	(0.00292)	(0.00347)	(0.00165)	(0.00301)	(0.00394)
$D.log\_GDP_{ct}$	0.809***	0.791***	0.0941	0.196	0.200
	(0.145)	(0.172)	(0.0650)	(0.134)	(0.175)
Observations	56,960	34,354	56,580	55,730	55,730
Number of firms	10,415	6,846	10,361	10,308	10,308
R^2	0.021	0.019	0.016	0.088	0.044
Dummies	f y	f y	f y	f y	f y
Clusters	1683	1489	1668	1670	1670

Notes: The table presents reduced-form estimates of the free movement effect on firm performance. The sample is restricted to firms with foreign capital. All specifications are estimated with firm fixed effects and time dummies. Dependent variables: TFP index, TFP LP - tfp estimated with Levinsohn-Petrin procedure, ROA - return on assets, Pers. costs - personnel costs per employee, C/L - capital-labour ratio. L.FM - Free Movement variable (distance-weighted, interacted with skill shortages in destination industries), 1 year lag. Standard errors (in parentheses) are clustered on country-industry (NACE 4-digit) level.
*** p<0.01, ** p<0.05, * p<0.1

reverse knowledge flows and increased research networks through their former employees (Braunerhjelm et al., 2015; Kaiser et al., 2015).

Table 1.6: Free Movement Effect on Firm Performance, Reduced Form, Firms with Patents

	(1) TFP index	(2) TFP LP	(3) ROA	(4) Pers. costs	(5) C/L
FM	-0.0702	-0.0883	-0.104***	0.256*	0.604***
	(0.127)	(0.109)	(0.0363)	(0.144)	(0.132)
$Mark-up_{ict}$	0.0460	0.125***	0.0103	-0.0768	-0.0220
	(0.0380)	(0.0385)	(0.0126)	(0.0563)	(0.0638)
$L.log\_investment_{ict}$	-0.0156	-0.0245*	-0.00272	0.0251**	0.0440**
	(0.0121)	(0.0139)	(0.00479)	(0.0117)	(0.0184)
$L.log\_FDI\_inward_{ict}$	-0.000839	-0.00584***	-0.000692	0.000865	-0.000384
	(0.00184)	(0.00203)	(0.000740)	(0.00213)	(0.00277)
$Log\_total\_sales_{it}$	-0.0147	-0.0304	0.00435	0.00534	0.00906
	(0.0180)	(0.0189)	(0.00715)	(0.0162)	(0.0252)
$Mean\ skill\ sh._{it}$	0.0703	0.284	0.0201	0.239	-0.110
	(0.154)	(0.193)	(0.0644)	(0.166)	(0.226)
$L.log\_FDI_{ct}$	0.00130	0.00194	0.00176	0.0128***	0.0151***
	(0.00351)	(0.00450)	(0.00167)	(0.00292)	(0.00398)
$D.log\_GDP_{ct}$	0.383	0.263	0.0656	-0.00291	-0.474*
	(0.253)	(0.252)	(0.0700)	(0.250)	(0.274)
Observations	20,526	13,276	20,507	19,694	19,694
Number of firms	2,812	2,165	2,812	2,769	2,769
R^2	0.113	0.037	0.120	0.128	0.266
Dummies	f y	f y	f y	f y	f y
Clusters	843	729	843	832	832

Notes: The table presents reduced-form estimates of the free movement effect on firm performance. The sample is restricted to firms with patents. All specifications are estimated with firm fixed effects and time dummies. Dependent variables: TFP index, TFP LP - tfp estimated with Levinsohn-Petrin procedure, ROA - return on assets, Pers. costs - personnel costs per employee, C/L - capital-labour ratio. L.FM - Free Movement variable (distance-weighted, interacted with skill shortages in destination industries), 1 year lag. Standard errors (in parentheses) are clustered on country-industry (NACE 4-digit) level.
*** p<0.01, ** p<0.05, * p<0.1

1.6.3 Skill Shortages Due to Emigration: 2SLS Regressions

The reduced form regressions represent the "intention-to-treat" effect. Furthermore, it is of interest to estimate the effects for those firms that were effectively constrained by the outflow of skilled workers. We consider skill shortages as an indicator for this problem. If changes in EU15+4 labour mobility legislation indeed induce higher emigration rates of the qualified workforce, we will observe increasing skill shortages as reported by firms in NMS. The measure of skill shortages is described in Section 1.4.

Table 1.7 shows the OLS results of different firm outcomes regressed on skill shortages. We find that only one measure of TFP is significantly negative, while for other measures the association appears to be zero or even positive. We believe that these OLS results are upward biased due to reverse causality and omitted variable bias. For instance, those firms that experience a positive shock are likely to be more productive and thus need more labour. They are consequently more likely to report skill shortages. In the following, we perform a 2SLS analysis, which confirms the upward bias of the OLS regression.

Table 1.8 presents 2SLS estimates with the Free Movement variable serving as an instrument for skill shortages. Comparable to the reduced form estimations, we estimate the reported models using the distance-weighted instrument. The first-stage details (FM coefficient with the standard error) are presented below the main regression results.[9]

The measure of skill shortages (share of firms in an industry, reporting to be constrained by labour) ranges from 0 to 1. The coefficient of interest thus represents the log point change in the dependent variables when skill shortages increase by 1 unit (or 100%). A one percentage point increase in skill shortages caused by the EU15+4 labour market opening thus leads to a 1.6-3.1% drop in firm TFP (depending on the measure) and a 3.0% increase in personnel costs. Comparable to the reduced-form estimates, innovating and foreign-owned companies do not experience significant decreases in TFP, but raise their personnel costs and increase their capital intensity.

[9]The reported first-stage coefficients might differ slightly from those reported in Table A.2, since some industry-year observations were dropped due to missing firm-level data.

Table 1.7: Skill Shortages and Firm Performance, OLS Regressions

	(1) TFP index	(2) TFP LP	(3) ROA	(4) Pers. costs	(5) C/L
L.skill sh.	-0.0183	-0.0751***	0.0287***	0.0964***	-0.181***
	(0.0300)	(0.0258)	(0.00856)	(0.0323)	(0.0482)
$Mark-up_{ict}$	0.336***	0.302***	0.120***	-0.0254	0.0132
	(0.0817)	(0.0537)	(0.0205)	(0.0297)	(0.0474)
$L.log\_investment_{ict}$	0.00636	-0.00531	-0.00668***	0.00391	0.0238**
	(0.00698)	(0.00573)	(0.00198)	(0.00704)	(0.00955)
$L.log\_FDI)ct$	-0.00158	-0.000693	-0.000471	0.000983	0.00242
	(0.00153)	(0.00119)	(0.000351)	(0.00128)	(0.00200)
$Log\_total\_sales_{it}$	-0.00339	-0.0161	-1.11e-06	0.0504***	0.00719
	(0.0109)	(0.0107)	(0.00232)	(0.00819)	(0.0123)
$Mean\ skill\ sh._{it}$	-0.0205	0.106	0.158***	0.128	0.0451
	(0.160)	(0.124)	(0.0365)	(0.110)	(0.167)
$L.log\_FDI_{ct}$	0.0141***	0.0130***	0.00213***	0.00772***	0.0117***
	(0.00269)	(0.00245)	(0.000669)	(0.00199)	(0.00246)
$D.log\_GDP_{ct}$	1.549***	1.294***	0.239***	-0.0605	-0.285*
	(0.171)	(0.148)	(0.0370)	(0.0940)	(0.163)
Observations	501,277	291,346	497,393	486,190	486,190
Number of firms	88,370	54,965	87,651	86,960	86,960
Dummies	y f	y f	y f	y f	y f
Robust	yes	yes	yes	yes	yes
Clusters	2377	2210	2345	2361	2361

Notes: The table presents estimations of the skill shortages effect on firm productivity. All specifications are estimated with firm fixed effects and time dummies. Dependent variables: TFP index, TFP LP - tfp estimated with Levinsohn-Petrin procedure, ROA - return on assets, Pers. costs - personnel costs per employee, C/L - capital-labour ratio. Standard errors (in parentheses) are clustered on country-industry (NACE 4-digit) level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.8: Skill Shortages as the Consequence of the Free Movement, 2SLS Regressions

	(1) TFP index	(2) TFP LP	(3) ROA	(4) Pers. costs	(5) C/L
L.skill sh.	-3.071*	-1.635***	-0.281	3.042**	2.127
	(1.631)	(0.595)	(0.187)	(1.315)	(1.872)
$Mark-up_{ict}$	0.330***	0.308***	0.119***	-0.0106	0.0247
	(0.106)	(0.0647)	(0.0230)	(0.0594)	(0.0670)
$L.log\_investment_{ict}$	0.0298	0.00582	-0.00426	-0.0161	0.00814
	(0.0216)	(0.0101)	(0.00263)	(0.0179)	(0.0224)
$L.log\_FDI\_inward_{ict}$	-0.00338	-0.00171	-0.000652	0.00275	0.00381
	(0.00331)	(0.00193)	(0.000475)	(0.00271)	(0.00362)
$Log\_total\_sales_{it}$	0.0102	-0.00712	0.00135	0.0361*	-0.00400
	(0.0226)	(0.0140)	(0.00370)	(0.0211)	(0.0200)
$Mean\ skill\ sh._{it}$	0.977*	0.654***	0.260***	-0.826**	-0.702
	(0.562)	(0.231)	(0.0665)	(0.403)	(0.706)
$L.log\_FDI_{ct}$	0.0311**	0.0238***	0.00384**	-0.0100	-0.00219
	(0.0136)	(0.00705)	(0.00150)	(0.0116)	(0.0147)
$D.log\_GDP_{ct}$	1.375***	1.236***	0.221***	0.211	-0.0718
	(0.237)	(0.162)	(0.0510)	(0.249)	(0.240)
Observations	501,277	291,346	497,393	486,190	486,190
Number of firms	88,370	54,965	87,651	86,960	86,960
Dummies	y f	y f	y f	y f	y f
Robust	yes	yes	yes	yes	yes
Clusters	2377	2210	2345	2361	2361
fs_coef	0.0988	0.147	0.0981	0.0985	0.0985
fs_se	0.0423	0.0463	0.0424	0.0435	0.0435

Notes: The table presents estimations of the skill shortages effect on firm productivity. All specifications are estimated with firm fixed effects and time dummies. Dependent variables: TFP index, TFP LP - tfp estimated with Levinsohn-Petrin procedure, ROA - return on assets, Pers. costs - personnel costs per employee, C/L - capital-labour ratio. L.FM - Free Movement variable (distance-weighted, interacted with skill shortages in destination industries), 1 year lag. Standard errors (in parentheses) are clustered on country-industry (NACE 4-digit) level. $Firststage\_coef$ is the first-stage coefficient of the instrument and $Firststage\_se$ is the standard error.
*** p<0.01, ** p<0.05, * p<0.1

We include a number of covariates to switch off demand-driven changes in the reported measure of skill shortages. Country GDP growth rates and FDI inflows (GDP_{ct}, FDI_{ct}) control for general country-specific shocks. Lagged investment ($investment_{ict}$) accounts for country-industry specific increases in skill shortages due to the expansion of existing companies or new entries. The measure of average skill shortages in a given industry in EU15 countries ($Mean\ skill\ sh_{it}$) controls for industry-specific labour demand shocks, which are common across all EU members.

For all the specifications, we report the coefficients from the first-stage regressions (where we regress skill shortages on the Free Movement variable). A complete opening (from FM=0 to FM=1) of one industry in all EU15+4 labour markets results in a 10% increase in skill shortages for firms in the corresponding industry in the NMS. The free movement coefficient is statistically significant, and the F-test rejects the null hypothesis of insignificance.

1.7 Robustness

1.7.1 Exogeneity Assumption

The identification of the skill shortages effect builds on the exogeneity assumption of the constructed instrumental variable. Variation in the Free Movement variable comes from changes in legislation, bilateral distances, and skill shortages in destinations. All three components are determined on the industry level for *old* EU member states and hence should be exogenous to country-industry-year productivity shocks or changes in other unobservables in *new* EU member countries. As a robustness check for the validity of our IV approach, we ran the first-stage regression (1.5) on another variable, which also varies at the country-industry-year level, but, in contrast to skill shortages, should not systematically react to changes in EU labour mobility legislation. In the EU Business Survey, apart from skill shortages, firms also report on financial constraints. Table A.3 presents first-stage regression results with financial constraints as a dependent variable. While for skill shortages all four IV modifications returned statistically significant coefficients, only one of them is weakly correlated with reported financial constraints. This, however, is not the modification we use in our regressions. This result reassures that the constructed IV captures labour supply shrinking due to emigration instead of other contemporaneous shocks.

1.7.2 Using Different Lags of the Instrument

In our main specification, we have looked at the effects of emigration on firm performance one year after the respective labour market opening. We have chosen a one-year lag because we expect the effects appear with a certain delay, for instance due to the decision making process to migrate, the migration preparation process and the notice period. In the following, we are looking at simultaneous effects as well as the effects up to three years before and after the sector opening.

Figure 1.2 shows firm TFP that is regressed on lagged (1, 2 and 3 year lag), simultaneous and forwarded (1, 2 and three year forward) FM values. One can see that the forwarded values are always insignificantly different from zero. This is reassuring for us, as we do not want the future sector openings to affect current firm outcomes (for instance due to anticipation). The Free Movement variable gains significance during the year of the opening but is only borderline significant. The effect becomes stronger and more significant after one year and then remains at this lower level during the following two years.

Figure A.2 in the appendix shows other firm outcomes that are regressed on lagged (1, 2 and 3 year lag), simultaneous and forwarded (1, 2 and three year forward) FM values.

Figure 1.2: Annual Treatment Effects of Free Movement on Firm TFP

Notes: Dependent variable - firm TFP, estimated with Levinsohn-Petrin procedure. The displayed coefficients correspond to the number of years before and after the largest increase in the Free Movement variable for a given industry. Year, industry, and country-fixed effects are included. Errors are clustered at the country-industry level.

1.8 Conclusion

This paper uses firm- and industry-level panel data to evaluate the effect of skilled emigration on firm productivity. To overcome the endogeneity problem, we exploit the natural experiment of the EU enlargements in 2004 and 2007. We argue that the gradual and industry-specific opening of the EU and EFTA labour markets to citizens from new member states throughout 2004-2014 has created exogenous variation in the emigration rates experienced by NMS. We show that an emigration-driven reduction in labour supply resulted in lower total factor productivity for firms in NMS. We also document an increase in personnel costs and training expenditures. This confirms the predictions of our model. Furthermore, we find that innovating and foreign-owned firms increased their personnel costs by more and experienced smaller drops in productivity. These firms have been more successful in circumventing the loss in TFP.

Our results are important both for firms and for policy-makers. Being aware of the problem helps firms to react timely and in an adequate way. Firms can benefit from active human resource strategies, focusing, for instance, on providing training and retention measures. For policy makers, the effects of migration 'are not a matter of fate, [but] to a large extent, they depend on the public policies adopted in the receiving and sending countries'[10]. The prevalence of skill shortages, for instance, justifies the need to invest in the skills of their local labour force and to mitigate search frictions. A skill upgrading of the local labour force can in the short term be addressed by providing specific training courses by public institutions and in the long term by adjusting the education system to labour market needs. Knowing that those skilled people are needed can justify the investment. An increase in local human capital might also happen in the long term due to increased incentives to invest in education, which rise with the prospect to emigrate (Beine et al. (2001)).

While the outflow of skilled workers leads to deteriorating firm performance in the short term, emigration can also create opportunities and countries can experience brain gain if they put the right policies in place. One possibility for brain gain is return migration. If companies and politicians in the new EU member states succeed in bringing back their skilled workers after some time abroad, then firms could benefit from even more experienced workers. These workers can create knowledge spillovers and bring their firms closer to the technological frontier. Another opportunity is to attract workers from other EU member states. An efficient labour agency, and especially harmonised EU-wide labour agencies, could inform workers within the EU of all EU-wide job vacancies. This might encourage unemployed workers in other EU states to search for a job in countries and industries that experience shortages. By attracting workers from other EU countries and incentivising return migration, firms in new member states could also reap the benefits of labour migration in an enlarged Europe.

This paper does not take General Equilibrium effects or aggregate effects at the European level into consideration. It might well be that while emigration reduces firm productivity in the source country, the better allocation of workers and jobs increases efficiency and welfare at the European level. We leave these questions for further research.

[10]Docquier and Rapoport (2012).

Chapter 2

Knowledge Remittances: Does Emigration Foster Innovation?*

2.1 Introduction

Remittances, the money international migrant workers are sending back from the country of employment to their home country, represent an important source of income for developing countries and hence constitute a direct benefit of emigration.[1] Furthermore, apart from financial contributions, skilled migrants can "send" back the knowledge they have acquired while working in other countries. This remittance of knowledge has the potential to increase innovation in the origin countries and bring them closer to the technology frontier, thus mitigating the negative effects of the loss of human capital due to emigration.

The number of highly educated foreigners in the OECD area now exceeds 31 million, accounting for 45 percent of the increase in the foreign born population over the last decade (OECD Database on Immigrants in OECD Countries, 2016). The number of skilled migrants has especially increased within Europe since many members of the European Union (EU) and the European Free Trade Association (EFTA) have introduced

*This chapter is based on joint work with Thomas Fackler und Yvonne Giesing.

[1] Russell (1986).

free movement for citizens of the partner countries. Given the strong increase in labour mobility and raising concerns in countries experiencing net outflows of skilled people, it is important to understand the consequences of migration. Should firms and policy-makers think and act in the context of a "global war for talent" or can the international mobility of skilled individuals make everyone better off, in particular, by stimulating cross-border knowledge flows?

In this project, we establish a causal link between labour mobility, knowledge flows, and innovation activities. By exploiting changes in the European labour mobility legislation as a quasi-experimental setting, we evaluate the effect of skilled emigration on innovation. We find that the emigration of skilled individuals increases patenting in source countries and argue that knowledge remittances can explain this positive effect. Using data on patent citations and migration flows from 32 European countries, we find that emigration increases cross-border knowledge flows. Industries that are exposed to a higher mobility of their workers start to cite patents from the emigrants' destinations more frequently than before. The international mobility of skilled workers seems to enlarge R&D networks and promote the transfer of tacit knowledge. In this way, migration enables a faster diffusion of knowledge from more to less technologically advanced countries and helps the latter to catch up.

We embed these results within the following conceptual framework. We assume a knowledge production function, where innovation (here, for instance, measured by the number of patents) is produced with the inputs of capital and labour and a certain production technology. Emigration leads to a reduction in labour and thus has a direct negative effect on innovation production. However, there might also be an indirect effect, which has often been overlooked in this discussion. International migration can increase the flow of ideas and knowledge across borders. Migrants might share knowledge about new technologies, processes, and products with their former colleagues and friends at home. This increases the stock of knowledge in the source countries and, through the recombination of ideas, positively affects innovation. The production technology thus improves and patent production can grow even if the available skilled labour is reduced. Our conceptual considerations thus suggest that migration has a negative direct and a positive indirect effect on patenting levels in source countries. Although we cannot disentangle these effects with our data, we provide empirical evidence on the total effect.

The main challenge in the empirical analysis is the endogeneity of migration flows. This could be due to reverse causality or omitted variables. To establish causality, we construct an instrumental variable (IV) for migration, using changes in labour mobility laws within Europe. These laws are adopted and enforced by the destination countries and hence can be treated as exogenous to economic conditions in migrants' source countries.

The aim of our estimations are twofold. Combining several data sources, we do not only establish a link between emigration and innovation in the source country, but also shed light on the effect on knowledge remittances, potentially driving innovation. We start by analysing the effects of international labour mobility on total patenting activity in source countries. The IV estimate suggests that a one percent increase in the number of emigrants increases patent applications by 0.64 percent in the following two years. This result is statistically significant at the one percent level and robust to controls, fixed effects, and varying lags. The effect is quantitatively more pronounced when we consider only the flows of migrants with patenting potential.

We complement the analysis of innovation activity by looking at the convergence in patenting between migrants' origin and destination industries. We limit the sample to pairs where the destination is more technologically advanced than the origin and analyse whether the difference in patenting levels changes with migration flows. This is a highly policy-relevant question, especially in the context of the European Union: Some countries may block the initiatives aimed at enhancing within-EU labour mobility by arguing that the outflow of skilled people will further augment the asymmetries between richer and poorer member states. Contrary to this argument, though, our results show that patenting differences between origins and destinations decrease in the number of emigrants. Hence, emigration can promote convergence to the innovation level of more advanced economies.

To establish the channel for the positive impact of emigration on innovation, we link emigration to reverse knowledge flows, that is the transfer of knowledge from migrants' destinations back to their origins. While skilled emigrants do not patent in their home country anymore, they can stimulate knowledge and technology diffusion, thus improving the production technology in the origin country. Common to the innovation literature, we use cross-border patent citations as a proxy for knowledge flows. The regression analysis relates the number of citations to a particular destination country with the number of migrants that currently work there. We find evidence that knowledge flows from destination to origin indeed increase with migration: the 2SLS regressions yield an elasticity of knowledge flows to emigration equal to 0.59.

Our project relates to two broad strands of the literature. The first one investigates the effects of labour mobility on innovation. Several papers have established a positive effect of migration on patenting in destination countries. Kerr and Lincoln (2010) use random visa allocations to find causal effects for the US. Bosetti et al. (2015), Parrotta et al. (2014), Ozgen et al. (2014) and Niebuhr (2010) focus on European countries and establish cultural diversity as one of the main channels to generate new ideas and innovation. The effect of migration on source countries received less attention. Kerr (2008) and Choudhury (2015) find that source countries benefit from knowledge flows and return migration and consequently increase patenting and innovation. Kaiser et al. (2015) provide firm-level evidence by looking at worker mobility within Denmark. They find that hiring new knowledge workers increases a firm's patenting activity. Interestingly, the former employers of these workers also increase patenting, which can be explained by reverse knowledge flows. Braunerhjelm et al. (2015) conduct a similar analysis with a matched employer-employee dataset from Sweden and also show that both the receiving and the sending firms benefit from the mobility of knowledge workers. The effects are stronger for interregional mobility. We contribute to this literature by providing causal evidence that emigration leads to an increase in patenting. We thereby confirm what Kerr (2008) and Choudhury (2015) showed for China and India in a very different context and using another methodology. As we have comparable patenting data for source and destination countries, we can extend this result and show that emigration leads to a catch-up process.

The second strand of the literature analyses the determinants of knowledge flows. Starting with the seminal contribution by Jaffe et al. (1993), these studies have established that knowledge is localised beyond the effects of agglomeration. Later studies focused on international knowledge spillovers (Hu and Jaffe, 2003; Jaffe and Trajtenberg, 1999), showing that knowledge takes time to cross country borders. Thompson and Fox-Kean (2005) challenge the approach by Jaffe et al. (1993) and point out that intra-national localization effects are not robust to a finer technology classification. However, even with their more conservative estimations, the international localization remains significant. Singh and Marx (2013) investigate whether advances in communication technologies and lower costs of travelling reduce the localisation of knowledge over time. While they find evidence for a reduction in the significance of state borders in the US, their results show that the effect of international borders has even strengthened over time. Few studies so far analysed the impact of international migration on cross-border knowledge flows.[2] Kerr (2008), for instance, studies the role of skilled immigrants in the U.S. and finds that immigrants form ethnic scientific networks that enhance the technology transfer to source countries.

We extend this literature on knowledge flows to the European context using an identification strategy that allows for a causal interpretation. We build a unique dataset by merging comparable migration data for 32 European countries with European patent data and find evidence for knowledge flows. Due to our unique European enlargement setting, we are able to estimate causal effects of labour mobility independently of other integration events by exploiting different opening times for trade, FDI and migration. We find that the positive effect of mobility on knowledge remittances is particularly high for migrants with patenting potential and is robust to a variety of specifications and samples.

The paper is organised as follows. The next section describes a conceptual framework to guide our empirical analysis. Section 3 outlines the data, followed by Section 4 that presents the empirical specification and describes the instrument. Section 5 discusses the results. Section 6 suggests knowledge flows as the channel. Section 7 provides robustness checks and Section 8 concludes.

[2] Prior literature on the international knowledge flows has focused on trade, foreign direct investment and R&D accessibility (MacGarvie, 2005, 2006; Peri, 2005).

2.2 Conceptual Considerations

This paper analyses the effects of emigration on innovation in source countries. As there are two opposing effects, our storyline becomes clearer if we support it with some conceptual considerations. The considerations are based on a classical knowledge production function as introduced by Griliches (1979) and further developed by Jaffe (1986) and Jaffe (1989). We augment the knowledge production function with emigration. The concept illustrates two opposing effects: a reduction in knowledge production due to a decreasing skilled labour force vs. an increase due to a better production technology induced by knowledge flows and technological spillovers.

We assume a simplified knowledge production function of the form

$$Y = Af(K, L_s). \tag{2.1}$$

K is a measure of relevant capital available for research and development such as laboratories and equipment. L_s stands for skilled labour and A measures total factor productivity (efficiency of knowledge production). In our case A describes how well labour and capital can be combined to produce the knowledge output Y and captures factors that are not explicitly modelled, such as the knowledge stock on which researchers can build. To measure the output Y, we refer to patents, as is common to the literature.

The direct effect of emigration, in this setting, is a reduction in L_s. Due to the outmigration of skilled people, less workers are available for the production of innovation in the source country. The innovation output Y should thus decrease.

However, there is a second indirect effect of emigration that works through the total factor productivity A. After emigration, workers send back knowledge to their home countries. For instance they may transmit technological information and ideas back to their previous employer through communication with former colleagues. This employer becomes better at producing innovation, which is reflected in an increasing A.

Theoretically, it is unclear whether the negative direct or the positive indirect effect prevails. This depends on several other characteristics such as the industry, the technology, and the innovation process. Consequently, it is even more important to gain this knowledge from a rigorous empirical assessment of the question. Using patent data as a measure of innovation output Y and controlling for various other factors corresponding to K and components of A that are unrelated to the stock of knowledge, our empirical specification is able to identify this net effect.

2.3 Data Description

We create a unique dataset by merging comparable migration data for 32 European countries with European patent data. The dataset has four dimensions: origin region[3], destination country, industry (two-digit, NACE Rev. 2), and year. The dependent variables of interest are the number of patent applications (by origin-industry-year) as a proxy for innovation and the number of cross-border citations (by origin-destination-industry-year) as a proxy for knowledge flows. The main explanatory variable is the annual number of emigrants from a given origin currently employed in a given destination industry.

