

# Empirical essays on wage setting and immigrant labor market opportunities

**Tove Eliasson** 



Presented at Department of Economics, Uppsala University

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala Visiting address: Kyrkogårdsgatan 6, Uppsala Phone: +46 18 471 70 70 Fax: +46 18 471 70 71 ifau@ifau.uu.se www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the Department of Economics, Uppsala University, September 5, 2014. Essay 1 is a revised version of IFAU Working paper 2013:7, Essay 2 is published as IFAU Working Paper 2014:19 and Essay 3 is published as IFAU Working paper 2014:10.

ISSN 1651-4149

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Uppsala, Friday, 5 September 2014 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Tor Eriksson (Aarhus university).

#### Abstract

Eliasson, T. 2014. Empirical Essays on Wage Setting and Immigrant Labor Market Opportunities. 134 pp. Uppsala: Department of Economics, Uppsala universitet. ISBN 978-91-85519-55-2.

#### This thesis consists of three self-contained essays.

Essay 1: This essay estimates wage assimilation among non-western immigrants in Sweden, controlling for selection into employment by including individual fixed effects. Furthermore, using matched employer-employee panel data covering the complete Swedish labor market, this essay decomposes wage catch-up into relative wage growth within and between workplaces and occupations. The results show that failing to control for selection into employment is likely to underestimate relative wage growth of immigrants, as early entrants in the labor market differ from later entrants along unobservable dimensions. Even after 30 years in the country, the group of non-western immigrants still earns substantially lower wages than natives. Wages catch up mainly within workplaces and occupations, are of importance for the wage growth of non-western immigrants.

Essay 2: Earlier research has shown that immigrant- and minority entrepreneurs have difficulties accessing capital through the formal financial markets. This essay studies what role immigrant employees within the local bank sector have for the probability of immigrants to run their own businesses. I use linked employer-employee data covering the whole Swedish labor market for the years 1987 to 2003 and utilize a nationwide refugee dispersal policy to get exogenous variation in the exposure to co-ethnic bank employees. Results suggest that there is a positive relation between co-ethnic bank employees and the probability of being self-employed. This effect is most pronounced for immigrants who arrived with low education, for males and for those residing in metropolitan regions. The effects are substantial and robust to a wide set of controls for labor market characteristics of the ethnic group at the local level. These results provide evidence of an ethnic component in the formal credit markets.

Essay 3 (with Oskar Nordström Skans): This essay investigates the impact of a collective agreement stipulating a one shot increase in establishment-specific wage levels in a public-sector setting where wages otherwise are set according to individualized wage bargaining. The agreement stipulated that wages should increase in proportion to the number of low-paid females within each establishment. We find that actual wages among incumbents responded to the share of females with a wage below the stipulated threshold, conditional on the separate effects of the share of low wage earners, and the share of females. We find clear evidence of path-dependence in wages, covered workers remained on higher wage levels 4 years after the agreement took effect. The increase in wages resulted in a reduced probability of exit among young workers with relatively good grades and a lower frequency of new hires at the establishment level.

*Keywords:* Firm sorting, mobility, wage assimilation, host country specific human capital, employer learning, self-employment, immigrant entrepreneurs, capital access, information asymmetry, minority representation, collective bargaining, turnover, hours of work, labor input, labor costs

Tove Eliasson, Department of Economics, Box 513, Uppsala University, SE-75120 Uppsala, Sweden. Institute for Housing and Urban Research, Box 514, Uppsala University, SE-75120 Uppsala, Sweden.

© Tove Eliasson 2014

ISBN 978-91-85519-55-2 urn:nbn:se:uu:diva-226084 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-226084)

# Acknowledgements

The academic environment is one of constant learning, a learning accompanied by a constant feeling of not knowing enough. To deal with this uncertainty, it is crucial to be surrounded by good people, and I certainly have been.

First of all, I would like to extend my greatest gratitude to my supervisor Oskar Nordström Skans, without whom I would not be anywhere near the researcher that I am today. Thank you for struggling with my never ending questions and for being the greatest teacher of economic theory and method I have ever had. Thank you also for forcing me to keep my eyes on the ball. Luckily, I learned how to stay focused despite the endless number of new ideas, and side tracks, which popped into my mind. I would also like to thank my coadvisor Per-Anders Edin, whose critical eye and intelligence have improved my essays remarkably. I am thankful for all the times you have asked me what my results really mean. Finding an answer to that question has pushed me forward, again and again. It has been truly inspiring to have meetings with the two of you, at least after I realized that the moments when you were both quiet, staring straight into the wall, were not moments of awkward silence, but moments of great thinking. In the end, I think we sat there silent, all three of us.

I would further like to thank Matz Dahlberg, who supervised me my first years as a PhD student, and who encouraged me to apply for the PhD program in the first place. I also wish to thank Fredrik Heyman and Kristiina Huttunen who were discussants of these papers during my licentiate seminar, as well as final seminar. Thank you both for a thorough reading and many creative comments. I am also very grateful for the support from all of you who started the PhD program together with me, and in particular from Gabriella, with whom I shared office, newfound knowledge, laughs and Friday evening power music for years. Also Mattias, Linna, Arizo, Anna and several other were great companions during the time at the Department of Economics.

While writing this thesis, I have also been working at the Institute for Housing and Urban Research (IBF), where Ann, Jon, Helen and Kati become the people with whom I could air all my research related, as well as personal, issues. I am so grateful for all the workdays which you have made highly enjoyable, and for all the Chinese take-out we have enjoyed together. The list of people at IBF to thank is long, but I hope you all understand that I have learned a lot from you and that I have found IBF to be a very stimulating and heartwarming work environment. A special thanks to Cecilia, though, for your endless support and help. The administrative staff at both the Department of Economics and the Institute for Housing and Urban Research has given me invaluable help so many times, thank you for that.

In terms of academic possibilities, I want to thank Lori Beaman who gave me the chance to visit the Department of Economics at Northwestern University for a semester. The discussions I had there, and the seminars I could follow, made me a more grown-up economist. Thank you also, Daniel, who shared these months in Evanston with me. I have also had the opportunity to spend some research time at the Institute for Evaluation of Labour Market and Education Policy (IFAU) in Uppsala, and I thank Olof Åslund for this opportunity. Thank you also to everyone at IFAU for being so kind and share so much of your knowledge regarding the labor market with me.

Academia is one thing, but while working with the thesis, I have also tried to maintain a private life. Even though I have too often bailed on my friends to stay at home and run regressions, I am overjoyed that you are still around, and I do hope we have more time for dancing now when this thesis is done. Thank you Emma and Emeli for being such good friends, no matter what, and a special thanks to Linda for always being there, and always offering a couch to sleep on when trains were canceled. To everyone else, I am grateful for you not giving up on me, and I am very sorry for all the times I have been a bore.

My family has done so much for me during these years, mainly by filling my life with great things, far from Economics. For that I am forever grateful. And Jonatan, thank you for making me laugh all the time.

Årsta, June 2014 Tove Eliasson

# Contents

1	Immigrant wage assimilation - the importance of unobserved hetero-						
	gen	geneity, workplaces and occupations 1.					
	1	Introduction	14				
	2	Data	16				
		2.1 Descriptive statistics	18				
	3	Describing relative wage growth	20				
	4	Modeling wage catch-up	22				
		4.1 Compositional bias of wage catch-up estimates	24				
		4.1.1 Controls for local labor market conditions .	27				
	5	Decomposing wage catch-up	28				
		5.1 Wage catch-up within and between workplaces	30				
		5.1.1 Relaxing the assumption of equal workplace					
		effects	34				
		5.2 Wage catch-up within and between occupations	35				
	6 Discussion						
	References						
	Appendix A Data and variable description						
	Appendix B Robustness: choice of wage measure and sample 4						
	Appendix C Analysis divided by sex						
	Ap	pendix D Detailed regression results	52				
2	Immigrant entrepreneurship and the origin of bankers 5'						
	1	Introduction	58				
	2	Capital access for immigrant entrepreneurs	60				
		2.1 Lower rates of bank loans	60				
		2.2 Co-ethnic bankers and potential capital access	61				
	3	Data and sample selection	63				
		3.1 Defining the explanatory variable	64				
		3.2 Descriptive statistics	64				
	4	Econometric strategy	65				
		4.1 OLS specification	65				
		4.2 Municipal group level covariates	67				

1

	5	OLS results		68
	6	Analysis based on initial	exogenous variation	71
		6.1 Placement of ref	fugees: the policy	71
		6.2 Modeling the in	npact of placement	72
		6.3 Instrumental var	iable analysis	73
		6.4 Results based on	initial exogenous variation	74
		6.5 The importance	of local group quality	77
	7	Extensions		79
		7.1 Own-group emp	ployees of a similar sector	79
		7.2 Dynamics of sel	f-employment: entry and exit	80
	8	Sensitivity		82
	9	Concluding remarks		85
	Refe	rences		87
	App	endix A Variable descr	iption	90
	App	endix B List of munici	ipal group level covariates	92
	App	endix C Description; ł	oank, insurance and self-employment .	94
	App	endix D Further regres	ssion results	96
2	Nag	stisted wars in anosco and	the labor meriliest outcomes of low was	
3	Ineg	or a cridence from the S	undial public costor (with Oslice Nord	e
	woir	steps)	wedish public sector (with Oskar Nord	- 101
	1	Introduction		101
	2	Institutional background		102
	4	2.1 The 2007 agreen		105
	3	Data and descriptive stat	tietice	100
	<u>J</u>	The first stage impact: w	rage growth during 2007-2008	111
	т	4.1 Restrictions		113
		4.2 Effect of individ	ual treatment status	113
		4.3 Effect of the est	ablishment level treatment intensity	115
	5	Post-agreement outcome	es for the covered individuals	118
	5	5.1 Wage trajectorie	s	118
		5.2 Hours of work		120
		5.3 Separations		122
		5.4 Establishment-l	evel responses	122
		5.5 Robustness		124
	6	Conclusions		126
	Refe	rences		129
	App	endix A Descriptives		131
	App App	endix A Descriptives endix B Further regres	sion result	131 134

# Introduction

This thesis consists of three self-contained essays, all contributing to our understanding of the determinants of individuals' labor market outcomes. People who enter the labor market do not only care about finding a job, but also about the wages, hours and working conditions that the job entails. These are all examples of what economists refer to as labor market outcomes, and studying the determinants of these outcomes is one of the main focuses within Labor Economics.

As labor economists it is our job to explain, describe and analyze the mechanisms of the labor market, so that individuals, policy makers and social partners can make informed decisions. While two essays in this thesis study the labor market outcomes of non-western immigrants, the third instead focuses on blue collar public sector workers, mainly women. Both these groups have a relatively weak position in the labor market, and for policy makers who are concerned with the welfare of the least advantaged individuals in the society, knowledge about how the labor market functions for these groups is important.

The literature proposes a wide range of explanatory factors to understand individual differences in labor market outcomes. These include individual level factors such as investments made by the individual in education, health and social capital; factors typically referred to as human capital investments (Becker, 1962; Mincer, 1970). The human capital perspective can explain why the labor market attachment is weaker for young workers, with less experience and less accumulated human capital, and why the position in the labor market improves as experiences and human capital accumulate.

But there are also factors at the group level, as well as the local and national level, which influence the employment and earnings prospects of individuals. Local level factors, such as the characteristics of neighbors, networks and peers also matter for the way individuals partake in the labor market. One example is that the use of friends and relatives is a particularly efficient job-search channel (Ioannides and Loury, 2004). Thus, informal contacts, as provided by schools, workplaces and neighborhoods may influence labor market outcomes. Furthermore, social norms and institutions matter, and in societies plagued by prejudices, individuals with a minority background may experience difficulties which the majority group does not (Becker, 1971).

In addition, the role played by labor market institutions, such as the degree of unionization, bargaining coverage and employment protection, for aggregate and individual economic outcomes has been extensively studied. A general conclusion from this literature is that countries with a more heavily regulated labor market, which Sweden is definitely defined as, exhibit lower levels of earnings inequality than countries in which the labor market is deregulated, such as for example the US (see e.g. Freeman (2007) for an overview).

The efficiency of the labor market is constantly hampered by informational deficiencies. Neither employers nor individuals have perfect information regarding each other, and thus all employment decisions are associated with uncertainty. The employer observes a noisy signal of productivity, and the noisier the signal is, the less certain can she be about the outcome of hiring the worker (Oettinger, 1996; Spence, 1973). Therefore, imperfect information may be detrimental both for the employment prospects of individuals, but also for the aggregate matching process in the labor market.

Although this is not an exhaustive overview of the factors which affect the observed patterns of employment and earnings, it stands clear that many labor market outcomes at the individual and aggregate level are determined by a complex mix of factors. This thesis aims at studying the relevance which individual abilities, uncertainty, personal interactions and institutions have for individual labor market outcomes. Both essay 1 and 2 focus on understanding the process through which non-western immigrants establish a position in the host country labor market. Essay 1 studies the wage growth of non-western immigrants and the roles which mobility and sorting in the labor market have for relative wage increases compared to native workers. The focus of essay 2 is on how owngroup members employed in local banks influence self-employment probability among non-western immigrants. Bankers may influence self-employment probability through an increasing access to small business credit. A common factor discussed in both these essays is the potential difficulty for immigrants to properly signal skills and experiences in a host country labor market.

Essay 3 (co-authored with Oskar Nordström Skans) instead focuses on collective bargaining and its role for actual wage growth, and subsequent labor market outcomes, in the blue collar, predominantly female, public sector. Empirical studies on how negotiated wages are implemented in practice are scarce, and most of the existing studies have focused on the impact of changes in the negotiated minimum wages. Here we shed light on how a centrally negotiated wage agreement affects workers locally, thus contributing with knowledge regarding the effects of bargaining institutions.

As a researcher it is crucial to distinguish the effect of interest from confounding effects of factors which cannot be controlled for. All essays in this thesis is concerned with this issue in one way or another. Finally, all essays answer questions with the help of population-wide matched employer-employee data. By using these types of data, through which characteristics of individuals and their workplaces can be studied simultaneously, labor economists have in recent years been able to study questions which could not be answered earlier. This thesis contributes to this literature. The essays in the thesis are described more in detail below.

## Immigrants' labor market position

Internationally, the position of migrants in host country labor markets has been a concern for policy makers for many years. The process of integration is different, though, for immigrants who come from parts of the western world, and for those who do not. Western immigrants experience good economic outcomes in many countries, or at least much better economic outcomes than the group of immigrants from non-westerns countries. Non-western immigrants start out at low employment rates and earnings, and/or experience slow catch-up with the native populations (see for example Sarvimäki (2011) for Finland and Hammarstedt and Shukur (2006) for Sweden).

Sweden, which is studied in this thesis, is a country where immigrants' employment rate relative to natives' is very low, by international standards. According to OECD (2012), Sweden is the country where recent immigrants face the lowest relative employment rate of all countries studied. The employment rate is increasing over time since migration, but the employment gap (relative to natives) is not fully closed after 25 years in the country (Nekby, 2002).

It is of great policy interest to understand what causes these poor labor market outcomes for immigrants from low income source countries. To some extent, group differences can be explained by factors which we can observe and control for, and one often considered difficulty for immigrants is the lack of host country-specific human capital, such as cultural specific knowledge or language skills (Lalonde and Topel, 1997). If the outcome gap is influenced by the lack of human capital, a fruitful policy would be to improve on, or facilitate access to, language training and schooling for the immigrants. But generally, after controlling for observed factors, a large outcome gap between immigrants and natives remains unexplained (Hammarstedt, 2003; Le Grand and Szulkin, 2002). This unexplained part is often attributed to discrimination.

Discrimination can be based on the preferences of co-workers, employers, as well as customers (Becker, 1971), and indeed we have seen evidence of ethnic discrimination in hires in Sweden (see for example Carlsson and Rooth (2007)). But discrimination could also be based on imperfect information regarding the individual, which leads employers to make decisions based on expectations of the group as a whole, rather than of the individual. This is what is usually called statistical discrimination (Altonji and Blank, 1999; Phelps, 1972). An employer with little knowledge of an immigrant group, might find it difficult to evaluate differences in experiences and human capital among the groups' members, and will therefore not differentiate wages between group-members according to individual productivity.

There are of course several other factors related to the poor labor market outcomes of immigrants in western labor markets, many of which have been extensively studied within labor economics. These include the lack of labor market networks, segregation, and spatial mismatch between jobs and residence, to mention some.

## Wage assimilation

In essay 1 of this thesis I study how wages of non-western immigrants converge towards wages of comparable natives with time spent in Sweden. I furthermore decompose relative wage growth into a part which takes place within workplaces and occupations and a part associated to movements between workplaces and occupations. From this analysis we learn how mobility and sorting between firms influence the relative wage growth of natives and immigrants.

The analysis of immigrant wage assimilation is associated with a methodological problem as the immigrant employment rate is low, and thus the group of employed immigrants represents a selected subsample of the immigrant group. This selection process becomes less important with time since immigration, due to increasing employment rates, and therefore it is likely that the group of immigrants employed the first years in the country differs from the individuals who get their first job a couple of years later. If these individuals differ from each other in some dimension which we cannot properly control for this selective employment rate will bias the estimates of wage catch-up. It is possible, for example, that individuals with high ability to communicate in the host country labor market manage to get a job early, and thus estimated relative wage is high the first years after immigration. As more individuals manage to get a job, these individuals might be hired at lower paying positions, and the relative wage will therefore be lower when more individuals are employed. This scenario would lead to an underestimation of the relative wage increases, if we could not account for these unobserved differences between the individuals.

With the increasing access to panel data, where individuals can be followed over a period of time, researchers have developed ways to control for unobserved characteristics of individuals which are constant over time. This could be factors such as innate ability to learn, motivation, or as in the above example; a communicative ability. In the estimations performed in essay 1, I therefore control for the part of an individuals' wage which can be explained by the timeinvariant characteristic of the individual, and by doing so I can estimate wage assimilation, netting out the compositional bias of the group of employed workers.

The results show that we are likely to underestimate wage catch-up if we do not control for these unobservable individual level factors. When we do take these factors into account we see that non-western immigrants in Sweden experience a relative wage catch-up of between 6 and 13 percentage points (depending on education level) but they do not catch-up with the wages of the native population, even after 30 years in the country.

In many countries, a large part of wage differences comes from wage differences within firms, but during the last decades wage differences between firms have become increasingly important (Lazear and Shaw, 2005). This is also true for Sweden (Edin, Holmlund, and Skans, 2009). In practice this means that where you work matters more for your wage today than it did during the early 1990's. Furthermore, large parts of minority-majority wage gaps can be attributed to segregation between firms and occupations (see for example McTague et al. (2006)). These two facts taken together imply that the workplaces and occupation affiliation may matter for the wages of immigrants, as well as for the relative wage growth.

Using a matched employer-employee data set, I decompose relative wage growth into a part which takes place within workplaces and occupations and a part associated to movements between workplaces and occupations. The results show that relative wages of immigrants grow to large extent through higher wage growth within employment spells. One interpretation of this is that employers, who were initially uncertain of the productivity of the immigrant workers, learn about the skills of the worker, and thereby update the wages of immigrants more than they do for native employees (Altonji and Pierret, 2001; Oettinger, 1996). This can be seen as indications of that information regarding skills are of importance for immigrants' position in the labor market.

#### Immigrants' access to business capital

Imperfect, or asymmetric, information is a concern also when banks decide on who to grant a loan. The banks do not have exact information regarding the default risk of the loan applicant and therefore base their decisions on both objective measures such as a credit score, and the bank official's personal evaluation of the potentiality of the loan repayment rate (Committee on Foreign Born Entrepreneurship, 1999; Fraser, 2009). In other situations where there is room for personal discretion, the characteristics of the clerks have often shown to be important for minority outcomes (Bradbury and Kellough, 2011), why the bank official's background may be important for individual applicants' outcomes.

In essay 2, I study how representation of own-group bank clerks affect the probability of being self-employed, for non-western immigrants in Sweden. Credit access is important for the possibility to start small businesses, and immigrant entrepreneurs in Sweden have been shown to struggle in terms of accessing business credit (Agency for Economic and Regional Growth, 2007, 2011). This could potentially be explained by the lack of experiences from the host country which the bank official can evaluate, but banks may also find it

more difficult to understand business ideas posed by individuals from other than the mainstream culture (Committee on Foreign Born Entrepreneurship, 1999). Internationally, survey studies on credits between business partners have shown that credit is facilitated by shared language (Raijman and Tienda, 2010). However, the only study focusing on shared background between employees in formal banks and applicants is a study on personal loans in India (Fisman, Paraviini, and Vig, 2012). Their study shows that bank managers from the same group is better to perform ex-ante evaluations of loan repayments, something which indicates that information may be important for the capital allocation.

The main methodological concern associated with studying a question like this is to distinguish a causal mechanism from an observed correlation. Banks may employ individuals from a particular group in response to a credit demand from that group, which would mean that we cannot interpret a positive relation between bankers from the own country and self-employment probability as an effect. Immigrants may also locate in municipalities where the community's entrepreneurial activity is high, and the group is represented in the bank. Both these things could be driven by factors that are difficult to quantify and control for in empirical studies, such as preferences or the quality of the group's labor market attachments locally. If these factors are omitted from the analysis we may erroneously conclude that own-group bankers matter for self-employment, even though the observed co-variation could actually be caused by the omitted variable.

Therefore I make use of a placement policy of refugee immigrants which was in place between 1987 and 1991 in Sweden. Through the policy, arriving refugees were placed in municipalities all over Sweden, placements which rarely were affected by the wishes of the individuals (Andersson, Musterd, and Robinson, 2003; Borevi and Myrberg, 2010). Thus the placement policy provides settlement municipalities which are unaffected by the individual's expectations regarding the benefits of residence in the municipality, and therefore we can study the effect of initial conditions in this municipality on subsequent outcomes. In this essay I study the effect of exposure to a banker from the own country of origin in the municipality of arrival the first year in Sweden on self-employment probability years later.

The results from this analysis clearly show that, for low educated and for males, there is a positive effect of own-group banker exposure on the likelihood of running a business. This study thus furthers our understanding on the process which immigrants go through in establishing themselves in a host country labor market. It shows that shared background is a factor which may help overcome issues associated with capital access. The fact that we see most effect for the low educated is non-surprising as the highly educated have better prerequisites for obtaining business capital even in the absence of own group bank employees.

## Collective agreements and wage setting

Employers' behavior is another important part of understanding labor market outcomes. Compared to studies on workers' behavior, studies on the effects of institutions on employers' behavior have been scarce. In many countries wages are set in a complex system of negotiations between employer associations and labor unions (Freeman, 2007). Despite of this, we have very little knowledge of how collectively bargained agreements affect actual wage setting and other decisions of the employers.

To some extent the lack of studies on the subject is due to methodological concerns, as it is difficult to distinguish the effect of an agreement from changes that would have taken place even without the agreement. The reason for this is that the bargaining parties are likely to incorporate possible effects in their bids, thereby taking the outcome into account (Holden, 1998).

What has been extensively studied, on the other hand, is the effects of minimum wages (see Neumark and Wascher (2008) for an overview) on employment and wages. Most minimum wage studies have focused on the effects of changes in legally binding minimum wages, but some have studied minimum wages which are negotiated through collective bargaining. In contrast to the studies on effects of minimum wage changes, the type of agreement studied in essay 3 of this thesis concerns wage increases for all covered workers, and not only the marginal workers employed at minimum wage.

In this essay we focus on blue collar public sector workers, for whom the labor union and the employer organization in 2007 struck an agreement which stipulated extra wage growth in establishments with a high share of low-wage female workers. This meant that the local employer was induced to provide larger wage increases as a function of the composition of the workforce. As we can control for the direct impact of the share of females, and the share of low-paid at each workplace, we here study very similar establishments which differ only in respect of having more low paid female workers. By doing so we can estimate effects of the agreement on individual labor market outcomes.

How employers react when faced with increased labor costs is an empirical question. The reaction from the employers depends on the type of work performed by the workers, the market which the employer operates in and the economy as a whole. The employer could either increase job separations or reduce hires, as labor costs are increasing, or they could reduce hours worked. But they can also invest in the productivity of the workforce and reduce labor turnover (Hirsch, Kaufman, and Zelenska, 2011; Salop and Salop, 1976).

We study whether this agreement had an effect on the actual wage paid to the covered workers, and show that workers in those establishments with a higher share of low wage women had a higher wage growth than those in otherwise comparable establishments. The results furthermore show that these establishment-level wage increases survive up to at least four years, and that the employers react with adjustments of labor input, both through reduced working hours and fewer new hires (compared to other establishments). Among the younger workers we also see that the composition of workers remaining at the establishment is altered, as more productive (measured by high school grades) workers are more likely to stay at the establishments with increasing wages, while this is not true for the young workers with lower productivity. Taken together, we can understand these results as evidence of employers reducing their use of labor inputs as wages increase, and moving towards a relative skill upgrading of the workforce.

# Concluding comment

To sum up, the essays in this thesis all contribute to our understanding of the complex processes which affect individuals in the labor market. By studying where relative wages of immigrants grow we improve our understanding of the process which immigrants go through when establishing a position in the labor market. Similarly, the fact that shared background mediates the capital access process for immigrant entrepreneurs teaches us about the barriers which needs to be overcome in situations with asymmetric information and uncertainty. Lastly, we see that negotiated agreements have direct (as well as indirect) effects on the distribution of wage increases for covered workers. This contributes to our understanding both of how collective bargaining affects workers' outcomes, and of how employers adjust to increasing labor costs.

## References

- Agency for Economic and Regional Growth. 2007. "Finansieringssituationen vid företagande for utrikes födda kvinnor och män".
  - —— 2011. "Företagare med utändsk bakgrund".
- Altonji, J. G. and R. M. Blank. 1999. "Race and gender in the labor market". In *Handbook of Labor Economics*. Ed. by O. C. Ashenfelter and D. Card. Vol. 3. Elsevier, 3143–3259.
- Altonji, J. G and C. R Pierret. 2001. "Employer learning and statistical discrimination". *The Quarterly Journal of Economics* 116, 313–350.
- Andersson, R., S. Musterd, and V. Robinson. 2003. Spreading the 'burden'? A review of policies to disperse asylum seekers and refugees. The Policy Press, Bristol, UK.
- Becker, Gary S. 1962. "Investment in human capital: A theoretical analysis". *The Journal of Political Economy* 70, 9–49.
- Becker, Gary. S. 1971. The Economics of Discrimination. University of Chicago Press.
- Borevi, K. and G. Myrberg. 2010. "Välfardsstaten och de nyanlända: En flyktingplaceringspolitisk probleminventering". MIM Working Paper 3.
- Bradbury, M. and J. E. Kellough. 2011. "Representative Bureaucracy: Assessing the Evidence on Active Representation". *The American Review of Public Administration* 41, 157–167.
- Carlsson, M. and D-O. Rooth. 2007. "Evidence of ethnic discrimination in the Swedish labor market using experimental data". *Labour Economics* 14, 716–729.
- Committee on Foreign Born Entrepreneurship. 1999. Invandrare som företagareför lika möjligheter och ökad tillväxt- Betänkande av Utredningen om företagande för personer med utländsk bakgrund. SOU 1999:49. Stockholm fritzes.
- Edin, P-A., B. Holmlund, and O. Nordström Skans. 2009. "Wage dispersion between and within plants: Sweden 1985-2000". In *The Structure of Wages: An International Comparison*. Ed. by E. P. Lazear and K. L Shaw. University of Chicago Press, 217–260.
- Fisman, R., D. Paraviini, and V. Vig. 2012. "Cultural proximity and loan outcomes". NBER Working Paper 18096.
- Fraser, S. S. 2009. "Is there Ethnic Discrimination in the UK Market for Small Business Credit?" *International Small Business Journal* 27, 583–607.

- Freeman, R. B. 2007. "Labor market institutions around the world". NBER Working Paper 13242.
- Hammarstedt, M. 2003. "Income from work among immigrants in Sweden". *Review of Income and Wealth* 49, 185–203.
- Hammarstedt, M. and G. Shukur. 2006. "Immigrants' Relative Earnings in Sweden - A Cohort Analysis". *Labour* 20, 285–323.
- Hirsch, B., B. Kaufman, and T. Zelenska. 2011. "Minimum wage channels of adjustment". IZA Working Paper.
- Holden, S. 1998. "Wage drift and the relevance of centralised wage setting". *The Scandinavian Journal of Economics* 100, 711–731.
- Ioannides, Y. M. and L. Datcher Loury. 2004. "Job information networks, neighborhood effects, and inequality". *Journal of Economic Literature*, 1056– 1093.
- Lalonde, R. J. and R. H. Topel. 1997. "Economic Impact of International Migration and the Economic Performance of Migrants". In *Handbook of Population and Family Economics*. Ed. by M. R. Rosenzweig and O. Stark. Vol. 1. Elsevier, 799–850.
- Lazear, E. P. and K. L. Shaw. 2005. "Wage structure, raises and mobility". In *The Structure of Wages: An International Comparison*. Ed. by E. P. Lazear and K. L Shaw. University of Chicago Press.
- Le Grand, C. and R. Szulkin. 2002. "Permanent Disadvantage or Gradual Integration: Explaining the Immigrant Native Earnings Gap in Sweden". *Labour* 16, 37–64.
- McTague, T., C. Robinson, K. Stainback, T. Taylor, D. Tomaskovic Devey, and C. Zimmer. 2006. "Documenting Desegregation: Segregation in American Workplaces by Race, Ethnicity, and Sex, 1966-2003". *American Sociological Review* 71, 565–588.
- Mincer, J. 1970. "The distribution of labor incomes: a survey with special reference to the human capital approach". *Journal of Economic Literature* 8, 1–26.
- Nekby, L. 2002. "How long does it take to integrate?: employment convergence of immigrants and natives in Sweden". FIEF Working Paper.
- Neumark, D. and W. Wascher. 2008. Minimum wages. The MIT Press.
- OECD. 2012. "Labour market outcomes". In Settling In: OECD Indicators of Immigrant Integration 2012. OECD Publishing.

- Oettinger, G. S. 1996. "Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap". *Journal of Labor Economics* 14, 52– 78.
- Phelps, E. S. 1972. "The Statistical Theory of Racism and Sexism". *The Ameri*can Economic Review 62, 659–661.
- Raijman, R. and M. Tienda. 2010. "Ethnic foundations of economic transactions: Mexican and Korean immigrant entrepreneurs in Chicago". *Ethnic and Racial Studies* 26, 783–801.
- Salop, J. and S. Salop. 1976. "Self-selection and turnover in the labor market". *The Quarterly Journal of Economics* 90, 619–627.
- Sarvimäki, M. 2011. "Assimilation to a Welfare State: Labor Market Performance and Use of Social Benefits by Immigrants to Finland\*". *The Scandina*vian Journal of Economics 113, 665–688.
- Spence, Michael. 1973. "Job market signaling". The Quarterly Journal of Economics 87, 355-374.

# Essay 1

Immigrant wage assimilation - the importance of unobserved heterogeneity, workplaces and occupations

## 1 Introduction

Due to growing evidence of poor labor market integration of recent immigrant cohorts, the economic integration of immigrants has become a cause of concern for policy makers in many western countries (for an overview, see OECD (2007)). This has led to a wide range of studies on the earnings-, employmentand wage assimilation of immigrants in host country labor markets. Adding to the literature, this essay estimates the rate of wage catch-up among nonwestern immigrants in Sweden, controlling for potential selection into employment. Furthermore, this essay decomposes the estimated wage catch-up into relative wage increases taking place within or between workplaces and occupations. The decomposition offers insights into the potential mechanisms which contribute to, and impede, immigrant wage catch-up.

Throughout the western world, the labor market position and assimilation pattern differ between groups of immigrants. Western migrants have experienced high employment- and wage rates in most European labor markets, while this has not been the case for non-western immigrants. Many studies show evidence of both a poor starting position and slow earnings catch-up for non-western immigrants.<sup>1</sup> Le Grand and Szulkin (2002) use a cross-sectional data for Sweden and estimate an 18 percent wage gap for immigrant men from non-European countries six years after immigration. The gap is decreased to about 12 percent after 20 years. Cross-sectional analysis is constrained though, as the assimilation effect cannot be distinguished from potentially different wages among immigrants arriving at different points in time.

Studying wage assimilation differs from studying convergence of earnings, as wage earners are a selected sample of individuals in the labor market (Husted et al., 2002). The employment rate is low among recent immigrants in many OECD countries (OECD, 2012), in particular among those from low income source countries (Hammarstedt and Shukur, 2006), and selection into employment may therefore be particularly pronounced for this group.<sup>2</sup> This selection is a problem if individuals who enter employment the first years in the country differ from those who do so years later, along dimensions which we cannot properly control for.

Using a longitudinal linked employer-employee dataset covering the complete Swedish labor market between 1995 and 2008, I estimate a wage assimilation model controlling for individual unobserved heterogeneity. This essay

<sup>&</sup>lt;sup>1</sup>See Barth, Bratsberg, and Raaum (2012), Hayfron (1998), and Shields and Wheatley Price (1998) for recent European estimates and Arai, Regnér, and Schröder (2000) and Åslund, Edin, and Lalonde (2000) for Swedish estimates. Sarvimäki (2011) finds for Finland that most of the closing of the earnings gap can be attributed to increased employment rates among the immigrants, while Husted et al. (2002) attribute some of the relative earnings growth in Denmark to increasing relative wages.

<sup>&</sup>lt;sup>2</sup>See Lubotsky (2007) for a discussion on delayed earners.

shows that the wage catch-up is underestimated if the selection into employment is not accounted for, which has implications for how wage assimilation estimates from repeated cross-sectional data should be interpreted. When controlling for individual fixed effects, estimated wage catch-up is between 6 and 13 percentage points over 30 years, depending of the education level of the individual.

The second part of this essays aims at explaining how wage catch-up is affected by sorting over, and mobility between, workplaces and occupations. This is of major importance as racial segregation explains a large part of racial wage differences (see for example McTague et al. (2006) for results for US). Similarly, immigrants in Sweden are over-exposed to other immigrant colleagues, something which is associated with lower average labor earnings (Åslund and Skans, 2010). In addition to this, the increasing wage dispersion in later years has to a large extent been driven by increased wage dispersion between firms,<sup>3</sup> which further motivates a focus towards the role that firm and occupation affiliation have for group difference in wages.<sup>4</sup>

Aydemir and Skuterud (2008) and Pendakur and Woodcock (2010) show for Canada that recent immigrants work in more low-paying firms than non-recent immigrants, which indicates that with longer time spent in the host country immigrants move into better paying firms. But these studies do not follow individuals, or cohorts, over time, and thus cannot distinguish the effect of time spent in the country from the potentially different wage effects of arriving at different points in time. A number of studies have therefore used repeated cross-sections to study the same question, decomposing the wage catch-up rate by comparing estimates from models with and without firm fixed effects. Barth, Bratsberg, and Raaum (2012) show, for Norway, that wage catch-up is slow, and that this can partly be explained by lack of movements into workplaces with higher wage levels. For Portugal, Damas de Matos (2012), on the other hand, shows that sorting into firms with higher wages explains about one third of wage catch-up for immigrant workers.

This essay contributes to the literature by decomposing wage catch-up into wage catch-up within and between workplaces and occupations. In line with Damas de Matos (2012), I also control for individual fixed effects. Furthermore, I present estimates when controlling for the match between the individual and the workplace (occupation), thus studying how the wage assimilation rate within workplaces and occupations compare with the total wage catch-up rate. This yields novel insights as to how different aspects of worker mobility contribute to the wage catch-up of immigrants. The results show that most of

<sup>&</sup>lt;sup>3</sup>See for example Card, Heining, and Kline (2013), Edin, Holmlund, and Skans (2009), and Lazear and Shaw (2005).

<sup>&</sup>lt;sup>4</sup>Ethnic occupational segregation is documented in US (Catanzarite, 2000) as well as in UK (Elliott and Lindley, 2008).

the wage catch-up can be attributed to immigrants having higher wage growth within workplaces and occupations than natives do. Neither cross-occupation nor cross-workplace mobility contributes to narrowing the wage gap. The only indication of a positive sorting effect is that highly educated immigrants experience a small, but positive, wage catch-up from sorting into occupations which pay better wages to them as a group.

As wage catch-up mainly takes place within workplaces and occupations, and not from mobility or sorting between them, the lack of host country-specific capital is unlikely to be the sole barrier explaining immigrants' poor labor market attachment. Improved language skills and acquired host country specific human capital should contribute both to within workplace wage growth and wage growth from finding better labor market matches and better paying employers and professions. The results thus indicate that other factors influence the relative wage growth as well.

The lack of improved relative wage from sorting into better paying firms suggests that the outcomes cannot be explained by initially poor knowledge about the labor market. Within the workplaces, wages grow relative to natives, something which may be interpreted as that employers, who are initially uncertain regarding the immigrant worker's productivity, update beliefs as they observe the worker, and thereby raise wages. Similarly, within-occupation wage catchup might be understood as reduced uncertainty regarding occupation-specific skills.

In the next section the data and variables used are described. Section 3 describes the relative wage growth non-parametrically, while section 4 presents the model for estimating wage catch-up and studies how the estimates are influenced by unobserved individual heterogeneity. The wage catch-up rates are then decomposed in section 5. Section 6 discusses the results, and section 7 concludes the paper.

# 2 Data

In this essay I use the Register Based Labor Market Statistics (RAMS) from Statistics Sweden, which includes all working age individuals in Sweden. This dataset links individuals to workplaces through tax records on annual income. I study the years 1995 to 2008. For individuals with multiple jobs in the same year, the employment with the highest total income (by year) is used for the analysis.

The data contains information on annual income from each specific employment and information on number of months worked in the employment each year. From this information I construct a measure of approximate monthly wages. The data does not contain information on hours of work, why I exclude all individuals with a approximate wage below 75 percent of the mean wage for a publicly employed janitor (see Edin, Holmlund, and Skans (2009)). This cut-off is chosen since it can be seen as a minimum wage for a full-time employed worker, and thus excluding individuals with approximate wages below this cut-off reduces the risk of including part-time workers.<sup>5</sup> This variable will, however, still be an imperfect measure of monthly wages, as it is affected by hours worked.

For a sampled data a more precise measure of monthly wages is available (Wage and Salary Structure Data (WSSD)). This wage measure is standardized full-time monthly wage equivalence, which implies that it is adjusted to take differences in hours worked into account.<sup>6</sup> This data also contains information on occupations, based on 3-digit ISCO coding. The WSSD data covers all public sector workplaces and a stratified sample of private sector workplaces in the Swedish labor market. About 50 percent of all employees in the private sector are included in the data set, with an over-sampling of large firms.<sup>7</sup> Due to the sampling structure of this data, it is non-representative for the full population and therefore it will only be used as a sensitivity analysis. The correlation between the approximate wage measure available for the full population and the register based wage measure available in the sample is 0.86.