The ideal migration dataset would contain precise data on migration flows, disaggregated by origin and destination (countries and employing industries), skill level, and occupation. In the absence of such a dataset, we use the second-best data from Eurostat Labour Force Surveys (2000 - 2014). These are harmonised surveys, which take place annually in all EU countries, Iceland, Norway and Switzerland and cover around 5% of national populations. The surveys provide demographic information on individuals, including their current country of residence, region of origin (EU15+4, NMS10, NMS2 or Other),

[3]Here and in the following text "region" refers to the region to which Eurostat's LFS data aggregate migrants' origin countries: EU15+4 (EU15 and EFTA), NMS10 (new member states in 2004), NMS3 (Bulgaria, Romania and Croatia) and all other countries. The fact that the EU3 region consists of Bulgaria and Romania, which joined the EU in 2007, and Croatia, which followed only in 2013, adds further imprecision, as we cannot tell from the data how many emigrants from this region came from Bulgaria and Romania and were able to take advantage of the EU's right to free movement already.)

education level, occupation, and currently employing industry.[4] We thus obtain the stock of migrants by year, region of origin, destination country, and destination industry. In addition, we can use the information by education level (university degree, vocational degree, or below) and by occupation (two-digit, ISCO) to identify the stock of migrants with patenting potential.[5] The available dataset has several limitations. We can only observe the region of migrants' origin instead of the country. This means that we cannot differentiate between different 2004 accession countries but have to treat them as one region (NMS10). Similarly we have to treat Romania and Bulgaria as one region (NMS2). Furthermore, as we do not observe the origin industry of a migrant, we assume that it is the same as the current industry at the destination. Besides, we cannot identify flows of return migrants. These limitations result in high observational noise and might bias our estimations towards zero.

To construct the instrument for migration flows we use changes in the European labour mobility legislation. We obtain the relevant information from the Labour Reforms database, prepared by the European Commission, which we complement with information from national legislations of the destination countries. Our baseline dataset covers the years from 2000 to 2012, this period encompassed several changes to European labour mobility as described in more detail in Subsection 2.4.2.

The data on innovative activity and knowledge flows come from the EPO's Worldwide Patent Statistical Database (PATSTAT, 2014 Autumn Edition).[6] We are able to assign patents to industries (two-digit NACE Rev. 2) via the International Patent Classification

[4] EU15+4 include 15 pre-2004 EU member countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom) + 4 EFTA countries (Iceland, Liechtenstein, Norway, Switzerland). NMS10 include countries that joined the EU in 2004 (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, Slovenia, Malta, and Cyprus) and NMS2 include countries that joined the EU in 2007 (Bulgaria and Romania).

[5] We assign a dummy called *patenting potential* to migrants working in occupations "Managers" and "Professionals" (ISCO codes: 11, 13, 21, 22, 23, 25, 31, 32, 35).

[6] Patents and patent citations are imperfect measures for innovation and knowledge flows and have been criticised for example by Duguet and MacGarvie (2005). Yet, these are the best proxies, which are available over long periods of time and comparable across the countries we study.

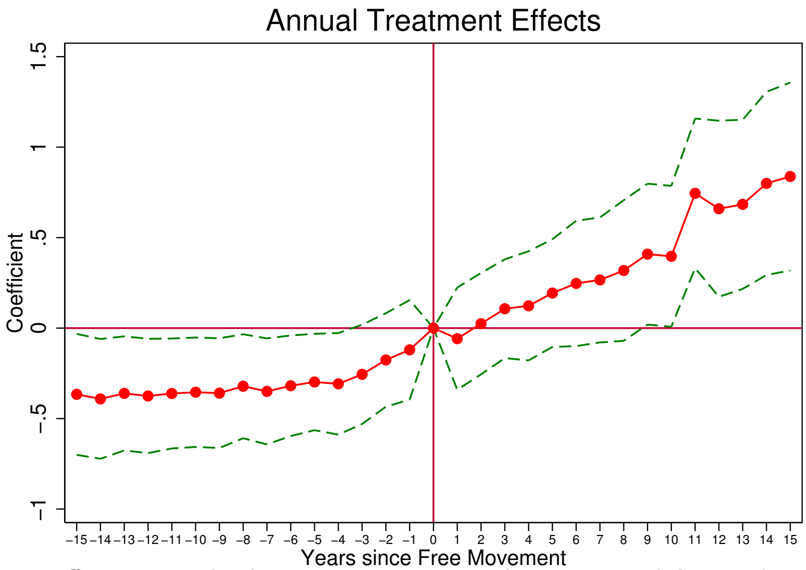
(IPC) of patents.[7]. We then aggregate patent applications by country, industry, and year and patent citations by patenting country, cited country, industry, and year. In our dataset, *patenting country* corresponds to the origin country of migrants, while cited country corresponds to their current destination. To assign patents to countries, we use the PATSTAT information about the location of patent inventors and applicants, which are usually the organisations employing the inventors. Since a patent can have several inventors, it may be assigned to multiple countries if it is the result of an international collaboration. In these cases, we assign a share of the patent to each country that is proportional to the share of co-inventors from that country. The causes and consequences of such collaborations have been studied by Kerr and Kerr (2015). Through this assignment of patents to the inventors' countries it is possible to link a patent with the location of all the patents that cite it.

Figure 2.1 motivates the subsequent econometric analysis: cross-border patent citations (a proxy for knowledge flows) significantly increase following the introduction of free labour mobility between a pair of countries. This figure mirrors the response of migration flows to changes in labour mobility regulation within Europe (Figure B.1 in the Appendix).

We complement the dataset with several important control variables: bilateral industry-specific FDI flows (provided by Eurostat), GDP and bilateral trade flows (from CEPII). By combining these different data sources, we can draw conclusions about the effects of international migration on patenting in the origin countries and establish reverse knowledge flows as the channel, while controlling for possible fixed and time-varying confounders.

[7] In order to assign four-digit IPC classes to industries, we use the concordance table provided by Eurostat in Appendix 1 of the publication "Patent Statistics: Concordance IPC V8 - NACE REV.2", published in October 2014 and last accessed on 21 November 2016.
https://circabc.europa.eu/sd/a/d1475596-1568-408a-9191-426629047e31/2014-10-16-Final%20IPC_NACE2_2014.pdf

Figure 2.1: Cross-Border Patent Citations, Annual Treatment Effects of Free Labour Mobility



Notes: Annual treatment effects on cross-border citations in patent applications around the introduction of free movement (1965-2014). The regression includes year and country-pair fixed effects. Standard errors are clustered at the citing country level.
Source: PATSTAT, European Commission, own calculations.

2.4 Econometric Specification

In our empirical analysis we first provide causal evidence for the effect of emigration on patenting in origin countries. Second, we link this effect to the increase in knowledge flows. We obtain the elasticities of patenting and cross-border citations to migration using OLS and 2SLS approaches. In the latter, the variation in migration flows is generated only by the exogenous changes in labour mobility laws over 2000-2012. Our baseline regressions include the sample of all patenting European countries. Besides, we provide separate estimates for a sub-sample of Eastern European countries, which were affected the most by the changes in labour mobility over the analysed period.

2.4.1 Baseline Regressions

Patenting in Origin Countries

We start by analysing the effects of emigration on total patenting in the origin countries. For this, we aggregate the data at the origin, industry, and year level. Because we do not have detailed country-of-origin data, we use the region of migrants' origin: EU15+4, NMS10 and NMS2. The dependent variable is the number of patent application in a specific origin, industry, and year. The explanatory variable is the number of emigrants from a specific region that work in the same industry but in other European countries. We estimate the following fixed-effects regression:

$$Y_{oiy} = \beta_1 M_{oiy-l} + \beta_2 X_{oiy} + \phi_y + \phi_{oi} + \epsilon_{oiy} \tag{2.2}$$

where o denotes the region of origin, i the two-digit industry, and y the year. Y_{oiy} is the log number of patent applications in a given region and industry. M_{oiy-l} is the log number of emigrants from an origin region, currently working in a given industry.[8] l stands for the lag between migration and patenting. The coefficient β_1 captures the elasticity of patenting to migration. X_{oiy} contains time-varying controls: a dummy for EU membership, trade inflows, and FDI inflows. ϕ_y and ϕ_{oi} denote time and origin-industry fixed effects. ϵ_{oiy} is the error term. The identifying variation thus comes from the within origin-industry changes in the number of emigrants and patent applications. To account for a possible endogeneity bias, we complement the OLS estimations with the 2SLS results, where we instrument migration with changes in labour mobility legislation. We describe the instrument in more detail in Section 2.4.2 below.

Patenting Asymmetries between More and Less Advanced Countries

We go one step further and analyse whether migration increases or, on the contrary, lowers patenting asymmetries between more and less advanced economies. On the one hand, agglomeration effects and the resources available for research could lead to richer destinations specializing even more on their comparative advantage, thus hindering convergence. If we assume that skilled migrants move from less innovative to more innovative places, labour mobility can increase patenting asymmetries despite some positive effects on the origin. On the other hand, through the migrants working abroad, industries at origins can get access to the frontier knowledge from more advanced economies. This can

[8] Here and in all other specifications, before taking logs we add 1 to each observation. This transformation ensures that we do not lose observations with zero values.

increase innovation efficiency in origin industries and can allow a faster catch-up process with the technology leaders. Hence, patenting asymmetries between destinations and origins of migrants might decrease. We empirically evaluate the effect of migration on patenting asymmetries with the following regression:

$$log(\frac{P_{diy}}{P_{oiy}}) = \beta_1 M_{odiy-l} + \beta_2 X_{1oy} + \beta_3 X_{2dy} + \beta_4 X_{3odiy} + \phi_y + \phi_{odi} + \epsilon_{odiy} \tag{2.3}$$

The level of observation is origin-destination (od) pair, industry (i), and year (y). The dependent variable $log(\frac{P_{diy}}{P_{oiy}})$ is the log difference in patent applications between the destination and origin industries. The main explanatory variable is M_{odiy-l} - the log number of migrants from origin o working in industry i in destination d. l stands for the lag between migration flows and patenting. The coefficient β_1 shows whether the patenting asymmetries increase or decrease in migration. In this specification we can also control for time-varying origin- and destination-specific effects ($X_{1oy}, X_{2dy}, X_{3odiy}$): the total number of patents at origin, the total number of patents at destination, the total number of patents in a given industry, a within EU dummy (equals one when both origin and destination are EU members), the ratio of GDP per capita between destination and origin, bilateral industry-level FDI, and trade flows. ϕ_y and ϕ_{odi} denote time and origin-destination-industry fixed effects. ϵ_{doiy} is the error term. The coefficient β_1 is thus identified solely through the variation in the number of emigrants within an origin-destination-industry. General changes in patenting at origin and destination cannot confound the results. As with specification 2.2, we estimate OLS and 2SLS regressions.

Knowledge Flows

Further, we investigate one potential channel behind the effect of migration on innovation: knowledge flows. One speaks of knowledge flows whenever a researcher or an inventor builds on the work done by others to create ideas or to solve a specific technological problem. A common way to track knowledge flows is to use citations data (Jaffe et al., 1993). This approach assumes that a citation to a particular patent or a publication reflects the usefulness of the knowledge contained therein for further work. To determine the effect of migration on knowledge flows we estimate the following empirical model:

$$Y_{odiy} = \beta_1 M_{odiy-l} + \beta_2 X_{1oiy} + \beta_3 X_{2diy-l} + \beta_4 X_{3odiy} + \phi_y + \phi_{odi} + \epsilon_{odiy} \tag{2.4}$$

As in specification 2.3, the level of observation is origin-destination (od) pair, industry (i), and year (y). The outcome of interest Y_{odiy} represents the log number of cross-border citations. M_{odiy-l} is the log number of migrants from origin o working in industry i at destination d. l stands for the lag between migration flows and patenting. We focus on reverse knowledge flows, i.e. knowledge flowing from destination to origin countries of migrants. Hence, Y_{odiy} represents citations to patents from destination countries by new patents at origin.[9] For example, $Y_{PL/BEiy}$ counts citations by Polish patents in industry i, filed in year y, to existing Belgian patents. It proxies the knowledge flows from Belgium to Poland. $M_{PL/BEiy-l}$ represents the number of Polish migrants in Belgium, currently working in industry i. The coefficient β_1 captures the elasticity of citations to migration. In our example, it shows the percent change in the number of citations from Poland to Belgium if the number of emigrants from Poland to Belgium increased by 1 percent.

To avoid mechanic effects from the general increase in patenting at origin or destination industries, we control for the number of patent applications in the origin industry (X_{1oiy}) and for the lagged number of patent applications in a destination industry X_{2diy-l}. X_{3odiy} denote other controls: a within EU dummy (equals one when both origin and destination countries are EU members), the total number of patents in a given industry, the bilateral FDI, and trade flows. ϕ_y and ϕ_{odi} denote time and origin-destination-industry fixed effects. ϵ_{doiy} is the error term. We again run both OLS and 2SLS regressions.

2.4.2 Instrument for Migration Flows

Even though we control for many observable factors and have a number of fixed effects in the baseline OLS regressions, an endogeneity problem might still arise. Estimates could be biased, for instance, if reduced patenting at the origin forces inventors to leave. To

[9] We consider citations in patent publications and date patents with their application filing date.

avoid this problem, we use changes in the labour mobility laws in Europe as a source of exogenous variation for migration flows.

The freedom of movement for workers is a policy chapter of the acquis communautaire of the European Union and represents one of the four economic freedoms: free movement of goods, services, labour and capital. According to the Article 45 of the Treaty on the Functioning of the EU, "freedom of movement shall entail the abolition of any discrimination based on nationality between workers of the Member States as regards employment, remuneration and other conditions of work and employment." In practice, it means that there are no restrictions (such as quotas on foreign workers) or additional bureaucratic procedures (such as obtaining a work permit or a permission from the local authorities) related to the employment of foreign citizens. This right primarily concerns the citizens of the EU and EEA member states who, starting from 1958, have gradually introduced free labour mobility towards their partner countries.[10]

In our project, we exploit two episodes of changes in the free labour mobility in Europe. First, in 2004 all EEA countries introduced free movement for the citizens of Switzerland. Switzerland responded with a symmetric measure in 2007.[11] Second, a special scheme has been in force following the EU enlargements in 2004 and 2007. For up to seven years after the accession, old EU members could restrict the access to their labour markets for citizens of new member states. While some countries kept the restrictions for the whole period, some provided easier labour market access only in certain industries, and some opened up their entire labour markets directly upon the accession. When imposing restrictions the countries had to apply them to the whole group of NMS from the same entry year. Therefore, they could not target labour mobility laws at the citizens of some particular states. Iceland, Liechtenstein, Norway, and Switzerland applied the transitional provisions towards the accession countries in the same way. These labour mobility laws created variation in the migration flows between European countries on the country, industry, and year level. Table B.1 in the Appendix provides an overview of the

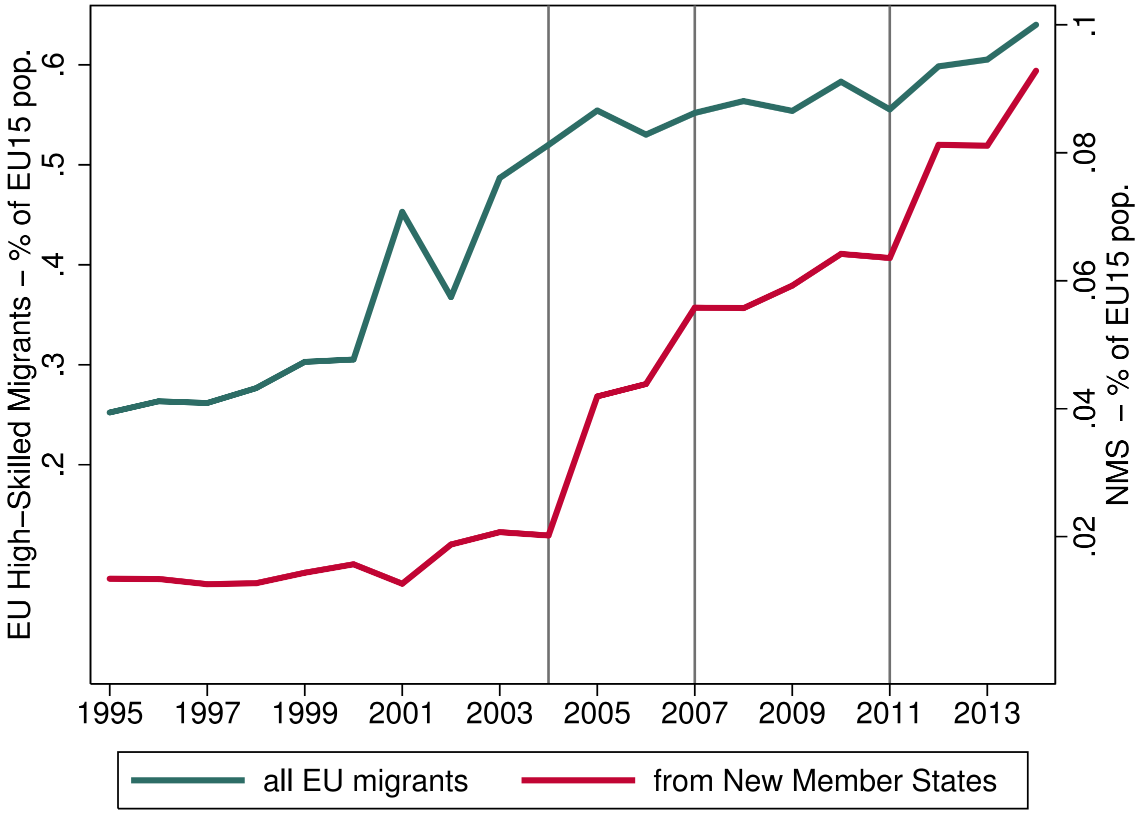
[10]Norway and Iceland exert this right since 1994. Liechtenstein exerts this right since 1995, but imposes a permanent quota for all EEA citizens.

[11]However, as a result of the "Against mass immigration" initiative, Switzerland is scheduled to impose permanent quotas on residence/work permits for citizens of all EEA countries except Liechtenstein, starting from 2017.

precise opening dates of countries and industries. Importantly for the identification, these changes to labour mobility did not coincide with other integration events (free movement of capital and goods).

Figure 2.2 shows the spikes in migration from NMS during the initial opening in 2004, when countries such as the UK, Sweden, and Ireland opened their labour markets and in 2011 when all transitional provisions for the 2004 accession countries where abolished and Germany, for instance, fully opened its labour market.

Figure 2.2: High-skilled Migration in Europe



Notes: The graph shows the share of high-skill migrants (born in other European countries) in the EU15 population. Source: Eurostat.

We can thus instrument real migration with exogenous labour mobility legislation. The first-stage regression takes the following form:

$$M_{odiy} = \gamma_1 FM_{odiy-1} + \gamma_2 FM_{odiy-2} + \gamma_3 FM_{odiy-3} + \gamma_4 X_{odiy} + \nu_y + \nu_{odi} + u_{odiy} \quad (2.5)$$

FM_{odiy-l} is an indicator variable, which is equal to one if a specific industry i in a destination country d is open for labour migrants from a country o in a given year y. We include a one, two and three year lag to allow for the delayed effect. In our sample this indicator changes only for origin and destination pairs with either Switzerland or new EU member states. As these migration flows might be different, we show separate results for migration from only Eastern Europe in every case. X_{odiy}, ν_y, and ν_{odi} are the same controls and fixed effects as used in the baseline OLS specifications. When using the instrument for the patenting regressions (specification 2.2), we aggregate the values of the free movement variable by origin, industry, and year.[12] In this case, the FM variable can be interpreted as the exposure of a given origin-industry (oi) to free labour mobility of its workers.

When constructing the free movement dummies, we take into account the fact that many old EU members did not explicitly specify which industries are open to migrants from the NMS, but rather allowed for special job schemes in sectors that experienced labour shortages. In case of such implicit exceptions, we set the free movement dummy equal to 1 and multiply it by a measure of labour shortages in a given industry of an old EU member state. As such measure, we use the share of firms (in the destination industries) reporting to be constrained by the factor labour. These data are available from the European Commission Business Survey. To account for possible endogeneity (arising, for instance, when labour shortages are reported in industries that grow faster in all EU countries), we control for the overall number of patent applications in a given two-digit industry (aggregate over all European countries).

2.5 Results

In this Section, we first show the effects of migration on total patenting at the origin. Second, we provide evidence that emigration can reduce patenting asymmetries between less and more advanced economies. We show OLS as well as 2SLS results. First-stage and

[12]For each origin region we have 31 free movement indicators corresponding to 31 possible destinations. We aggregate them to one measure by using proximity weights (the inverse log distances between the two largest cities of two countries.)

reduced form regressions are provided in the Appendix. Our baseline sample includes all patenting European countries. In addition, we show separate estimations for the sub-sample of Eastern European countries.

2.5.1 Migration and Patenting

This Section shows that the emigration of labour increases overall innovation, measured by the number of patent applications per year in a region. As the migration data only allow us to estimate the effect of emigration at the region level, we aggregate the free movement variable by industry and region of origin: EU15+4, NMS10, and NMS2. The aggregated FM measure approximates the number of countries to whose labour markets an inventor in a certain industry and region of origin had access to and is normalised to be between 0 and 1, where 1 corresponds to full access to all EU15+4 countries.

The first three columns of table 2.1 show the baseline OLS regressions and the last three columns show 2SLS regressions, which use the labour mobility legislation as an instrument for migration.[13] Columns 1 and 4 estimate the relationship between the overall number of emigrants and the number of patent applications from inventors in that region. These regressions show that workers' migration to other EU member states has a significant and positive effect on patenting in the regions of origin. As both variables are measured in logarithms, the coefficient can be interpreted as the elasticity: the effect in the IV estimation in column 4 suggests that a one percent increase in the number of emigrants in an industry causes patent applications in the region of origin to increase by 0.6 percent. The 95% confidence interval for the elasticity is between 0.37 and 0.91. If we consider the average number of emigrants in the year 2004 (2459 emigrants) and the average number of patent applications 2 years later (255 applications) for new EU member states per industry, this implies that about 1 to 2 additional applications result from 25 additional

[13]Note that the right to free movement was not symmetric due to a one-sided transition period, e.g. workers of old EU member states have been able to move to new EU member states as a rule earlier than the other way round. Thus the instrument varies also with the direction of migration and we observe variation in emigration and patenting over time for pairs of origin region and industry. We cluster on the origin-industry level to account for autocorrelation in the regressions in table 2.1. When we consider asymmetries and citations, there is additional variation depending on the destination country, such that we cluster on the origin-destination-industry level.

Table 2.1: Patent Applications and Migration, OLS and 2SLS

	(1) OLS Patents	(2) OLS cit. weighted	(3) OLS Patents	(4) 2SLS Patents	(5) 2SLS cit. weighted	(6) 2SLS Patents
L2.Migrants	0.0994***	0.0949**		0.637***	0.903***	
	(0.0259)	(0.0420)		(0.139)	(0.199)	
L2.Migr.pat.potential			0.0572			1.175***
			(0.0420)			(0.332)
in EU	-0.262***	-0.298***	-0.296***	-0.112	-0.0729	-0.406**
	(0.0903)	(0.0752)	(0.0844)	(0.157)	(0.205)	(0.164)
L2.Trade flow	1.634***	2.535***	2.124***	-0.679	-0.945	3.325***
	(0.348)	(0.432)	(0.342)	(0.607)	(0.877)	(0.724)
L2.FDI inflow	2.03e-05**	3.15e-05**	2.10e-05**	1.16e-05	1.84e-05	2.34e-06
	(9.82e-06)	(1.22e-05)	(8.06e-06)	(2.07e-05)	(2.90e-05)	(1.12e-05)
Observations	383	383	383	383	383	383
Region industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	53	53	53	53	53	53
F				6.517	11.29	7.285

Notes: The regressions in this table estimate the relationship between the migration flow out of a country and innovation in that country. The first three columns are estimated with OLS and the last three column use a 2SLS estimation with our instrument based on free movement legislation. The dependent variables are the number of patent applications in an industry and origin region in a year or, in columns 2 and 5, the citation-weighted patent applications (i.e. patent applications + forward citations to these patents). Patent application numbers and citation-weighted counts, number of migrants and trade flows are taken in natural logarithms (more precisely, for variable x we use $\log(x+1)$ to include observations where $x = 0$). The sample includes all EU members and countries in the European Free Trade Association. All specifications include year and region-industry fixed effects. Robust standard errors are clustered at the region-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

emigrants.[14] Note however, that this number only includes migrants in industries that were matched to the patent data, i.e. in which there is patenting. Furthermore, the number of patent applications in 2006 we have used for this calculation already includes the additional applications, such that the number of additional patents is likely to be lower. Despite the noise and the level of aggregation in our data the regressions are able to reject that there is a negative effect.

The second and fifth columns of table 2.1 use citation-weighted patents as the treatment variable, i.e. the number of patent applications plus the number of citations to these patents in a region, industry, and year. The number of later patents building on and therefore citing a patent is often used as a measure of quality.[15] The citations for more recent cohorts in our sample are subject to truncation, which is controlled for through

[14]One percent of 2459 emigrants is about 25 and 0.37% (0.91%) of 255 applications is 0.94 (2.32).

[15]The relationship between citations and the social value of an invention has been documented in a case study on Computed Tomography scanners in Trajtenberg (1990). A more recent study by Moser et al. (2016) finds a robust correlation between citations of hybrid corn patents and the improvement in yield reported in field trial data.

year fixed effects. As the coefficients are similar, we conclude that the quality of patenting has not deteriorated. Thus, merely a higher propensity of inventors in origin regions to file patents as a result of European integration does not seem to be the driver of the effect. Of course, the number of later patents citing a patent (forward citations) is only a rough measure of quality and may be affected by emigrants spreading information about their home countries' latest technologies abroad as well. Nonetheless, a higher number of forward citations would likely be associated with a greater benefit of source countries' innovations, since they indicate that more follow-on innovation built on them.

Columns 3 and 6 differ from the other regressions in table 2.1 in the migration variable, which here includes only emigrants with patenting potential. Whereas the OLS regression shows a smaller and insignificant partial correlation, the coefficient in the IV regression is larger than the corresponding coefficient for all migrants in column 4.