Individual data contains information such as age, sex, country of origin, year of entry to Sweden and highest level of education. Year of entry is recorded as the year when permanent residence permit was obtained. Country of origin is used to classify whether an immigrant comes from a non-western country or not. This study will only focus on the labor market outcomes of the group of non-western immigrants, as this group highly overlaps with the group of refugees who have not migrated for labor market reasons. This makes them a suitable group for studying labor market progress, reducing the risk of selective in-migration based on employment prospects. Small source countries share the same code due to confidentiality reasons.<sup>8</sup> All future data description and analysis focus on immigrants from non-western countries and natives.

To restrict the sample towards refugee immigrants, all immigrants arriving before 1975 are excluded. I exclude all immigrants arriving before age 20, as the experiences from the host country differs substantially between those arriving as grown-ups and those migrating at a young age (see Friedberg (1992) for a

<sup>&</sup>lt;sup>5</sup>One consequence of using this cut-off is that it will exclude low-paid part time workers from the data, while more highly paid part-time workers are more likely to remain in the data. Also the 1 percent highest approximated monthly wages are excluded.

<sup>&</sup>lt;sup>6</sup>Standardization is performed by Statistics Sweden and based on the hours which constitute full time in each sector. This means that this measure is capturing hourly wages, aggregated to monthly wages.

<sup>&</sup>lt;sup>7</sup>Stratification is based on the combination of firm size and industry. All employees of the sampled firms at the time of the data collection are included in the data. The data collection is in November, and thereby seasonal workers are excluded.

<sup>&</sup>lt;sup>8</sup>See list in Appendix for countries which are classified as non-western in this analysis.

discussion). Also workers who change educational status during the observed years are excluded, as this reduces the risk of measurement error.<sup>9</sup> A result of this restriction is that one avenue through which success in the labor market can be obtained is eliminated from this analysis.

After these restrictions are imposed, the data contains about 34 million observations, more or less evenly distributed over 14 years. In total 173,000 nonwestern immigrants are observed in connection to a workplace, compared to 3,8 million natives. The analysis includes individuals between 20 and 65 years of age. Over the whole 14 years the number of workplaces observed is 509,000, and I observe 124 occupations.<sup>10</sup>

Many firms in the data are very large (in the public sector for example every municipality is coded as a unique firm) and consists of several workplaces, why firms are a potentially unsuitable unit for studying labor market sorting. Therefore the analysis will be performed on the level of the establishment/workplace. This implies that any estimates of mobility between workplaces also include mobility between workplaces within the same firm. As a sensitivity analysis, this choice of unit of analysis will be varied.

Throughout the text the terms *Wage catch-up* and *relative wage growth* are used interchangeably. These terms describe the parameter of interest, which captures the rate at which the wage of an immigrant increases over time compared to the wage increases of a native with similar background characteristics. A formal definition is presented in section 4.

#### 2.1 Descriptive statistics

Table 1 presents some basic descriptive statistics in terms of mean values for the group of natives and non-western immigrants for the year 2002, using the RAMS data of the full labor market.<sup>11</sup> Table 5 in the appendix shows the same statistics for the WSSD data which is used for the decomposition of the wage catch-up within and between occupations, as well as for a robustness analysis.

From table 1 we see that the non-western immigrants in the labor market is a heterogeneous group with about one third of the group in each educational category. This is not surprising as recent cohorts of immigrants are highly educated (Eriksson, 2007). The immigrants are younger in general, and earn lower

<sup>&</sup>lt;sup>9</sup>The education variable represents the highest attained education level, so individuals changing status to a lower education are certain to be measurement errors. About 13 percent of both the immigrant and native group change education level. These individuals are excluded.

<sup>&</sup>lt;sup>10</sup>See table 4 in the appendix for an overview of the different datasets used for the different analyses.

<sup>&</sup>lt;sup>11</sup>For the decomposition analysis, the data has to be grouped to identify both individual and occupation or workplace effects (see discussion under empirical specification), and therefore the data used in the respective empirical analyses will differ slightly from this data I describe here.

	Natives	Non-Western born
Male	0.525	0 519
Age (Mean)	44 374	41 756
< 25 years	0.064	0.023
>25 years $& <35$ years	0.185	0.209
>35 years & $<45$ years	0.240	0.393
>45 years & $<55$ years	0.247	0.278
>55 years & $<65$ years	0.264	0.097
Less than high school	0.200	0.284
High school	0.519	0.379
University	0.281	0.336
Approx monthly wage (RAMS)	22,839	19,860
Register based monthly wage* (WSSD)	22,819	19,320
Income from employment	231,384	166,375
Employment rate	0.809	0.527
Years since migration (Mean)		10.608
<5 years in Sweden		0.205
>5 years & $<10$ years in Sweden		0.294
<10 years & $<20$ years in Sweden		0.386
>20 years in Sweden		0.115
N	3,905,884	220,187

## Table 1: Descriptive statistics RAMS - individual details - year 2002

*Note:* Mean values for individuals between 20 and 65 in RAMS data, year 2002. \* Register based monthly wages are only available for the restricted sample. Mean income and mean wage are calculated among individuals with positive values. Approximate monthly wage is set to missing when below the defined minimum wage threshold.



Figure 1: Monthly wage gap and employment rate

Note: Both employment rate and wage gap is identified in the RAMS data set, which covers the full labor market. Approximate wage measure is used. The level of the wage gap is plotted for the cohort arriving between 1995 and 1999, for which the wage gap is most similar to the mean wage gap over all cohorts. The wage gap is defined as the immigrant's wage compared to the mean wage among natives of the same sex, age the same year.

wages and incomes than natives, regardless which wage measure is used. The immigrant employment rate is also much lower than the native.

## 3 Describing relative wage growth

In figure 1 the wage gap between immigrants and natives is outlined. This is defined as the ratio of the immigrants' wage to the mean wage of natives of the same age and sex, in the same year. The wage gap for each year since migration is then estimated, controlling for arrival cohort through 7 dummies covering the time span from 1975 to 2008, and the predicted wage gap at each year since migration plotted. The intercept is defined by the cohort of immigrants arriving between 1995 and 1999, for which the mean wage gap is most similar to the mean wage gap over all cohorts.

The wage gap decreases from about -16% of natives two years after immigration to around -13% after 25 years in Sweden, indicating that wage assimilation is virtually non-existent. The dashed line in figure 1 shows the employment rate of the studied group over years since migration. The low employment rate the first years in Sweden tell us that the initial wage gap is identified by much fewer observations than it is in later years, and during these years we also see that the wage gap is smaller than in the succeeding years.

Figure 2 describes the wage gap by educational attainment. The observed pattern of a relatively high wage in the first years is mainly driven by the group of highly educated individuals. The wage gap for this group two years after immigration is about -23% and this is reduced to about -16% after 30 years. The relative wage of the high school graduates increases gradually from about -18% one year after immigration to -13% in the first 10 years. For the least educated the first year represents a true outlier in terms of the wage gap, and already one year after immigration the wage gap is around -13%, and thereafter there is virtually no increase in the relative wage.



Figure 2: Monthly wage gap over educational attainment

Note: See note in previous figure. The wage gap is defined as the immigrant's wage compared to the mean wage among natives of the same sex, age, education level the same year.

The main thing to note from these descriptions is the low employment rate the early years. This indicates that the individuals in employment early on might have better prerequisites for employment in the Swedish labor market than those individuals entering employment later have. This, along with the peculiarity of high relative wages the first years for the highly educated, implies that there might be a positive selection into employment, which will bias estimates of wage assimilation if not accounted for. One way to understand this is that individuals with high earnings potential enter employment shortly after arrival to the host country, and do so on relatively high wages. As time passes, individuals with less earnings potential gain access to employment, but do so on lower paying positions. This means that the average relative wage may decrease over time, as the group of employed immigrants becomes more heterogeneous. In the next section I present a model which accounts for this selection into employment, and test how unobservable time-invariant characteristics of the individuals employed at different points in time influence the wage assimilation estimates.

Figure 1 and 2 are presented using the data over the full labor market, RAMS, and thus use the approximated wages constructed from annual incomes and months of work. The relative wage growth patterns are generally invariant to both the wage measure and sample used (results available on request).<sup>12</sup> When estimating the assimilation models in the following sections, I present sensitivity checks of how the choice of wage measure influence the results.

This same description as above is performed separately for women and men, and the results are presented in figure 9 in the appendix. Men start out on lower relative wages than do the immigrant women, but experience an increase of the relative wage over time, while women in principal do not catch up. The pattern of high initial wages for the highly educated is found for both groups.

# 4 Modeling wage catch-up

How to best estimate immigrant's success in the host country has been subject to debate among researchers for decades. This is mainly due to the methodological issues which arise when comparing individuals arriving at different points in time to each other. The first model was developed by Chiswick (1978), who used a single cross-section and compared the earnings of recently arrived immigrants with those who had stayed in the country longer. The disadvantage of this approach is that it cannot distinguish the effect of assimilation from the effect of different entry cohorts having different opportunities in the host country labor market.

The model was then developed by Borjas (1985), who used repeated crosssections, and who thereby could control for cohort of arrival. In the most basic version of Borjas' model immigrants' wages are estimated as a function of years since migration and age while the corresponding natives' wages are estimated as a function of age. The wage catch-up parameter estimated by this set-up captures the rate of wage convergence between immigrants and natives, and is composed by the differential return to aging for immigrants and natives and

<sup>&</sup>lt;sup>12</sup>The level of the wage gap is 2 percentage point smaller when using the restricted WSSD data, for which we can also use the register based measure of wage. If instead using the register based wage measure, the wage gap is 2 percentage points larger. As the WSSD is not fully representative of the full labor market, the wage gap estimated in the RAMS data is likely to provide a better approximation of the average wage gap in the full economy.

the return to one extra year in the host country for the immigrants (see Borjas (1999)).

Extensions to this basic model include controlling for year of observation and/or age at migration, as both these are important determinants of wage assimilation (Friedberg, 1992). Methodologically, including these controls introduces perfect linearities in the model, as observation year equals the sum of the year of arrival and the years since migration (YSM) for the immigrants. Similarly age is a perfectly linear combination of years since arrival and age at migration:

 $Year \equiv Arrival Year + YSM$ 

Age  $\equiv$  Age at Migration + YSM

For this reason, it is not possible to control for all the factors influencing wage catch-up rates in the same model. With access to panel data, one way to deal with these co-linearities is to control for individual fixed effects. This indirectly controls for the effect of time-invariant individual level characteristics, and thus controls for the composition, along unobservable dimensions, of individuals employed at different years since immigration. As the earlier description gave reason to suspect an initial positive selection into employment, it is likely that the estimates of wage catch-up will be downwardly biased if not taking the composition of employed individuals into account. The individual fixed effect model reduces the identifying variation to within individual variation, and therefore a drawback of the model is that it cannot identify different wage gaps for different entry cohorts.

$$\ln \operatorname{real} \operatorname{wage}_{it} = \delta_I \operatorname{Age}_{it} + \theta_I \operatorname{YSM}_{it} + \mu_i + \varepsilon_{it} \tag{1}$$

$$\ln \operatorname{real} \operatorname{wage}_{it} = \delta_N \operatorname{Age}_{it} + \mu_i + \varepsilon_{it} \tag{2}$$

Here equation (1) estimates the log real wage for the individual i in year t for immigrants, indicated by the sub index I. Equation (2) estimates the log real wage for natives (sub index N). The wages of immigrants and natives are estimated simultaneously by interacting equation (1) and (2).

Individual fixed effects ( $\mu_i$ ) implicitly control for age at migration, year of birth and year of entry. This means that in addition to the previously mentioned co-linearities, when controlling for individual fixed effects, the effect of year of observation and age will be perfectly co-linear (since Year  $\equiv$  Age + Year of Birth). For this reason I do not include year fixed effects in the model (Borjas, 1999; Pischke, 1992).<sup>13</sup> To control away for trends in nominal wages, log of real

<sup>&</sup>lt;sup>13</sup>This set-up implies the assumption that period effects are equal for natives and immigrants, an assumption which is questioned by for example Barth, Bratsberg, and Raaum (2006).

wages are used as outcome variable.<sup>14</sup>

The assumption that needs to be imposed for the model to correctly identify the parameter of interest is the assumption of equal age effects. This constrains the coefficient for age to be the same for natives and immigrants (Borjas, 1999), which means that the age effects are identified by natives, and imposed on the immigrants. Thus, imposing this restriction implies that the coefficient of interest,  $\theta_I$ , should be interpreted as *the differential return to aging plus the return to spending time in the host country*. Henceforth this is what I will refer to as the wage catch-up parameter, and it is comparable to the parameter presented in Borjas (1999). The interacted model is outlined below, with *I* being an indicator variable for being an immigrant.

$$\ln \operatorname{real} \operatorname{wage}_{it} = \delta \operatorname{Age}_{it} + \theta_I \operatorname{YSM}_{it} I + \mu_i + \varepsilon_{it}$$
(3)

In the estimation of this model, age is introduced as a third degree polynomial. The year since migration variable, on the other hand, is introduced as three splines, and is thus estimated linearly in three intervals; 0-10, 10-20, 20-30 years since migration (YSM). The splines are used since a polynomial function will be affected by the distribution of observations over years since migration distribution, the polynomial function will be less precise in the tails (Husted et al., 2002).<sup>15</sup> The model allows for the different education groups to have different returns to both age and years since migration, but these interaction effects are excluded from the equations for notational purposes.

## 4.1 Compositional bias of wage catch-up estimates

In this section I present estimates of the catch-up parameter and study how unobserved individual time-invariant characteristics influence these estimates. I do this by comparing estimates from model (3) with estimates from an otherwise identical model which controls only for the observable parts of timeconstant individual characteristics ( $\psi_k$ ). In practice this means that I replace the individual fixed effect  $\mu_i$  with  $\psi_k$ . I call this fixed effect a quasi-individual fixed effect. This variable takes on unique values for each combination of the observable time-invariant components of the individual fixed effect: sex, education level, country of origin, year of birth and year of immigration (set to 0 for natives).

Section 4.1.1 presents results when relaxing this assumption and estimates from the model which includes individual fixed effects are unaffected by this assumption.

<sup>&</sup>lt;sup>14</sup>Here CPI for year 1990 is used as the reference year for deflating the monthly wages.

<sup>&</sup>lt;sup>15</sup>Both models in section 4.1, section 5.1 and section 5.2 are performed using third degree polynomials in YSM instead, and the results are robust to this change in model specification.

$$\psi_k = \begin{bmatrix} Sex \\ Education level \\ Country of Origin \\ Year of Birth \\ Year of Immigration \end{bmatrix}$$

 $\ln \operatorname{real} \operatorname{wage}_{it} = \delta \operatorname{Age}_{it} + \theta_I \operatorname{YSM}_{it} I + \psi_k + \varepsilon_{it} \tag{4}$ 

This set-up implies that the difference between the estimates from equation (3) and (4) should be interpreted as the compositional bias of unobserved time-invariant individual characteristics on the wage catch-up rate.

For simplicity, the estimates are primarily presented as figures, using the estimated coefficients of wage catch-up to plot predicted relative wage at different years since migration. But as these models control for all time-invariant factors we only identify the changes in the immigrant-native wage gap, and not the initial level of the gap. For illustrative purposes, though, I impute the level of the wage gap presented in the description in figure 2 in the following figures. The choice of which wage gap to impute affects the intercept, but not the estimated rate of wage convergence, which is represented by the slopes in the following figures. The wage gap imposed is the gap at 30 years after immigration.<sup>16</sup> This year is chosen since the selection into employment, which might influence the described wage gaps in the earlier figures, becomes less important the more time is spent in the country. With an imposed wage gap, the figures show the predicted immigrant-native wage gap, conditional on covariates, at different times since entry.

The fully drawn line of figure 3 represents the wage catch-up controlling for individual fixed effects. We see the largest catch-up, from an initial gap of around -29% to the imputed gap of about -16% after 30 years, for the group of university educated immigrants. For the least educated and those with high school education, wage catch-up is high the first 10 years (about 6 percentage points) and then the relative wage remains relatively stable (at about -13% gap). The initial wage gap for the highly educated, when controlling for individual fixed effects, is much lower than the earlier described initial wage gap. This is also what could be expected if the early entrants in the labor market are highwage earners.

It is clear from figure 3 that when instead including the quasi-individual fixed effects, wage catch-up is lower, and the wage gap seems actually to be increasing with time in Sweden for the least educated. This result implies that the average wage of the immigrant group decreases over time as individuals with a lower

<sup>&</sup>lt;sup>16</sup>The intercept of the plotted curves are set so that each curve reaches the described gap at 30 years since migration. This is 13, 12 and 16 percentage points for the groups respectively.



Figure 3: Wage catch-up and bias from selection into employment

Note. Estimates from models (3) and (4), using RAMS data and approximated wage measure. Predicted wage gap over Years Since Migration. The gaps used at 30 years since migration are -13% (Less than High School), -12% (High School), -16% (More than high school). Slopes of prediction lines represent estimated wage catch-up parameters. Estimates presented in column 1 and 2 in table 7 in the appendix.

earnings potential enter the labor market, and do so on lower paying positions. We see this most clearly for the university educated, and mainly during the first ten years in the country, which is expected as this is the group where a substantial number of workers enter on high wages during the first years.

A robustness analysis is performed to study to what extent these results are influenced by the wage measure used. These results are presented in section B in the appendix, and the above results are invariant to the wage measure, as well as the sample, used. The total wage catch-up rate, though, for the least educated, changes slightly depending on which wage measure is used (see section B in the appendix), why it is reasonable to conclude that part of wage catch-up estimated for this group captures changes in hours worked.<sup>17</sup> I also re-estimate the models excluding each individual's first two years in the country, as there might be institutional features which keep individuals from entering the labor market during these years. The conclusion remains robust to this restriction.

<sup>&</sup>lt;sup>17</sup>The wage catch-up rate is also slightly lower for the highly educated when using the registerbased wage measure, but the pattern of relative wage growth is invariant to the wage measure used.



Figure 4: Wage catch-up with controls for local unemployment

Note. Estimates from models (3) and (4) with included controls for group specific effects of the natural log of the local unemployment rate. Estimated on RAMS data with approximated wage measure. Predicted wage gap over Years Since Migration. The gaps used at 30 years since migration are -13% (Less than High School), -12% (High School), -16% (More than high school). Slopes of prediction lines represent estimated wage catch-up parameters.

## 4.1.1 Controls for local labor market conditions

Another way to control for the selection into employment may be to add controls of the labor market situation, which has large implications for the group of individuals who enter, and exit, employment, to the model. Therefore, we here re-estimate models (3) and (4), introducing local level (municipal level) controls for the unemployment rate and allow for natives and immigrants to be differently affected by the unemployment rate.<sup>18</sup> Results from this estimation is shown in figure 4.

As estimates from the specification with individual fixed effects are unaffected by introducing these local level labor market controls, it is clear that the controls do not affect estimates of wage catch-up, once the selection into employment is controlled for. We do see, though, that with these controls the

<sup>&</sup>lt;sup>18</sup>Here we follow a set-up introduced by Barth, Bratsberg, and Raaum (2006) where the equal period effect assumption can be relaxed by controlling for local level time variant labor market conditions. We use the natural log of the municipal level unemployment rate as a measure of local level labor market conditions, and it turns out that with controls for individual fixed effects, the results are invariant to including these controls. Thus differential responses to labor market conditions is not affecting the empirical results.

wage catch-up estimates in the quasi-fixed effect model are now very similar to estimates in the individual fixed effect model for the least educated groups. This means that local labor market conditions influence the group of individuals who are employed at a particular point in time, and thus controlling for the local business cycle may offer a way of dealing with the bias of selection into employment for these groups. For the highly educated, though, we see that the wage catch-up is underestimated the first 10 years in the country if not using individual fixed effects, even when controlling for local labor market conditions. Thus, the selection problem for the highly educated remains when controlling for the group-specific effects of local unemployment.

## 5 Decomposing wage catch-up

The second purpose of this study is to disentangle how movements between, and wage growth within, workplaces and occupations contribute to wage catchup for the immigrant workers. The wage catch-up estimates of model (3) consists of three parts; catch-up from sorting into better paying workplaces, catchup from mobility into workplaces with a higher match quality, and catch-up from wage growth within the workplace. Thus the decomposition of wage catch-up can be performed by estimating two additional models, including different sets of fixed effects (FE).

$$\ln \operatorname{real} \operatorname{wage}_{it} = \delta \operatorname{Age}_{it} + \theta_I \operatorname{YSM}_{it} I + \mu_i + \varepsilon_{it}$$
(5)

$$\ln \operatorname{real} \operatorname{wage}_{it} = \delta \operatorname{Age}_{it} + \theta_I \operatorname{YSM}_{it} I + \mu_i + \psi_j + \varepsilon_{it}$$
(6)

$$\ln \operatorname{real} \operatorname{wage}_{it} = \delta \operatorname{Age}_{it} + \theta_I \operatorname{YSM}_{it} I + \gamma_{ij} + \varepsilon_{it}$$
(7)

Model (5) presented here is the same as model (3) in section 4.1, and it includes only individual fixed effects. We will refer to the estimates from this model as *Total catch-up*. The second model, on the other hand, uses both individual- and workplace fixed effects, thereby controlling for the general wage level of the workplace where the individual is employed (Abowd, Kramarz, and Margolis, 2003). This means that this specification controls for the part of wage catch-up which is due to sorting across workplaces with different wage levels, and it thus captures wage catch-up among workplaces of similar wage levels. The estimates of this model will be referred to as *Double FE catch-up*.

For estimating model (6) we add a workplace fixed effect  $(\psi_j)$ , which is equal for both immigrants and natives, to the individual fixed effect model. Here the individual effects can be interpreted as the effect on wages of the innate human capital of the worker regardless of which workplace she is working in. The fixed workplace effects can be seen as the time-constant wage premium for workers in
that specific workplace, regardless of their own ability, motivation or earnings potential (Abowd and Kramarz, 1999). This simultaneous identification of individual and workplace effects requires that there are workers who have changed employer, and that there are other employees in both workplaces to contrast the wage outcome with (Abowd, Creecy, and Kramarz, 2002). To ensure that the data fulfills this requirement, I use an algorithm which creates groups of connected workplaces and individuals. The largest group of connected workers and workplaces, which contains around 98 percent of all employed individuals, constitutes the estimation sample.<sup>19</sup>

In addition to the above models, model (7) controls for the match between an employer and an individual, by inclusion of a fixed effect for each individualworkplace combination ( $\gamma_{ij}$ ). This model thus controls for both sorting into better paying workplaces and improved match quality, and therefore estimates the wage catch-up within the given workplace (referred to as *Match FE catchup*).

As the models control for different parts of the wage catch-up, we can use the estimates from the three models to learn which factors are associated to wage catch-up. The differences between the model parameters can be understood in the following ways:

- If *Double FE catch-up* is smaller than *Total catch-up*, then sorting into workplaces which pay higher wage premiums contributes to total wage catch-up.
- If *Match FE catch-up* is smaller than *Total catch-up*, then mobility between workplaces, or sorting into better paying workplaces, contributes to total wage catch-up.
- If *Match FE catch-up* is smaller than *Double FE catch-up*, then job mobility (into better matches) between workplaces with similar wage levels contributes to wage catch-up.

The double FE model assumes that the workplace fixed effects are the same for immigrants and natives. But the immigrant group might be paid differently from natives in the same firm, for example if the employer is more uncertain regarding the productivity of the immigrant worker. Therefore it is also useful to study if sorting across workplaces is related to the overall wage levels of the workplace or the wage level for the immigrant group specifically. This is studied in section 5.1.1.

<sup>&</sup>lt;sup>19</sup>See table 4 in the appendix for mean number of workplaces and occupations of natives and immigrants in the different data sets. The workplaces which are excluded by the grouping algorithm are very small. Most workplaces are excluded by the restriction of having more than two employees, which is imposed prior to the grouping, as it is necessary for the algo-

For those who do not change workplace, wage catch-up will be the same regardless which of these three models is used. Thus the difference between the estimates from the models are identified by those who do change workplace. But the relative size of the group of job-changers and non-changers will also influence how large the difference will be.<sup>20</sup>

The decomposition outlined above will also be performed using data on occupations instead of workplaces. The interpretation of the parameters will be the same, replacing workplace with occupation.

#### 5.1 Wage catch-up within and between workplaces

The decomposition of the wage gap is performed through estimating equations (5), (6) and (7). Estimates from each model are used to plot a separate line in Panel A of figure 5, showing the predicted wage gap over years since migration.<sup>21</sup>

The fully drawn line is the exact same line as in figure 3, showing the total wage catch-up rate. The dashed line shows estimates from the double FE model, capturing the wage catch-up within workplaces of similar wage levels. The estimates from the match FE model correspond to the dotted line in panel A of figure 5, and these represent the wage catch-up rate within workplaces. Again, the initial gap used here is the gap from the earlier estimations with individual fixed effects (see figure 3).

In terms of where the wage catch-up takes place, there are small differences in the specification with and without workplace fixed effects.<sup>22</sup> This means that sorting into better paying workplaces does not significantly affect the wage gap between natives and immigrants.<sup>23</sup> For those with university education the match FE catch-up is also very similar to total wage catch-up. This means that neither mobility between workplaces nor sorting into higher-paying workplaces

rithm to run properly.

<sup>&</sup>lt;sup>20</sup>As job-changers on average are observed fewer times with each employer, the catch-up from the match FE model is to a higher degree identified by individuals who do not change workplace. Therefore, the difference between total and Match FE catch-up can to some extent also reflect differences in total wage catch-up between those who do change, and those who do not change, workplace.

<sup>&</sup>lt;sup>21</sup>Slopes of plotted lines represent regression estimates in column 2-4 in table 7 in the appendix. The level of the wage gap is imputed.

<sup>&</sup>lt;sup>22</sup>For all groups the wage catch-up in the double FE model is slightly higher during some time period, but the differences are small and for all estimates but one, the confidence intervals of the total catch-up and the double FE catch-up overlaps, see table 2.

<sup>&</sup>lt;sup>23</sup>As relative wages of immigrants and natives can also be affected by firm's wage setting behavior (as opposed to the workplace's), I have performed the same analysis as presented in figure 5 using the firm as the unit of analysis. The only difference here is that the within firm wage growth is now higher for the highly educated than when studying within workplace wage growth. This implies that this group catch-up through mobility between workplaces within the same firm.



Figure 5: Wage catch-up; Within and between workplaces

Note. Estimates from using RAMS data and approximated wage measure. Predicted wage gap over Years Since Migration. The initial gap is the same initial gap as in the individual fixed effect model, plotted in figure 3. Initial gaps are -19% (Less than High School), -18% (High School), -29% (More than high school). Slopes of prediction lines represent estimated wage catch-up parameters. Estimates for Panel A presented in column 2, 3 and 4 in table 7 in the appendix. Total catch-up represents the estimates from the specification with individual fixed effects, model (5). Double FE represents estimates when also including workplace fixed effects, model (6). Match FE catch-up is the parameter from the specification with a control for the match between workplace and individual, model (7). Estimates in Panel B are obtained through versions of model (8) and (9).

contributes to the relative wage growth for this group.

For the less educated, on the other hand, we see that during the first 10 years, relative wages grow substantially faster within workplaces (Match FE) than in total. This means that mobility in the labor market widens the wage gap during the first ten years for these groups. The fact that within-workplace wage catch-up is higher than estimated catch-up from the double FE model implies that mobility between workplaces of similar wage levels also contributes to widening the wage gap for the low educated.<sup>24</sup>

Barth, Bratsberg, and Raaum (2012) specify similar models when studying immigrant wage catch-up in Norway. They present an assimilation framework similar to model (5) and model (6), but they do not control for individual fixed effects. Their results suggest that while natives continuously move into better paying firms over the life cycle, there are no such indications for the immigrants in Norway. This contributes to the wage gap between natives and immigrants widening over time. My results are in line with these results, as they show no indications of improved access to better paying firms with time spent in the country. This contradicts with earlier results from Canada (Aydemir and Skuterud, 2008; Pendakur and Woodcock, 2010), and results for Portugal (Damas de Matos, 2012).

Again, this decomposition analysis is re-estimated using the restricted data set for which the more precise wage measures is available (see section B in the appendix). Here we see that within-workplace wage catch-up is somewhat lower when using the register based wage measure than when using the approximated wage measure. Despite this, we find that within-workplace wage catch-up is at least as high, or higher, than total wage catch-up, regardless which wage measure is used.

The decomposition results are similar for both men and women, but low educated women experience a very small, but negative wage catch-up (results in figure 10 in appendix). This negative wage catch-up rate of the low educated women is however sensitive to the choice of wage measure (results available on request).<sup>25</sup>

<sup>&</sup>lt;sup>24</sup>With the empirical set-up I am using, I cannot rule out that some workplaces offer better wage returns than others, and that immigrants are sorted into these workplaces. As a tentative test, I have limited the analysis to establishments which employ a fair number (at least five each) of both groups, finding an identical role for within-establishment catch-up, suggesting that differential wage growth within the same workplaces is a key element in the process which narrows the wage gap.

<sup>&</sup>lt;sup>25</sup>Using the register based wage measure, this group does catch-up with native women, albeit slowly. The under estimation of wage catch-up using the approximate wage measure may be related to native women increasing hours of work to a higher extent than the immigrant low educated women do.

	Quasi-ind f	ixed effects	Total ca	ttch-up	Double FI	C catch-up	Match FE	catch-up
Figure 3 & Panel A of figure 5	$B_L$	$B_U$	$B_L$	$B_U$	$B_L$	$B_U$	$B_L$	$B_U$
Less than high school		01000	11000	1700 0	0,00,40		100 0	
U - 10 Y SIVI	-0.001	0.0018	0.0044	0.006/	0.0048	0.0062	0.00/	0.0098
11 - 20 YSM	-0.0021	-0.001	-0.0008	0.0004	-0.0002	0.0009	-0.0006	0.0008
21 - 30 YSM	-0.0016	0.0005	-0.0007	0.0014	-0.0001	0.0012	-0.0008	0.0016
High school								
0 - 10 YSM	0.0023	0.0034	0.0056	0.0069	0.0057	0.0066	0.0076	0.0091
11 - 20 YSM	-0.002	-0.0011	-0.001	-0.0001	0.0000	0.0006	-0.0007	0.0004
21 - 30 YSM	-0.0014	0.0001	-0.0007	0.0007	0.0002	0.0011	-0.0008	0.001
More than high school								
0 - 10  YSM	-0.0029	-0.0014	0.0064	0.008	0.0076	0.0086	0.008	0.0099
11 - 20 YSM	0.0007	0.0019	0.0028	0.0039	0.0033	0.0039	0.0006	0.002
21 - 30 YSM	0.0006	0.0027	0.0019	0.004	0.0023	0.0036	0.0009	0.0033
Note: Standard errors of all estin	mates but the	Double FE estir	nates are cluster	ed at the level o	f the individual.	For the Doubl	e FE estimates s	tandard errors are

Table 2: Confidence intervals of wage catch-up parameters; Workplace decomposition

obtained through bootstrapping with 15 repetitions. The confidence intervals are based on 95 % confidence level.

#### 5.1.1 Relaxing the assumption of equal workplace effects

As stated earlier, model (6), which estimates the part of wage catch-up taking place within workplaces with similar wage levels, assumes that the wage premium of entering a given workplace is the same for natives and immigrants. But employers may pay the groups differently, in which case immigrants could catch-up through sorting into workplaces which pay on average higher wages for them as a group.

To study this, I re-estimate the double fixed effect of model (6) separately for natives and immigrants, thus estimating separate workplace effects for the groups. As we need to sustain the equal age effect of the baseline model we perform this estimation in two steps. First, the double fixed effects model is estimated for natives, with controls only for age, individual and workplace (equation (8)).<sup>26</sup> In the second step, the part of the predicted wage for natives which can be attributed to their age will be subtracted from the wage level for the immigrants, thereby eliminating the part of the immigrant's wage which is due to age. The revised wages for the immigrants will then be used as outcome variables in the double fixed effects model for the immigrants (equation (9)). This way the coefficient  $\theta_I$  in model (9) can be interpreted as wage catch-up controlling for the group-specific wage premium of entering a specific firm.

$$\ln \operatorname{real} \operatorname{wage}_{it} = \delta_N \operatorname{Age}_{it} + \mu_i + \psi_i^N + \varepsilon_{it}$$
(8)

$$(\ln \operatorname{real} \operatorname{wage}_{it} - \widehat{\delta}_N \operatorname{Age}_{it}) = \theta_I \operatorname{YSM}_{it} + \mu_i + \psi_j^I + \varepsilon_{it}$$
(9)

The results from the regressions allowing for differential workplace effects for natives and immigrants are shown in Panel B of figure 5. Qualitatively the results do not change much. We do see, though, indications of higher wage catch-up when controlling for immigrant specific workplace wage level for the low educated. This indicates that the low educated immigrants fall behind natives in terms of benefiting from sorting into workplaces which pay them better. The difference between the double FE catch-up and the total catch-up is, however, not significant (95 percent confidence interval of estimates overlap each other, see table 6 in the appendix).

These estimations are performed on a restricted data, including only those workplaces which employ at least two native (equation (8)) or at least two immigrant (equation (9)) workers, as this restriction is necessary to identify the individual and workplace effects simultaneously. For the 21,313 workplaces where I can identify both a native and immigrant workplace wage premium the correlation between these premiums is 0.41. This rather low correlation could

<sup>&</sup>lt;sup>26</sup>As with the earlier models, the age parameters are estimated separately for each education group.

be interpreted as group-specific wage setting of the firm, but it might also reflect the fact that the immigrant workplace wage effects are identified by few observations, and thus lack precision. An indication in favor of the latter interpretation is that the variance of workplace premiums for the immigrants is larger than the variance for natives.

This analysis shows us that the wages of the immigrant group do not catchup with natives' wages through movements into better paying establishments, regardless if the wage level of the firm is defined separately for the groups or in a pooled regression.

#### 5.2 Wage catch-up within and between occupations

Occupational codes are only available for the WSSD, and not for the full RAMS data. This means that for the decomposition over occupations we need to focus the analysis on this restricted sample. The advantage of using this restricted data is that we can use the register based wage measure in these estimations. By studying how wages of immigrants and natives grow within occupations and from mobility between them, we can better understand how specific skills and experiences are utilized and signaled in the host country labor market. Here I perform an analogous analysis as the one performed above for workplaces, but employ occupation- instead of workplace fixed effects. Panel A of figure 6 shows the results when assuming equal occupation effects for natives and immigrants.

Again, total wage catch-up is represented by the fully drawn line and this differs from the earlier estimated total wage catch-up only because of the sample and wage measure used (as discussed in section 4.1). The level of the wage gap at 30 years since migration is imputed from the description in section 3, which means that the arrival year intercept in this figure will differ from the intercept in figure 5.<sup>27</sup> The double FE model here captures wage catch-up when controlling for individual-, as well as occupation, fixed effects. The estimates from this model are similar to the total catch-up for all education groups. This indicates that sorting into higher-paying occupations does not contribute to wage catch-up (the confidence intervals of the estimates overlap, see table 3).

The match FE model estimates wage catch-up by controlling for the individual match between an individual and an occupation. The estimates from this model show that within the specific occupation, the highly educated immigrants are experiencing a high catch-up, but that occupational mobility is associated with a substantial decrease of the total catch-up.

<sup>&</sup>lt;sup>27</sup>Here we can only use the WSSD data, and only the years 1996-2008 for which we have occupation codes. The restriction of excluding individuals with approximate wages lower than the cut-off for a full time minimum wage is not applied here, as these estimates are not compared to estimates in the full data set. See section B for an analysis of how the wage measure and data set used affects the catch-up rate.



Figure 6: Wage catch-up; Within and between occupations

Note. Estimates using WSSD data and register based wage measure. Predicted wage gap over Years Since Migration. The gap at 30 years since migration, for the total wage catch-up estimates in panel A, is imputed from figure 3. This yields initial gaps of -17% (Less than High School), -17% (High School), -25% (More than high school), which are then used as the first year intercept for all predicted wages in both panel A and Panel B. Slopes of prediction lines represent estimated wage catch-up parameters. Estimates in Panel A presented in column 2,3 and 4 in table 8 in the appendix. Total catch-up represents the estimates from the specification with individual fixed effects, model (5). Double FE represents estimates when also including occupation fixed effects, model (6). Match FE catch-up is the parameter from the specification with a control for the match between occupation and individual, model (7). Estimates in Panel B are obtained through versions of model (8) and (9).

Results from relaxing the assumption of the equal occupation effect are outlined in Panel B of figure 6. Here it is clear that the wage gain from sorting is no different when assuming equal or differential occupation effects for the poorly educated. But there is a slightly higher total catch-up than the catch-up within occupations of similar wage levels (double FE model) for the highly educated the first ten years. This means that sorting into occupations which are better paying for immigrants contributes to narrowing the wage gap for this group.

The correlation between wage levels in different occupations for natives and immigrants is 0.72, which indicates that relatively well paying occupations for natives are also relatively well paying occupations for immigrants. Despite this, it does seem like the highly educated immigrants sort into occupations for which their specific premium is higher, rather than occupations where the general wage level is better.

As the WSSD data, which we are constrained to use when studying occupations, is a restricted data set with an over-representation of public sector firms and larger firms, it is possible that these results are influenced by the data being non-representative. Therefore, these conclusions should be seen as valid for this restricted subsample of the labor market. The high within-occupation wage catch-up is mainly seen for men (see figure 11 in the appendix).

	Total c	atch-up	Double FE catch-up		Match FI	E catch-up
Panel A of figure 6	$B_L$	$\dot{B_U}$	$B_L$	$B_U$	$B_L$	$B_U$
Less than high school						
0-10 YSM	0.0058	0.0073	0.0056	0.0073	0.0052	0.0069
11-20 YSM	-0.0001	0.0008	0.0007	0.0013	0.0002	0.0012
21-30 YSM	-0.0032	-0.0018	-0.0021	-0.00133	-0.0028	-0.0012
High school						
0-10 YSM	0.0043	0.0053	0.0044	0.0052	0.0037	0.0049
11-20 YSM	0.0007	0.0014	0.00133	0.0017	0.0015	0.0022
21-30 YSM	-0.0007	0.0003	-0.0003	0.0005	-0.0004	0.0007
More than high school						
0-10 YSM	0.0042	0.0056	0.0042	0.005	0.0054	0.0072
11-20 YSM	0.0017	0.0027	0.0022	0.0028	0.0026	0.0037
21-30 YSM	0.0007	0.0022	0.0011	0.0019	0.001	0.0026

 Table 3: Confidence intervals of wage catch-up parameters; Occupation decomposition

*Note:* Standard errors of all estimates but the Double FE estimates are clustered at the level of the individual. For the Double FE estimates standard errors are obtained through bootstrapping with 30 repetitions. The confidence intervals are based on 95 % confidence level.

### 6 Discussion

We have seen that initial wage catch-up is underestimated if not accounting for time-invariant unobservable characteristics of the individuals employed at different points in time. This effect is most pronounced for the highly educated immigrants. As employment rates the first years in the country are low, this can be interpreted as a bias arising from a positive selection into employment among recent immigrants. This conclusion is robust to the choice of data, as well as the wage measure, used.