The OLS estimate is likely to be downward biased due to omitted variables and reverse causality. If there is an omitted variable bias in the OLS regressions that is negatively correlated with emigration and positively with patenting levels, then the OLS estimate is downward biased. This is very likely and could be driven, for instance, by management quality. A good manager might lead to a good work and research environment. This results both in high patenting levels and low emigration from this firm and consequently biases the OLS estimate downward. Moreover, we might encounter reverse causality in the OLS regressions. If higher patenting levels lead to less migration, then we observe a negative relationship between the two variables that goes in the other direction. As a consequence, the OLS estimator is smaller than it should be and thus downward biased.

Tables B.2 and B.3 in the Appendix provide the first stage results and the reduced form that complement the 2SLS results analysis. One can see that the instrument is highly relevant in the first stage and that the overall effect of the three lags for the free movement variables sum up to a positive effect.

Table B.4 in the Appendix provides the same table with the restricted sample of NMS10 countries (2004 accession years). Due to the level of aggregation in the migration data, the 2SLS effects are not significant. Importantly we find no evidence of a significant negative effect, which would be expected if the loss of human capital dominated.

2.5.2 Migration and Convergence

While the results of the previous Section suggest that emigration can positively affect innovation at the origin, this Section investigates whether this positive effect is enough to reduce patenting asymmetries between less and more advanced economies or whether international migration still benefits knowledge production at destination countries more. This analysis is relevant for policy discussions about benefits and costs of free labour mobility in Europe. Furthermore, the results in this Section serve as a robustness check for the effects found above. When analysing asymmetries we use all four dimensions of our dataset: origin, destination, industry, and year, and can therefore control for unobserved origin- and destination-specific time-varying changes, which could bias our estimates of patenting elasticity to migration in Section 2.5.1.

To have a clear direction of migration flows from less to more advanced economies, we restrict the sample to the origin-destination pairs, where destinations are EU15+4 countries and origins are new EU member states. In addition, in our baseline sample we consider origin and destination pairs with Switzerland as a destination and other EU15+4 countries as origins. We also show the results for migration from Eastern Europe only, and the results are consistent. For each industry and year, we construct an asymmetry measure as the log difference between the amount of patent applications at destinations and origins.

On average, destination industries file more than three times the amount of patent applications compared to origins. As expected, the patent quality of the former is also higher. We then regress the asymmetry measure on the number of migrants. Table 2.2 presents OLS (columns 1-3) and 2SLS (columns 4-6) results. The OLS coefficient of migration is slightly positive, but is not statistically significant. This may be caused by the bias due to higher migration outflows from more problematic industries. Another reason is that once we move to the more disaggregated level, we introduce more noise in the migration data (more missing and zero observations). This especially concerns already disaggregated migration data by skill and occupation. 2SLS estimates, however, suggest that emigration allows origin industries to catch up to the patenting level of destinations: a one percent increase in the number of migrants reduces patenting asymmetries by 0.30

percent (column 4 and 5 in 2.2). The coefficient for migrants with patenting potential is much larger in magnitude, but is imprecisely estimated (see column 6). Overall, the regressions' results fit the framework of a patent production function with decreasing returns to skilled labour: a marginal increase in patent production at destination (due to the immigration of skilled labour) is smaller than the marginal increase in patenting at origins (due to the increase in patenting efficiency).

Table B.6 in the Appendix presents the results from the same specifications but estimated on a restricted sample with new EU member states as origins and EU15+4 as destinations (thus excluding emigration from EU15+4 to Switzerland). The obtained coefficients are slightly smaller in magnitude, but still significant. Table B.7 in the Appendix shows the reduced form results, where instead of migration figures we use the bilateral free movement dummies. One of the drawbacks of our migration data is the large amount of missing observations, which could be either due to the effective absence of migrants or to misreporting.[16] This raises external validity issues to our estimations in terms of a generalisation to all European countries. Therefore, the most interesting results of Table B.7 are in columns 5 and 6 where we present the coefficients from the regressions over the whole sample of origin and destination pairs. The number of observations increases multiple times, yet the coefficients for the free movement dummies are very close to the estimates from the baseline sample. Moreover, most coefficients are more precisely estimated due to improved power: we note that EU membership, higher bilateral trade flows and FDI also help the convergence.

While interpreting the regression coefficients, we implicitly assume that migrants stay within the same industry. This is reasonable, as for skilled migrants the losses associated with changing the industry are substantial. Hence, they are more likely to seek employment in the same sector in the destination countries. If the assumption would not hold for some industries, how would this affect our estimations?[17] Suppose there are two industries: L and M in Poland and Belgium. The Polish migrants from industry L move

[16]For example, due to missing migration data we have to drop all observations with Germany as a destination country.

[17]There are pairs of NACE industries, between which inventors may indeed be likely to move, for example between "26 Manufacture of computer, electronic and optical products" and "27 Manufacture of electrical equipment".

Table 2.2: Convergence in Patenting Levels and Migration, OLS and 2SLS

	(1) OLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents	(2) OLS $log(\frac{P_{diy}}{P_{oiy}})$ cit. weighted	(3) OLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents	(4) 2SLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents	(5) 2SLS $log(\frac{P_{diy}}{P_{oiy}})$ cit. weighted	(6) 2SLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents
L2.Migrants	0.0319	0.0376		-0.305**	-0.334**	
	(0.0223)	(0.0276)		(0.146)	(0.158)	
L2.Migr.pat.potential			0.117**			-1.831
			(0.0575)			(2.212)
Patents, origin	-1.220***	-1.391***	-1.206***	-1.207***	-1.376***	-1.419***
	(0.0762)	(0.0817)	(0.0753)	(0.0883)	(0.0946)	(0.281)
Patents, dest	1.066***	1.105***	1.069***	1.058***	1.096***	1.021***
	(0.0713)	(0.0908)	(0.0717)	(0.0777)	(0.0978)	(0.0894)
Within EU	0.00806	-0.0884*	0.0109	0.0194	-0.0759	-0.0180
	(0.0483)	(0.0531)	(0.0487)	(0.0520)	(0.0572)	(0.0635)
GDP_d/GDP_o	-0.173	0.400	-0.197	-0.188	0.384	0.173
	(0.316)	(0.367)	(0.319)	(0.338)	(0.394)	(0.530)
L3.Trade flow	-0.0791	-0.0236	-0.0718	-0.0281	0.0326	-0.113
	(0.0629)	(0.0799)	(0.0622)	(0.0679)	(0.0867)	(0.0827)
L3.FDI flow	0.000575	-0.000443	0.000380	-0.000116	-0.00120	0.00254
	(0.00668)	(0.00668)	(0.00671)	(0.00783)	(0.00786)	(0.00926)
Observations	2,946	2,946	2,946	2,864	2,864	2,864
R-squared	0.486	0.551	0.486	0.424	0.500	0.325
Origin-dest-industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	582	582	582	500	500	500
F				83.92	122.8	76.50

Notes: The dependent variable is the natural logarithm of $Patents_{dest}/Patents_{origin}$. Number of migrants, number of patents (in origin and destination countries), GDP ratio between destination and origin, FDI, and trade flows are in natural logarithms. The sample includes all EU and EFTA members. All specifications include year and origin-destination-industry fixed effects. Robust standard errors are clustered at the origin-destination-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

to Belgium to work in industry M. Empirically, we observe $M_{BE/PL/M/y}$ to increase. The inflow of the skilled Polish workers in the Belgian industry M raises its innovation activities (or in the worst case, does not affect them). The performance of the Polish industry M is likely to remain unchanged. The asymmetry measure $log(\frac{P_{BE/M/y}}{P_{PL/M/y}})$ either increases or at most stays the same, which goes in the opposite direction of the reported effect. We thus might underestimate the magnitude of the effect.

2.6 The Channel: Knowledge Flows

Having established that emigration leads to an increase in patenting, we want to analyse one potential channel in more detail: knowledge flows. This Section shows that migrants stimulate knowledge flows from their new destinations to their countries of origin.

Table 2.3 presents the baseline OLS and 2SLS results. The dependent variable is the log count of citations by patents in the origin to the destination country. This dependent variable proxies the knowledge flows due to emigration. In the baseline estimations, we allow for two-year lags between the time of migration and the citations in the patent applications. The results are similar for a one-year lag but slightly weaker. Importantly, given the structure of the dataset, we can account for origin-industry and destination-industry shocks. A possible threat to identifying the coefficient of interest would arise if destination industries, which experienced a positive patenting shock, started to attract more workers from other countries. A higher supply of patents from this destination would also mechanically increase the amount of citations to this country. We can control for such an effect by including the number of patent applications in the destination industry (with a three year lag).[18] In a similar way, we control for the number of patent applications in the source country. The migration effect is identified from the within origin-destination variation in the migration stocks and the count of cross-border citations. Since both dependent and explanatory variables are in natural logs, the coefficient represents the elasticity of cross-border citations to the number of migrants.

[18] As a rule of thumb, it takes about three years for a patent to be granted.

Table 2.3: Citations to Destination Industries, OLS and 2SLS

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
L2.Migrants	0.0334*	0.0269		0.799***	0.588***	
	(0.0170)	(0.0167)		(0.213)	(0.225)	
L2.Migr.pat.potential			0.0638*			2.916
			(0.0348)			(2.302)
Patents, origin		0.191***	0.192***		0.174***	0.192***
		(0.0237)	(0.0238)		(0.0268)	(0.0310)
L3.Patents, dest		0.0435***	0.0431***		0.0427***	0.0219
		(0.0145)	(0.0145)		(0.0158)	(0.0236)
Within EU		-0.0501	-0.0471		-0.0698*	0.0468
		(0.0378)	(0.0379)		(0.0416)	(0.0845)
L3.Trade flow		0.00665	0.0119		-0.104*	0.00902
		(0.0392)	(0.0390)		(0.0617)	(0.0440)
L3.FDI flow		0.00780	0.00711		0.0126**	-0.0134
		(0.00493)	(0.00495)		(0.00570)	(0.0203)
Observations	7,299	7,287	7,287	7,136	7,124	7,124
R-squared	0.080	0.095	0.095			
Origin-dest-industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	1322	1320	1320	1159	1157	1157
F				20.29	22.20	14.98

Notes: The dependent variable is number of citations from a region and industry to another country per year. Citation counts, number of migrants, total number of patent application in origin and destination industries, FDI and trade flows are taken in natural logarithms. The sample includes all EU and EFTA members. All specifications include year and origin-destination-industry level fixed effects. Robust standard errors are clustered at the origin-destination-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

In the first column, we regress the citations on the overall number of migrants M_{odiy}, year, and origin-destination-industry fixed effects; in column 2 we add additional time-varying controls; in column 3 we use the number of migrants with patenting potential as the main independent variable. OLS results suggest a positive association between migration and cross-border citations. The estimated coefficient for migrants with patenting potential is robust to all controls and is twice as large compared to the overall migration stock.

Columns 4 to 6 of table 2.3 show the 2SLS results that yield quantitatively larger elasticities than the OLS. A one percent increase in emigrants induces a 0.59 percent growth in cross-border citations to their origins. Table B.8 in the Appendix summarises the results for the sub-sample where new EU member states are origins and EU15+4 are destinations. Despite the reduction in the sample size, the main 2SLS coefficients remain positive and significant. The reduced form regressions (Table B.9 in the Appendix) are also consistent. When we estimate the reduced form for the whole sample of origins and destinations, the free movement coefficients gain significance and quantitatively remain almost identical to those from the baseline sample. This indicates that some of the insignificant results in the baseline regressions (as, for example, the imprecise coefficient for migrants with patenting potential) are mainly due to power problems with noisy migration data.

Previous research has emphasised the role of communication between moving researchers and their former colleagues at the previous employers (e.g. Braunerhjelm et al., 2015; Kaiser et al., 2015). To test whether the channel they have found for inventors moving between firms within a country is also the primary channel of international knowledge flows in our setting, we exclude the inventor's network. To do this, we exclude citations between inventors and all employers (applicants) and other inventors they are listed with on a patent application at any point in time. Table 2.4 reports the results for the restricted sample. While the coefficients change slightly, they remain positive and significant. Thus only a small part of the effect seems to be driven by the inventors' close network. Knowledge flows that this method could not capture include, for example, if a student at an Eastern European university moves on to work in Western Europe, filing patents for the first time and citing her professors' research. However, the sizable effect that remains suggests that wider spillovers play an important role.

Citations are not always added by the inventor himself but can also be added by the examiner. One worry might thus be that examiners become more aware of research done in other European countries and that they consequently are more likely to add citations from these countries. Alternatively, the effect might be driven by the fact that more patents are filed at the European Patent Office, where examiners may be more likely to add references to foreign patents than at the national offices.[19] This concern is addressed by Table B.11, which shows the results only with citations that were added by the applicant (rather than the examiner or a third party) according to PATSTAT and we can see that there are no qualitative changes.[20]

[19] The latter concern is also addressed in table B.10, where only citations among patents filed with the USPTO are included, such that European institutional changes should not affect the results.

[20] In unreported regressions, we limit citations further to only include those that are marked in PATSTAT as applicant-added and, additionally, where citing and cited patents are both priority patents filed at the USPTO. The results are qualitatively similar despite the fact that only less than 1% of citations remain.

Table 2.4: Citations to Inventor's Network Excluded

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
L2.Migrants	0.0346**	0.0276*		0.797***	0.631***	
	(0.0171)	(0.0167)		(0.197)	(0.203)	
L2.Migr.pat.potential			0.0464			3.878*
			(0.0314)			(2.355)
Patents, origin		0.174***	0.175***		0.155***	0.177***
		(0.0226)	(0.0226)		(0.0262)	(0.0358)
L3.Patents, dest		0.0353***	0.0350***		0.0344**	0.00732
		(0.0134)	(0.0134)		(0.0150)	(0.0247)
Within EU		-0.0496	-0.0471		-0.0711*	0.0787
		(0.0355)	(0.0356)		(0.0403)	(0.0867)
L3.Trade flow		0.0296	0.0350		-0.0893	0.0312
		(0.0389)	(0.0387)		(0.0597)	(0.0484)
L3.FDI flow		0.00982**	0.00925*		0.0150***	-0.0179
		(0.00481)	(0.00482)		(0.00569)	(0.0224)
Observations	7,299	7,287	7,287	7,136	7,124	7,124
R-squared	0.077	0.091	0.091			
Origin-dest-industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	1322	1320	1320	1159	1157	1157
F				19.18	19.85	10.82

Notes: In this table, citations within the network of the inventor are excluded, i.e. citations from applicants and inventors with whom the cited inventor has patented at any point in time. The dependent variable is the number of citations from a region and industry to another country in a year. Citation counts, number of migrants, total number of patent application in origin and destination industries, FDI and trade flows are taken in natural logarithms. The sample includes all EU and EFTA members. All specifications include year and origin-destination-industry level. Robust standard errors are clustered at the origin-destination-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

There are a number of ways for the knowledge flows to occur in practice. One possibility is that emigrants increase the awareness of new knowledge or technologies. This could happen, for example, if emigrants inform their former colleagues or if they meet at conferences. Another possibility is that researchers in the source countries are aware of new knowledge or technologies but need to learn how to use the tacit knowledge embedded in them. A close contact among former colleagues might spur the transfer of tacit knowledge. Additionally return migration can increase innovation in source countries. Often, emigrants return to their home countries after several years abroad and create start-ups or contribute to innovation in other ways.[21]

[21] Our time frame of analysis is more likely to reflect the increasing awareness of new technologies or the transfer of tacit knowledge.

2.7 Robustness

To confirm the validity of the results, we conducted a number of robustness checks. We find that the increase in patenting activities as a result of emigration is not driven by different pre-trends or institutional changes in the European patenting system.

One way to check the validity of the results is to examine pre-trends. If our results are valid, the coefficient of interest should be zero if we regress citation patterns on future labour market openings. Figure 2.1 in Section 2.3 shows the annual treatment effects for the regression of cross-border citations on the free movement variable. We look specifically at bilateral citations during the time period 15 years before and 15 years after free movement between two countries has been established. The data we use for this graph are based on patent applications over the 50 year period from 1965 to 2014. The regression includes year dummies and country-pair fixed effects to take out trends. The figure shows that there is no significant change in cross-border citations in the years prior to the establishment of free labour mobility.[22] This is reassuring and increases the credibility of our results. It becomes clear that the effect only starts to gain momentum at the time of the introduction of free movement and builds up over the following years.

One might also worry that the institutional framework of registering patents has changed in the EU, especially in the context of EU enlargement and the European Patent Convention. We thus restrict the sample to patents that have been registered at the United States Patent and Trademark Office (USPTO). Table B.10 in the Appendix shows the results. While we have fewer observations, the qualitative results remain the same. The results thus do not seem to be driven by institutional changes in Europe.

[22] Note that this graph uses country-level data, such that the free movement indicator only switches to 1 once all sectors are open. Some of the (insignificant) increase before time 0 may thus be due to the partial openings during the transition periods, which we exploit in the main part of the paper for identification.

2.8 Conclusion and Policy Implications

This paper analyses the effects of emigration on patenting levels in source countries. We find that countries that experience emigration increase their level of patenting. We further suggest that this has led to a catch-up process that brought origin countries closer to the technology frontier. We also find that the international mobility of people has increased technology and knowledge spillovers as evidenced by cross-border patent citation in the respective countries. Specific channels that could have fostered the knowledge spillovers are the transfer of tacit knowledge, the increased and improved network of inventors and return migration.

One policy recommendation that directly follows from these findings is that the EU could benefit from further facilitating migration within Europe. As there are no more legal barriers to free labour mobility, hindering factors are mostly language and administrative barriers. The EU could reduce these barriers by ensuring the recognition of foreign qualifications and the promotion of language courses at all age levels. In this way, the EU can exploit the full potential of migrants both for destination and source countries.

Another policy implication is to ease skilled migration to Europe from outside the European Union. This could be achieved by easing the access to European labour markets and the recruitment of highly qualified foreign workers. While the Blue Card has been a step in this direction, its scope could be increased to obtain a higher impact and administrative barriers should be reduced. For those skilled migrants that are already in Europe, for instance skilled refugees, labour market restrictions should be lifted to ease labour market integration. If these people can be integrated fast into qualified positions without a loss in human capital, the innovation system would greatly benefit.

We have shown in this paper that source countries can benefit from emigration through knowledge flowing back into the country. These benefits of knowledge flows can be maximised by facilitating research networks with emigrated inventors, for example by organising conferences in the origin countries. Furthermore, governments can design programmes to actively keep the diaspora engaged and by encouraging and facilitating return migration. Return migrants bring back the newly gained knowledge and many times create their own start-ups which can foster development in the countries of origin.

While this paper establishes that knowledge flows mitigate the negative consequences of emigration, further research is needed to shed light on the precise way these knowledge flows are created and characterised. Do migrants possess tacit knowledge that flows with people but cannot be transferred by other means? Or do migrants enlarge the R&D network and create better awareness of technologies in other countries? Do migrants have a competitive advantage in negotiating licensing fees with their country of origin? These open questions may guide further research in order to better understand how to increase knowledge flows and maximise their benefits.

Chapter 3

From Friends to Foes: National Identity and Team Performance

3.1 Introduction

In a globalised world, with lower communication costs and fewer obstacles to labour mobility, diverse teams have become common in many organisations. These teams exist because the benefits of pooling diverse individuals together outweigh higher communication and coordination costs. Yet, managing a diverse team is challenging. The team members might identify themselves as being part of different social groups, for instance, based on nationality or political preferences. What if, suddenly, these social groups find themselves on the opposite sides of the barricades, for example, due to a political or an ethnic conflict? Even if the conflict does not directly influence the team's environment, it can change identity prescriptions of the team members, spur stronger in-group/out-group feelings, lower trust, and thus reduce the team's productivity. Is there causal evidence for such effects in real-life teams? Do identity-related conflicts exert the economic impact by harming the performance of previously efficient teams and by preventing otherwise profitable future collaborations? Which channels transmit the identity effect?

This project uses a novel and relevant setting to address the above questions. I study the consequences of an exogenous political conflict for the performance of teams on GitHub, the world's largest open-source platform. In particular, I exploit the unexpected conflict between Russia and Ukraine following the annexation of Crimea in March 2014 and analyse the impact of this conflict on the online collaboration between Ukrainian and Russian programmers. Using microdata from GitHub, I show that the conflict exerted a strong negative effect on Ukrainian-Russian collaboration and argue for the causal role of identity. I then investigate how the conflict affected specifically the performance of projects with mixed Russian-Ukrainian teams. Further, I discuss and link to the empirical evidence two potential channels, through which identity could affect team performance: first, by hindering communication and coordination within a team and second, by changing the taste of some team members for projects initiated by a "hostile" social group.

The GitHub setting provides several advantages for my analysis. First, GitHub was specifically designed as an online platform for collaboration. It offers powerful infrastructure that allows multiple users from different locations to coordinate their efforts while working on the same project. Therefore, GitHub represents an ideal laboratory to study team interactions and performance. Second, GitHub features an environment with very low information asymmetries: individual contributions of team members are directly observable by other GitHub users, thus, alleviating free-riding concerns in teams. From a researcher's perspective, this also limits the number of available interpretations for the observed results. Third, since its launch in 2007, GitHub has grown to include more than 14 million users. Many technology companies use GitHub for both open-source and commercial projects. Hence, a virtual team on GitHub should be representative of a real high-skilled team working on complex tasks. Fourth, prior to the 2014 conflict, both Ukrainian and Russian programmers were well represented at the platform. Because of a similar (or the same) language and technical backgrounds, Ukrainians and Russians had often worked together on various projects. After March 2014, there were no major interruptions in the access to the Internet in Ukraine or Russia nor could the introduced bilateral sanctions between the countries prevent programmers from working on the US-based online platform. For my analysis, therefore, I can treat this conflict as exogenous to the virtual working environment on GitHub. Finally, I work with a detailed dataset, generated from GitHub databases. The dataset contains information about all users, projects, and related activities ever registered on the platform over 2012-2015.

My empirical results show that the negative effect of the conflict on Ukrainian and Russian collaboration on GitHub has been mainly driven by the drop of Ukrainian contributions to Russian projects. Over the first six months of 2014, the amount of contributions decreased sharply relative to the reference period in 2013. The empirical approach comprises the difference-in-difference and triple-difference methods and controls for the time-varying quantity and quality of Russian projects and for the activity of Ukrainian programmers. I conduct several robustness checks, confirming that the economic reasons, such as higher transaction costs or career expectations, cannot fully rationalise the observed changes.

Rather, the drop in collaboration concords with the identity-based explanation (following Akerlof and Kranton (2000)). The conflict with Russia created a new prescription for the Ukrainian identity: "to not collaborate with the enemy country". For Ukrainian programmers, collaborating with Russians now means violating this new prescription and results in additional identity costs. Hence, the conflict could shift Ukrainian programmers' taste for teammates or projects and thus distort existing and future collaborations, profitable from a purely economic perspective.

Intuitively, the identity effect should be larger for Ukrainian programmers with initially stronger national identity. Although the strength of the national identity cannot be observed, I can construct a proxy for it by identifying the preferred language of Ukrainian programmers. Upon registration on GitHub, users report their name and location. I exploit this information to associate a stronger Ukrainian identity to programmers, who used the Ukrainian rather than the Russian version of their name and location.[1] In line with the identity hypothesis, programmers identified as having a stronger national identity reduce the amount of their contributions to Russian projects significantly more relative to other Ukrainian programmers.

[1]In practice, most Ukrainians speak both Russian and Ukrainian fluently, so this choice cannot signal possibly higher communication costs with Russians. Furthermore, at least prior to the 2014 conflict, there was no negative stigma attached to those speaking Russian.

The shock to Ukrainian identity could affect team performance through two channels. First, it could increase (psychological) communication costs between Ukrainian and Russian team members, leading to lower coordination within a team and, thus, lower amount of contributions from both Ukrainian and Russian programmers. I find some evidence for this channel by looking at the project-level data. I note an almost symmetric reduction in the amount of contributions from both Ukrainian and Russian programmers that used to work together on projects from other countries. However, for projects owned by either Russians or Ukrainians the response is asymmetric. While Ukrainian programmers sharply reduce their contributions to Russian projects, Russian programmers keep their activity level on Ukrainian projects almost intact. This asymmetry can point to the presence of identity-related costs, other than poorer peer interaction. The second possible "identity" channel could make Ukrainian programmers less willing to identify with projects owned by Russians. On the intensive margin, Ukrainian programmers would decrease their contributions to Russian projects, as they start valuing the success of these projects less. On the extensive margin, fewer Ukrainian programmers would now want to become members of Russian projects. The data confirm presence of both effects.

This project contributes to several strands of literature. First, it relates to the literature on team performance. Bandiera et al. (2005, 2009) and Mas and Moretti (2009) emphasised the importance of social preferences and provided evidence that socially-connected teams may be more resilient, for instance, to free-riding or to coordination problems. Lazear (1999) and Prat (2002) discussed the optimality conditions for team diversity. Together these studies focused on the importance of peer interaction and outlined the trade-off related to diverse or socially-distant teams: potential productivity gains against collaboration challenges. My project adds to this literature by providing empirical evidence for one of the risks that diverse teams face: external events can exacerbate social differences within teams and hence inhibit the performance. Second, the project also contributes to the theoretical and experimental literature on social identity and group performance (Charness et al., 2007; Chen et al., 2014; Chen and Li, 2009). These works have established the importance of a common group identity for the individual decision-making and suggested it as a tool for improving cooperation and coordination in diverse teams. When people, apart from their own utility, value the group's or, in my context, the project's payoff, it increases the overall effort. Programmers in my sample reduce their contributions to or never join projects with a conflicting national identity. This effect seems to be stronger than the drop in collaboration due to poorer interactions between team members, as the first strand of literature would suggest.