The highly educated experience a larger initial wage gap and a larger subsequent wage catch-up than other groups do. The higher wage catch-up for the more highly educated could be explained by the greater importance of gaining the human capital, predominantly language skills, needed to succeed in the part of the labor market which the highly educated enter (Lalonde and Topel, 1997).

There are many factors which may constrain immigrants' labor market integration, in addition to the lack of host country human capital. These include discriminatory practices (both preference based and information based) which limit both employment opportunities and wages (Becker, 1971; Phelps, 1972), but imperfect information regarding the host country labor market may also serve as a barrier for recent immigrants (Barth, Bratsberg, and Raaum, 2012).

The empirical patterns displayed in the second part of the paper clearly show that immigrants' wage growth (relative to natives') mainly is achieved within workplaces and occupations. This seems to contradict stories based on the accumulation of country-specific human capital and the hypothesis that immigrants' starting wages are affected by poor information regarding their labor market options, since that would have implied that relative wages should have grown through job mobility towards better matches, workplaces and occupations.

Instead, the results seem more in line with explanations based on the accumulated firm-specific human capital or reduced firm-specific uncertainty regarding skills and experiences. We can think of this in the following way: Employers only observe imperfect signals of workers' productivity, and this uncertainty (signal noise) is likely to be larger for immigrants, for example due to difficulties in evaluating foreign work experience and education. As a consequence, immigrant workers are hired based on (expectations regarding) the productivity of the group, rather than that of the individual, i.e. they face statistical discrimination (Phelps, 1972). If employers are risk-averse this could lead to low initial wages for the immigrants (Oettinger, 1996). This initial wage penalty can be revised as the employer learns about the actual productivity of the worker, and in a setting where the current employer learns more than alternative employers relative wage increases would primarily occur within workplaces.

If, for example, revealed productivity is match-specific (Oettinger, 1996), or

if the worker has acquired human capital which is utilized specifically at the firm, <sup>28</sup> employers may be inclined to differentiate between wages of the employees even if other employers are not updating their beliefs about the worker's productivity accordingly. Employers may also relocate workers to positions in which their skills are better utilized, or they could lay off less productive workers, thus increasing outside options for remaining workers (see Altonji and Pierret (1997) for a discussion of this). Therefore, despite the potential lack of outside pressure from other employers, it is likely that current employers to some extent update wages based on new information, and thus employer learning may contribute to the observed empirical patterns.

Similarly, the high within-occupation wage catch-up can be understood as signs of occupational experience serving as a clear signal of a specific skill, which is rewarded within the occupation, but which does not have the same return in other occupations. We see this very clearly for the highly educated, for which variance in occupation-specific wage is much larger than for the other groups.

Overall, the decomposition results of this essay are consistent with findings from Norway, where Barth, Bratsberg, and Raaum (2012) show that the lack of wage catch-up is to a large extent explained by a lack of sorting into highpaying workplaces. These results contradict with the results of Damas de Matos (2012), who shows that sorting into more high-paying firms can explain about one third of the closing of the wage gap among migrant workers in Portugal. This can possibly be explained by the studied group of workers, since Damas de Matos (2012) only studies low-skilled labor migrants in the private sector and follows them for up to 10 years.

# 7 Conclusions

To conclude, this essay shows that estimates of wage catch-up will be biased if the initial positive selection into employment is not taken into account. The immigrant-native wage gap is unchanged, or increasing, with time spent in Sweden, if not accounting for selection into employment. The total wage catchup for the most educated is about 13 percentage points over 30 years, while the other groups experience a wage catch-up of about 6 percentage points, which levels off after about 10 years. Still after 30 years, all groups have substantially lower wages than comparable natives.

The decomposition of the wage catch-up into wage catch-up within and between workplaces and occupations shows that immigrants' wages grow relative to natives' almost entirely because of higher wage growth within workplaces and occupations. This suggests that other barriers than the lack of host country

<sup>&</sup>lt;sup>28</sup> Lazear (2003) presents a way to think of the return to seniority not as a function of firmspecific human capital, but rather as a function of general human capital which is utilized in particular combinations within a firm.

human capital are impeding the integration process of the immigrants. The results can be interpreted as signs of employers reducing their uncertainty regarding the immigrant employees, thus increasing wages.

This has been a first step in trying to improve our understanding of the wage assimilations for non-western immigrants, who as a group has experienced difficulties in many western labor markets. By studying individuals, workplaces and occupations simultaneously we learn where immigrants' and natives' wage growth differ. The results of this essay motivate further use of matched employer-employee data to provide more in-depth analysis of how matches with employers and occupations influence initial labor market position and careers paths for different groups of workers.

# References

- Abowd, J. M., R. H. Creecy, and F. Kramarz. 2002. "Computing person and firm effects using linked longitudinal employer-employee data." Technical Paper TP-2002-06. Longitudinal Employer-Household Dynamics, Center for Economic Studies, US Census Bureau.
- Abowd, J. M. and F. Kramarz. 1999. "The analysis of labor markets using matched employer-employee data". In *Handbook of Labor Economics*. Ed. by O. C. Ashenfelter and D. Card. Vol. 3. Elsevier, 2629–2710.
- Abowd, J. M., F. Kramarz, and D. N. Margolis. 2003. "High wage workers and high wage firms." *Econometrica* 67, 251–333.
- Altonji, J. G. and C. R. Pierret. 1997. "Employer Learning and Statistical Discrimination." NBER Working Paper.
- Arai, M., H. Regnér, and L. Schröder. 2000. "Invandrare på den svenska arbetsmarknaden". Ds 2000:69.
- Åslund, O., P-A. Edin, and R. J. Lalonde. 2000. "Emigration of Immigrants and Measures of Immigrant Assimilation: Evidence from Sweden". Swedish Economic Policy Review 7, 163–204.
- Åslund, O. and O. Nordström Skans. 2010. "Will I See You at Work? Ethnic Workplace Segregation in Sweden, 1985-2002". *Industrial and Labor Relations Review* 63, 471–493.
- Aydemir, A. and M. Skuterud. 2008. "The immigrant wage differential within and across establishments". *Industrial and Labor Relations Review* 61, 334– 352.
- Barth, E., B. Bratsberg, and O. Raaum. 2006. "Local unemployment and the Relative wages of Immigrants: Evidence from the Current Population surveys". *The Review of Economics and Statistics* 88, 243–263.
- Barth, E., B Bratsberg, and O. Raaum. 2012. "Immigrant wage profiles within and between establishments". *Labour Economics* 19, 541–556.
- Becker, Gary. S. 1971. *The Economics of Discrimination*. University of Chicago Press.
- Borjas, G. 1985. "Assimilation, changes in cohort quality, and the earnings of immigrants". *Journal of Labor Economics* 3, 463–489.
- Borjas, G. 1999. "The economic analysis of immigration". In *Handbook of Labor Economics*. Ed. by O. C. Ashenfelter and D. Card. Vol. 3. Elsevier, 1697–1760.

- Card, D., J. Heining, and P. Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality". *The Quarterly Journal of Economics* 128, 967–1015.
- Catanzarite, L. 2000. "Brown-Collar Jobs: Occupational Segregation and Earnings of Recent-Immigrant Latinos". *Sociological Perspectives* 43, 45–75.
- Chiswick, B. R. 1978. "The Effect of Americanization on the Earnings of Foreign-born Men". *Journal of Political Economy* 86, 897–921.
- Damas de Matos, A. 2012. "The careers of immigrants". CEP Discussion Paper.
- Edin, P-A., B. Holmlund, and O. Nordström Skans. 2009. "Wage dispersion between and within plants: Sweden 1985-2000". In *The Structure of Wages: An International Comparison*. Ed. by E. P. Lazear and K. L Shaw. University of Chicago Press, 217–260.
- Elliott, R. J. R. and J. K. Lindley. 2008. "Immigrant wage differentials, ethnicity and occupational segregation". *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171, 645–671.
- Eriksson, S. 2007. "Arbetsutbud och sysselsättning bland personer med utländsk bakgrund : en kunskapsöversikt". DS 2007:4.
- Friedberg, R. M. 1992. "The Labor Market Assimilation of Immigrants in the United States: The Role of Age at Arrival". Brown University.
- Hammarstedt, M. and G. Shukur. 2006. "Immigrants' Relative Earnings in Sweden - A Cohort Analysis". *Labour* 20, 285–323.
- Hayfron, J. E. 1998. "The performance of immigrants in the Norwegian labor market". *Journal of Population Economics* 11, 293–303.
- Husted, L., M. Rosholm, H. Skyt Nielsen, and N. Smith. 2002. "Employment and Wage Assimilation of Male First-Generation Immigrants in Denmark". *International Journal of Manpower* 22, 39–71.
- Lalonde, R. J. and R. H. Topel. 1997. "Economic Impact of International Migration and the Economic Performance of Migrants". In *Handbook of Population and Family Economics*. Ed. by M. R. Rosenzweig and O. Stark. Vol. 1. Elsevier, 799–850.
- Lazear, E. P. 2003. "Firm-specific human capital: A skill-weights approach". NBER Working Paper 9679.
- Lazear, E. P. and K. L. Shaw. 2005. "Wage structure, raises and mobility". In *The Structure of Wages: An International Comparison*. Ed. by E. P. Lazear and K. L Shaw. University of Chicago Press.

- Le Grand, C. and R. Szulkin. 2002. "Permanent Disadvantage or Gradual Integration: Explaining the Immigrant Native Earnings Gap in Sweden". *Labour* 16, 37–64.
- Lubotsky, D. 2007. "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings". *Journal of Political Economy* 115, 820–867.
- McTague, T., C. Robinson, K. Stainback, T. Taylor, D. Tomaskovic Devey, and C. Zimmer. 2006. "Documenting Desegregation: Segregation in American Workplaces by Race, Ethnicity, and Sex, 1966-2003". *American Sociological Review* 71, 565–588.
- OECD. 2007. "Integration of immigrants in OECD countries: Do Policies Matter?"
- ——— 2012. "Labour market outcomes". In *Settling In: OECD Indicators of Immigrant Integration 2012*. OECD Publishing.
- Oettinger, G. S. 1996. "Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap". *Journal of Labor Economics* 14, 52– 78.
- Pendakur, K. and S. Woodcock. 2010. "Glass ceilings or glass doors? Wage disparity within and between firms". *Journal of Business and Economic Statistics* 28, 181–189.
- Phelps, E. S. 1972. "The Statistical Theory of Racism and Sexism". *The Ameri*can Economic Review 62, 659–661.
- Pischke, J-S. 1992. "Assimilation and the Earnings of Guestworkers in Germany." ZEW Working Paper.
- Sarvimäki, M. 2011. "Assimilation to a Welfare State: Labor Market Performance and Use of Social Benefits by Immigrants to Finland\*". *The Scandina*vian Journal of Economics 113, 665–688.
- Shields, M. A. and S. Wheatley Price. 1998. "The earnings of Male Immigrants in England: Evidence from the Quarterly LFS". *Applied Economics* 30, 1157– 1168.

# Appendix A Data and variable description

#### Full data- RAMS (Register based labor market statistics)

Connections between firms and individuals. Available for the years 1995-2008. Includes individuals between the age of 20 and 65.

#### Firm data - WSSD (Wage and salary structure data)

A stratified sample of firms and their employees, for 1995-2008. Contains all firms in the public sector, and a sample of firms in private sector. Firms with more than 500 employees have sampling probability equal to 1. In total, about 50 % of employees in the private sector are covered.

- Log real register based wage Monthly wage (full time equivalents) Registered wages from WSSD. Deflated by CPI, using 1990 as the reference year.
- Log real approximated wage Annual income divided by number of months in employment for a specific firm-individual match. Deflated by CPI, using 1990 as reference year. Individuals with less than 75 % of minimum full time wage are excluded, as they are likely to be part-time employed. Also wages above the 99th percentile in the wage distribution is excluded.
- **Education** Less than High School=Up to 10 years of schooling; High school=11-12 years of schooling; Post High School Education=At least some tertiary education, includes vocational training as well as university education.
- Years since migration Years since recieving residence permit in Sweden. Based on calender year. May be recorded with a lag.
- **Immigrant** The variable is defined from the information of country of origin. It takes on the value 1 if the individual origins from any of the non-European states, with the exceptions listed here:
  - European countries included as non-Western countries: Yugoslavia, Croatia, Macedonia, Slovenia, Bosnia-Herzegovina, Albania, Bulgaria, Romania, Russia, Ukraine, Belarus.
  - Non-European countries not included as non-Western countries: USA, Canada, Australia, New Zealand and other countries in Oceanic region, Japan, China, South Korea, Hong Kong.
- **Firm** Identifiers for the firm. Large organizational units can be coded as one firm, in the public sector all municipal workplaces belong to the same firm

Establishment Identification numbers for the workplace

Occupation 3-digit ISCO coding for occupation, available for the WSSD data

CIATEN	MS RAMS - grouped	WSSD - grouped Workplaces	WSSD - grouped Occupations
# Natives 3,805,67 # Immigrants 173,282 # Workplaces (Occupations) 509,819	5,670 3,675,788 282 161,745 819 257,940	2,700,727 108,781 97,231	3,074,402 140,582 124
# Observations/individual (mean- native)	.6 8.3	6.7	6.8
# Observations/individual (mean-immigrant) 5.4	.4 5.5 	4.8	ν,
# Workplaces (Occupations)/individual (mean- native) 2.1 # Workplaces (Occupations)/individual (mean- immigrant) 1.8	.1 2.1 .8 1.8	1.7	1.6 $1.4$
# Observations/workplace (Occupation) (mean- native) 4.1	.1 4.1	4	4.2
# Observations/workplace (Occupation) (mean- immigrant) 3	3.1	3.1	3.5

Table 4: Description of data sets

values are for occupation instead for workplace as in the other columns. In column 3 the data set which is used for section B in the appendix is presented. This is smaller than the WSSD grouped over occupations, mainly because this dataset excludes all individuals with an approximate monthly wage below the minimum wage threshold.

# Appendix B Robustness: choice of wage measure and sample

The baseline analysis uses an approximate wage measure, constructed from information on annual incomes and months of work. For this reason the estimates may be affected by systematic differences in hours worked between natives and immigrants. Therefore, the above analysis is performed also for the WSSD data, for which we have both the approximate wage measure and the register based wage. This analysis will help us understand how the wage measure and sample used influence the results. Results from both section 4.1 and 5.1 are re-estimated using this restricted data set.

Table 5 presents descriptive statistics of mean individual characteristics for natives and non-western immigrants using the WSSD data which is used for the decomposition over occupations as well as the robustness analysis.<sup>29</sup> When comparing mean values from table 5 with mean values from table 1 we see that the highly educated are oversampled in the WSSD data. Women are also overrepresented in the WSSD, potentially due to the full coverage of the public sector. Annual incomes are substantially higher in the WSSD data, which could reflect the higher share of highly educated or the fact that the data collection of the WSSD is in November each year, thereby excluding seasonal workers. As the RAMS consists of tax registers, all employments during the year is covered in RAMS. Overall, immigrants in the WSSD data have been longer in the country, which has to do with the low employment rate the early years. Thus, the WSSD is not fully representative for the full labor market, but the sampling probabilities for employed natives and immigrants are rather similar (57 and 59 percent respectively).

Panel A of figure 7 is directly comparable to figure 3, but plots estimates using the restricted data set. It is clear that the main result of an underestimation of wage catch-up when not including individual fixed effects is invariant to the wage measure, as well as the data source, used. But the total wage catchup (solid line) for the least educated is affected by the choice of wage measure used (compare Panel A and Panel B). In particular, the wage catch-up is negative after ten years when using the register based wage measure instead of the approximated. In total numbers, also the wage catch-up for the most highly educated is lower when using the register based wage measure, even if the pattern of relative wage growth is identical, regardless which wage measure is used. The slightly higher wage catch-up when using approximated wage measure may reflect increases in hours worked among the immigrants. For the high school

<sup>&</sup>lt;sup>29</sup>This data set, without further restrictions, is used for the decomposition of wage catch-up within and between occupations. For the robustness analysis, the restriction of excluding individuals with an approximate wage below the threshold for full time minimum wage are excluded from the WSSD data.

	Natives	Non-Western born
Male	0.457	0.491
Age (Mean)	44.561	42.079
<25 years	0.043	0.008
>25 years & <35 years	0.169	0.173
>35 years & $<45$ years	0.262	0.439
>45 years & $<55$ years	0.294	0.314
>55 years & $<65$ years	0.232	0.067
Less than high school	0.136	0.193
High school	0.498	0.402
University	0.366	0.405
Approx monthly wage	22,792	20,193
Register based monthly wage	22,819	19,320
Income from employment	257,923	214,089
Employment rate	0.987	0.956
Years since migration (Mean)		11.971
<5 years in Sweden		0.103
>5 years & <10 years in Sweden		0.328
<10 years & <20 years in Sweden		0.432
>20 years in Sweden		0.137
N	1,684,213	58,532

#### Table 5: Descriptive statistics WSSD - individual details - year 2002

*Note:* Mean values for individuals between 20 and 65 in WSSD data, year 2002. Mean income and mean wage are calculated among individuals with positive values. Approximate monthly wage is set to missing when below the defined minimum wage threshold.



Figure 7: Bias from time-invariant unobservable characteristics, different wage measures

Note: Estimates from models (3) and (4), using the WSSD data. Panel A uses approximate wage measure, and Panel B uses register based wage measure. Gap at 30 years since migration imputed from figure 3. Estimates in panel A presented in column 1-2 in table 9 in the appendix. Estimates in panel B presented in column 1-2 in table 10 in the appendix.

educated, wage catch-up is slightly lower in the WSSD data than in the full labor market (compare Panel A with figure 3). This might reflect the sampling structure of the WSSD data with public sector, and large firms, being overrepresented.

Panel A of figure 8 is directly comparable to Panel A in figure 5, but represents estimates using the restricted data set. In terms of where the wage catch-up takes place, approximated wages consistently shows a slightly larger within-workplace (Match FE) catch-up than does the register based wage measure (Panel B). As the approximated wage also captures wage catch-up in terms of changes in hours worked, an interpretation of this may be that changes along this intensive margin are more common within the workplace than from job mobility. But the estimated within-workplace wage catch-up is also much lower in the restricted data than in the full data (compare panel A of figure 5 and panel A of figure 8). Thus, the lower within-workplace wage catch-up than total catch-up seen here in figure 8 is affected by the fact that the restricted data



Figure 8: Decomposition of wage catch-up with different wage measures

Note. Estimates from regressions using WSSD data. Panel A uses the approximate wage measure, and Panel B uses the register based wage measure. Predicted wage gap over Years Since Migration. The intercept is set so that the total predicted wage catch-up of the individual fixed effect model (fully drawn line) at 30 years since migration is the same as in figure 3. Slopes of prediction lines represent estimated wage catch-up parameters. Estimates in panel A presented in column 2-4 in table 9 in the appendix. Estimates in panel B presented in column 2-4 in table 10 in the appendix. Total catch-up represents the estimates from the specification with individual fixed effects, model (5). Double FE represents estimates when also including workplace fixed effects, model (6). Match FE catch-up is the parameter from the specification with a control for the match between workplace and individual, model (7).

samples private firms each year, and thus we do not identify within-workplace wage catch-up for individuals who are not sampled with the same employer more than one year. This means that the decomposition results are sensitive to the sampling structure of the WSSD data, while the conclusion of a withinworkplace wage catch-up as high as, or higher, than total wage catch-up is invariant to the wage measure used.

# Appendix C Analysis divided by sex



Figure 9: Women and men - Relative wage growth



#### Figure 10: Women and men - workplace decomposition

Figure 11: Women and men - occupational decomposition



# Appendix D Detailed regression results

	Total c	atch-up	Double Fl	E catch-up	Match FE	catch-up
Panel B of figure 5	$B_L$	$\hat{B}_U$	$B_L$	$B_U$	$ $ $B_L$	$B_U$
Less than high school						
0-10 YSM	0.0033	0.005	0.0041	0.0063	0.006	0.008
11-20 YSM	-0.0012	-0.0002	-0.0006	0.0003	-0.0008	0.0002
21-30 YSM	-0.0004	0.0012	-0.0001	0.0016	-0.0005	0.0012
High school						
0-10 YSM	0.0043	0.0052	0.0051	0.0061	0.0061	0.0072
11-20 YSM	-0.0015	-0.0008	-0.0007	0.0001	-0.0011	-0.0004
21-30 YSM	-0.0011	0.0000	-0.0006	0.0006	-0.0011	0.0002
More than high school						
0-10 YSM	0.0064	0.0073	0.0072	0.0084	0.0076	0.0087
11-20 YSM	0.0029	0.0036	0.0029	0.0038	0.001	0.0019
21-30 YSM	0.0023	0.0035	0.0018	0.0032	0.0012	0.0027
Panel B of figure 6						
Less than high school						
0-10 YSM	0.006	0.007	0.0059	0.0068	0.0054	0.0066
11-20 YSM	0.0001	0.0006	0.0005	0.001	0.0004	0.001
21-30 YSM	-0.0029	-0.0019	-0.0024	-0.0014	-0.0025	-0.0015
High school						
0-10 YSM	0.0045	0.0051	0.0043	0.005	0.004	0.0047
11-20 YSM	0.0011	0.0015	0.0014	0.0018	0.0018	0.0022
21-30 YSM	-0.0002	0.0004	-0.0001	0.0005	-0.0001	0.0006
More than high school						
0-10 YSM	0.0049	0.0055	0.0038	0.0048	0.0061	0.0068
11-20 YSM	0.0025	0.0029	0.0024	0.003	0.0032	0.0036
21-30 YSM	0.0016	0.0023	0.0015	0.0022	0.0017	0.0025

**Table 6:** Confidence intervals of estimated wage catch-up parameters 

 Group specific workplace and occupation effect

*Note:* Standard errors of all estimates but the Double FE estimates are clustered at the level of the individual. For the Double FE estimates standard errors are obtained through bootstrapping with 30 repetitions. The confidence intervals are based on 95 % confidence level.

	(1)	(2)	(3)	(4)
	Quasi-fixed	Individual-	Double fixed	Match fixed
	effect	fixed	effect	effect
		effect		
Corresponds to lines in figure 3 and Pa	anel A in figur	e 5		
Less than high school*Age	0.0537***	0.0696***	0.0712***	0.0806***
0 0	(0.001)	(0.001)	(.001)	(0.001)
Less than high school* $Age^2$	-0.0005***	-0.0008***	-0.0008***	-0.0011***
6	(0.000)	(0.000)	(.000)	(0.000)
Less than high school* $Age^3$	0.0000***	0.0000***	0.0000***	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)
High school*Age	0.0569***	0.0583***	0.0568***	0.0647***
8	(0.000)	(0.000)	(.000.)	(0.000)
High school* $Aae^2$	-0.0005***	-0.0004***	-0.0004***	-0.0006***
	(0.000)	(0.000)	(0.000)	(0.000)
High school* $Aae^3$	0.0000***	0.0000***	0.0000***	0.0000***
	(0.000)	(0.000)	0.000)	(0.000)
More than high school*Age	0.0795***	0.0589***	0.0478***	0.0322***
	(0.001)	(0.001)	(.000)	(0.001)
More than high school* $Age^2$	-0.0006***	0.0001***	0.0003***	0.0006***
	(0.000)	(0.000)	(0.000)	(0.000)
More than high school* $Age^3$	0.0000***	-0.0000***	0.0000***	-0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*Less than high school*YSM1	0.0008*	0.0055***	0.0055***	0.0084***
	(0.000)	(0.001)	(.000.)	(0.001)
Imm*Less than high school*YSM2	-0.0016***	-0.0002	0.0003	0.0001
8	(0.000)	(0.000)	(.000)	(0.000)
Imm*Less than high school*YSM3	-0.0006	0.0003	0.0005	0.0004
8	(0.001)	(0.001)	(.000)	(0.001)
Imm*High school*YSM1	0.0028***	0.0062***	0.0061***	0.0083***
8	(0.000)	(0.000)	(.000)	(0.000)
Imm*High school*YSM2	-0.0016***	-0.0005**	0.0003**	-0.0001
8	(0.000)	(0.000)	(.000)	(0.000)
Imm*High school*YSM3	-0.0007*	-0.0000	0.0006***	0.0001
0	(0.000)	(0.000)	(.000)	(0.000)
Imm*More than high school*YSM1	-0.0022***	0.0072***	0.0081***	0.0089***
8	(0.000)	(0.000)	(.000)	(0.000)
Imm*More than high school*YSM2	0.0013***	0.0033***	0.0036***	0.0013***
8	(0.000)	(0.000)	(.000)	(0.000)
Imm*More than high school*YSM3	0.0017***	0.0029***	0.003***	0.0021***
0	(0.001)	(0.001)	(.000)	(0.001)
	(	(	()	(/
Observations	31,568,955	31,568,955	31,568,955	31,568,955
R-squared	0.344	0.809	0.827	0.883

Table 7: Full regression output; Decomposition over workplaces

*Note:* Estimations performed using RAMS data. Robust standard errors in parentheses \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1. Standard errors in column 1, 2, and 4 are clustered at the level of the individual. Standard errors in double FE estimations are obtained through bootstrapping (15 repetitions). YSM parameters are estimated as splines, each capturing a ten year period.

	(1)	(2)	(3)	(4)
	Quasi-fixed	Individual-	Double fixed	Match fixed
	effect	fixed	effect	effect
		effect		
Corresponds to lines in Panel A in figu	ıre 6			
Less than high school*Age	0.0427***	0.0538***	0.0525***	0.0494***
8	(0.001)	(0.001)	(.001)	(0.001)
Less than high school* $Age^2$	-0.0003***	-0.0006***	-0.00058***	-0.0005***
8	(.000)	(.000)	(.000)	(.000)
Less than high school* $Aae^3$	0.0000***	0.0000***	0.0000	0.0000***
2000 than ingh benoor rige	(.000)	.000)	(.000)	(.000)
High school*Age	0.0616***	0.0628***	0.0612***	0.0583***
	(0.000)	(0.000)	(.000)	(0.000)
High school* $Aae^2$	-0.0007***	-0.0007***	-0.0007***	-0.0007***
	(.000)	(.000)	(.000)	(.000)
High school* $Aae^3$	0.0000***	0.0000***	0.0000	0.0000***
riigii teritoti rige	.000)	(.000)	(.000)	(.000)
More than high school*Age	0.113***	0.109***	0 1048***	0.0986***
infore than high behoof rige	(0.001)	(0.001)	(001)	(0.001)
More than high school* $Aae^2$	-0.0013***	-0.0012***	-0.0012***	-0.0011***
	(000)	(000)	(000)	(000)
More than high school* $Aae^3$	0.0000***	0.0000***	0.0000***	0.0000***
filore than high behoof rige	(000)	(000)	(000)	(000)
Imm*Less than high school*YSM1	0.0036***	0.0065***	0.0064***	0.006***
finin Dess than high school Tolvil	(0,000)	(0,000)	(000)	(0,000)
Imm*Less than high school*YSM2	-0.0008***	0.0003	0.001***	0.0007***
Thin Dess than high school 101012	(0,000)	(0,000)	(000)	(0,000)
Imm*Less than high school*YSM3	-0.0024***	-0.0025***	-0.0017***	-0.002***
Thin Dess than high school Total	(0,000)	(0,000)	(000)	(0,000)
Imm*High school*VSM1	0.0021***	0.0048***	0.0048***	0.0043***
	(0,000)	(0,000)	(000)	(0,000)
Imm*High school*YSM2	-0.0002	0.0011***	0.0015***	0.0019***
inini Tiigii School Toivi2	(0,000)	(0,000)	(000)	(0,000)
Imm*High school*YSM3	-0.0001	-0.0002	0.0001	0.0001
	(0,000)	(0.0002)	(000)	(0,000)
Imm*More than high school*YSM1	-0.0032***	0.00493***	0.0046***	0.0063***
	(0,000)	(0,000)	(000)	(0,000)
Imm*More than high school*YSM2	-0.0005*	0.0022***	0.0025***	0.0031***
	(0,000)	(0.000)	(.000)	(0.000)
Imm*More than high school*VSM3	0.0009*	0.0015***	0.0015***	0.0018***
min trole than high school 10003	(0,000)	(0,000)	(000)	(0,0010)
	(0.000)	(0.000)	(.000)	(0.000)
Observations	21.695 159	21.695 159	21.695 159	21.695 159
R-squared	0.414	0.922	0.925	0.949

Table 8: Full regression output; Decomposition over occupations

*Note:* Estimations performed using WSSD data. Robust standard errors in parentheses \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1. Standard errors in column 1, 2, and 4 are clustered at the level of the individual. Standard errors in double FE estimations are obtained through bootstrapping (30 repetitions). YSM parameters are estimated as splines, each capturing a ten year period.