Third, by investigating the consequences of an interstate conflict on workplace behaviour, the project contributes to the literature on microeconomic effects of international or ethnic tensions (Fisman et al., 2013; Hjort, 2014; Ksoll et al., 2010; Marx et al., 2015; Rohner et al., 2013a,b). The research in this area, apart from documenting the economic effects, has tried to identify the underlying mechanisms: the external political pressure, trust, national or ethnic preferences and social norms. The evidence so far has been mixed. Fisman et al. (2013) evaluate the consequences of Sino-Japanese political tensions and find a strong negative stock-market response for firms that depended on bilateral trade relations. Their evidence points to the large role of political pressure rather than that of potential consumer animosity in shaping the effect. Rohner et al. (2013a) argue for another channel: they show that a military conflict in Uganda enforced ethnic identity and decreased generalized trust toward "out-group" people, which in its turn inhibited inter-ethnic cooperation and slowed down the subsequent economic recovery. Hjort (2014) examines the effect of ethnic divisions on team productivity at a Kenyan flower firm. Using micro-level data and a convincing identification strategy, he argues that an external political conflict increased the taste for discrimination of a rival ethnic group, thus resulting in the misallocation of resources and lower productivity of heterogeneous teams.[2] The analysis performed in my paper complements the existing empirical work. Direct political pressure, though possible in other economic settings, could hardly impose restrictions on the open-source cooperation on GitHub. It is also unlikely that the conflict altered the beliefs of Ukrainian programmers regarding trustworthiness or quality of Russian colleagues. As in Hjort (2014), it seems that the shift in social preferences (the need to comply with the new identity prescription) can explain the drop in the Ukrainian-Russian collaboration.

[2] In recent work, however, Berge et al. (2016) conduct a series of "lab-in-the-field" experiments in Kenya and find no evidence for ethnic preferences. The authors call for more careful linking of the observed ethnic bias in behaviour and the actual ethnic bias in preferences.

The paper is organised as follows. Section 2 describes the setting: it outlines the conflict between Russia and Ukraine, discusses the particularities of online collaboration on GitHub, and presents the dataset. Sections 3 and 4 explain the empirical strategy and report the main results. Section 5 discusses the channels of the identity effect. Section 6 concludes.

3.2 The Setting

3.2.1 The Russian-Ukrainian Conflict

After the collapse of the Soviet Union, Ukraine had still preserved close ties to Russia. In 2012, Russia was Ukraine's largest trading partner accommodating 25% of all Ukrainian exports and accounting for 32% of all Ukrainian imports.[3] As of the latest census in 2001, Russians, constituting 17% of Ukrainian population, represented the second largest ethnic group. Almost 30% of all Ukrainian citizens considered Russian their primary language.[4] Practically, however, the vast majority of the Ukrainian population has been bilingual, and since the Ukrainian independence in 1991, no major ethnicity- or language-based conflict had ever occurred.

The internal crisis in Ukraine burst out in November 2013, when the former president Viktor Yanukovych suspended preparations for an association agreement with the European Union. This unilateral decision, which contradicted the previously promised policy, became the impetus for the series of so-called Euromaidan protests in November - February 2014. The protesters called for the resignation of Yanukovych and the government, with primary accusations being corruption and abuse of power. The protests culminated in February 2014 with intensive fights between protesters and police. In late February 2014, Yanukovych fled the country and a new government led by the opposition leaders came into office.

[3] http://wits.worldbank.org/CountryProfile/en/Country/UKR/Year/2012/Summary.

[4] According to the latest available census data: http://2001.ukrcensus.gov.ua/eng/results/general/.

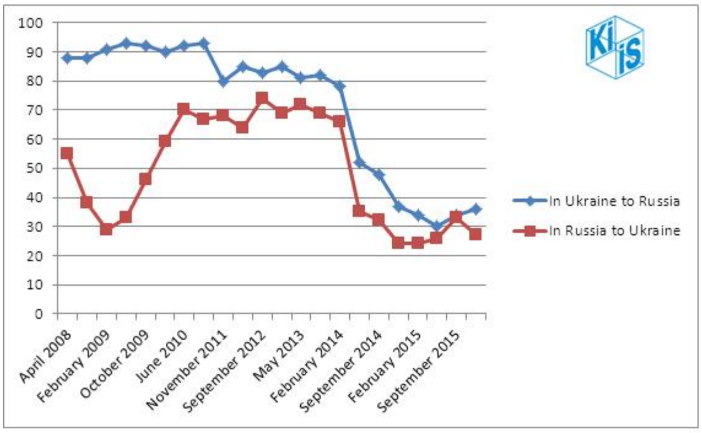
Meanwhile, anti-Euromaidan demonstrations started in Crimea and Eastern regions of Ukraine. There, the apparent impetus was given by the Parliament's vote (on the second day after Yanukovych fled) to abolish the Law on the State Language Policy. This law was adopted under the rule of Yanukovych and effectively gave Russian the status of the second state language in regions where at least 10% of the population reported Russian as their mother tongue. Although this decision never came in force, it easily became one of the headlines of the pro-Russian propaganda. In the end of February 2014, during clashes between pro- and anti-Euromaidan supporters in Crimea, troops without insignia (who turned out to be Russian military) silently took control over the Supreme Council of Crimea and the major military bases on the peninsula. After that, the Crimean parliament voted for a new government led by Sergey Aksyonov, the leader of the Russian Unity party. On March 16, 2014, the new Crimean government held a Referendum. According to the officially reported figures, the turn-out reached 83.1%, and 96.77% of the voters supported the government's proposal first, to declare independence from Ukraine and second, to become part of Russia.

The Referendum marked the beginning of the severe crisis between Ukraine and Russia. Shortly after the Crimean events, pro-Russian governments in Luhansk and Donetsk regions declared their independence from Ukraine. In April 2014, active military actions (Donbass War) between the separatists and the regular Ukrainian army began. In February 2015, after the peace negotiations in Minsk, both parties agreed to seize fire. Effectively, however, irregular fighting still takes place with both sides accusing each other of violating the peace agreements. According to the report of UNCHR, the war has already resulted in over 10,000 casualties, more than 22,000 wounded, and 1.4 million displaced persons.[5] Although the international institutions classify the war in the Eastern Ukraine as an "internal conflict", Russia is often accused of sending own troops and of supplying the separatists. In January 2015, the Ukrainian Parliament voted on officially calling Russia the "aggressor country".

For the Ukrainian population, the conflict with Russia came unexpectedly. Figure 3.1 illustrates changing attitudes of Ukrainians and Russians toward each other. While in

[5] http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=20496LangID=E

Figure 3.1: Dynamics of the Positive Attitudes of the Russian and the Ukrainian Populations Towards Each Other



Notes: The graph plots the percent of the Ukrainian and the Russian Population who have positive and very positive attitudes towards the other country. Source: Kiev International Institute of Sociology

February 2014, more than 80% Ukrainians had still reported to have a positive attitude toward Russia, in the next poll conducted in June 2014, this number had dropped to 50%. As of the latest report available (November 2015), the share of Ukrainians with a positive attitude toward Russia stays below 40%.

Unsurprisingly, the conflict had direct economic consequences. In 2015, Ukrainian exports to Russia dropped to constitute 12.1% of all Ukrainian exports, whereas the share of Russian imports declined to 20%.[6] Bilateral economic sanctions and increased tariff protection introduced and reinforced throughout 2014-2015 had directly contributed to the decline. Hence, for general economic collaboration, it is impossible to separate the effects related to changes in people's attitudes (identities) from those imposed by political decisions. I try to overcome this problem by exploiting the setting where external factors, such as sanctions, should not play a role - the online open-source platform GitHub.

3.2.2 GitHub

GitHub is a web-based repository hosting service, launched in 2007. It offers the infrastructure to store and to share software projects and provides several collaboration features such as code review, bug tracking, feature requests, task management, and wikis. As of

[6]http://stat.wto.org/CountryProfile/WSDBCountryPFView.aspx?Country=UA

April 2016, GitHub reports having more than 14 million users and more than 35 million repositories (projects), making it the largest host of source code in the world. The projects on GitHub can be initiated by individuals as well as by companies. The products may be aimed at end users or at developers. Users can choose between public and private repositories. GitHub provides the latter on the paid basis and allows restricting access to the general public. Project's founders can choose from a variety of licences to protect their code and to stipulate sharing rules.[7]

To begin working on GitHub, a person first registers as a user. During the registration, users choose their login and optionally report their name, location (in most cases, city), company, and biography. To start a project, a user creates a repository to store all source codes and related materials. Every project can have only one owner. The project owner then invites other GitHub users to collaborate, by offering them to become project members. Alternatively, an interested user can signal his willingness to become a project member by starting to contribute to the project (McDonald and Goggins, 2013). Project members have the right to copy (in GitHub slang: fork) the code to their own repositories, directly modify the source code (in GitHub slang: commit) in the master repository, open and close issues related to the project. They can also remove themselves from the project without the agreement of the project owner. All other users, who are not project members, can copy the project's source code, report bugs and other issues, and *suggest* their own modifications (in GitHub slang: pull requests). The project owner or other project members then review these proposals and decide on whether to accept them or not. If a pull request is approved, the proposed modification is merged to the source code. Every GitHub user can observe who, when, and how much contributed to a project, provided the project is public. The profile page of every user displays projects he/she contributes to, together with the timeline of commits. Moreover, upon interest, users can directly examine written codes of each other. This feature makes GitHub a collaboration environment with low information asymmetry and low free-riding opportunities. Quality and quantity of contributions by each project member or external contributer can be directly observed. Apart from the collaboration environment, GitHub also offers some

[7] The available licences range from GNU General Puplic License, which literally make the source code open to anyone to the more restrictive Apache Licenses.

features of a social network. Users can get updates on the activities of other users they choose to "follow". Similarly to the "like" button on the Facebook, users can "star" an interesting project on GitHub or "watch" it to receive project updates.

Motivations of GitHub users differ.[8] Economic research, starting with the works by Lerner and Tirole (2001, 2005) and followed, among others, by Belenzon and Schankerman (2008) and Hergueux and Jacquemet (2015) aimed at identifying motivation of open-source contributors. Programmers can be driven by pure economic incentives (software companies pay for working on the open-source or the platform is used to launch own new product), career concerns (contributing to the open-source increases programmer's visibility), utilitarian needs (developing software for own purposes), reciprocity, or pure altruism. Whereas, motivations of open-source contributors vary, the utility from their work should relate to the success of the projects they are contributing to. The commonly accepted measures of success are the number of commits, the number of committers, and the number of forks (Kalliamvakou et al., 2014; McDonald and Goggins, 2013). Apart from that, GitHub combines a variety of activity indicators including stars, forks, commits, follows, and page views to construct ranks of projects and users. The ranks are updated in real-time and are visible to every one interested.[9]

3.2.3 Dataset

The main dataset is the data dump from the GitHub platform. I work with a snapshot of the GitHub database as of June 2016.[10] The database records all activities, which happen on the platform, such as user registration, project creation, adding commits, issues, comments, etc.[11] In the analysis, I use several tables from this database. The table "Users" contains all users ever registered on the GitHub. The table "Projects" records all repositories created at the GitHub. The tables "Commits" and "Issues" track activities

[8] It should be noted, that substantial amount of GitHub users do not really collaborate, but rather use the platform for storage purposes. Consequently, many GitHub projects have only one committer. I exclude these cases from my analysis

[9] https://github.com/trending

[10] downloaded from http://ghtorrent.org/

[11] The activities are only observable for public repositories.

related to projects and users, such as modifications of source code, adding comments, exchanging messages, bug tracking and fixing. I can also observe pull requests and their approval/decline, stars, and followers. Through the unique identifiers of users, projects, commits, and other events I can link the tables together. As the dataset is huge (over 300 GB), I make several restrictions to construct the working data sample. First, I consider only projects with at least two contributors and only those users who have at least one commit. Second, I select only those projects and events, which are related to users from the countries of interest (Ukraine, Russia, and several countries for the control group, which will be discussed later). I determine nationality based on self-reported user locations and country codes provided by GitHub.[12] Third, I aggregate all activity events (per user and project) by month. In order to control for overall activity levels of particular projects and users, I also calculate their total monthly activity levels (number of commits, issues, etc.). Table C.1 in the Appendix summarizes the main data tables used in the analysis.

Most events in my dataset are automatically recorded by GitHub. Therefore, the measurement error related to the timing of different events (user registration, project creation, commits, etc.) is small. For each activity, GitHub generates a timestamp. Sometimes, dumps of huge databases such as GitHub suffer from "holes" in data (for example, due to a connectivity problem during the data dump, some observations may be missing). However, I do not expect this problem to affect activities of Ukrainian and Russian programmers differently from all other GitHub users. The main measurement error for my analysis can occur due to misreporting of user locations, which I use to identify the programmers' nationalities. I observe locations as of the data dump (June 2016). Part of the attenuation bias will come from users reporting wrong or non-existing locations. It is more problematic if some users adjusted their locations after the conflict (for example, Ukrainian programmers who still wanted to work with Russians without being "blamed" by others could have opened additional accounts or Russian users could have masked their real locations). Under such selective mis-reporting, I would get an upper-bound estimate of the effect's magnitude due to misclassifying users with the highest

[12]To clean self-reported locations (as they can be in any language/spelling), I use Google geocoding API, which generates the standard English name and geographical coordinates.

benefits from the collaboration. Yet, given the nature of GitHub, such misreporting is unlikely. First, reputation building is important for most users, and it accrues through the cumulated number of commits, followers, and stars, thus lowering incentives to hold multiple accounts. Second, complete activity history is usually observable, making it difficult for users to hide their information. In addition, I can make use of country codes, automatically generated by GitHub and not subject to user control.[13]

3.2.4 Descriptives

This section presents the descriptive statistics on activities of Ukrainian and Russian programmers on GitHub. Figure 3.2 below and Figure C.1 in the Appendix illustrate that projects from Ukraine and Russia follow the dynamics of other GitHunb projects very closely. Figure 3.2 shows the monthly count of newly registered projects by region. I assign projects to regions: Ukraine, Russia, EU, or Overseas (US, Canada, Japan) based on the location of the project owner. Figure C.1 presents the count of "star" events.[14] To generate these measures, I count how many stars projects from Ukraine, Russia, EU, and Overseas regions received per month. New projects create demand for potential contributions, while the amount of stars proxies projects' quality. There seems to be no Russia- or Ukraine-specific demand or quality shock during the analysed period. The inclusion of month fixed effects can absorb general changes in project activity level.[15]

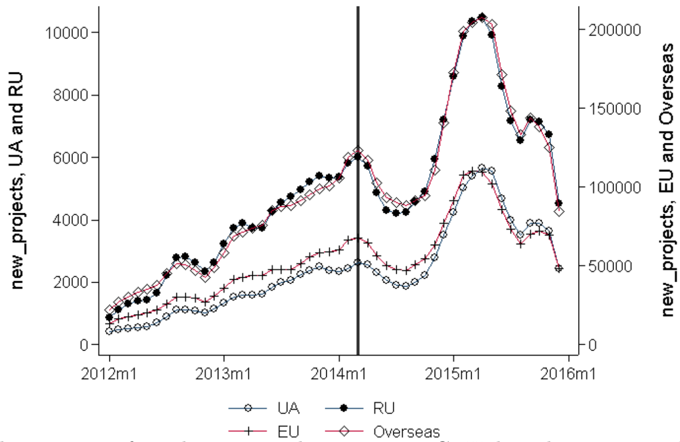
Figures C.2 and C.3 in the Appendix and Figure 3.3 below refer to collaborative activities of Ukrainian and Russian users on GitHub. I use the number of commits (source code modifications) as the main activity measure of GitHub users. Both Russian and Ukrainian users do not seem to be less active after March 2014: their number of commits increases almost every month. While the share of Russian commits to Ukrainian projects stays rather stable throughout 2013-2015 (commits to Ukrainian projects represent about 0.6% of all Russian commits), the share of Ukrainian commits to Russian projects drops

[13] The country code information is, however, not available for all users.

[14] A "star" event is recorded whenever a user puts a star (a like) on a particular project.

[15] One may wonder about the observable drops in the time-series around May-December 2014 and July-December 2015. I am not aware of GitHub-specific changes in policy or technology, which could have provoked it. Most likely, these holes are due to technical problems during the data dump.

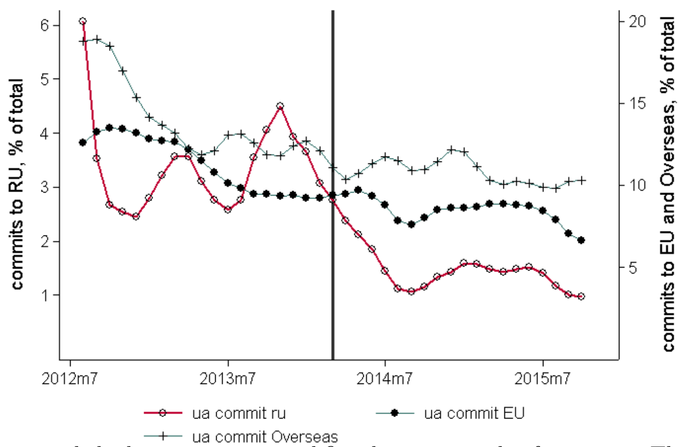
Figure 3.2: Registration of New Projects on GitHub



Notes: The graph plots the amount of newly registered projects on GitHub. The region is determined by the location of the project owner. The data is smoothened with a 3-month moving average filter. To accommodate different activity levels, the left y-axis accounts for Ukrainian and Russian projects, the right y-axis - for EU and Overseas (US, Canada, Japan) projects. The vertical black line corresponds to March 2014 (The Crimean Referendum).

substantially (from, on average, 4.5% before the conflict to 1.5% afterwards). Until the end of 2015, there had been no signs of recovery. I will further validate this descriptive

Figure 3.3: Commits by Ukrainian Users to Russian, EU, and Overseas Projects



Notes: A "commit" event is recorded whenever a user modifies the source code of a project. The graph shows the commits by Ukrainian users to projects owned by Russian, EU, and Overseas users as a share of total Ukrainian commits. The monthly data on commits is smoothened with a 3-month moving average filter. The vertical black line corresponds to March 2014 (The Crimean Referendum).

result with a regression analysis and will argue that changes in identity prescriptions could have led to this drop in the collaboration of Ukrainian and Russian users.

3.3 Empirical Strategy

The empirical analysis comprises two parts. First, I analyse overall changes in the collaboration between Ukrainian and Russian programmers. Second, I focus on the Github projects that prior to the conflict received contributions from Russian and Ukrainian programmers (projects with mixed teams) and investigate how their performance changed after the conflict relative to a comparable group of other GitHub projects.

3.3.1 Estimating the Effect of the Conflict: Change in the Collaboration between Ukrainian and Russian Programmers

To estimate the effect of the conflict on the collaboration between Ukrainian and Russian programmers, I focus on changes in the number of bilateral commits. A commit means any modification to the source code of a project. In comparison to other contributions (reporting issues, tracking bugs, adding comments), it is of the greatest value to any project and serves as the direct measure of the project's progress (McDonald and Goggins, 2013). From a user perspective, commits represent their most visible activity and serve as the measure of their productivity. GitHub summarizes user commits by project directly on user profile pages (see Figure C.6 in the Appendix). To create the baseline sample, I select all commits done by Ukrainian and Russian programmers over 2012-2015. I also add commits done by programmers from Belarus, Kazakhstan, Poland, and Czech Republic, who will later constitute the control group. Through the unique identifiers I link commits to user and project data. For each project and user, I also generate several monthly performance measures. For projects, these include total number of commits from all users, number of project members, and received stars. For users, I obtain total number of commits, total number of project memberships, and total number of followers. On GitHub, both internal (project members) and external users can submit commits to a particular project. In the baseline sample, I pool both internal and external commits together.[16] I aggregate the dataset by summing commits on month, user, and project level. I further restrict the sample to include only those commits where user ID and project owner ID are different to focus on collaborative activities (the costs of contributing to own rather than to others' projects are very different and are not directly comparable). I thus drop more than 60% of all observations, which leaves me with 85,138 user-project-month observations for Ukrainian programmers and 202,434 observations for the Russian programmers. I have 1650 observations (user-project-month) for commits by 600 unique Ukrainian programmers to Russian projects and 1484 observations for commits by 611 unique Russian programmers to Ukrainian projects. As the main outcome variable, I employ the number of Ukrainian-Russian collaborations, i.e. the number of commits by all Ukrainian users to projects owned by Russians and vice versa. To this end, I further aggregate the dataset by month, committer country, and project region.

My baseline empirical strategy builds on a difference-in-difference approach and evaluates changes in the Ukrainian-Russian collaboration patterns caused by the conflict over Crimea and the Donbass region. I use monthly data from 2012 to 2015 and set March 2014 (the month of the Crimean Referendum) as the start of the conflict. I run separate regressions for the Ukrainian and for the Russian sample of commits (based on committer country).[17] Below, I describe the estimated regressions for the Ukrainian sample. The regressions for the Russian sample are identical.

The baseline difference-in-difference regression has the following form:

$$Commits_{pt} = \beta_0 + \beta_1 * RU_p * POST_t + \beta_2 * X_{pt} + \gamma_t + \delta_p + \epsilon_{pt} \tag{3.1}$$

[16] The technical difference between the two is that internal commits can appear in the master project, whereas external commits must be linked to the local copies (forks) of the master project (the external users do not have access rights to directly modify the source code). Because I can observe the "parent" for each forked project, I can link all the commits to the actual master projects.

[17] Alternatively, I could pool the commits together, but given cross-country differences, the separation allows for more precise estimation of fixed effects and controls. Besides, from the descriptive evidence, the effect of the conflict on Ukrainian and Russian commits is not symmetric.

In (1), the independent variable $Commits_{pt}$ is the natural logarithm of commits by Ukrainian users to projects from a given region p in a month t.[18] Project regions (Ukraine, Russia, Post-USSR, EU, Overseas, Other) correspond to the location of project owners. Commits to Russian projects represent the "treatment" group. The coefficient β_1 is the classic difference-in-difference estimator (as in Duflo (2001) or Moser and Voena (2012)). X_{pt} is the vector of variables, which control for region-specific time-varying demand effects: number of active projects and number of stars. γ_t are month fixed effects, which control for time shocks in the activity of Ukrainian programmers and possible "holes" in the dataset. δ_p are project-region fixed effects.

The difference-in-difference estimator β_1 is consistent provided common pre-trends and the absence of time-varying unobservable factors, which independently from the conflict could influence the collaboration patterns between Ukrainian and Russian programmers. To account for the first requirement, I center the data around quarter 1, 2014 and allow treatment coefficients to vary across quarters before and after the conflict.

$$Commits_{pt} = \beta_0 + \sum_{t=-6, t\neq -1}^{6} \beta_{1,t} * RU_p * Q_t + \beta_2 * X_{pt} + \gamma_t + \delta_p + \epsilon_{pt} \tag{3.2}$$

In (2), $\beta_{1,t}$ are the interaction coefficients between quarterly time dummies and the indicator for commits to Russian projects. The reference period (omitted interaction) is quarter 4, 2013. All other controls and fixed effects are the same as in the Specification 1. In the absence of pre-trends, the pre-conflict interaction coefficients should be zero. The Specification 2 also allows to see how the effect evolves over time: do Ukrainian programmers immediately respond to the conflict? Does the effect persist?

Another threat to the consistency of β_1 is that the collaboration patterns with Russia, relative to other regions, in the absence of the conflict would change anyway, for example, due to a quality shock to Russian projects. To account for this, I estimate a triple difference regression. I select a control group of commits by users from several other countries (Belarus, Kazakhstan, Poland, and the Czech Republic) with activity levels similar to Ukrainian programmers. Whereas the time-varying quality and the eco-

[18] Here and further in the text, whenever I refer to the natural logarithm of a variable x, I mean $ln(x + 1)$. Adding 1 allows to keep observations with zeros.

nomic considerations regarding Russian projects should be similar across all countries, the conflict-related effects are likely to be stronger for Ukrainian users. The estimated equation has the following form:

$$\begin{aligned} Commits_{cpt} = \beta_0 + \beta_{11} * UA_c RU_p * POST_t + \beta_{12} * RU_p * POST_t + \\ + \beta_2 * X_{pt} + \gamma_{ct} + \delta_{cp} + \epsilon_{cpt} \end{aligned} \tag{3.3}$$

In (3), the level of observation is at the committer-region (c), project-region (p), and month (t). c includes Ukraine and the above control countries. p includes all the c countries, Russia, EU, Overseas, and Other regions. β_{11} is the triple-difference estimator, measuring the effect of the conflict on Ukrainian commits to Russian projects relative to other non-Ukrainian commits. β_{12} captures (if any) general changes in commits to Russian projects after the conflict. X_{pt} are time-varying controls of the project-regions. $\gamma_c t$ - committer-region by month fixed effects. δ_{cp} committer-region by project-region specific effects.

3.3.2 Estimating the Effect of the Conflict on Project Performance

Did the change in collaboration result in real productivity effects? In the next step, I evaluate the consequences of the conflict for project performance. The difference-in-difference approach compares the performance of the affected projects before and after the conflict relative to the control group. As affected (treated), I consider projects that received both Russian and Ukrainian commits before March 2014. I start with a sample of GitHub projects that received at least one commit by users from Ukraine, Russia, and the above control countries in 2012-2015. I further restrict the sample: first, I drop all projects with only one committer (to focus on collaborative work only) and, second, I consider only project that were active throughout February 2013-February 2014. This leaves a sample of 30,808 unique projects. I then assign treatment status if at least one Russian and one Ukrainian programmer collaborated on a project within the same month.[19] With the disaggregated data, I can identify all the projects with Russian and Ukrainian collaborations, without restricting the nationality of project owners. In the treated sample, owners from EU, Russia, and Ukraine account each for about 10% of projects, Overseas users own 18% of the projects, and about 50% of owners are located in other countries.

As Table C.2 in the Appendix illustrates, the treated projects are very different from an average GitHub project: on average, they are twice as old, have more project members, and over February 2013-February 2014 they received several times more commits and stars. This is due to the fact that despite seemingly low collaboration costs, the majority of GitHub projects, similar to scientific research or patent production, feature strong localisation bias. Only projects of the highest quality manage to attract committers from a different location or company. Therefore, average local GitHub projects cannot constitute a reasonable control for the treatment group. In order to identify comparable projects, I apply the coarsened exact matching procedure (see Blackwell et al. (2009)) to match projects on age, programming language, region of owner (Russia, Ukraine, EU, Overseas, Other), number of project members, and number of commits over February 2013-February 2014. I manage to match 690 treated projects (from 272 bins). Columns 3 and 4 in the Table C.2 illustrate that the pre-conflict descriptives for the matched sample are now well balanced.