Quasi-fixed effect         Individual- fixed effect         Double fixed Match fixed effect           Corresponds to lines in Panel A in figure 7 and figure 8           Less than high school*Age $0.0246^{***}$ $0.0349^{***}$ $0.0336^{***}$ $0.0315^{***}$ Less than high school*Age $0.0002^{***}$ $0.00000^{***}$ $0.0000^{***}$ $0.0000^{***}$ Less than high school*Age $0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ Less than high school*Age $0.0323^{***}$ $0.0287^{***}$ $0.0266^{***}$ $0.0000^{***}$ High school*Age $0.000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $0.0002^{***}$ $0.0002^{***}$ High school*Age $0.000^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ High school*Age <sup>2</sup> $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ High school*Age <sup>3</sup> $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{****}$ $0.0000^{****}$ More than high school*Age <sup>3</sup> $0.0002^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{****}$		(1)	(2)	(3)	(4)
Letter         Linear         Circlet           Corresponds to lines in Panel A in figure 7 and figure 8           Less than high school*Age $(0.001)$ $(0.001)$ $(0.001)$ Less than high school*Age <sup>2</sup> $0.0002^{***}$ $-0.0000$ $(0.000)$ Less than high school*Age <sup>3</sup> $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ Less than high school*Age $0.0323^{***}$ $0.0265^{***}$ $0.0266^{***}$ (0.000) $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school*Age $0.000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school*Age <sup>2</sup> $0.0000^{***}$ $0.0000^{***}$ $-0.0000^{***}$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ More than high school*Age <sup>2</sup> $0.0002^{***}$ $0.0000^{***}$ $-0.0000^{***}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ More than high school*Age <sup>3</sup> $-0.0000^{***}$ $0.0000^{***}$ $-0.0000^{***}$		Quasi-fixed	Individual-	Double fixed	Match fixed
Corresponds to lines in Panel A in figure 7 and figure 8           Less than high school*Age $0.0246^{***}$ $0.0336^{***}$ $0.0315^{***}$ Less than high school*Age <sup>2</sup> $0.0000^{***}$ $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ Less than high school*Age <sup>3</sup> $-0.0000^{***}$ $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ Less than high school*Age $0.0323^{***}$ $0.0265^{***}$ $0.0265^{***}$ $0.0265^{***}$ High school*Age <sup>2</sup> $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ High school*Age <sup>2</sup> $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ High school*Age <sup>3</sup> $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ More than high school*Age $0.0423^{***}$ $0.0011^{***}$ $0.0000^{***}$ $0.0000^{***}$ More than high school*Age <sup>3</sup> $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ More than high school*Age <sup>3</sup> $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ More than high school*YSM1 $0.0001^$		cheet	effect	cifect	chect
Less than high school*Age $0.0246^{***}$ $0.0349^{***}$ $0.0336^{***}$ $0.0315^{***}$ Less than high school* $Age^2$ $0.0002^{***}$ $-0.0000$ $0.0000$ $0.0000$ Less than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ High school*Age $0.323^{***}$ $0.025^{***}$ $0.026^{***}$ $0.000^{***}$ High school*Age $0.0323^{***}$ $0.025^{***}$ $0.026^{***}$ $0.000^{***}$ High school*Age $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ High school*Age <sup>2</sup> $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ High school*Age <sup>3</sup> $-0.000^{***}$ $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ More than high school*Age <sup>2</sup> $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ More than high school*Age <sup>3</sup> $-0.000^{***}$ $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ More than high school*YSM1 $0.000^{**}$ $0.000^{***}$ $0.0000^{***}$ $0.0000^{***}$	Corresponds to lines in Panel A in figu	ire 7 and figure	e 8		
Less than high school*Age $0.0246^{***}$ $0.0339^{***}$ $0.0336^{***}$ $0.0315^{***}$ Less than high school*Age <sup>2</sup> $0.0002^{***}$ $-0.0000$ $0.0000$ $(0.000)$ Less than high school*Age <sup>3</sup> $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ Less than high school*Age $0.0323^{***}$ $0.0287^{***}$ $0.0266^{****}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school*Age $0.0232^{***}$ $0.0267^{***}$ $0.0266^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ High school*Age $0.000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ More than high school*Age $0.0423^{***}$ $0.0102^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ More than high school*Age <sup>3</sup> $-0.000^{***}$ $-0.000^{***}$ $-0.000^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ Imm*Less than high school*YSM1 $0.0006^{***}$ $0.0002^{***}$ $0.0002^{***}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*Less than high school*YSM1 $-0.0005^{***}$ $0.0003^{***}$ $0.0002^{***}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ Imm*High school*YSM1 $-0.0005^{***}$ $0.0003^{***}$ $0.0002^{***}$ $(0.000)$ $(0.000)$ $(0.000)$					
$u_{1}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ Less than high school* $Age^2$ $0.0002^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $0.0265^{***}$ $0.0265^{***}$ $0.0265^{***}$ $0.0265^{***}$ $0.0265^{***}$ $0.0002^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0000^{***}$	Less than high school*Age	0.0246***	0.0349***	0.0336***	0.0315***
Less than high school* $Age^2$ $0.000^{2***}_{(0.000)}$ $-0.0000$ $0.0000_{(0.000)}$ $(0.000)_{(0.000)}$ Less than high school* $Age^3$ $-0.000^{***}_{(0.000)}$ $-0.0000^{***}_{(0.000)}$ $-0.0000^{***}_{(0.000)}$ $-0.0000^{***}_{(0.000)}$ High school*Age $0.0323^{***}_{(0.000)}$ $0.0265^{***}_{(0.000)}$ $0.0266^{***}_{(0.000)}$ $0.0002^{***}_{(0.000)}$ High school*Age <sup>2</sup> $0.0000^{***}_{(0.000)}$ $0.0000^{***}_{(0.000)}$ $0.0000^{***}_{(0.000)}$ $0.0000^{***}_{(0.000)}$ High school*Age <sup>3</sup> $-0.0000^{***}_{(0.000)}$ $0.0000^{***}_{(0.000)}$ $0.0000^{***}_{(0.000)}$ $0.0000^{***}_{(0.000)}$ More than high school*Age $0.0423^{***}_{**}$ $0.0162^{***}_{**}$ $0.0000^{***}_{(0.000)}$ $0.0000^{***}_{(0.000)}$ More than high school*Age <sup>2</sup> $0.0000^{***}_{**}$ $0.0000^{***}_{**}$ $0.0000^{***}_{**}$ $0.0000^{***}_{**}$ More than high school*Age <sup>3</sup> $-0.0000^{***}_{**}$ $0.0000^{***}_{**}$ $0.0000^{***}_{**}$ $0.0000^{***}_{**}$ More than high school*YSM1 $0.0001^{***}_{**}$ $0.0000^{***}_{**}$ $0.0000^{***}_{**}$ $0.0000^{***}_{**}$ Mum*Less than high school*YSM2 $-0.0003$ $0.0003$ $0.0006^{***}_{**}$ $0.0002^{***}_{**}$ $0.0001^{***}_{**}$ Mum*High school*YSM2 $-0.0010^{***}_{**}$ $0.0003^{***}_{**}$ $0.0036^{***}_{**}$ $0.0036^{***}_{**}$ Mum*High school*YSM3 $-0.0005^{***}_{**}$ $0.0003^{***}_{**}$ $0.0036^{***}_{**}$ $0.0036^{***}_{**}$ Mum*High school*YSM3 $-0.0002^{****}_{**}$ $0.0003^{***}_{**}$ <	6 6	(0.001)	(0.001)	(0.001)	(0.001)
$(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Less than high school* $Age^3$ $-0.000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ High school* $Age$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school* $Age^2$ $0.0000^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school* $Age^3$ $-0.0000^{***}$ $0.0000^{***}$ $-0.0000^{***}$ $0.0000^{***}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ More than high school* $Age^2$ $0.002^{***}$ $0.0011^{***}$ $0.0001^{***}$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ More than high school* $Age^2$ $0.000^{***}$ $-0.0000^{***}$ $0.0000^{***}$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*Less than high school*YSM1 $-0.0003$ $-0.0003$ $-0.0001$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0005^{***}$ $-0.0003^{***}$ $0.003^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0005^{***}$ $0.0003^{***}$ $0.0002^{***}$ $(0.001)$ $(0.000)$ <td< td=""><td>Less than high school*<math>Age^2</math></td><td>0.0002***</td><td>-0.0000</td><td>0.0000</td><td>0.0000</td></td<>	Less than high school* $Age^2$	0.0002***	-0.0000	0.0000	0.0000
Less than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $0.0265^{***}$ $0.0266^{***}$ $0.002^{***}$ $0.002^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0000^{***}$ $0.$	-	(0.000)	(0.000)	(0.000)	(0.000)
High school*Age $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school*Age $0.0237^{***}$ $0.0265^{***}$ $0.0266^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ High school*Age <sup>2</sup> $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school*Age <sup>3</sup> $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ More than high school*Age $0.0423^{***}$ $0.0162^{***}$ $0.001^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.001)$ More than high school*Age <sup>2</sup> $0.000^{***}$ $0.000^{***}$ $0.000^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ More than high school*YSM1 $0.0006^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ Imm*Less than high school*YSM2 $-0.0003$ $-0.0003$ $-0.0001^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ Imm*High school*YSM1 $-0.0011^{***}$ $0.0021^{***}$ $0.0036^{****}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ Imm*High school*YSM1 $-0.0011^{***}$ $0.0000$ $(0.000)$ Imm*High school*YSM1 $-0.0011^{***}$ $0.0000$ $(0.000)$ Imm*More than high school*YSM1 $-0.0003^{***}$ $0.00$	Less than high school* $Age^3$	-0.0000***	-0.0000***	-0.0000***	-0.0000***
High school*Age $0.0323^{***}$ $0.0287^{***}$ $0.0265^{***}$ $0.0266^{***}$ High school*Age2 $0.000$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ High school*Age3 $-0.000^{***}$ $-0.000^{***}$ $-0.000^{***}$ $-0.000^{***}$ More than high school*Age $0.0423^{***}$ $0.0162^{***}$ $0.0079^{***}$ $0.003^{***}$ More than high school*Age2 $0.0002^{***}$ $0.00162^{***}$ $0.000^{***}$ $0.000^{***}$ More than high school*Age2 $0.0002^{***}$ $0.0010^{***}$ $0.0011^{***}$ $0.0012^{***}$ More than high school*Age3 $-0.0000^{***}$ $-0.000^{***}$ $-0.000^{***}$ $-0.000^{***}$ More than high school*Age3 $-0.0000^{***}$ $-0.000^{***}$ $-0.000^{***}$ $-0.000^{***}$ More than high school*Age3 $-0.0000^{***}$ $-0.000^{***}$ $-0.000^{***}$ $-0.000^{***}$ More than high school*YSM1 $0.0006$ $0.046^{***}$ $0.0049^{***}$ $0.0060^{***}$ Imm*Less than high school*YSM2 $-0.0003$ $-0.0003$ $-0.0001$ $(0.000)$ Imm*High school*YSM1 $-0.001^{***}$ $0.003^{***}$ $0.0033^{***}$ $0.0000$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ Imm*High school*YSM2 $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ Imm*High school*YSM3 $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$		(0.000)	(0.000)	(0.000)	(0.000)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	High school*Age	0.0323***	0.0287***	0.0265***	0.0266***
High school* $Age^2$ $0.0000^{***}$ $0.0002^{***}$ $0.0003^{***}$ $0.0002^{***}$ High school* $Age^3$ $0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ More than high school*Age $0.0423^{***}$ $0.0162^{***}$ $0.0079^{***}$ $0.0000^{***}$ More than high school* $Age^2$ $0.0002^{***}$ $0.0011^{***}$ $0.0011^{***}$ $0.0012^{***}$ More than high school* $Age^3$ $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ More than high school*YSM1 $0.0006$ $0.046^{***}$ $0.0049^{***}$ $0.0060^{***}$ More than high school*YSM2 $-0.0000^{***}$ $-0.0000^{***}$ $0.0000^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ Imm*Less than high school*YSM2 $-0.0003$ $0.0003$ $0.0006^{***}$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ $(0.001)$ Imm*High school*YSM1 $-0.0010^{***}$ $0.003^{****}$ $0.0036^{****}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0008^{***}$ $-0.0003$ $-0.0000^{****}$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0026^{****}$ $0.0029^{****}$ $0.0028^{****}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.001)$ Imm*More than high		(0.000)	(0.000)	(0.000)	(0.001)
High school* $Age^3$ (0.000)(0.000)(0.000)(0.000)More than high school*Age $0.0423^{***}$ $0.0162^{***}$ $0.0007^{***}$ $0.0000^{***}$ More than high school*Age $0.002^{***}$ $0.0162^{***}$ $0.0079^{***}$ $0.0030^{***}$ More than high school* $Age^2$ $0.0002^{***}$ $0.0011^{***}$ $0.0011^{***}$ $0.0012^{***}$ More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ More than high school*YSM1 $0.0006$ $0.046^{***}$ $0.0049^{***}$ $0.0060^{***}$ Mm*Less than high school*YSM2 $-0.0003$ $0.0003$ $0.0006^{***}$ $0.0002$ Imm*Less than high school*YSM3 $-0.0003$ $-0.0003$ $-0.0001$ $-0.0002$ Imm*High school*YSM1 $-0.0010^{**}$ $0.0021^{***}$ $0.003^{***}$ $0.003^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ $(0.001)$ $(0.001)$ Imm*High school*YSM2 $-0.0011^{***}$ $0.0021^{***}$ $0.003^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0026^{***}$ $0.0029^{***}$ $0.0028^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*High school*YSM2 $-0.0011^{***}$ $0.0029^{***}$ $0.0028^{***}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.001)$ <t< td=""><td>High school*<math>Age^2</math></td><td>0.0000***</td><td>0.0002***</td><td>0.0003***</td><td>0.0002***</td></t<>	High school* $Age^2$	0.0000***	0.0002***	0.0003***	0.0002***
High school* $Age^3$ $-0.0000^{***}$ $-0.0002^{***}$ $-0.0001^{**}$ $-0.0001^{***}$ $-0.0001^{***}$ $-0.0001^{***}$ $-0.0001^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{***}$ $-0.0003^{$		(0.000)	(0.000)	(0.000)	(0.000)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	High school* $Age^3$	-0.0000***	-0.0000***	-0.0000***	-0.0000***
More than high school*Age $0.0423^{***}$ $0.0162^{***}$ $0.0079^{***}$ $0.0030^{***}$ More than high school* $Age^2$ $0.0002^{***}$ $0.0010^{***}$ $0.0011^{***}$ $0.0012^{***}$ More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ More than high school*YSM1 $0.0006$ $0.0046^{***}$ $0.0049^{***}$ $0.0000^{***}$ Imm*Less than high school*YSM2 $-0.0003$ $0.0003$ $0.0006^{***}$ $0.0002$ Imm*Less than high school*YSM3 $-0.0003$ $-0.0001$ $0.0001$ $(0.000)$ Imm*Less than high school*YSM1 $-0.0005$ $-0.0003$ $-0.0001$ $-0.0002$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.001)$ $(0.001)$ $(0.001)$ Imm*High school*YSM2 $-0.0011^{***}$ $0.0021^{***}$ $0.003^{***}$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ $(0.001)$ Imm*High school*YSM2 $-0.0011^{***}$ $0.0021^{***}$ $-0.0003$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0026^{***}$ $0.0028^{***}$ $-0.0003$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0026^{***}$ $0.0029^{***}$ $0.0028^{***}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.000)$ $(0.001)$ Imm*More than high school*YSM2 $-0.$		(0.000)	(0.000)	(0.000)	(0.000)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	More than high school*Age	0.0423***	0.0162***	0.0079***	0.0030***
More than high school* $Age^2$ $0.0002^{***}$ $0.0010^{***}$ $0.0011^{***}$ $0.0012^{***}$ More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ Mm*Less than high school*YSM1 $0.0006$ $0.0046^{***}$ $0.0049^{***}$ $0.0000^{***}$ Mm*Less than high school*YSM2 $-0.0003$ $0.0003$ $0.0006^{***}$ $0.0002^{***}$ Mm*Less than high school*YSM3 $-0.0003$ $0.0003$ $0.0006^{***}$ $0.0002^{***}$ Mm*Less than high school*YSM3 $-0.0005$ $-0.0003$ $-0.0001$ $0.0001$ Mm*High school*YSM1 $-0.0015^{***}$ $0.0021^{***}$ $0.003^{***}$ $0.0002^{***}$ Mm*High school*YSM1 $-0.0010^{***}$ $0.0021^{***}$ $0.003^{***}$ $0.0003^{***}$ Mm*High school*YSM2 $-0.0011^{***}$ $0.0000$ $(0.000)$ $(0.001)$ Imm*High school*YSM3 $-0.0011^{***}$ $0.0005^{***}$ $-0.0003^{***}$ Mm*High school*YSM2 $-0.0011^{***}$ $0.0000^{***}$ $0.0003^{***}$ Mm*High school*YSM3 $-0.0026^{****}$ $0.0000^{***}$ $0.0000^{***}$ Mm*High school*YSM3 $-0.0026^{****}$ $0.0025^{***}$ $0.0022^{****}$ More than high school*YSM2 $-0.001^{***}$ $0.0028^{***}$ $0.0022^{***}$ More than high school*YSM3 $0.0026^{****}$ $0.0022^{***}$ $0.0011^{***}$ More than high school*YSM3 $0.0008^{***}$ $0.0022^{***}$ $0.0012^{***}$ More than high school*YSM3 $0.0008^{***}$ $0.0021^{***}$		(0.001)	(0.001)	(0.000)	(0.001)
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	More than high school* $Age^2$	0.0002***	0.0010***	0.0011***	0.0012***
More than high school* $Age^3$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $-0.0000^{***}$ $0.0060^{***}$ $0.0060^{***}$ $0.0060^{***}$ $0.0060^{***}$ $0.0060^{***}$ $0.0006^{***}$ $0.0060^{***}$ $0.0000^{***}$ $0.0060^{***}$ $0.0000^{***}$ $0.0060^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0002^{***}$ $0.0000^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0002^{***}$ $0.0000^{***}$ $0.0002^{***}$ $0.0000^{***}$ $0.0002^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0002^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0003^{***}$ $0.0003^{***}$ $0.0003^{***}$ $0.0003^{***}$ $0.0003^{***}$ $0.0003^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ $0.0000^{***}$ <th< td=""><td></td><td>(0.000)</td><td>(0.000)</td><td>(0.000)</td><td>(0.000)</td></th<>		(0.000)	(0.000)	(0.000)	(0.000)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	More than high school* $Age^3$	-0.0000***	-0.0000***	-0.0000***	-0.0000***
$ \begin{array}{llllllllllllllllllllllllllllllllllll$		(0.000)	(0.000)	(0.000)	(0.000)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Imm*Less than high school*YSM1	0.0006	0.0046***	0.0049***	0.0060***
$ \begin{array}{llllllllllllllllllllllllllllllllllll$		(0.001)	(0.001)	(0.001)	(0.001)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Imm*Less than high school*YSM2	-0.0003	0.0003	0.0006***	0.0002
Imm*Less than high school*YSM3 $-0.0005$ $-0.0003$ $-0.0001$ $-0.0002$ Imm*High school*YSM1 $-0.0010^{***}$ $0.0021^{***}$ $0.003^{***}$ $0.003^{***}$ Imm*High school*YSM2 $-0.0011^{***}$ $0.0000$ $(0.000)$ $(0.001)$ Imm*High school*YSM3 $-0.0001^{****}$ $0.0000$ $(0.000)$ $(0.000)$ Imm*High school*YSM3 $-0.0008^{***}$ $-0.0003$ $-0.0000$ $-0.0004$ Imm*More than high school*YSM1 $-0.0026^{***}$ $0.0098^{***}$ $0.0029^{***}$ $0.0082^{***}$ Imm*More than high school*YSM2 $-0.0001$ $0.0009^{***}$ $0.0029^{***}$ $0.0011^{***}$ Imm*More than high school*YSM2 $-0.0001$ $0.0030^{***}$ $0.0029^{***}$ $0.0011^{***}$ Imm*More than high school*YSM3 $0.0008$ $0.0021^{***}$ $0.0029^{***}$ $0.0011^{***}$ Observations $18,745,201$ $18,745,201$ $18,745,201$ $18,745,201$ $18,745,201$ R-squared $0.385$ $0.880$ $0.887$ $0.919$		(0.000)	(0.000)	(0.000)	(0.000)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Imm*Less than high school*YSM3	-0.0005	-0.0003	-0.0001	-0.0002
Imm*High school*YSM1 $-0.0010^{**}$ $0.0021^{***}$ $0.003^{***}$ $0.0036^{***}$ Imm*High school*YSM2 $-0.0011^{***}$ $0.000$ $(0.000)$ $(0.001)$ Imm*High school*YSM3 $-0.0011^{***}$ $0.0000$ $(0.000)$ $(0.000)$ Imm*More than high school*YSM1 $-0.0026^{***}$ $0.0098^{***}$ $0.0022^{***}$ Imm*More than high school*YSM2 $-0.0026^{***}$ $0.0098^{***}$ $0.0102^{***}$ Imm*More than high school*YSM2 $-0.0001$ $(0.001)$ $(0.001)$ $(0.000)$ Imm*More than high school*YSM2 $-0.0001$ $0.0030^{***}$ $0.0029^{***}$ $0.0011^{***}$ Imm*More than high school*YSM3 $0.0008$ $0.0021^{***}$ $0.0029^{***}$ $0.0011^{***}$ Imm*More than high school*YSM3 $0.0008$ $0.0021^{***}$ $0.0021^{***}$ $0.0014^{***}$ (0.000) $(0.001)$ $(0.001)$ $(0.000)$ $(0.001)$ $(0.001)$ Imm*More than high school*YSM3 $0.0028$ $0.0021^{***}$ $0.0014^{***}$ (0.001) $(0.001)$ $(0.000)$ $(0.001)$ $(0.001)$ Observations $18,745,201$ $18,745,201$ $18,745,201$ $18,745,201$ R-squared $0.385$ $0.880$ $0.887$ $0.919$		(0.001)	(0.001)	(0.000)	(0.001)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Imm*High school*YSM1	-0.0010**	0.0021***	0.003***	0.0036***
Imm*High school*YSM2 $-0.0011^{***}$ $0.0000$ $0.0005^{***}$ $-0.0003$ Imm*High school*YSM3 $-0.0008^{**}$ $-0.0003$ $(0.000)$ $(0.000)$ $(0.000)$ Imm*More than high school*YSM1 $-0.0026^{***}$ $0.0098^{***}$ $0.0102^{***}$ $0.0082^{***}$ Imm*More than high school*YSM2 $-0.0001$ $(0.001)$ $(0.000)$ $(0.001)$ Imm*More than high school*YSM2 $-0.0001$ $0.0030^{***}$ $0.0029^{***}$ $0.0011^{***}$ Imm*More than high school*YSM3 $0.0008$ $0.0021^{***}$ $0.0021^{***}$ $0.0011^{***}$ Imm*More than high school*YSM3 $0.0008$ $0.0021^{***}$ $0.0021^{***}$ $0.0014^{***}$ Observations $18,745,201$ $18,745,201$ $18,745,201$ $18,745,201$ $18,745,201$ R-squared $0.385$ $0.880$ $0.887$ $0.919$		(0.000)	(0.000)	(0.000)	(0.001)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Imm*High school*YSM2	-0.0011***	0.0000	0.0005***	-0.0003
Imm*High school*YSM3       -0.0008***       -0.0003       -0.0000       -0.0004         (0.000)       (0.000)       (0.000)       (0.000)       (0.000)         Imm*More than high school*YSM1       -0.0026***       0.0098***       0.0102***       0.0082***         (0.001)       (0.001)       (0.000)       (0.001)       (0.001)       (0.000)         Imm*More than high school*YSM2       -0.0001       0.0030***       0.0029***       0.0011***         (0.000)       (0.000)       (0.000)       (0.000)       (0.000)         Imm*More than high school*YSM3       0.0008       0.0021***       0.0014**         (0.001)       (0.001)       (0.000)       (0.001)         Observations       18,745,201       18,745,201       18,745,201         R-squared       0.385       0.880       0.887       0.919		(0.000)	(0.000)	(0.000)	(0.000)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Imm*High school*YSM3	-0.0008**	-0.0003	-0.0000	-0.0004
Imm*More than high school*YSM1       -0.0026***       0.0098***       0.0102***       0.0082***         (0.001)       (0.001)       (0.000)       (0.001)         Imm*More than high school*YSM2       -0.0001       0.0030***       0.0029***       0.0011***         (0.000)       (0.000)       (0.000)       (0.000)       (0.000)         Imm*More than high school*YSM3       0.0008       0.0021***       0.0021***       0.0014**         (0.001)       (0.001)       (0.000)       (0.001)       (0.001)         Observations       18,745,201       18,745,201       18,745,201       18,745,201         R-squared       0.385       0.880       0.887       0.919		(0.000)	(0.000)	(0.000)	(0.000)
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	Imm*More than high school*YSM1	-0.0026***	0.0098***	0.0102***	0.0082***
Imm*More than high school*YSM2       -0.0001       0.0030***       0.0029***       0.0011***         Imm*More than high school*YSM3       (0.000)       (0.000)       (0.000)       (0.000)         Imm*More than high school*YSM3       0.0008       0.0021***       0.0021***       0.0011***         Observations       18,745,201       18,745,201       18,745,201       18,745,201         R-squared       0.385       0.880       0.887       0.919		(0.001)	(0.001)	(0.000)	(0.001)
Imm*More than high school*YSM3 $(0.000)$ $(0.000)$ $(0.000)$ $(0.000)$ $(0.001)$ $0.0021^{***}$ $0.0021^{***}$ $0.0014^{**}$ $(0.001)$ $(0.001)$ $(0.000)$ $(0.001)$ Observations $18,745,201$ $18,745,201$ $18,745,201$ R-squared $0.385$ $0.880$ $0.887$ $0.919$	Imm*More than high school*YSM2	-0.0001	0.0030***	0.0029***	0.0011***
Imm*More than high school*YSM3       0.0008       0.0021***       0.0021***       0.0014**         (0.001)       (0.001)       (0.000)       (0.001)         Observations       18,745,201       18,745,201       18,745,201       18,745,201         R-squared       0.385       0.880       0.887       0.919		(0.000)	(0.000)	(0.000)	(0.000)
(0.001)(0.001)(0.000)(0.001)Observations18,745,20118,745,20118,745,20118,745,201R-squared0.3850.8800.8870.919	Imm*More than high school*YSM3	0.0008	0.0021***	0.0021***	0.0014**
Observations18,745,20118,745,20118,745,20118,745,201R-squared0.3850.8800.8870.919		(0.001)	(0.001)	(0.000)	(0.001)
R-squared 0.385 0.880 0.887 0.919	Observations	18,745,201	18,745,201	18.745.201	18,745,201
	R-squared	0.385	0.880	0.887	0.919

*Note:* Estimations performed using WSSD data. Robust standard errors in parentheses \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1. Standard errors in column 1, 2, and 4 are clustered at the level of the individual. Standard errors in double FE estimations are obtained through bootstrapping (30 repetitions). YSM parameters are estimated as splines, each capturing a ten year period.

	(1)	(2)	(3)	(4)
	Quasi-fixed	Individual-	Double fixed	Match fixed
	effect	fixed	effect	effect
	enteet	effect	onocc	enreet
Corresponds to lines in Panel B in figu	re 7 and figure	: 8		
Less than high school*Age	0.0394***	0.0518***	0.0514***	0.0486***
	(0.001)	(0.001)	(0.001)	(0.001)
Less than high school* $Age^2$	-0.0003***	-0.0005***	-0.0005***	-0.0005***
-	(0.000)	(0.000)	(0.000)	(0.000)
Less than high school* $Age^3$	0.0000***	0.0000***	0.0000***	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)
High school*Age	0.0636***	0.0658***	0.0639***	0.0618***
	(0.000)	(0.000)	(0.000)	(0.000)
High school* $Age^2$	-0.0007***	-0.0007***	-0.0007***	-0.0007***
	(0.000)	(0.000)	(0.000)	(0.000)
High school* $Age^3$	0.0000***	0.0000***	0.0000***	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)
More than high school*Age	0.1375***	0.1270***	0.1175***	0.1156***
	(0.001)	(0.001)	(0.000)	(0.001)
More than high school* $Age^2$	-0.0020***	-0.0016***	-0.0014***	-0.0014***
	(0.000)	(0.000)	(0.000)	(0.000)
More than high school* $Age^3$	0.0000***	0.0000***	0.0000***	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*Less than high school*YSM1	0.0030***	0.0072***	0.0073***	0.0076***
	(0.001)	(0.001)	(0.000)	(0.001)
Imm*Less than high school*YSM2	-0.0019***	-0.0009***	-0.0006***	-0.0009***
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*Less than high school*YSM3	-0.0028***	-0.0029***	-0.0027***	-0.0029***
	(0.000)	(0.000)	(0.000)	(0.001)
Imm*High school*YSM1	-0.0005	0.0029***	0.0034***	0.0036***
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*High school*YSM2	-0.0013***	-0.0000	0.0004***	-0.0001
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*High school*YSM3	-0.0011***	-0.0009***	-0.0008***	-0.0015***
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*More than high school*YSM1	-0.0060***	0.0059***	0.0063***	0.0040***
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*More than high school*YSM2	-0.0008**	0.0026***	0.0025***	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)
Imm*More than high school*YSM3	0.0008	0.0019***	0.0015***	-0.0002
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	18,745,201	18,745,201	18,745,201	18,745,201
R-squared	0.413	0.918	0.926	0.949
. 1				

Table 10: Full regression output; Register based wage measure

*Note:* Estimations performed using WSSD data. Robust standard errors in parentheses \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1. Standard errors in column 1, 2, and 4 are clustered at the level of the individual. Standard errors in double FE estimations are obtained through bootstrapping (30 repetitions). YSM parameters are estimated as splines, each capturing a ten year period.

# Essay 2

# Immigrant entrepreneurship and the origin of bankers

# 1 Introduction

Immigrants from low-income source countries struggle to find employment in many western countries. Self-employment may offer an alternative entry way into the labor market for these individuals (Andersson Joona and Wadensjö, 2008; Blume et al., 2009), but starting a business is not without its own constraints. In particular, small businesses are often liquidity constrained,<sup>1</sup> and the relationship with lenders can therefore be crucial for the start-up and survival of small firms. Banks base their credit decisions on both objective credit scores and subjective assessments of the business idea. This has been suggested to contribute to immigrant- and minority entrepreneurs facing discrimination from lenders (Bates, 1989; Blanchflower, Levine, and Zimmerman, 2003; Cavaluzzo and Cavaluzzo, 1998).

The aim of this essay is to contribute to the literature by studying the role of local co-ethnic bank employees for the probability of immigrants to run their own businesses. To this end, I use a linked employer–employee data set covering the whole Swedish labor market for the years 1987 to 2003, from which I can define the country of origin of both bank employees and small business owners. This analysis adds to our understanding of the relationship between banks and entrepreneurs. It also provides novel evidence of the role of minority representation in a sector where there is room for subjective personal assessments. To the extent that credit is allocated according to factors unrelated to the potential prospects of the business idea, society loses out on investment and individual entrepreneurs lose a potential income source. For this reason it is important to understand which factors determine capital access for small business owners.

Following Bates (1989), a number of articles have tried to explain group differences in credit acquisition for small firm start-ups and survival.<sup>2</sup> A more recent strand of this literature focuses on how social ties between lenders and entrepreneurs affect the accessed credit, with a specific focus on shared ethnicity or language. Using survey data, both Aaronson et al. (2004) and Raijman and Tienda (2010) show that same-language suppliers are important for trade credit.

The credit decision performed by the bank is also similar to decisions made in a range of areas in which there is room for discretion. These types of decisions have been extensively studied by scholars within the field of minority representation, focusing on contexts such as educational administration, the police and the judicial system. Typically, these studies show that minority representation

<sup>&</sup>lt;sup>1</sup> Blanchflower and Oswald (1998), Evans and Jovanovich (1989), Evans and Leighton (1989), and Lindh and Ohlsson (1996, 1998).

<sup>&</sup>lt;sup>2</sup>See for example Blanchflower, Levine, and Zimmerman (2003), Cavaluzzo and Cavaluzzo (1998), Cavaluzzo and Wolken (2005), and Fraser (2009).

is positively related to minority outcomes (see Bradbury and Kellough (2011) for an overview). For Sweden it has also been shown that immigrant managers are more prone to hire immigrant employees than native managers are, indicating a relatively favorable behavior towards the non-majority group from minority managers (Åslund, Hensvik, and Skans, forthcoming).

There are, however, very few previous studies on how representation in banks affects capital allocation. Fisman, Paraviini, and Vig (2012) is the main exception to this. They study the personal loan market in India and make use of a rotation of bank managers between regions, which implies a near random assignment of a customer to either a bank official from the own group or from another group. Their results indicate that in-group bank officials perform better ex ante screening, as in-group borrowers have higher repayment rate as well as larger variance of the loan amount. My aim is to contribute with new knowledge regarding the relationship between entrepreneurs and formal banks in an industrialized western context, focusing on credit access for a group which has been shown to have difficulties in many labor markets.

Banks may employ individuals from a specific country in response to the groups' demand for credit. Therefore OLS estimates imply the existence of an ethnic component in the capital market, but they do not tell us the direction of causality. This, along with individuals potentially sorting into municipalities with good banking and business opportunities, may bias the estimates. In an attempt to remedy this bias, I use a nationwide refugee dispersal policy in place in the second half of the 1980s as a source of exogenous settlement.<sup>3</sup> During this time the refugees who arrived to Sweden were placed in municipalities all over the country, thus providing settlement patterns that are exogenous to the outcome. This means that, for the group of refugees who were placed through this policy, we can estimate the effect of the exogenous exposure to co-ethnic bankers their first year in Sweden, as well as use the initial exposure as an instrument for exposure in later years.

The OLS results show a positive co-variation between co-ethnic bank employees and the probability of being self-employed. Results from specifications using the exogenous arrival year exposure to co-ethnic bankers are in line with the OLS estimates. These results imply that self-employment is facilitated by the exposure to local own-group bank employees. The estimates are largest for low educated, for males and for those residing in metropolitan areas. Highly educated individuals potentially have more relevant experience and credentials which facilitate capital access and entry into the regular labor market. Thus it is likely that the low educated immigrants benefit more from the community's resources in this regard. To ensure that what we estimate is not the effect of a good labor market position of the group at the local level, a wide range of

<sup>&</sup>lt;sup>3</sup>It was in place between 1985 and 1994 but its strictest application was between 1987 and 1991 (see institutional background).

municipal group level covariates are controlled for in the model. A number of sensitivity checks are performed for the whole group and for the group of low educated, and the results are robust to changes in the definitions of the main variables as well as other variables.

The remainder of this essay is structured as follows. The next section describes immigrant entrepreneurship and capital allocation, while section 3 describes the data and the sample, and presents some descriptive statistics. The econometric model is outlined in section 4 and section 5 presents the baseline results. Section 6 outlines an analysis based on initial exogenous variation in coethnic bank representation, and presents results from this analysis. Section 7 presents two extensions of the baseline model, and section 8 presents a sensitivity analysis. Section 9 concludes.

# 2 Capital access for immigrant entrepreneurs

#### 2.1 Lower rates of bank loans

To start-up and run a small business is often entailed with difficulties, many of which are related to the large information asymmetries which complicate banks' credit decisions. The bank clerks have imperfect knowledge regarding the productivity of the entrepreneur, and they therefore carefully assess the default risk before granting a loan. In this process the banks use information on past defaults or employment history, as well as less objective measures such as the bank clerks' personal assessment of the business idea (Committee on Foreign Born Entrepreneurship, 1999; Fraser, 2009).

Regardless the evaluation method, there are factors which might affect the capital decision negatively for immigrants. When asked about barriers for running business, immigrants from countries outside of Europe state that they find it difficult to access the necessary capital far more often than natives and European immigrants do (Agency for Economic and Regional Growth, 2007, 2011). However, part of these differences can be explained by factors such as lack of own business capital and an inclination to start firms in over-established sectors. Also, the formal requirements might act as a barrier for recently arrived immigrants who do not have credit histories on record in the host country. It is not clear to what degree these factors explain the larger perceived difficulty of loan access (Committee on Foreign Born Entrepreneurship, 1999).

Even immigrants with strong credentials and relevant experiences from the country of origin might face other difficulties than natives do when applying for a loan. Immigrants may not always have proper information regarding the host country financial system. Furthermore, bank clerks may find it more difficult to evaluate a business idea posed by a person with foreign background, both due to cultural differences and the potential language barrier (Committee on Foreign Born Entrepreneurship, 1999). This uncertainty may lead to statistical discrimination, where immigrants face credit decisions based more on the perceived productivity of the group, than of the individual (Phelps, 1972). In addition, trust built through personal contacts with credit suppliers are important for reducing credit uncertainty (Fraser, 2009) and immigrants' short history within the host country system makes them less likely to have a well-established contact with bank employees.

Potentially as a consequence of the poor access to financial markets, immigrants finance their businesses to a higher degree with capital obtained from family and friends than native entrepreneurs do. Abbasian and Yazdanfar (2012) show, for a sample of entrepreneurs in Sweden, that the share of immigrant entrepreneurs who has used bank loans as a means of start-up finance is 29%, compared to 35% for native born entrepreneurs. A governmental report shows that the main financial sources for small business owners differs substantially depending on country of origin (Agency for Economic and Regional Growth, 2007).<sup>4</sup> Immigrants' loan applications also seem to be rejected far more often than natives' applications. Between 8 and 10 percent (depending on year studied) of native entrepreneurs state that their application has been rejected while for immigrants from refugee countries (i.e. the group I study), between 30-50 % state that they have applied for a loan and been rejected. Interviews with small business owners indicate that some individuals also refrain from applying for loans as they expect to be rejected (Agency for Economic and Regional Growth, 2007).

#### 2.2 Co-ethnic bankers and potential capital access

There are two main reasons why bank employees from the own ethnic group may be positively related to the probability of obtaining capital for small businesses, and hence also for the probability of being self-employed. These individuals may improve information used by both the entrepreneur and the bank officials and provide the trust or contacts needed for a positive credit decision. But they might also make relatively favorable decisions towards members of their own group, what is usually referred to as active representation (Bradbury and Kellough, 2011).<sup>5</sup>

<sup>&</sup>lt;sup>4</sup>Amongst immigrant entrepreneurs from former Yugoslavia, Romania and Poland between 44 and 49 percent have used bank loans as start-up funding, while the same number for entrepreneurs from Turkey, Africa, Iran, Lebanon and Chile is between 31 and 36 percent. Only 26 percent of entrepreneurs from the Middle East seem to have used bank loans when starting up their business.

<sup>&</sup>lt;sup>5</sup>It might also be the case that own-group bank clerks behave the opposite way and reject applications from their own group in fear of being accused of favoritism. Both the positive and negative effect of active representation are likely of very small importance here since the number of own-group bank clerks is small.

Notably, the average share of the bank sector which the own group constitutes is around 0.1%, which indicates that there would be virtually no chance to meet with a co-ethnic bank employee when applying for a loan if the choice of bank office was random. But we do not expect the applicants to randomly approach a bank, instead they are likely turn to the bank where the likelihood of a positive credit decision is the largest.

If there is local representation of the own ethnic group in a local bank office the applicant is expected to turn to the bank where this co-ethnic banker is employed. The likelihood of a positive credit decision at this particular bank may be larger than in other banks regardless if the applicant meets with the owngroup employee or not. The reason for this is that the presence of a minority banker may alter the perception of the particular group among other clerks and thereby affect the decisions made by the whole office (Bradbury and Kellough, 2011).

Furthermore, the information that the potential entrepreneurs have regarding the financial system may be improved by residing nearby a banker from the own country. Once there is someone in the group working in a bank at the local level the community has better access to proper information regarding financial businesses. Again, as co-ethnic bankers are very rare, an individual would most likely not benefit from this knowledge without actively searching for it within the community. Therefore the relevant factor is whether the knowledge regarding banking is available in the community, rather than the number of individuals who will be able to provide this knowledge.

As we expect individuals to act non-randomly when choosing bank office, and when searching for financial information, the explanatory variable of interest will be defined as a dummy variable in the empirical model (see section 3.1). This will indicate whether or not an individual is exposed to at least one coethnic banker locally. This variable will capture both the probability of improved knowledge, and own-group representation, in the local bank, and the improved financial system knowledge within the local group. Thus we will not be able to distinguish between these mechanisms.

As the banks might act upon a local demand from the entrepreneurs through employing individuals from particular countries, we cannot interpret a positive correlation as a causal effect of local co-ethnic bank employees on the probability of being self-employed. Instead a positive OLS estimate should be seen as a market outcome where banks and entrepreneurs act in response to one another. In this process the shared background would function as a mediator for reducing uncertainty and providing information related to credit access.

## 3 Data and sample selection

For this study I use register data over the Swedish labor market for the years 1987 to 2003, where individuals are linked to their workplace through tax records on annual income. In this data I can see if anyone is self-employed.<sup>6</sup> I also have information of the industry of each firm or workplace, from which I can define those working in the bank sector.<sup>7</sup> The data contains information on municipality of residence as well as municipality of workplace, which means that I can study co-ethnic bank employees both residing and working locally.

Apart from this, I also have information on individual characteristics, including country of origin and the immigration year, as well as year of birth, sex, and level of education.<sup>8</sup> I restrict the analysis to the group of non-western immigrants in Sweden, since this group highly overlaps with the group of refugee immigrants for which I can use the placement policy as an source of exogenous settlement. Own-group individuals is quite broadly defined (17 groups) as small source countries are grouped due to confidentiality reasons.<sup>9</sup> Throughout the text I will use the term co-ethnic or own-group individual as a short term for an individual from the same country/region of origin.<sup>10</sup>

I study individuals who were between 18 and 55 when arriving to Sweden, and who arrived between 1987 and 1991 when the placement policy was most effective. I further restrict the sample to individuals who did not have an adult relative already residing in Sweden at the time of arrival, since tied movers were not subject to the governmental placement.<sup>11</sup> This will be referred to as

<sup>&</sup>lt;sup>6</sup>Until 1993 self-employment only included sole traders but from 1993 there is information on both sole traders and limited liability companies. For a further description of the definition of self-employment, see the variable description and figure 3 in the appendix.

<sup>&</sup>lt;sup>7</sup>Industry coding for 1987 and 1988 are imputed from the industry coding in 1989 due to missing information in the earlier years. See graphical explanation of this in the variable description in the appendix.

<sup>&</sup>lt;sup>8</sup>For the individuals immigration before 1990 many have missing educational information the first years, so for everyone for whom there is no data indication of further education acquired in Sweden I impute the early observations with the first observed non-missing educational level.

<sup>&</sup>lt;sup>9</sup>See country of origin definitions in variable description in the appendix, and distribution over countries of origin in table 9 in the appendix.

<sup>&</sup>lt;sup>10</sup>I am aware that the crude country of origin measure that is used here does not by any means capture an individuals' ethnicity, why the term may be misleading. A similar term which could be used would be co-national, but it also has its limitations. It should be noted that, regardless which concept is used, the variable captures relations between individuals who share the same country/region of origin.

<sup>&</sup>lt;sup>11</sup>A comparison is made with Immigration Boards numbers on residence permits for refugee of asylum reasons, and for each of the studied immigration years my data includes a slightly smaller number than those of the immigration board. My sample only includes individuals arriving at an age between 18 and 55, while their numbers include all immigrants. Another comparison is made with Åslund and Fredriksson (2009) who use the same placement policy for identification. For the immigrations years which they study the sample I use here

the main sample, and in an extension, I reestimate the model using a larger set of non-western immigrants.

As outcome variable I use a dummy variable capturing self-employment status at a particular point in time, based on individuals for whom self-employment is the main income source.<sup>12</sup> Both individuals who are unemployed and employed in regular employment are thus defined as not being self-employed. In an extension I also study entry into, and exits from, self-employment.

#### 3.1 Defining the explanatory variable

It follows from the discussion in section 2.2 that the processes determining the access to financial information and the choice of local bank office is likely to be non-random. This means that if there is a local own-group bank representative the applicant might turn to the bank where she is employed. Therefore the key explanatory variable is defined as a dummy, *co-ethnic banker*, indicating the exposure to at least one co-ethnic banker in the municipality of residence.

From the data, the exposure to a co-ethnic banker can be defined based on bankers' municipality of residence, or municipality of work. These alternatives would capture slightly different mechanisms, but defining the variable *co-ethnic banker* based on where bankers live or work yields variables which are highly correlated (0.73).<sup>13</sup> This means that the variable will capture both the possibility of meeting a co-ethnic banker in the community and the possibility of meeting a co-ethnic banker in the local bank, whichever of these definitions is used. For this reason I will perform the baseline analysis on the variable based on bankers' municipality of work as a sensitivity check.<sup>14</sup>

I also perform a robustness check where the explanatory variable is instead defined as the share of the local co-ethnics who are employed in banks.

#### 3.2 Descriptive statistics

Table 1 shows descriptive statistics for the data containing all non-western immigrants arriving to Sweden between 1987 and 1991, for which I can use the

includes slightly less individuals (20,307 compared to 22,556) and the distribution of countries/regions of origin looks very similar. Therefore I conclude that the data used for instrumental variable analysis here is a reasonable approximation of refugee immigrants entering during the studied years.

<sup>&</sup>lt;sup>12</sup>An individual is only coded as self-employed if annual incomes from self-employment is positive.

<sup>&</sup>lt;sup>13</sup>The variables differ mostly in the metropolitan regions, where commuting over municipality borders are common.

<sup>&</sup>lt;sup>14</sup>In commuting regions bank customers might turn to a bank in a neighboring municipality when applying for a loan. If this is the case, it is not certain what the explanatory variable based on bankers' municipality of work captures. With a co-ethnic banker residing in the municipality, an individual might turn to the bank where this banker is employed, even if
government placement as a source of exogenous settlement. Mean values are also presented divided by treatment status; i.e the value of the *co-ethnic banker* variable.

The data consists of a large share of men, due to the exclusion of tied movers. The studied group consists of both low and highly educated individuals. About half of the individuals in the data do not have a positive annual income, and the earnings are low.<sup>15</sup>

Among the studied individuals, 4 percent are self-employed, and the likelihood of self-employment is increasing with time spent in Sweden (see figure 4 in the appendix). The share employed in a bank is about 0.1 percent.<sup>16</sup> Despite the very low share employed in banks, about 46 percent of the individuals have someone from their own country/region of origin residing in the same municipality who works in a bank. This can partly be explained by the fact that the bankers primarily work in larger municipalities, and municipalities where the group constitutes a larger share of the local population. This means that many countrymen are exposed to each banker's presence. 63 percent of those treated were placed in a metropolitan region at arrival to Sweden and this is true for only 36 percent among those non-treated. The treated group resides in municipalities where the own group consists of more highly educated individuals with higher annual incomes.

The appendix shows mean values for the group municipality level covariates which are not shown here. It also shows the distribution over country of origin. The largest groups in the data are individuals from states in the Middle East, Iran, Chile and Eastern Africa.

## 4 Econometric strategy

## 4.1 OLS specification

I estimate the effect of the presence at least one local co-ethnic banker on the self-employment probability. Here *i* is the index for individual, *c* is country of origin, *m* is municipality of residence and *t* is year of observation. I also control for individual characteristics such as level of education when arriving to Sweden, sex and age  $(X_{it})$ . When controlling for both country of origin fixed effects

that bank is located in a neighboring municipality. Therefore defining the dummy by bankers who reside locally is the preferred choice.

<sup>&</sup>lt;sup>15</sup>Description is presented for the year 1996, which is chosen because it is in the middle of the studied period. For most variables the mean values in 1996 is very similar to the mean values in the pooled data, but for the indicator for having positive annual earnings and mean earnings it is clear that 1996 represents a year with very low employment rate and earnings. This has to do with the economic crisis in Sweden in the middle of the 1990s.

<sup>&</sup>lt;sup>16</sup>When calculating the exposure to co-ethnic bankers a sample consisting of all immigrants from non-western countries is used, regardless of their year of arrival or whether they arrived

	All	Co-ethnic	Co-ethnic
		banker=1	banker=0
Age	37.3991	37.2377	37.5373
Share male	0.5829	0.5793	0.5860
Education Missing	0.0124	0.0125	0.0124
No more than 8 years	0.1895	0.1678	0.2080
9 years	0.1854	0.1800	0.1900
2 year high school	0.1548	0.1558	0.1539
3 year high school	0.2024	0.2174	0.1896
University, at most 2 years	0.1339	0.1405	0.1283
University, more than 2 years	0.1099	0.1126	0.1075
PhD	0.0117	0.0134	0.0102
Positive annual earnings	0.5291	0.5418	0.5182
Earnings (1,000 SEK)	20.0761	20.2303	19.9442
Self-employed	0.0400	0.0369	0.0426
Employed in bank	0.0010	0.0017	0.0005
Own-group banker	0.4612	1.0000	0.0000
Placed in metropolitan area	0.4858	0.6338	0.3590
Municipality size (in 1,000s)	149.4033	251.6320	61.8947
Municipal group level covarariates			
Employment rate	0.4959	0.5187	0.4763
Share self-employed	0.0352	0.0364	0.0342
Mean earnings (1,000 SEK)	38.9110	40.9626	37.1548
Share with university education	0.2422	0.2601	0.2269
Share of local population	0.0106	0.0142	0.0076
Share with above median income	0.2650	0.2