I estimate the following difference-in-difference regression:

$$\begin{aligned} Y_{jt} = \beta_0 + \beta_{11} * RU * TREAT_j * POST_t + \beta_{12} * UA * TREAT_j * POST_t + \\ + \beta_{13} * TREAT_j * POST_t + \delta_{pt} + TREAT_j * \kappa_p + \beta_2 * X_{jt} + \epsilon_{jt} \end{aligned} \tag{3.4}$$

In Specification (3.4), Y_{jt} represents one of a project's j monthly performance measures: total number of commits (project's progress) or number of stars (project's popularity), both in natural logarithms. The main coefficients of interest are β_{11}, β_{12}, and β_{13}. β_{13} measures whether performance of all treated projects changed relative to the matched control group after the conflict. For example, if some users decided to leave their projects

[19] Alternatively, I define the continuous treatment by calculating the shares of Ukrainian and Russian commits to total project commits within one year preceding the conflict and construct the treatment variable as the $min\{share_{ua}, share_{ru}\}$.

to not work together with people from a "hostile" country or if coordination costs within a Russian-Ukrainian team increased, this could have negatively impacted the flow of commits and, consequently, the project's popularity. β_{11} and β_{12} allow for a different treatment effect for projects owned by Russians or by Ukrainians. δ_{pt} are project-region by month fixed effects. As previously, region p is defined by the location of the project's owner. $TREAT_j * \kappa_p$ is the interaction between the treatment indicator and the owner's region fixed effect to account for possible region-specific differences of treated projects at the baseline. X_{jt} are project-specific controls, such as owner type (company or individual) or the pre-conflict number of commits. Provided the matching of the treatment and control projects was successful, the inclusion of terms $TREAT_j * \kappa_p$, and $\beta_2 * X_{jt}$ should not affect the coefficients of interest.

3.4 Results

The section presents the results of the empirical analysis. First, I show that the conflict negatively impacted the collaboration between Ukrainian and Russian programmers. Second, I provide evidence for the role of national identity behind this result. Third, I investigate the effect of the conflict on performance of existing projects with mixed Russian-Ukrainian teams. The regressions confirm the descriptive evidence that the number of Ukrainian commits to Russian projects dropped after March 2014. Moreover, the conflict negatively affected the overall performance of Russian projects, which prior to the conflict received contributions from both Ukrainian and Russian programmers.

3.4.1 Change in Collaboration Patterns

The regression analysis aims to illustrate that the observable drop in the number of Ukrainian commits to Russian projects (Figure 3.3) is directly caused by the conflict rather than by some other Ukraine- or Russia-specific shocks. In this Sub-Section I focus on changes in Ukrainian contributions to Russian projects. The relative number of Russian commits to Ukrainian projects did not change after March 2014. While I cannot establish the true reason for this asymmetric effect with the observational data, it could be that the identity shock affected only Ukrainian programmers. Possibly, Russian programmers, unlike an average Russian (who participated in the survey from Figure 3.1), did not change the attitudes toward Ukraine after the burst-out of the conflict.

Table 3.1 reports the results of the baseline difference-in-difference estimations. The outcome of interest is the natural logarithm of commits by users from one region (committer-region) to projects owned by users in another region (project-region). The data covers January 2012-December 2015, with the post period starting in March 2014. All specifications include month- and committer-region by project-region fixed effects. The main explanatory variable is the dummy UA_cRU_pPOST, which equals one for Ukrainian commits to Russian projects after March 2014. Columns 1 and 2 present the results of the baseline difference-in-difference regressions, where I consider only commits made by Ukrainian programmers to projects from different regions: Ukraine, Russia, EU, Overseas, and Other. Specification 1 includes only fixed effects and Specification 2 additionally controls for changes in demand for commits across different regions. Both specifications yield very similar estimates of the UA_cRU_pPOST dummy (-0.8 log points) suggesting that the number of Ukrainian commits to Russian projects in the post period was about 60% smaller than what Russian projects would have received if the conflict did not take place.[20] To ensure that this result is not driven by the general decline of Russian projects I estimate triple-difference regressions where I compare changes in Ukrainian commits to those from similar countries. It is important to keep in mind that the triple-difference coefficient may be overstated if Russian project owners started to more actively attract users from the control countries to compensate for a possible loss of Ukrainian collaborators. On another side, it may represent a lower-bound magnitude, if the Russian-Ukrainian conflict also negatively affected the attitudes toward Russia in other countries. Specification 3 includes only Belarusian programmers as the control group. The estimate of the UA_cRU_pPOST dummy effectively does not change. Specification 4 extends the control group to Kazakhstan, Poland, and the Czech Republic. The magnitude of the

[20] $1 - e^{-0.8} \sim 0.55$

coefficient drops to 0.6, but remains highly significant.[21] Specification 5 repeats the same regression as in 4 but drops all commits from the Eastern regions of Ukraine, which were directly affected by war. If before the conflict these regions accounted for the large portion of the collaborations with Russia, the observed drop in commits could be caused by mechanic destruction of the infrastructure. Dropping Luhansk, Donetsk, and Crimea regions, however, leaves the coefficient intact.

Table 3.1: The Effect of the Conflict on Ukrainian Commits to Russian Projects

	(1) Commits Diff-in-diff	(2) Commits Diff-in-diff	(3) Commits Triple diff	(4) Commits Triple diff	(5) Commits Triple diff
UA_cRU_pPOST	-0.838***	-0.820***	-0.807***	-0.607***	-0.609**
	(0.167)	(0.166)	(0.298)	(0.232)	(0.240)
RU_pPOST			-0.00648	-0.214	-0.214
			(0.247)	(0.161)	(0.161)
New projects, 3m		0.461	0.714**	0.528**	0.515**
		(0.326)	(0.279)	(0.233)	(0.233)
Stars, 3m		0.710	0.0687	-0.206	-0.180
		(0.600)	(0.495)	(0.480)	(0.481)
Observations	288	288	575	863	863
R^2	0.953	0.954	0.938	0.955	0.955
Month FE	yes	yes			
Project-region FE	yes	yes			
Committer-region*project-region FE			yes	yes	yes
Committer-region*month FE			yes	yes	yes
Extended control group				yes	yes
Without Crimea and Donbass					yes
Robust	yes	yes	yes	yes	yes

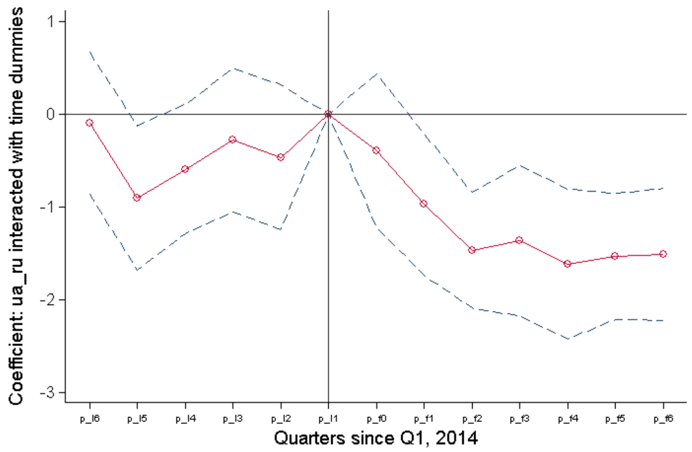
Notes: The dependent variable in all specifications is $ln(commits + 1)$, where *commits* is the number of commits from programmers in one region (c) to projects from another region (p). UA_cRU_pPOST - a dummy for commits by Ukrainian programmers to Russian projects. *POST* period starts after March 2014. *New projects, 3m* and *Stars, 3m* - the number of new and "starred" projects from a project-region over the last 3 months. In columns 1 and 2, the sample includes commits by Ukrainian programmers to projects from different regions (Ukraine, Russia, EU, Overseas, Other). In columns 3, the sample includes commits by Ukrainian and Belarusian programmers; columns 4-5 in addition include commits from other control countries. Column 5 drops commits from programmers in Crimea and Donbass before aggregating Ukrainian commits.

Table C.3 in the Appendix presents the same regressions but for a different outcome: the monthly number of unique user-project collaborations between two particular regions. $Y_{UA/RU/t}$ for example, corresponds to the number of Ukrainian users committing to Russian projects in a month t. This measure, hence, can serve as an extensive margin of collaboration. The estimated coefficients in all specifications remain negative and

[21]The obtained coefficient may include both biases discussed above: Belarusian users could easily substitute for Ukrainian programmers on Russian projects, while the attitudes toward Russia could also significantly worsen in Poland and the Czech Republic after March 2014, given already low popularity of Russia in these countries.

strongly significant, but are statistically significantly smaller in magnitude compared to the results in Table 3.1. This suggests that the conflict had a stronger effect on the intensity of collaboration.[22] One explanation for such result is that the conflict increased communication costs between Ukrainian and Russian programmers (i.e. psychological costs of communicating with someone who shares opposing views). Given the collaborative nature of GitHub (in many cases commits have to be approved by another team member before being integrated) these additional communication costs would inhibit coordination and, hence, reduce the amount of commits. Alternatively, the conflict could have lowered the preference of Ukrainian programmers for the identification with Russian projects and made them, for example, less willing to accept project membership. In the open-source environment, where non-monetary incentives (among them, membership commitments) play an important role, this change could lead to lower average commits. I will provide additional evidence for this explanation in Section 3.5.

Figure 3.4: Commits of Ukrainian Programmers to Russian Projects: Quarterly Treatment Effect of the Conflict



Notes: The graph plots the interaction coefficients between quarterly time dummies and the indicator for Ukrainian commits to the Russian projects. Dependent variable: ln(commits+1) aggregated by month and project-region. Controls: month and project-region fixed effects. The vertical black line corresponds to the reference period (Q4 2013).

Finally, Figure 3.4 reports the check for the absence of pre-trends in Ukrainian commits to Russian projects. I use a similar specification as presented in Table 3.1, but instead of using a single dummy UA_cRU_pPOST, I interact the indicator for the Ukrainian-Russian commits with the quarterly time dummies, which I center around the first quarter of 2014.

[22]The mean of commits is equal to 2,540, whereas the mean of user-project collaborations is 350.

Figure C.4 in the Appendix presents a similar graph but for a triple difference (relative to the commits from the control group countries). Both figures show a clear absence of pre-trends. They are also illustrative about the timing of the effect: collaboration drops rapidly over the first two quarters following the start of the conflict, and over the analysed period (until December 2015) there had been no signs of recovery.

3.4.2 Evidence for the Identity Effects

The above results establish that the conflict indeed reduced the Ukrainian-Russian collaboration on GitHub. It is, yet, not enough to argue that this effect was due to the identity. In this Sub-Section, I, first, investigate the possibility of alternative (pure economic) explanations that could have led to the observed results. Second, I provide additional evidence, which supports the identity hypothesis.

It might be that the main motivation to commit to a particular project is to increase own visibility in front of a potential employer. If Russian companies lose their attractiveness due to the economic crisis that burst out in 2014, fewer Ukrainian programmers would seek a job in Russia. The incentives to contribute to Russian projects would diminish, leading to a smaller number of commits. This could still represent the indirect effect of the conflict (introduced sanctions against Russia account for some of the economic decrease), however, it is not related to the identity. The triple-difference estimations, with users from other countries as the control group, can partly alleviate these concerns. Career opportunities in Russia for Belarusian or Kazakh programmers due to language similarity and the intensity of previous relations should be very similar to those of their Ukrainian colleagues. Therefore, if the bad economic situation makes Russian companies and projects less attractive, it should also influence the flow of contributions from the control countries. The data, however, does not confirm it: the estimated coefficient RU_pPOST, which measures changes in commits from the control countries to Russian projects after March 2014, is not significant (see Tables 3.1 and C.3).

Another possibility is that Ukrainian programmers, who used to work on GitHub for Russian companies, were affected by real or anticipated increase in bilateral transaction costs. Higher uncertainty about future collaboration costs (for example, due to possible interruptions in banking services between Russia and Ukraine) could make Ukrainian programmers withdraw from Russian projects. If this were the case, the triple-difference check would not suffice. To alleviate this concern, I compare changes in Ukrainian commits to projects owned by Russian companies vs. those owned by individual users. If bilateral transaction costs mattered, I would expect a stronger decrease in commits to companies. I reestimate regressions from Table 3.1 for the sample of Ukrainian commits including an additional dummy $Company * RU_pPOST$. This dummy allows a different treatment effect for Ukrainian commits directed at Russian companies. Yet, as the results from Table C.4 in the Appendix indicate, there are no differential effects for Ukrainian collaborations with Russian companies compared to individuals.

To provide more support for the identity effect, I analyse differences among Ukrainian programmers, for whom the role of external factors should be the same, but the identity effects could be different. One dimension to vary is the strength of national identity. The larger the importance of the national identity for an individual, the more he or she will react to changing prescriptions.[23] While the true national identity cannot be observed, the bilingualism of Ukrainians provides an opportunity to construct a proxy for it. Practically, all Ukrainians are fluent in both Ukrainian and Russian languages. Both languages are widely applied in everyday and professional life. Moreover, at least, prior to March 2014 there was no negative stigma attached to speaking Russian. Therefore, if an individual reveals to prefer Ukrainian as the first language it could be considered as a signal of a stronger national identity, rather than a consequence of other strategic considerations.

Using GitHub data I can identify the preferred language of Ukrainian programmers. Upon registration on the platform, users report their names and locations and can do it in any language they choose. For a number of Ukrainian cities and first names, the Russian and the Ukrainian versions are slightly different and can be easily distinguished from each

[23]For instance, in a recent experimental study Mechtel et al. (2016) show that the revealed strength of identification predicts subsequent allocation choices.

other.[24] By using spelling differences of 10 male names and of 11 major Ukrainian cities, I manage to classify about 70% of Ukrainian users: 34% as having a stronger Ukrainian identity (prefer Ukrainian language), 38% with a weaker identity (prefer Russian), for 28% of Ukrainian programmers the preferred language could not be established.

Table 3.2: The Effect of the Conflict on Ukrainian Commits to Russian Projects: Strong Ukrainian Identity

	(1) Collab.	(2) Commits	(3) Collab.	(4) Commits
UA identity*RU_pPOST	-0.113	-0.760***	-0.0924	-0.709**
	(0.129)	(0.294)	(0.122)	(0.290)
RU_pPOST	-0.288***	-0.230	-0.310***	-0.292*
	(0.0996)	(0.180)	(0.0945)	(0.175)
New projects, 3m	1.076***	0.967***	1.019***	0.916***
	(0.122)	(0.219)	(0.123)	(0.219)
Stars, 3m	0.0701	0.239	0.108	0.291
	(0.179)	(0.315)	(0.178)	(0.313)
Observations	860	860	860	860
R^2	0.960	0.923	0.960	0.923
Month FE	yes	yes	yes	yes
Identity*project-region FE	yes	yes	yes	yes
Without Donbass and Crimea			yes	yes
Robust	yes	yes	yes	yes

Notes: The dependent variable in Columns 1 and 3 is $ln(collaborations + 1)$; in Columns 2 and 4 - $ln(commits + 1)$. $UAidentityRU_pPOST$ - a dummy for commits by Ukrainian programmers with a stronger Ukrainian identity to Russian projects. $POST$ period starts after March 2014. Baseline group - other Ukrainian programmers. $New\ projects$, $3m$ and $Stars$, $3m$ - the number of new and "starred" projects from a project-region over the last 3 months. All specifications include $identity * project - region$ fixed effects. Columns 3 and 4 drop commits from programmers in Crimea and Donbass before aggregating Ukrainian commits.

Table 3.2 repeats the baseline estimations by allowing a differential treatment effect for users with a stronger Ukrainian identity. According to the estimates, these users indeed react stronger to the conflict relative to other Ukrainian programmers. This effect is mostly pronounced at the intensive margin of collaboration: the average number of commits per user decreases.

3.4.3 Effect on Project Performance

This Sub-Section investigates the consequences of the conflict on the performance of projects with mixed teams. For this analysis I look at the project level of the Github

[24]For example, the spelling of Kiev (Russian) or Kyiv (Ukrainian) or of the name Alexander (Russian) or Olexander (Ukrainian).

dataset. This analysis complements the above results, by first, focusing on the effects for projects that already existed prior to the conflict and second, by being able to identify the effects for all projects (also administered by owners from countries other than Russia and Ukraine) where Ukrainian and Russian programmers used to work together.

Table 3.3 reports the estimation results. In all the specifications I use monthly project-level data from 690 "treated" projects and 690 matched "control" projects identified as described in the section 3.3.2. The control group comprises similar projects, to which either Ukrainian or Russian programmers contribute, but which did not have a Russian-Ukrainian collaboration within one year preceding the conflict. The dummy $TREAT_jPOST_t$ shows the difference between treated and untreated projects after the conflict. In addition, I allow for the differential treatment effects for projects owned by Russians and by Ukrainians. For the third (neutral) countries' projects, the conflict's effect should come mainly through additional communication costs and lower coordination between Russian and Ukrainian team members. For projects owned by either Russian or Ukrainian users, the conflict effect has an additional channel: it imposes the stigma costs of helping the "enemy". Moreover, GitHub displays all open projects, to which a user contributes, on the profile page, thus amplifying the public image concerns. All specifications include month by project-region fixed effect, treat by project-region fixed effect, and programming language fixed effect. I conservatively cluster standard errors by project-region and programming language.

Column 1 displays the effect for the total number of commits to a project. Total number of commits captures the progress of a project and can be treated as a proxy of success. While there seems to be in general no difference between treated and untreated projects, the treatment effect for projects owned by Russians is negative. While in the presented Specification the coefficient is not statistically significant (p-value = 0.153), the coefficient gains significance with less conservative standard errors. Figure 3.5 shows the absence of pre-trends and reveals that the negative effect for Russian projects becomes sizable in about a year after the conflict started. Column 2 shows that, consistently with the results in the previous section, Ukrainian programmers significantly stronger reduce their commits to the treated Russian projects. Although, in response, the amount of contri-

Table 3.3: The Effect of the Conflict on Project Performance

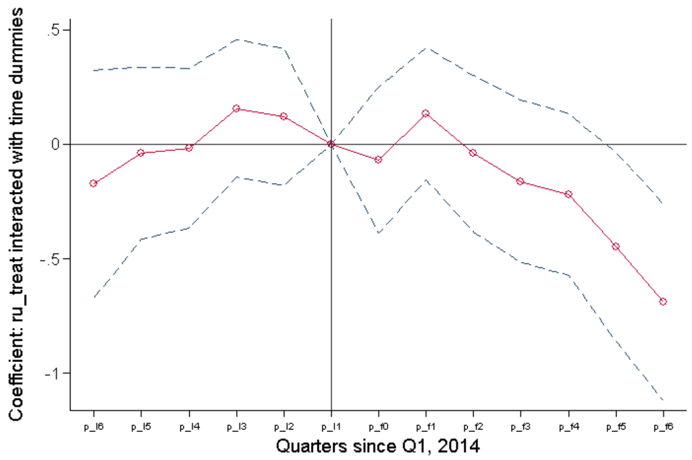
	(1) Total commits	(2) Commits by UA	(3) Commits by RU	(4) Stars
$RU\_TREAT_jPOST_t$	-0.270	-0.173***	0.219	-0.0869*
	(0.189)	(0.0589)	(0.190)	(0.0465)
$UA\_TREAT_jPOST_t$	0.140	0.253	-0.0124	-0.112
	(0.384)	(0.220)	(0.181)	(0.175)
$TREAT_jPOST_t$	0.0271	-0.0798**	-0.0830	0.0477
	(0.0950)	(0.0370)	(0.0537)	(0.0365)
Pre-confl. commits	0.701***	0.0584***	0.151***	0.0585***
	(0.0231)	(0.0111)	(0.0179)	(0.00984)
Company	-0.206***	0.0256	-0.0736	0.00903
	(0.0661)	(0.0372)	(0.0531)	(0.0534)
Observations	31,954	31,954	31,954	31,954
R^2	0.544	0.168	0.188	0.080
Treat*project-region FE	yes	yes	yes	yes
Month*project-region FE	yes	yes	yes	yes
Language FE	yes	yes	yes	yes
Robust	yes	yes	yes	yes
Clusters	83	83	83	83

Notes: The dependent variable are: Columns 1 *ln*(*total commits*+1); Columns 2 and 3 *ln*(*commits by UA or RU users*)+1; Column 4 *ln(stars+1). All variables are measured per month RU*(UA) $TREAT_jPOST$ - a post period dummy for treated projects owned by Russian or Ukrainian programmers; $TREAT_jPOST$ - a post period dummy for all treated projects. $Pre-confl.\ commits$ control for the total number of commits to a project before March 2014. *Company* - a dummy for projects owned by companies. All specifications include $month{*}project{-}region$, $treat{*}project{-}region$, and programming language fixed effects.

butions from Russian users increases, as column 3 reports, it does not fully compensate for the reduction.

Another observation, is that both Ukrainian and Russian users commit less to the treated projects from the third countries. This effect is not strong, but still indicates the presence of possible communication problems in mixed teams following the conflict. Column 4 compares treated and control projects in terms of their popularity, as measured by the amount of new stars, which a project receives from other users. As with the total performance, the coefficient is negative only for the treated Russian projects. This effect is estimated *beyond* possible general negative effects, which might arise for all Russian projects after the start of the conflict. It can be, therefore, attributed to the loss in value of the treated projects due to lower commits from Ukrainian team members.

Figure 3.5: Total Commits to the Affected Russian Projects: Quarterly Treatment Effect of the Conflict, Triple Difference



Notes: The graph plots the interaction coefficients between quarterly time dummies and the indicator for treated projects owned by Russian users. Dependent variable: ln(total commits) aggregated by month and project-region. The vertical black line corresponds to the reference period (Q4 2013).

3.5 Discussion: the Identity Channels

I have provided evidence that the political conflict between Ukraine and Russia has indeed modified collaboration patterns among programmers from these countries. Russian projects, which prior to the conflict received Ukrainian contributions, lost the most. In this section, I further discuss and provide suggestive evidence for channels, through which identity seems to affect collaboration between programmers.

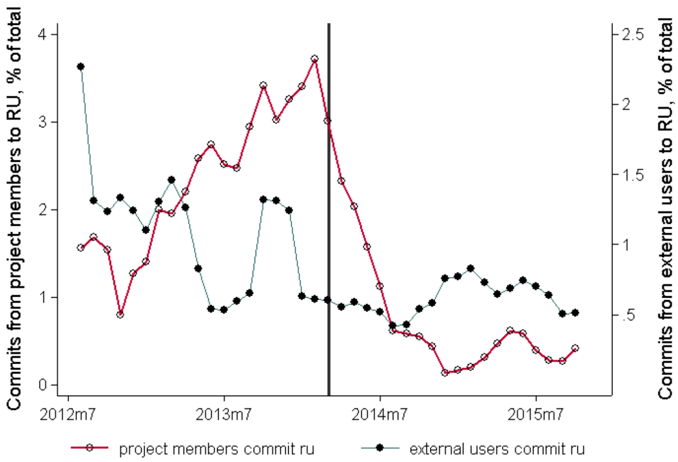
Possible channels, often discussed in the literature, include the change in bilateral trust (beliefs) and the increase in communication costs among team members from conflicting social groups. In the GitHub setting, trust is not likely to be the cause, as the programming quality and even collaborating skills are directly observable. Every registered user can view the code written by other programmers, their projects, number of followers, network, and upon deeper interest, comments, issues, etc. The trust channel would also result in lower number of commits mainly due to the reduction in new collaborations, however, the data show that the decrease on the extensive margin is significantly smaller than that of the intensive.

Another possibility is that the conflict imposed higher (psychological) communication costs between Ukrainian and Russian team members. Such costs would, for example, include poorer communication due to sudden animosity between team members. This would negatively affect coordination and create iceberg type of costs: for example, a Ukrainian programmer submits a commit, but due to the lack of proper communication, this commit is not approved by a Russian team member and not integrated into a project. This channel would hinder, thus, operations of all teams with Russian-Ukrainian collaborations. The magnitude of the effect should increase in the intensity of collaboration. Moreover, even if the identity shock is one-sided (for example, modifies preferences of only Ukrainian programmers), the communication channel should symmetrically affect the contributions of both Ukrainian and Russian programmers. This channel is consistent with the observed decrease in the intensity of collaboration. Although there is some evidence for this effect from the performance of mixed Ukrainian-Russian teams on projects from other countries (see Table 3.4, columns 2 and 3), it is not strongly pronounced and does not affect overall project performance.

The third possibility is that the political conflict with Russia changed the preference of Ukrainian programmers to identify themselves as contributors to Russian ("hostile") projects. This could result from the internal change in preferences or by the wish to follow the new social norm and to avoid public image concerns. If identification with a project plays a strong incentive role (i.e. caring for project's success as a whole directly enters a programmer's utility), the identity conflict is likely to reduce the flow of commits. The data seems to confirm it: across projects with mixed teams, the amount of Ukrainian commits falls significantly sharper if projects are owned by Russians. One can still argue that this result is driven by communication costs, as the intensity of interaction with Russian colleagues is likely to be higher on Russian-owned projects. With the observational data, I cannot fully rule-out this explanation. However, if the intensity of communication mattered, I would expect a symmetric reduction from Russian commits to Ukrainian projects (as they should also be affected by poorer communication), which does not happen.

A direct way to identify oneself with a project on GitHub is to become this project's member. In my sample, on average, project members make 12 modifications to a project's source code per month, while external contributors only make 6. Certainly, project

Figure 3.6: Commits of Ukrainian Programmers to Russian Projects: Project Members vs. External Contributors



Notes: The graph plots the shares of commits from Ukrainian programmers to Russian projects (in % of total commits by Ukrainian users). The left axis corresponds to commits by project members. The right axis - to commits by external users.

membership is not exogenous, but it could potentially represent a strong commitment device. Decomposition of all Ukrainian commits to Russian projects (Figure 3.6) on those by project members and by external contributors reveals an interesting picture: while the relative amount of commits by external contributors does not change and even seems to increase slightly in 2015, the amount of commits by project members drops sharply.

This result could, at first, be counter-intuitive: project members have the closest ties to projects and thus should experience the highest costs of leaving them. A closer inspection of the data, however, shows that this drop is not driven by lower contributions from the pre-conflict project members, but mainly occurs because fewer Ukrainian programmers have joined Russian projects as members after March 2014. This observation can then explain why the effect of the conflict is stronger on the intensive margin: the conflict changed the composition of Ukrainian contributors to Russian projects by lowering the share of internal project members, i.e. those with the highest commitment to contribute to a project. This result also accords with the observed differences in contributions by Ukrainian users with stronger vs. weaker national identity.

3.6 Conclusion

This project studies the role that the national identity plays in the performance of international teams. To conduct the analysis, I use data from GitHub - the world's largest open-source platform. For identification, I exploit the sharp aggravation in Russian-Ukrainian political relations due to the unexpected annexation of Crimea in March 2014. Following the event, collaboration between Ukrainian and Russian programmers fell significantly and was mainly driven by lower contributions of Ukrainian programmers to Russian projects relative to projects from other regions. This decrease cannot be explained by infrastructure problems, nor by general lower activity of Ukrainian and Russian programmers, nor by economic reasons such as higher bilateral transaction costs or changing career concerns. I provide additional evidence for the role of identity in shaping this effect: Ukrainian programmers identified as having a stronger national identity react stronger to the conflict relative to other Ukrainian programmers. I then try to empirically distinguish two channels that might transmit the identity effects. While I find some evidence that the conflict inhibited communication and coordination among Russian-Ukrainian team members, the strongest effect seems to come through changes in the taste of Ukrainian programmers to identify themselves with Russian projects.

This project has policy implications for firms managing or planning to attract diverse teams. The empirical evidence emphasises the risks of diverse teams due to their exposure to external factors. My results further show that the identity conflict not only hinders peer interaction within a diverse team, but also changes the preferences of some team members towards projects that are associated with a "hostile" social group. Therefore, in case of an identity-based conflict, having a third "neutral" party to lead team's work or enforcing a common project identity might be beneficial.

My results can be also relevant for open-source platforms, such as GitHub, that aim at facilitating international "barrier-free" collaboration. One option would be to hide the country of origin information before the real value of the collaboration is revealed.

Lastly, this project could be interesting for the general public by bringing to awareness the role of the national identity. Usually, people blame politicians and external factors for creating constraints to their collaboration. However, my results show that even in a setting with negligible legal or physical barriers, educated and informed people, still choose to follow the identity prescriptions at the cost of economically beneficial collaboration.

Appendices

Appendix A

Appendix to Chapter 1

A.1 Additional Tables and Graphs

Figure A.1: Example of Legislation Dummies

iso316612	nace2	year	DEU	ESP	GBR	FRA
PL	10	2004	0	0	.6666667	0
RO	10	2004	0	0	0	0
PL	43	2004	0	0	.6666667	0
RO	43	2004	0	0	0	0
PL	10	2009	0	1	1	1
RO	10	2009	0	1	1	0
PL	43	2009	0	1	1	1
RO	43	2009	0	1	0	0
PL	10	2012	1	1	1	1
RO	10	2012	0	0	1	0
PL	43	2012	1	1	1	1
RO	43	2012	0	0	0	0

Notes: The figure shows a print-screen from Stata to illustrate the construction of the Free Movement variable. 0 denotes that a given market is closed, 1 denotes that it is open for labour migrants from the new member states. If a country did not open from the beginning of a calendar year (for example, Great Britain opened in May 2004), the legislation dummy is weighted accordingly.

Table A.1: Summary of Variables

	(1) Full	(2) Main_sample	(3) Incumbent	(4) Foreign	(5) Hightech	(6) Innovator
Firm-level data:						
number of employees	40.5	65.5	80.0	195	87.1	238
	[305]	[296]	[344]	[554]	[367]	[727]
sales, 000 EUR	3,908	5,554	6,686	26,972	9,954	34,662
	[162,210]	[61,742]	[70,814]	[172,423]	[109,758]	[267,709]
assets, 000 EUR	3,251	4,854	5,735	21,440	7,629	28,953
	[60,287]	[61,480]	[65,025]	[145,629]	[90,648]	[226,421]
firm age	9.16	10.7	13.6	10.8	10.8	14.6
	[8.03]	[7.78]	[7.71]	[8.66]	[7.78]	[9.63]
labour productivity ($\frac{Y}{L}$)	3.47	3.67	3.69	4.28	3.85	4.22
	[1.35]	[1.10]	[1.09]	[1.16]	[1.07]	[0.90]
labour productivity ($\frac{Y}{WL}$)	2.08	1.97	1.94	1.91	1.71	1.80
	[1.08]	[0.89]	[0.85]	[0.92]	[0.89]	[0.71]
TFP index	-0.012	-0.046	-0.050	-0.059	-0.10	-0.15
	[0.89]	[0.67]	[0.68]	[0.63]	[0.63]	[0.55]
Industry-level data, 2 digits:						
FM	0.083	0.13	0.13	0.14	0.19	0.15
	[0.11]	[0.14]	[0.14]	[0.14]	[0.18]	[0.15]
human capital constraints	0.090	0.092	0.10	0.088	0.096	0.12
	[0.11]	[0.11]	[0.12]	[0.11]	[0.12]	[0.13]
financial constraints	0.25	0.18	0.17	0.15	0.13	0.12
	[0.17]	[0.14]	[0.14]	[0.13]	[0.12]	[0.12]
number of employees	11.6	15.4	16.6	22.2	21.7	22.3
	[20.9]	[25.0]	[25.9]	[38.7]	[38.8]	[34.8]
sales, 000 EUR	948	1,033	1,136	2,076	1,911	2,468
	[5,403]	[7,006]	[8,135]	[7,331]	[5,174]	[11,083]
labour productivity ($\frac{Y}{L}$)	4.03	3.79	3.78	4.03	4.05	4.23
	[0.84]	[0.70]	[0.73]	[0.71]	[0.67]	[0.56]
labour productivity ($\frac{Y}{WL}$)	2.21	1.91	1.92	1.91	1.78	1.91
	[0.62]	[0.40]	[0.40]	[0.46]	[0.50]	[0.41]
Observations	3.25e+06	532760	334693	55979	116540	19143
No. of firms	555072	108256	58245	10628	26224	2758

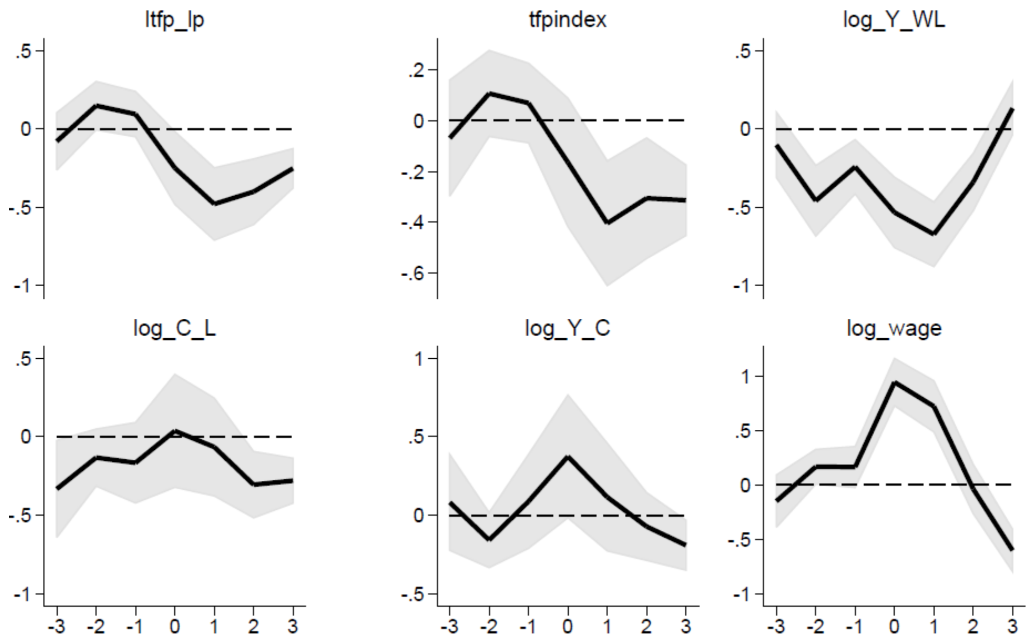
Notes: The table reports means and standard deviations (in brackets) of variables used in the regressions. 'Full' denotes the sample of all available observations. Further sub-samples do not include observations with missing variables. 'Main sample' is a sub-sample of firms used in the main regression. 'Incumbent' is a sub-sample of firms that existed prior to 2002. 'Innovator' denotes a sub-sample of firms with patents. 'High-tech' denotes a sub-sample of firms operating in high-tech industries according to the Statistical classification of economic activities in the European Community (NACE) at 2-digit level. 'Foreign' denotes a sub-sample of firms with foreign capital.
Productivity measures are reported in natural logarithms.
Constraints are measured as the shares of firms in a given industry-country-year reporting to be constrained.
FM is our preferred instrument: the sum of legislation dummies, weighted by proximity measures to a given old EU member-country.
Sources: Amadeus, EU Commission Business Survey, Eurostat Structural Business Statistics

Table A.2: First Stage Regression. Effect of Free Movement on Skill Shortages

	(1) FM, dist	(2) FM*skill sh., dist	(3) FM, migr	(4) FM*skill sh., migr
FM_{ict}	0.0522*	0.125***	0.0767***	0.112***
	(0.0298)	(0.0461)	(0.0144)	(0.0376)
$L.log\_investment_{ict}$	0.00644	0.00650	0.00400	0.00604
	(0.00451)	(0.00455)	(0.00447)	(0.00446)
$Log\_total\_sales_{it}$	0.0105	0.00993	0.0109	0.00912
	(0.00929)	(0.00914)	(0.00916)	(0.00885)
$Mean\ skill\ sh._{it}$	0.200**	0.165*	0.206**	0.148*
	(0.0905)	(0.0875)	(0.0891)	(0.0879)
$L.log\_FDI_{ct}$	0.00145***	0.00148***	0.00134***	0.00148***
	(0.000484)	(0.000483)	(0.000480)	(0.000482)
$D.log\_GDP_{ct}$	0.368***	0.372***	0.387***	0.377***
	(0.0807)	(0.0794)	(0.0812)	(0.0802)
Observations	2,069	2,069	2,069	2,069
Number of pp	428	428	428	428
R^2	0.349	0.352	0.357	0.355
Dummies	ci y	ci y	ci y	ci y
Clusters	428	428	428	428
Fstat	3.081	7.332	28.40	8.859
pval	0.0799	0.00704	1.60e-07	0.00308

Notes: The table presents reduced-form estimates of free movement on skill shortages. All specifications are estimated with industry·country fixed effects and time dummies. Dependent variable: % of firms reporting skill shortages. FM denotes the Free Movement variable. In specifications 1 and 2, we use distance-weighted FM dummies, in 3 and 4 - weights with migration stocks. In specifications 2 and 4, FM dummies are in addition interacted with skill shortages in destination industries. Standard errors (in parentheses) are clustered on the country-industry level.
*** p<0.01, ** p<0.05, * p<0.1

Figure A.2: Dynamic effects (lagging and forwarding the instrument)



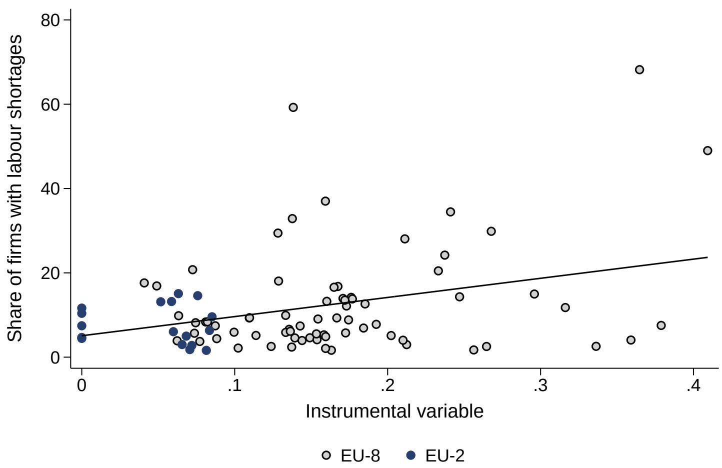
Notes: This graph shows the coefficients of the L&P TFP measure, the TFP index, wage-adjusted labour productivity, the capital labour ration, capital productivity and personnel costs when the instrument is lagged, simultaneous or forwarded by up to three years. The 0 value on the x-axis indicates the year of the labour market opening and the values 1,2,3 are the years following the opening, while the values -3,-2 and -1 are the years preceding the sector openings.

Table A.3: First Stage Regression (Robustness). Effect of Free Movement on Financial Shortages

	(1) FM, dist	(2) FM*skill sh., dist	(3) FM, migr	(4) FM*skill sh., migr
FM_{ict}	0.0791	-0.0414	-0.00399	-0.0800**
	(0.0535)	(0.0553)	(0.0274)	(0.0364)
$L.log\_investment_{ict}$	0.00126	0.00115	0.00131	0.00141
	(0.00638)	(0.00643)	(0.00654)	(0.00637)
$Log\_total\_sales_{it}$	0.00508	0.00522	0.00502	0.00601
	(0.0114)	(0.0115)	(0.0115)	(0.0113)
$Mean\ skill\ sh._{it}$	-0.0551	-0.0342	-0.0476	-0.00650
	(0.0970)	(0.0999)	(0.0967)	(0.102)
$L.log\_FDI_{ct}$	-0.000566	-0.000595	-0.000579	-0.000608
	(0.000508)	(0.000506)	(0.000512)	(0.000507)
$D.log\_GDP_{ct}$	-0.381***	-0.393***	-0.391***	-0.400***
	(0.0747)	(0.0730)	(0.0769)	(0.0734)
Observations	2,070	2,070	2,070	2,070
Number of pp	428	428	428	428
R^2	0.075	0.074	0.073	0.076
Dummies	ci y	ci y	ci y	ci y
Clusters	428	428	428	428
Fstat	2.184	0.561	0.0213	4.832
pval	0.140	0.454	0.884	0.0285

Notes: The table presents reduced-form estimates of free movement on skill shortages. All specifications are estimated with industry·country fixed effects and time dummies. Dependent variable: % of firms reporting skill shortages. FM denotes the Free Movement variable. In specifications 1 and 2, we use distance-weighted FM dummies, in 3 and 4 - weights with migration stocks. In specifications 2 and 4, FM dummies are in addition interacted with skill shortages in destination industries. Standard errors (in parentheses) are clustered on the country-industry level.
*** p<0.01, ** p<0.05, * p<0.1

Figure A.3: First stage illustration



Notes: Skill shortages and FM (instrumental) variable are aggregated on a country-level proportionally to the number of firms in each industry. Source: EU Commission Business Survey, own calculations.

A.2 Proof of the Comparative Statics and the Simulation of the Model

In this sub-section, we first present the proof of the comparative statics results using a general production function and then provide a numerical solution to the model using a Cobb-Douglas production function.

A.2.0.1 Comparative Statics

We assume a general production function with three variable inputs: skilled and unskilled labour (L_s, L_u) and training t. Capital is fixed in the short-term. The firm faces the output price P, wages w_s and w_u, and job separation rate of skilled labour δ_s. V_s denotes the number of posted vacancies, c_s denotes the cost of a skilled vacancy, and $c_t = 1$ denotes the costs per hour of training t. For simplicity search and training costs for unskilled labour are set to zero. The firm solves the following maximization problem:

$$\Pi = Pf(t, L_s, L_u) - L_s w_s - c_s V_s - tV_s - L_u w_u$$

s.t.

$$\frac{V_s}{L_s} = \delta_s.$$

$f(t, L_s, L_u)$ is increasing and strictly concave in t, L_s, L_u. We denote the first-order partial derivatives of f by f_i where $i = t, L_s, L_u$. $f_i > 0$. f_{ij} are the second-order derivatives. $f_{ii} < 0$. We assume that the cross-derivatives $f_{ij}, i \neq j$ are positive.

Firms maximise profits, by choosing the number of workers and the initial amount of training (which then affects the level of firm-specific knowledge t). The first order conditions give the implicit solution of the model.

FOC1: $\dfrac{\partial \Pi}{\partial L_s} = Pf_s - w_s - \delta_s c_s - \delta_s t = 0$

FOC2: $\dfrac{\partial \Pi}{\partial L_u} = Pf_u - w_u = 0$

FOC3: $\dfrac{\partial \Pi}{\partial t} = Pf_t - \delta_s L_s = 0$

We apply the implicit function theorem to determine the signs of $\dfrac{\partial L_s^*}{\partial \delta_s}$ - the effect of δ_s on the firm's demand for skilled labour and $\dfrac{\partial t^*}{\partial \delta_s}$ - the effect on the initial training and consequently the firm's TFP. We assume that the above system of equations has the unique internal solution L_s^*, L_u^*, t^*, which maximises the profit function.

$$\frac{\partial L_s^*}{\partial \delta_s} = \frac{|\tilde{D_{s\delta_s}}|}{|D|}$$

where $|D|$ is the determinant of the Hessian matrix:

$$D = \begin{bmatrix} Pf_{ss} & Pf_{su} & Pf_{st} - \delta_s \\ Pf_{su} & Pf_{uu} & Pf_{ut} \\ Pf_{st} - \delta_s & Pf_{ut} & Pf_{tt} \end{bmatrix}$$

To fulfil the second-order conditions, D has to be negative-definite, therefore, $|D| < 0$. $|\tilde{D_{s\delta_s}}|$ is the determinant of the following matrix:

$$|\tilde{D_{s\delta_s}}| = \begin{bmatrix} c_s + t & Pf_{su} & Pf_{st} - \delta_s \\ 0 & Pf_{uu} & Pf_{ut} \\ L_s & Pf_{ut} & Pf_{tt} \end{bmatrix}$$

The sign of $\dfrac{\partial L_s^*}{\partial \delta_s}$ depends on the term $(P^2 f_{uu} f_{tt} - P^2 f_{ut}^2)(c_s + t) + P^2 L_s f_{ut} f_{su} - P L_s f_{uu}(Pf_{st} - \delta_s)$. Under the assumption that the above profit maximization problem has a solution, this term will be positive.[1] Since $|D| < 0$, $\dfrac{\partial L_s^*}{\partial \delta_s} < 0$.

[1] $(P^2 f_{uu} f_{tt} - P^2 f_{ut}^2)(c_s + t) > 0$, $P^2 L_s f_{ut} f_{su} > 0$. $PL_s f_{uu}(Pf_{st} - \delta_s) > 0$, but should be smaller than the sum of the two first summands, otherwise the stationary point will be a saddle point.

Similarly, the sign of $\dfrac{\partial t^*}{\partial \delta_s}$ depends on the term $(P^2 f_{uu} f_{ss} - P^2 f_{su}^2)(L_s) + P^2(c_s + t) f_{ut} f_{su} - P(c_s + t) f_{uu}(P f_{st} - \delta_s)$, which is also positive. Therefore, $\dfrac{\partial t^*}{\partial \delta_s} < 0$.

A.2.0.2 Simulation

We further illustrate the effect of a increasing job separation rate δ_s on the firm's factor demand and TFP. For this exercise, we assume a Cobb-Douglas production function and simulate the model in Matlab.

$$\Pi = P t^{\gamma} L_s^{\alpha} L_u^{\beta} - L_s w_s - c_s V_s - t V_s - L_u w_u$$

s.t.

$$\frac{V_s}{L_s} = \delta_s.$$

The first-order conditions define the implicit solution of the problem.

FOC1: $\dfrac{\partial \Pi}{\partial L_s} = \alpha P t^{\gamma} L_s^{\alpha-1} L_u^{\beta} - w_s - \delta_s c_s - \delta_s t = 0$

FOC2: $\dfrac{\partial \Pi}{\partial L_u} = \beta P t^{\gamma} L_s^{\alpha} L_u^{\beta-1} - w_u = 0$

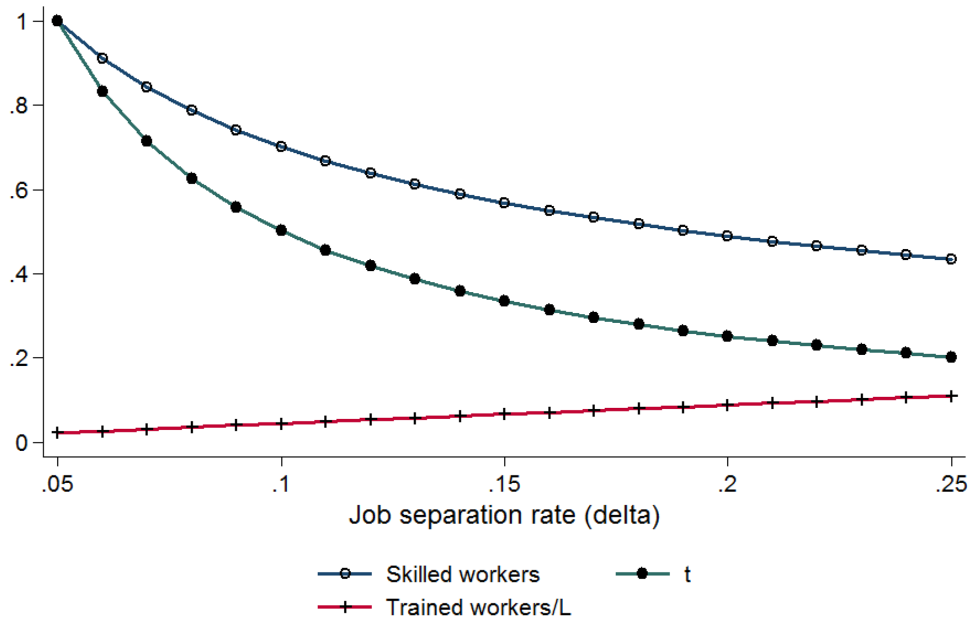
FOC3: $\dfrac{\partial \Pi}{\partial t} = \gamma P t^{\gamma-1} L_s^{\alpha} L_u^{\beta} - \delta_s L_s = 0$

Again, we assume that the solution to the above maximization problem exists: $\gamma + \alpha + \beta < 1$ and the second-order conditions are satisfied.

To simulate the model, we used the following parameter values: $\gamma = 0.1$, $\alpha = 0.5$, $\beta = 0.3$, $w_s = 1$, $w_u = 0.6$, $c_s = 0.05$.

Figure A.4 presents the simulation results. We investigate the effect of δ_s on the firm's demand for skilled workers L_s, on the firm-specific knowledge t, and on the share of trained workers $\dfrac{\delta_s L_s}{L}$. The graph illustrates that if δ_s increases, the firm's demand for the skilled labour decreases and so does the amount of training t that the firm decides to provide for the new hires. The latter eventually lowers the firm-specific knowledge and hence TFP. The share of trained workers to the total workforce, however, increases since more workers have to be trained.

Figure A.4: The Effect of the Job Separation Rate δ_s, Simulation Results



Note: This graph shows the simulation results of our theoretical framework for different values of delta (the job separation rate). "Skilled workers" - the amount of optimal skilled labour, "t" - the amount of of the optimal training (the firm-specific knowledge). Both "Skilled workers" and "t" are normalised to their optimal values at $\delta_s = 0.05$. "Trained workers/L" - the share of trained workers to the firm's total workforce.

A.3 TFP Index Calculation

We calculate the TFP index, following Gorodnichenko and Schnitzer (2013), according to the formula below:

$$TFP_{fict} = \hat{y}_{fict} - s_{ic}^{L}\hat{l}_{fict} - s_{ic}^{K}\hat{k}_{fict} - s_{ic}^{M}\hat{m}_{fict} \tag{A.1}$$

where $\hat{y}_{fict}, \hat{l}_{fict}, \hat{k}_{fict}, \hat{m}_{fict}$ are log deviations of a firm's output, labour, capital, and materials from industry's averages. The latter are calculated on a four-digit industry level (for each country), by taking geometric means across all firm-year observations.

By using deviations instead of levels, we exclude time-invariant country-industry fixed effects and make the index more comparable across different industries and countries.

$s_{ic}^{L}, s_{ic}^{M}, s_{ic}^{K}$ are cost shares of labour, materials, and capital, which are computed for each firm-year and then also aggregated on a four-digit industry level for each country.

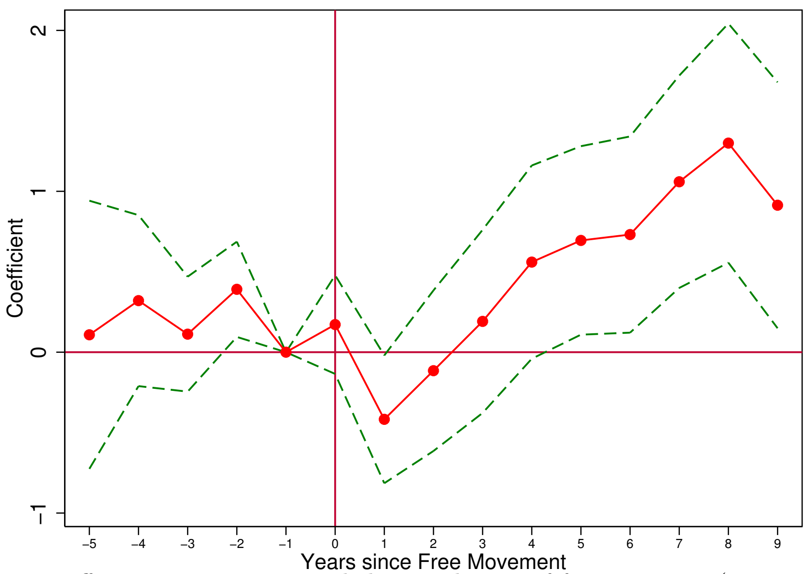
As a proxy of output we use firms' sales, labour (wages and salaries), capital (fixed assets), and material (material costs). We should note that the obtained TFP index contains not only firms' unobserved technology and management ability, but also firms' market power, and differences in their workforce composition.

Appendix B

Appendix to Chapter 2

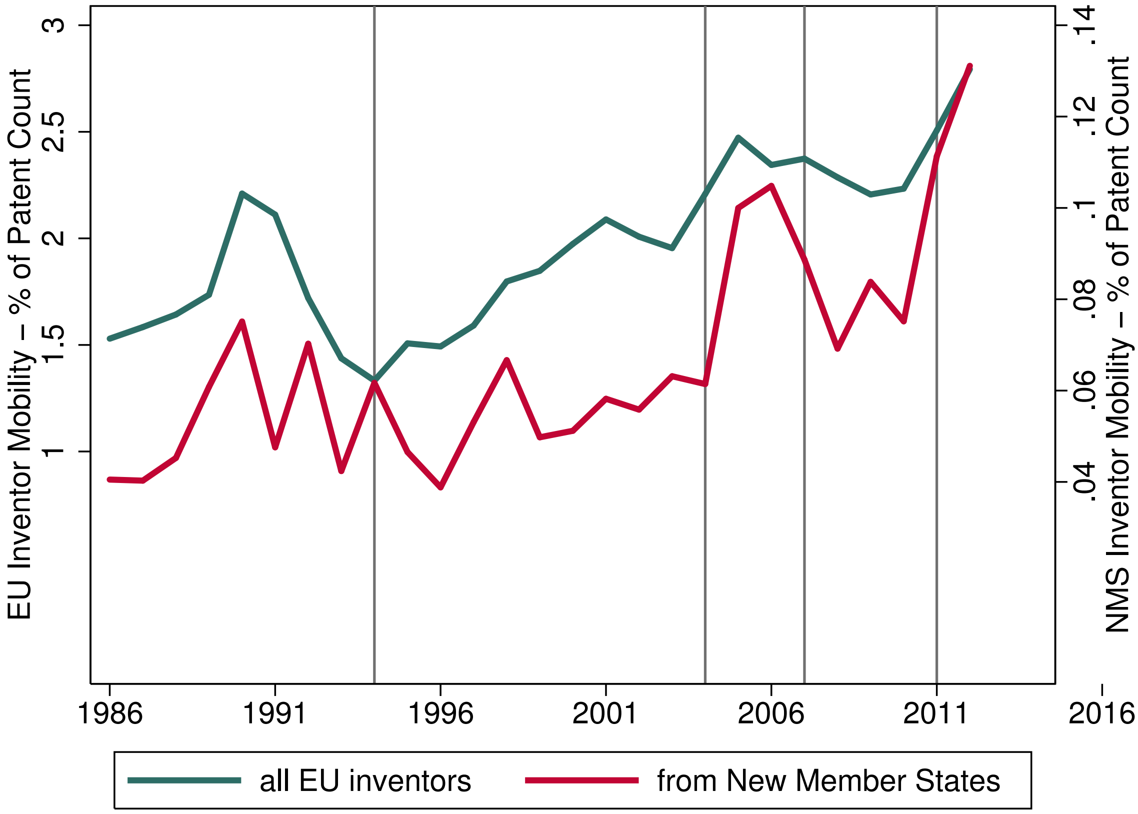
B.1 Additional Tables and Graphs

Figure B.1: Migration Flows, Annual Treatment Effects of Free Labour Mobility



Notes: Annual treatment effects on migration around the introduction of free movement (1986-2012). The regression includes destination-year and country-pair-industry fixed effects. Standard errors are clustered at the country-pair-industry level.
Source: PATSTAT, European Commission, own calculations.

Figure B.2: Inventor Mobility in Europe



Notes: The graph shows the number of mobile inventors normalised to the total number of patent applications. We count as mobile inventor and inventor who changes his country of residence compared to the previous patent application. Thus migrants can be identified only if they have at least one patent application in each country.
Source: PATSTAT.

Table B.1: Overview of the Gradual Opening of the EU15+4 Labour Markets

Country	NMS8 (2004 entry)	NMS2 (2007 entry)	Sectoral Exceptions
Austria	2011	2014	NMS8 (2007-2010), NMS2 (2007-2013): Construction, Manufacturing of Electronics and Metals, Food and beverage services (restaurant business), other sectors with labour shortages
Belgium	2009	2014	-
Denmark	2009	2009	-
Finland	2006	2007	-
France	2008	2014	NMS8 (2005-2007), NMS2 (2007-2013): Agriculture, Construction, Accommodation and food services (tourism and catering), other sectors with labour shortages
Germany	2011	2014	NMS8 (2004-2010), NMS2 (2007-2013): sectors with labour shortages
Greece	2006	2009	-
Iceland	2006	2012	-
Ireland	2004	2012	-
Italy	2006	2012	NMS8 (2004-2005): sectors with labour shortages; NMS2 (2007-2011): Agriculture, Construction, Engineering, Accommodation and food services (tourism and catering), Domestic work and care services, other sectors with labour shortages; Occupations: Managerial and professional occupations
Lichtenstein	2011	2016	-
Luxembourg	2008	2014	NMS2 (2007 - 2013): Agriculture, Viticulture, Accommodation and food services (tourism and catering)
Netherlands	2007	2014	NMS8 (2004-2006), NMS2 (2007-2013): International transport, Inland shipping, Health, Slaugther-house/meet-packaging, other sectors with labour shortages
Norway	2009	2012	NMS8 (2004-2008), NMS2 (2007-2011): sectors with labour shortages
Portugal	2006	2009	-
Spain	2006	2009	Reintroduction of restrictions for Romanians: 11/08/2011 - 31/12/2013
Sweden	2004	2007	-
Switzerland	2011	2014	-
United Kingdom	2004	2014	NMS2 (2007-2013): Agriculture, Food manufacturing

Notes: Column 2 shows the year of the labour market opening of the respective country for the NMS10 countries, column 3 shows the year of the labour market opening of the respective country for the NMS2 countries. Column 4 shows, which sectors were exempt from restrictions.
Source: European Commission.

Table B.2: Migration and Free Labour Mobility: First Stage

	(1) EU19 and NMS all migrants	(2) NMS all migrants	(3) NMS 2004 only all migrants	(4) EU19 and NMS patent potential
L3.FM	2.352***	5.039**	19.37*	-0.563
	(0.754)	(2.320)	(10.48)	(0.645)
L4.FM	1.860***	3.271*	4.298	1.156**
	(0.630)	(1.704)	(4.065)	(0.506)
L5.FM	-0.136	-0.0996	9.662	0.350
	(0.375)	(0.418)	(19.23)	(0.292)
in EU	0.447**	-4.541*		0.261
	(0.204)	(2.526)		(0.180)
L2.Trade flow	-1.077		-74.12	-2.072**
	(1.089)		(76.99)	(0.953)
L2.FDI inflow	1.14e-05	0.000161***	0.000185***	1.45e-05**
	(2.40e-05)	(4.31e-05)	(4.76e-05)	(6.90e-06)
Observations	383	186	163	383
R-squared	0.597	0.683	0.701	0.363
Region industry FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
F	10.20	48.57	660.0	10.11
Clusters	53	30	23	53

Notes: The regressions in this table estimate the first stage corresponding to table 2.1 in column 1 and 4: The dependent variable is the (second lag of the natural logarithm) of emigration in a region and outflow of migrants with patenting potential, respectively. The instruments are the free movement variables for the three previous years. The regressions include controls for EU membership, trade flows and FDI inflows. The first pair of columns includes all EU and EFTA countries, the third and fourth column limit the sample to new member states and the last two columns include only the 2004 accessions. All specifications include year and region-industry fixed effects. Robust standard errors are clustered at the region-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.3: Patent Applications and Free Labour Mobility (Reduced Form)

	(1) EU19 and NMS Patents	(2) EU19 and NMS cit. weighted	(3) NMS Patents	(4) NMS cit. weighted	(5) NMS 2004 only Patents	(6) NMS 2004 only cit. weighted
L3.FM	1.075*	1.309*	-0.276	-0.0181	1.758	2.047
	(0.576)	(0.717)	(2.315)	(2.193)	(3.247)	(4.016)
L4.FM	1.786***	2.206***	-0.606	-0.216	-4.447	-4.335
	(0.276)	(0.386)	(0.863)	(0.805)	(3.655)	(3.624)
L5.FM	-0.177	0.0565	-0.395	-0.264	3.418	4.612
	(0.392)	(0.526)	(0.545)	(0.710)	(3.751)	(3.579)
in EU	0.167	0.278**				
	(0.107)	(0.121)				
L2.Trade flow	-1.399**	-1.456*				
	(0.662)	(0.863)				
L2.FDI inflow	3.10e-05**	4.29e-05***	1.45e-05	4.15e-05	2.49e-05	5.34e-05
	(1.26e-05)	(1.28e-05)	(2.70e-05)	(2.75e-05)	(3.15e-05)	(3.20e-05)
Observations	496	496	209	209	184	184
R-squared	0.442	0.742	0.267	0.177	0.257	0.162
Region industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	56	56	32	32	24	24

Notes: The dependent variables in the regressions shown in this table are the number of patent applications (columns 1,3 and 5) and citation-weighted patent applications (columns 2,4 and 6). More precisely, the dependent variable is the natural logarithm of 1 plus these counts. The same transformation is applied to the trade flow regressor and for FDI inflows, the percentage change from the previous year is used as regressor. The first pair of columns includes all EU and EFTA countries, columns 3 and 4 include all countries which joined the EU in 2004 and later and the last two columns only includes those which joined in 2004. All specifications include year and region-industry fixed effects. Standard errors are clustered at the region-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.4: Patent Applications and Migration in NMS10, OLS and 2SLS

	(1) OLS Patents	(2) OLS cit. weighted	(3) OLS Patents	(4) 2SLS Patents	(5) 2SLS cit. weighted	(6) 2SLS Patents
L2.Migrants	0.0924**	0.0730*		0.115	0.212	
	(0.0350)	(0.0375)		(0.156)	(0.249)	
L2.Migr.pat.potential			0.203*			0.101
			(0.112)			(0.0950)
L2.Trade flow				0.482	-0.650	0.758***
				(0.518)	(0.820)	(0.251)
L2.FDI inflow	-1.41e-05	-5.82e-06	-3.23e-07	-1.80e-05	-3.03e-05	9.41e-07
	(1.87e-05)	(1.80e-05)	(1.72e-05)	(3.34e-05)	(4.83e-05)	(1.68e-05)
Observations	163	163	163	163	163	163
Region industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	23	23	23	23	23	23
F				16.81	26.31	65.74

Notes: The regressions in this table estimate the relationship between the migration flow out of a country and innovation in that country. The first three columns are estimated with OLS and the last three column use a 2SLS estimation with our instrument based on free movement legislation. The dependent variables are the number of patent applications in an industry and origin region in a year or, in columns 2 and 5, the citation-weighted patent applications (i.e. patent applications + forward citations to these patents). Patent application numbers and citation-weighted counts, number of migrants and trade flows are taken in natural logarithms. The sample includes only the 10 countries which joined the EU in 2004. All specifications include year and region-industry fixed effects. Robust standard errors are clustered at the region-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.5: Patent Applications and Migration, USPTO Patents Only, OLS and 2SLS

	(1) OLS Patents	(2) OLS cit. weighted	(3) OLS Patents	(4) 2SLS Patents	(5) 2SLS cit. weighted	(6) 2SLS Patents
L2.Migrants	0.0270	-0.0894		0.346**	0.503**	
	(0.0535)	(0.0694)		(0.171)	(0.232)	
L2.Migr.pat.potential			0.000889			0.702
			(0.0606)			(0.429)
in EU	0.0258	0.402	0.0182	0.115	0.567	-0.0508
	(0.204)	(0.305)	(0.206)	(0.238)	(0.374)	(0.222)
L2.Trade flow	1.623**	2.409**	1.740**	0.252	-0.144	2.493***
	(0.687)	(1.002)	(0.654)	(0.901)	(1.246)	(0.842)
L2.FDI inflow	1.24e-05	3.06e-05**	1.28e-05	7.25e-06	2.11e-05***	1.13e-06
	(1.02e-05)	(1.41e-05)	(9.91e-06)	(1.43e-05)	(8.16e-06)	(1.16e-05)
Observations	383	383	383	383	383	383
Region industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	53	53	53	53	53	53
F				32.56	273.0	26.87

Notes: The regressions in this table estimate the relationship between the migration flow out of a country and innovation in that country, counting only patents that were filed with the USPTO. The first three columns are estimated with OLS and the last three column use a 2SLS estimation with our instrument based on free movement legislation. The dependent variables are the number of patent applications in an industry and origin region in a year or, in columns 2 and 5, the citation-weighted patent applications (i.e. patent applications + forward citations to these patents). Patent application numbers and citation-weighted counts, number of migrants and trade flows are taken in natural logarithms. The sample includes all EU members and countries in the European Free Trade Association. All specifications include year and region-industry fixed effects. Robust standard errors are clustered at the region-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.6: Convergence in Patenting Levels ($Patents_{dest}/Patents_{origin}$) and Migration, NMS only, OLS and 2SLS

	(1) OLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents	(2) OLS $log(\frac{P_{diy}}{P_{oiy}})$ cit. weighted	(3) OLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents	(4) 2SLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents	(5) 2SLS $log(\frac{P_{diy}}{P_{oiy}})$ cit. weighted	(6) 2SLS $log(\frac{P_{diy}}{P_{oiy}})$ Patents
L2.Migrants	0.0289	0.0349		-0.254*	-0.259*	
	(0.0229)	(0.0285)		(0.141)	(0.151)	
L2.Migr.pat.potential			0.118			-1.809
			(0.0747)			(2.810)
Patents, origin	-1.080***	-1.052***	-1.081***	-1.083***	-1.055***	-1.065***
	(0.109)	(0.120)	(0.109)	(0.110)	(0.122)	(0.119)
Patents, dest	1.078***	1.128***	1.080***	1.071***	1.120***	1.044***
	(0.0713)	(0.0910)	(0.0717)	(0.0765)	(0.0962)	(0.0890)
Within EU	0.0435	-0.0278	0.0431	0.0572	-0.0136	0.0724
	(0.0529)	(0.0584)	(0.0533)	(0.0555)	(0.0612)	(0.0693)
GDP_d/GDP_o	-0.444	-0.0942	-0.446	-0.480	-0.132	-0.471
	(0.359)	(0.413)	(0.360)	(0.372)	(0.430)	(0.365)
L3.Trade flow	-0.0394	0.0516	-0.0355	0.00323	0.0959	-0.0277
	(0.0623)	(0.0792)	(0.0616)	(0.0662)	(0.0842)	(0.0633)
L3.FDI flow	0.00139	0.00150	0.00108	0.000924	0.00101	0.00544
	(0.00662)	(0.00662)	(0.00663)	(0.00750)	(0.00741)	(0.0112)
Observations	2,763	2,763	2,763	2,681	2,681	2,681
R-squared	0.499	0.565	0.499	0.458	0.535	0.406
Origin-dest-industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	559	559	559	477	477	477
F				90.89	137.0	81.10

Notes: The dependent variable is the natural logarithm of $Patents_{dest}/Patents_{origin}$. Number of patents (in origin and destination countries), number of migrants, FDI, and trade flows are in natural logarithms. The sample includes country-industry pairs, where origins are NMS and destinations - EU19 countries. All specifications include year and origin-destination-industry fixed effects. Robust standard errors are clustered at the origin-destination-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.7: Convergence in Patenting Levels ($Patents_{dest}/Patents_{origin}$) and Free Labour Mobility (Reduced Form)

	(1) EU19 and NMS Patents	(2) EU19 and NMS cit. weighted	(3) NMS only Patents	(4) NMS only cit. weighted	(5) EU19 and NMS (all) Patents	(6) NMS only (all) Patents
L3.FM	-0.0135	0.00186	-0.0150	-0.00166	0.0179	-0.0131
	(0.0368)	(0.0434)	(0.0412)	(0.0496)	(0.0122)	(0.0130)
L4.FM	-0.0631	-0.0573	-0.0534	-0.0554	-0.0403***	-0.0337**
	(0.0440)	(0.0469)	(0.0505)	(0.0554)	(0.0133)	(0.0146)
L5.FM	-0.0256	-0.0393	-0.0267	-0.0283	-0.0166	-0.00647
	(0.0419)	(0.0449)	(0.0495)	(0.0534)	(0.0127)	(0.0137)
Patents, origin	-1.242***	-1.407***	-1.094***	-1.067***	-0.640***	-0.618***
	(0.0797)	(0.0873)	(0.111)	(0.122)	(0.0113)	(0.0114)
Patents, dest	1.051***	1.090***	1.062***	1.112***	0.800***	0.813***
	(0.0725)	(0.0921)	(0.0729)	(0.0929)	(0.0216)	(0.0218)
Within EU	-0.00662	-0.100*	0.0241	-0.0442	-0.0781***	-0.0527***
	(0.0494)	(0.0549)	(0.0553)	(0.0620)	(0.0141)	(0.0149)
GDP_d/GDP_o	0.00771	0.555	-0.251	0.0737	0.183***	0.175***
	(0.331)	(0.391)	(0.384)	(0.451)	(0.0393)	(0.0402)
L3.Trade flow	-0.0450	0.00903	-0.0127	0.0766	-0.0499***	-0.0341***
	(0.0629)	(0.0810)	(0.0629)	(0.0807)	(0.00866)	(0.00878)
L3.FDI flow	0.00170	0.000342	0.00259	0.00241	-0.0140***	-0.0112***
	(0.00656)	(0.00665)	(0.00651)	(0.00659)	(0.00416)	(0.00418)
Observations	2,946	2,946	2,763	2,763	71,496	66,504
R-squared	0.487	0.552	0.500	0.565	0.217	0.225
Origin-dest-ind FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	582	582	559	559	5688	5304

Notes: The dependent variable is the natural logarithm of $Patents_{dest}/Patents_{origin}$. Number of patents (in origin and destination countries), number of migrants, FDI, and trade flows are in natural logarithms. All specifications include year and origin-destination-industry fixed effects. Robust standard errors are clustered at the origin-destination-industry level. Specifications 1-4 show the reduced form regressions for the sample used in the OLS/2SLS estimations (i.e. the sub-sample for which migration data are available), specifications 5-6 show estimates for the full sample of country-industry pairs in 2000-2012.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.8: Citations to Destination Industries, NMS only, OLS and 2SLS

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
L2.Migrants	0.00255	0.00895		0.427*	0.436*	
	(0.0281)	(0.0282)		(0.222)	(0.224)	
L2.Migr.pat.potential			0.124			5.695
			(0.133)			(5.457)
Patents, origin		0.124***	0.124***		0.146***	0.158**
		(0.0332)	(0.0331)		(0.0369)	(0.0622)
L3.Patents, dest		0.0118	0.0121		0.0183	0.0317
		(0.0224)	(0.0224)		(0.0248)	(0.0332)
Within EU		-0.00869	-0.00991		-0.0280	-0.0827
		(0.0608)	(0.0609)		(0.0637)	(0.0991)
L3.Trade flow		-0.0575	-0.0566		-0.122	-0.0791
		(0.0722)	(0.0724)		(0.0837)	(0.0895)
L3.FDI flow		0.00342	0.00306		0.00393	-0.0129
		(0.0122)	(0.0122)		(0.0127)	(0.0268)
Observations	2,763	2,763	2,763	2,681	2,681	2,681
R-squared	0.083	0.087	0.088			
Origin-dest-industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	559	559	559	477	477	477
F				11.64	8.404	6.418

Notes: The dependent variable is the number of citations from a region and industry to another country in a year. Citation counts, number of migrants, total number of patent application in origin and destination industries, FDI and trade flows are taken in natural logarithms. The sample is limited to new EU member states. All specifications include year and origin-destination-industry level. Robust standard errors are clustered at the origin-destination-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.9: Citations to Destination Industries and Free Labour Mobility (Reduced Form)

	(1)	(2)	(3)	(4)
			EU19 and NMS	NMS only
	EU19 and NMS	NMS only	(all)	(all)
L3.FM	0.00662	0.0785	0.0400**	0.0670**
	(0.0349)	(0.0516)	(0.0163)	(0.0306)
L4.FM	0.0734	0.0856	0.0431**	0.0903**
	(0.0451)	(0.0603)	(0.0182)	(0.0392)
L5.FM	0.0480	0.0753	0.0255	0.0406
	(0.0470)	(0.0559)	(0.0169)	(0.0337)
Patents, origin	0.138***	0.134***	0.0591***	0.0974***
	(0.0238)	(0.0308)	(0.00785)	(0.0119)
L3.Patents, dest	0.0478***	0.000519	0.0258***	-0.0130*
	(0.0137)	(0.0192)	(0.00541)	(0.00687)
Within EU	0.0154	0.163***	0.00274	0.180***
	(0.0361)	(0.0553)	(0.0144)	(0.0247)
L3.Trade flow	-0.152***	-0.0732	-0.0627***	0.0372**
	(0.0352)	(0.0596)	(0.00955)	(0.0144)
L3.FDI flow	-0.000418	0.0114	0.0257***	0.0235***
	(0.00520)	(0.0113)	(0.00357)	(0.00639)
Observations	7,279	3,498	29,604	11,851
R-squared	0.174	0.133	0.099	0.110
Origin-dest-industry FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Clusters	1322	592	2304	912

Notes: The dependent variable is the number of citations from a region and industry to another country in a year. Citation counts, number of migrants, total number of patent application in origin and destination industries, FDI and trade flows are taken in natural logarithms. All specifications include year and origin-destination-industry level. Robust standard errors are clustered at the origin-destination-industry level. Columns 1 and 2 show the reduced form regressions for the sample used in the OLS/2SLS estimations (i.e. the sub-sample for which migration data are available), columns 3 and 4 show estimates for the full sample of country-industry pairs in 2000-2012.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Sources: Patstat, Eurostat, CEPII

Table B.10: Citations to Destination Industries, USPTO Patents Only, OLS and 2SLS

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
L2.Migrants	0.0476**	0.0313*		0.679***	0.288	
	(0.0188)	(0.0178)		(0.197)	(0.184)	
L2.Migr.pat.potential			0.0512			0.745
			(0.0485)			(1.703)
Patents, origin		0.193***	0.194***		0.186***	0.195***
		(0.0221)	(0.0221)		(0.0227)	(0.0227)
L3.Patents, dest		0.0545***	0.0542***		0.0542***	0.0491**
		(0.0147)	(0.0147)		(0.0150)	(0.0195)
Within EU		0.00444	0.00724		-0.00468	0.0300
		(0.0332)	(0.0332)		(0.0343)	(0.0667)
L3.Trade flow		0.0797*	0.0858**		0.0294	0.0853**
		(0.0416)	(0.0417)		(0.0552)	(0.0418)
L3.FDI flow		-0.00960*	-0.0102*		-0.00738	-0.0152
		(0.00526)	(0.00527)		(0.00568)	(0.0134)
Observations	7,299	7,287	7,287	7,136	7,124	7,124
R-squared	0.132	0.150	0.149			
Origin-dest-industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	1322	1320	1320	1159	1157	1157
F				44.41	35.32	34.64

Notes: The dependent variable is the number of citations from a region and industry to another country in a year. Citation counts, number of migrants, total number of patent application in origin and destination industries, FDI and trade flows are taken in natural logarithms. The sample is limited to citations among US patents. All specifications include year and origin-destination-industry level. Robust standard errors are clustered at the origin-destination-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Sources: Patstat, Eurostat, CEPII

Table B.11: Only Citations Added by the Applicant

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
L2.Migrants	0.0234	0.0258		0.489***	0.336*	
	(0.0176)	(0.0172)		(0.170)	(0.192)	
L2.Migr.pat.potential			0.0759**			1.239
			(0.0342)			(1.915)
Patents, origin		0.149***	0.150***		0.139***	0.150***
		(0.0220)	(0.0220)		(0.0231)	(0.0233)
L3.Patents, dest		0.0253	0.0247		0.0249	0.0162
		(0.0161)	(0.0161)		(0.0165)	(0.0221)
Within EU		-0.0992***	-0.0958***		-0.110***	-0.0574
		(0.0335)	(0.0334)		(0.0354)	(0.0712)
L3.Trade flow		-0.0194	-0.0143		-0.0805	-0.0153
		(0.0369)	(0.0369)		(0.0542)	(0.0373)
L3.FDI flow		0.00811	0.00735		0.0108*	-0.000953
		(0.00506)	(0.00508)		(0.00554)	(0.0154)
Observations	7,299	7,287	7,287	7,136	7,124	7,124
R-squared	0.070	0.080	0.080			
Origin-dest-industry FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Clusters	1322	1320	1320	1159	1157	1157
F				22.90	20.46	18.95

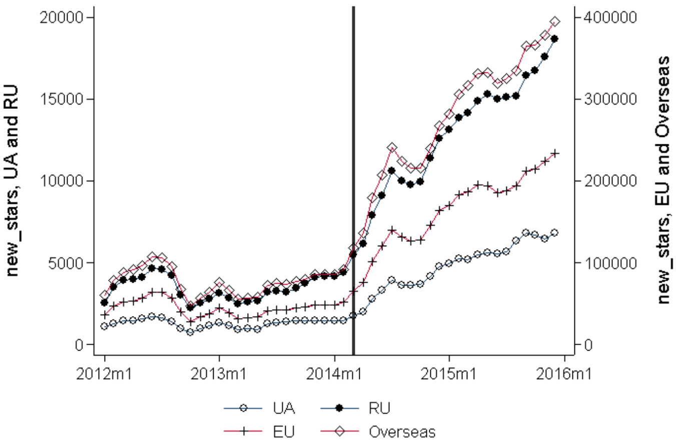
Notes: The dependent variable is the number of citations from a region and industry to another country in a year. Citation counts, number of migrants, total number of patent application in origin and destination industries, FDI and trade flows are taken in natural logarithms. The sample is limited to citations which have been added by the applicant according to PATSTAT. Robust standard errors are clustered at the origin-destination-industry level.
*** p<0.01, ** p<0.05, * p<0.1
Source: Eurostat and PATSTAT.

Appendix C

Appendix to Chapter 3

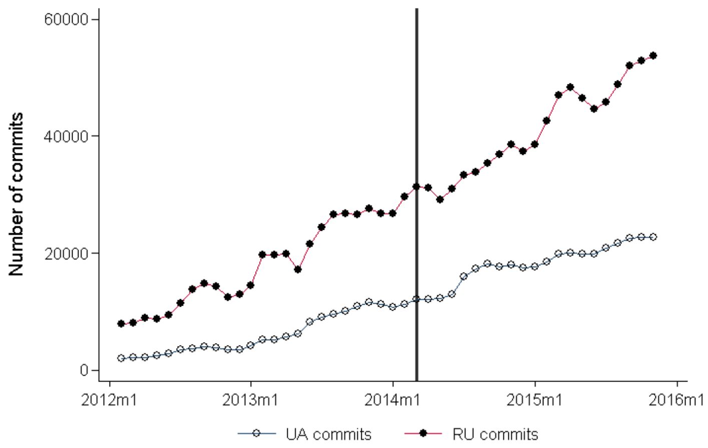
C.1 Additional Tables and Graphs

Figure C.1: New "Star Events" on GitHub



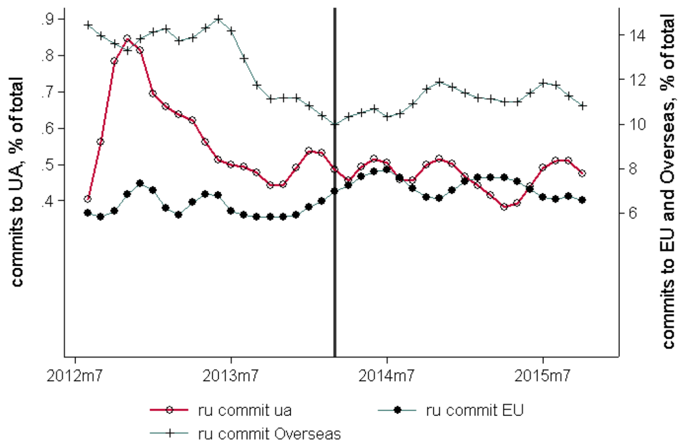
Notes: A "star" event is recorded whenever a user puts a star (a like) on a particular project. The amount of stars is used by GitHub as one of the quality measures. The region is determined by the location of the owner, whose project is "starred". The monthly data on new "star" events is smoothened with a 3-month moving average filter. The vertical black line corresponds to March 2014 (The Crimean Referendum).

Figure C.2: Commits by Ukrainian and Russian Users on GitHub



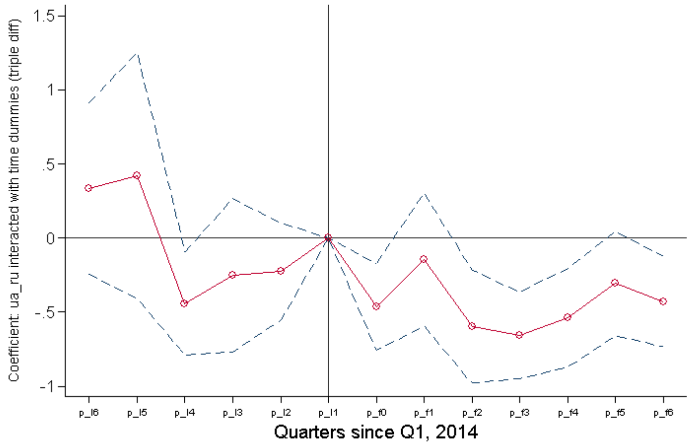
Notes: A "commit" event is recorded whenever a user modifies the source code of a project. The monthly data on commits is smoothened with a 3-month moving average filter. The vertical black line corresponds to March 2014 (The Crimean Referendum).

Figure C.3: Commits by Russian Users to Ukrainian, EU, and Overseas Projects



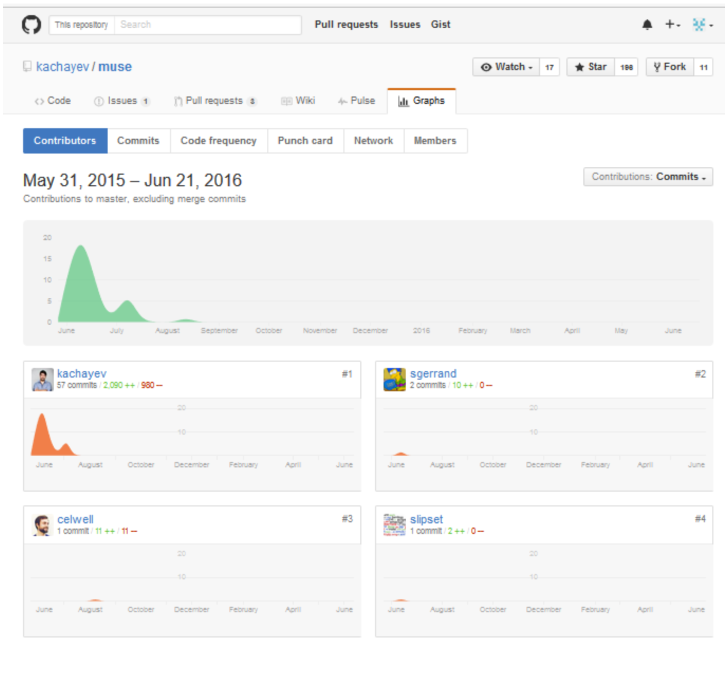
Notes: A "commit" event is recorded whenever a user modifies the source code of a project. The graph shows the commits by Russian users to projects owned by Ukrainian, EU, and Overseas users as a share of total Russian commits. The monthly data on commits is smoothened with a 3-month moving average filter. The vertical black line corresponds to March 2014 (The Crimean Referendum).

Figure C.4: Commits of Ukrainian Programmers to Russian Projects: Quarterly Treatment Effect of the Conflict, Triple Difference



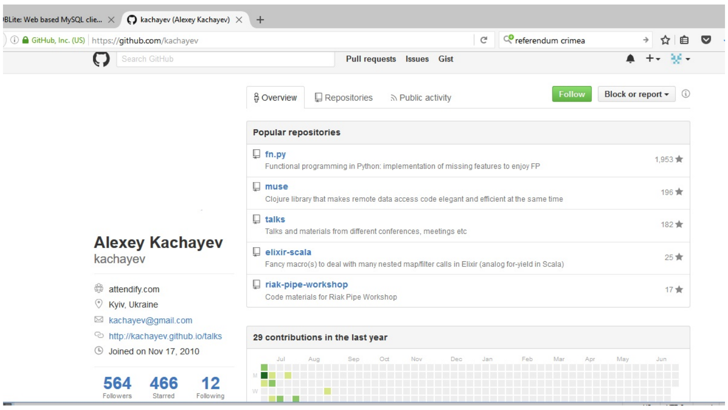
Notes: The graph plots the interaction coefficients between quarterly time dummies and the indicator for Ukrainian commits to Russian projects (estimated relative to the commits from the control countries to Russia). Dependent variable: ln(commits+1) aggregated by month, committer-region, and project-region. Controls: month and committing*receiving region fixed effects. The vertical black line corresponds to the reference period (Q4 2013).

Figure C.5: Profile of a Public Github Project



Notes: This view of a project is accessible to all registered Github users.

Figure C.6: Profile of a Github User



Notes: This view of a user profile is accessible to all registered Github users. Github links the user profile to all public projects, to which a user has contributed.

C.2 Tables

Table C.1: Summary of Github Tables

GitHub Table	Variables
Users	user id, registration date, deleted (dummy), fake (dummy), login, name, email, country code (generated), location (self-reported), company name, type (organisation or individual)
Projects	project id, registration date, deleted (dummy), url, owner id, name, description, language, forked from
Events (commits, issues, pull requests, stars, follows)	date, event type, user id, project id

Notes: GitHub Data Dump from http://ghtorrent.org/

Table C.2: Treated and Control Projects on GitHub

	(1) treat = 0 all	(2) treat = 1 all	(3) treat = 0 matched	(4) treat = 1 matched	(5) Diff. T-C (se)
Continuous treat	7.83e-05 [0.00450]	0.0426 [0.100]	2.09e-05 [0.000424]	0.0434 [0.101]	0.04*** (0.00)
Commits	83.33 [529.5]	540.6 [1,802]	446.0 [2,187]	483.0 [1,304]	37.00 (96.92)
Commits, weight.	35.71 [247.5]	248.0 [824.7]	216.5 [1,073]	223.5 [614.3]	7.04 (47.07)
Mean commits	16.15 [62.22]	59.65 [180.6]	55.48 [244.3]	54.19 [133.2]	-1.29 (10.59)
Commits, w/t RU and UA	74.93 [522.8]	476.4 [1,639]	428.1 [2,179]	426.2 [1,261]	-1.92 (95.83)
Number of members	0.282 [1.935]	0.637 [4.816]	0.351 [1.494]	0.457 [1.685]	0.11 (0.09)
Project age, m.	12.75 [14.01]	19.15 [16.94]	18.67 [16.51]	18.75 [16.65]	0.08 (0.89)
Stars	1.464 [15.84]	4.733 [25.95]	3.622 [19.63]	4.725 [26.16]	1.10 (1.25)
Stars, w/t RU and UA	1.456 [15.75]	4.703 [25.81]	3.591 [19.48]	4.694 [26.01]	1.10 (1.24)
Observations	30,105	703	690	690	1,380

Notes: The table compares projects with mixed Russian-Ukrainian teams (treated) to other projects on GitHub. Columns 1 and 2 compare all treated and non-treated projects. Columns 3 and 4 compared treated and non-treated projects in the matched sample. Column 5 shows the differences in values between Columns 4 and 3.

Table C.3: The Effect of the Conflict on Ukrainian Commits to Russian Projects: Extensive Margin

	(1) Collab. Diff-in-diff	(2) Collab. Diff-in-diff	(3) Collab. Triple diff	(4) Collab. Triple diff	(5) Collab. Triple diff
UA_cRU_pPOST	-0.450***	-0.417***	-0.384***	-0.355***	-0.338***
	(0.0776)	(0.0768)	(0.131)	(0.103)	(0.100)
RU_pPOST			-0.0432	-0.0781	-0.0785
			(0.107)	(0.0696)	(0.0696)
New projects, 3m		0.948***	0.644***	0.504***	0.490***
		(0.174)	(0.143)	(0.133)	(0.133)
Stars, 3m		0.231	0.236	0.0581	0.0692
		(0.388)	(0.253)	(0.289)	(0.289)
Observations	288	288	575	863	863
R^2	0.976	0.979	0.976	0.983	0.983
Month FE	yes	yes			
Project-region FE	yes	yes			
Committer-region*project-region FE			yes	yes	yes
Committer-region*month FE			yes	yes	yes
Extended control group				yes	yes
Without Crimea and Donbass					yes
Robust	yes	yes	yes	yes	yes

Notes: The dependent variable in all specifications is $ln(collaborations+1)$, where *collaborations* is the monthly number of unique user-project collaborations between a committer-region (c) and a project-region (p). UA_cRU_pPOST - a dummy for commits by Ukrainian programmers to Russian projects. $POST$ period starts after March 2014. *New projects*, $3m$ and *Stars*, $3m$ - the number of new and "starred" projects from a project-region over the last 3 months. In columns 1 and 2, the sample includes commits by Ukrainian programmers to projects from different regions (Ukraine, Russia, EU, Overseas, Other). In columns 3, the sample includes commits by Ukrainian and Belarusian programmers; columns 4-5 in addition include commits from other control countries. Column 5 drops commits from programmers in Crimea and Donbass before aggregating Ukrainian commits.

Table C.4: The Effect of the Conflict on Ukrainian Commits to Russian Projects: Company vs. Individual Projects

	(1) Collab.	(2) Commits	(3) Collab.	(4) Commits
Company*RU_pPOST	-0.0468	0.109	-0.0798	0.124
	(0.147)	(0.328)	(0.141)	(0.340)
RU_pPOST	-0.282***	-0.546***	-0.267***	-0.612***
	(0.107)	(0.193)	(0.0990)	(0.207)
New projects, 3m	0.924***	1.001***	0.909***	0.990***
	(0.153)	(0.269)	(0.155)	(0.272)
Stars, 3m	0.0710	0.361	0.0843	0.400
	(0.189)	(0.330)	(0.188)	(0.332)
Observations	575	575	575	575
R^2	0.967	0.928	0.968	0.926
Month FE	yes	yes	yes	yes
Company*project-region FE	yes	yes	yes	yes
Robust	yes	yes	yes	yes
Without Donbass and Crimea			yes	yes

Notes: The dependent variable in Columns 1 and 3 is $ln(collaborations + 1)$; in Columns 2 and 4 - $ln(commits + 1)$. $Company * RU_pPOST$ - a dummy for commits by Ukrainian programmers to Russian projects owned by companies. $POST$ period starts after March 2014. Baseline group - commits of Ukrainian programmers to individual projects. $New\ projects$, $3m$ and $Stars$, $3m$ - the number of new and "starred" projects from the project-region over the last 3 months. All specifications include $identity * project - region$ fixed effects. Columns 3 and 4 drop commits from programmers in Crimea and Donbass before aggregating Ukrainian commits.

List of Figures

List of Figures

List of Tables

Bibliography

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity". *The Quarterly Journal of Economics* 115 (3), pp. 715–753.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2005). "Social Preferences and the Response to Incentives: Evidence from Personnel Data". *The Quarterly Journal of Economics* 120 (3), pp. 917–962.

– (2009). "Social connections and incentives in the workplace: Evidence from personnel data". *Econometrica* 77 (4), pp. 1047–1094.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta (2013). "Cross-Country Differences in Productivity: The Role of Allocation and Selection". *American Economic Review* 103 (1), pp. 305–34.

Beine, Michel, Frederic Docquier, and Hillel Rapoport (2001). "Brain drain and economic growth: theory and evidence". *Journal of Development Economics* 64 (1), pp. 275–289.

Belenzon, Sharon and Mark Schankerman (2008). *Motivation and Sorting in Open Source Software Innovation*. CEPR Discussion Papers 7012. C.E.P.R. Discussion Papers.

Berge, Lars Ivar Oppedal et al. (2016). *How Strong are Ethnic Preferences?* NBER Working Papers 21715. National Bureau of Economic Research, Inc.

Blackwell, Matthew et al. (2009). "cem: Coarsened exact matching in Stata". *Stata Journal* 9 (4), pp. 524–546.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen (2012). "Americans Do IT Better: US Multinationals and the Productivity Miracle". *American Economic Review* 102 (1), pp. 167–201.

Bloom, Nicholas, Mirko Draca, and John Van Reenen (2016). "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity". *Review of Economic Studies* 83 (1), pp. 87–117.

Bosetti, Valentina, Cristina Cattaneo, and Elena Verdolini (2015). "Migration of skilled workers and innovation: A European Perspective". *Journal of International Economics*.

Braunerhjelm, Pontus, Ding Ding, and Per Thulin (2015). *Does Labour Mobility Foster Innovation? Evidence from Sweden*. Working Paper Series in Economics and Institutions of Innovation 403. Royal Institute of Technology, CESIS - Centre of Excellence for Science and Innovation Studies.

Brown, Charles and James Medoff (1978). "Trade Unions in the Production Process". *Journal of Political Economy* 86 (3), pp. 355–78.

Charness, Gary, Luca Rigotti, and Aldo Rustichini (2007). "Individual Behavior and Group Membership". *American Economic Review* 97 (4), pp. 1340–1352.

Chen, Roy et al. (2014). *Does team competition increase pro-social lending? Evidence from online microfinance*. Discussion Papers, Research Unit: Market Behavior SP II 2014-209. Social Science Research Center Berlin (WZB).

Chen, Yan and Sherry Xin Li (2009). "Group identity and social preferences". *The American Economic Review*, pp. 431–457.

Choudhury, Prithwiraj (2015). "Return migration and geography of innovation in MNEs: a natural experiment of knowledge production by local workers reporting to return migrants". *Journal of Economic Geography* 3 (16), pp. 585–610.

Clemens, Michael A. (2013). "Why Do Programmers Earn More in Houston Than Hyderabad? Evidence from Randomized Processing of US Visas". *American Economic Review* 103 (3), pp. 198–202.

Constant, Amelie F. (2011). *Sizing It Up: Labor Migration Lessons of the EU Enlargement to 27*. IZA Discussion Papers 6119. Institute for the Study of Labor (IZA).

Docquier, Frederic and Hillel Rapoport (2012). "Globalization, Brain Drain, and Development". *Journal of Economic Literature* 50 (3), pp. 681–730.

Duflo, Esther (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment". *American Economic Review* 91 (4), pp. 795–813.

Duguet, Emmanuel and Megan MacGarvie (2005). “How well do patent citations measure flows of technology? Evidence from French innovation surveys”. *Economics of Innovation and New Technology* 14 (5), pp. 375–393.

Dustmann, Christian and Albrecht Glitz (2015). “How Do Industries and Firms Respond to Changes in Local Labor Supply?” *Journal of Labor Economics* Vol. 33 (3).

Dustmann, Christian, Tommaso Frattini, and Anna Rosso (2015). “The Effect of Emigration from Poland on Polish Wages”. *Scandinavian Journal of Economics* 117 (2), pp. 522–564.

Elsner, Benjamin (2013). “Does emigration benefit the stayers? Evidence from EU enlargement”. *Journal of Population Economics* 26 (2), pp. 531–553.

Fisman, Raymond, Yasushi Hamao, and Yongxiang Wang (2013). “The impact of interstate tensions on economic exchange: Evidence from shocks to Sino-Japanese relations”. *Unpublished working paper. Columbia University.*

Fox, Jeremy T. and ValÃ©rie Smeets (2011). “Does Input Quality Drive Measured Differences In Firm Productivity?” *International Economic Review* 52 (4), pp. 961–989.

Freeman, Richard B. (2006). “People Flows in Globalization”. *Journal of Economic Perspectives* 20 (2), pp. 145–170.

Gorodnichenko, Yuriy and Monika Schnitzer (2013). “Financial constraints and innovation: Why poor countries don’t catch up”. *Journal of the European Economic Association* 11 (5), pp. 1115–1152.

Griliches, Zvi (1979). “Issues in Assessing the Contribution of Research and Development to Productivity Growth”. *Bell Journal of Economics* 10 (1), pp. 92–116.

Grossmann, Volker and David Stadelmann (2011). “Does international mobility of high-skilled workers aggravate between-country inequality?” *Journal of Development Economics* 95 (1), pp. 88–94.

– (2013). “Wage Effects of High-Skilled Migration: International Evidence”. *World Bank Economic Review* 27 (2), pp. 297–319.

Hergueux, Jérôme and Nicolas Jacquemet (2015). “Social preferences in the online laboratory: a randomized experiment”. *Experimental Economics* 18 (2), pp. 251–283.

Hjort, Jonas (2014). “Ethnic divisions and production in firms”. *The Quarterly Journal of Economics* 129 (4), pp. 1899–1946.

Hu, Albert G. Z. and Adam B. Jaffe (2003). “Patent citations and international knowledge flow: the cases of Korea and Taiwan”. *International Journal of Industrial Organization* 21 (6), pp. 849–880.

Jaffe, Adam B (1986). “Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits, and Market Value”. *American Economic Review* 76 (5), pp. 984–1001.

– (1989). “Real Effects of Academic Research”. *American Economic Review* 79 (5), pp. 957–70.

Jaffe, Adam B and Manuel Trajtenberg (1999). “International knowledge flows: evidence from patent citations”. *Economics of Innovation and New Technology* 8 (1-2), pp. 105–136.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson (1993). “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations”. *The Quarterly Journal of Economics* 108 (3), pp. 577–98.

Jäger, Simon (2016). *How Substitutable Are Workers? Evidence from Worker Deaths.* Job Market Paper.

Kahanec, Martin (2012). *Labor Mobility in an Enlarged European Union.* IZA Discussion Papers 6485. Institute for the Study of Labor (IZA).

Kahanec, Martin, Mariola Pytlikova, and Klaus F. Zimmermann (2014). *The Free Movement of Workers in an Enlarged European Union: Institutional Underpinnings of Economic Adjustment.* IZA Discussion Papers 8456. Institute for the Study of Labor (IZA).

Kaiser, Ulrich, Hans Christian Kongsted, and Thomas Ronde (2015). “Does the Mobility of R & D Labor Increase Innovation?” *Journal of Economic Behavior & Organization* 110 (336), pp. 91–105.

Kalliamvakou, Eirini, Georgios Gousios, and Blincoe Kelly (2014). “The promises and perils of mining GitHub (Extended Version)”. *MSR 2014 Proceedings of the 11th Working Conference on Mining Software Repositories* (1), pp. 92–101.

Kerr, Sari Pekkala and William R. Kerr (2013). “Immigration and Employer Transitions for STEM Workers”. *American Economic Review* 103 (3), pp. 193–97.

– (2015). *Global Collaborative Patents.* NBER Working Papers 21735. National Bureau of Economic Research, Inc.

Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln (2013). *Skilled Immigration and the Employment Structures of U.S. Firms*. NBER Working Papers 19658. National Bureau of Economic Research, Inc.

– (2014). “Firms and the Economics of Skilled Immigration”. *Innovation Policy and the Economy, Volume 15*. NBER Chapters. National Bureau of Economic Research, Inc.

Kerr, William R. (2008). “Ethnic Scientific Communities and International Technology Diffusion”. *The Review of Economics and Statistics* 90 (3), pp. 518–537.

– (2013). *U.S. High-Skilled Immigration, Innovation, and Entrepreneurship: Empirical Approaches and Evidence*. NBER Working Papers 19377. National Bureau of Economic Research, Inc.

Kerr, William R. and William F. Lincoln (2010). “The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention”. *Journal of Labor Economics* 28 (3), pp. 473–508.

Konings, Jozef and Stijn Vanormelingen (2015). “The Impact of Training on Productivity and Wages: Firm-Level Evidence”. *The Review of Economics and Statistics* 2 (97), pp. 485–497.

Ksoll, Christopher, Rocco Macchiavello, and Ameet Morjaria (2010). “DP8074 The Effect of Ethnic Violence on an Export-Oriented Industry”.

Lazear, Edward P (1999). “Globalisation and the Market for Team-Mates”. *Economic Journal* 109 (454), pp. C15–40.

Lerner, Josh and Jean Tirole (2001). “The open source movement: Key research questions”. *European Economic Review* 45 (4-6), pp. 819–826.

– (2005). “The Economics of Technology Sharing: Open Source and Beyond”. *Journal of Economic Perspectives* 19 (2), pp. 99–120.

Levinsohn, James and Amil Petrin (2003). “Estimating Production Functions Using Inputs to Control for Unobservables”. *Review of Economic Studies* 70 (2), pp. 317–341.

Lewis, Lafortune Tessada (2013). “People and Machines A Look at The Evolving Relationship Between Capital and Skill In Manufacturing 1850-1940 Using Immigration Shocks”. Working paper.

MacGarvie, Megan (2005). “The determinants of international knowledge diffusion as measured by patent citations”. *Economics Letters* 87 (1), pp. 121–126.

MacGarvie, Megan (2006). "Do Firms Learn from International Trade?" *The Review of Economics and Statistics* 88 (1), pp. 46–60.

Marx, Benjamin, Vincent Pons, and Tavneet Suri (2015). "Diversity and team performance in a kenyan organization".

Mas, Alexandre and Enrico Moretti (2009). "Peers at Work". *American Economic Review*, pp. 1996–2003.

Mayr, Karin and Giovanni Peri (2009). "Brain Drain and Brain Return: Theory and Application to Eastern-Western Europe". *The B.E. Journal of Economic Analysis & Policy* 9 (1), pp. 1–52.

McDonald, Nora and Sean Goggins (2013). "Performance and Participation in Open Source Software on GitHub". *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '13. Paris, France: ACM, pp. 139–144. ISBN: 978-1-4503-1952-2. DOI: 10.1145/2468356.2468382.

Mechtel, Mario, Florian Hett, and Markus Kröll (2016). *Endogenous Social Identity and Group Choice*. Annual Conference 2014 (Hamburg): Evidence-based Economic Policy 100307. Verein für Socialpolitik / German Economic Association.

Mitaritonna, Cristina, Gianluca Orefice, and Giovanni Peri (2014). *Immigrants and Firms' Productivity: Evidence from France*. IZA Discussion Papers 8063. Institute for the Study of Labor (IZA).

Moser, Petra and Alessandra Voena (2012). "Compulsory Licensing: Evidence from the Trading with the Enemy Act". *American Economic Review* 102 (1), pp. 396–427.

Moser, Petra, Joerg Ohmstedt, and Paul W Rhode (2016). "Patent Citations and the Size of the Inventive Step-Evidence from Hybrid Corn". *Management Science* (forthcoming).

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

OECD (2013). *Coping with Emigration in Baltic and East European Countries*. OECD Publishing, Paris.

Ottaviano, Gianmarco I.P., Giovanni Peri, and Greg C. Wright (2015). *Immigration, Trade and Productivity in Services: Evidence from U.K. Firms*. NBER Working Papers 21200. National Bureau of Economic Research, Inc.

Ozgen, Ceren et al. (2014). “Does Cultural Diversity of Migrant Employees Affect Innovation?” *International Migration Review* 48, pp. 377–416.

Parrotta, Pierpaolo, Dario Pozzoli, and Mariola Pytlikova (2014). “The nexus between labor diversity and firm’s innovation”. *Journal of Population Economics* 27 (2), pp. 303–364.

Paserman, M Daniele (2013). “Do high-skill immigrants raise productivity? Evidence from Israeli manufacturing firms, 1990-1999”. *IZA Journal of Migration* 2 (1), pp. 1–31.

Peri, Giovanni (2005). “Determinants of Knowledge Flows and Their Effect on Innovation”. *The Review of Economics and Statistics* 87 (2), pp. 308–322.

– (2012). “The Effect Of Immigration On Productivity: Evidence From U.S. States”. *The Review of Economics and Statistics* 94 (1), pp. 348–358.

Prat, Andrea (2002). “Should a team be homogeneous?” *European Economic Review* 46 (7), pp. 1187–1207.

Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti (2013a). “Seeds of distrust: conflict in Uganda”. *Journal of Economic Growth* 18 (3), pp. 217–252.

– (2013b). “War signals: A theory of trade, trust, and conflict”. *The Review of Economic Studies* 80 (3), pp. 1114–1147.

Russell, Sharon Stanton (1986). “Remittances from international migration: A review in perspective”. *World Development* 14 (6), pp. 677–696.

Shaw, Jason D. (2011). “Turnover rates and organizational performance: Review, critique, and research agenda”. *Organizational Psychology Review* 1(3), pp. 187–213.

Singh, Jasjit and Matt Marx (2013). “Geographic constraints on knowledge spillovers: political borders vs. spatial proximity”. *Management Science* 59 (9), pp. 2056–2078.

Strober, Myra H. (1990). “Challenges to Human Capital Theory: Implications for HR Managers”. *Industrial Relations* Spring.

Syverson, Chad (2011). “What Determines Productivity?” *Journal of Economic Literature* 49 (2), pp. 326–65.

Thompson, Peter and Melanie Fox-Kean (2005). “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment”. *American Economic Review* 95 (1), pp. 450–460.

Trajtenberg, Manuel (1990). “A Penny for Your Quotes: Patent Citations and the Value of Innovations”. *RAND Journal of Economics* 21 (1), pp. 172–187.

Yanadori, Yoshio and Takao Kato (2007). “Average employee tenure, voluntary turnover ratio, and labour productivity: evidence from Japanese firms”. *The International Journal of Human Resource Management* 18, (10).

Zaiceva, Anzelika (2014). “Post-enlargement emigration and new EU members’ labor markets”. *IZA World of Labor*.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Datum:

Unterschrift: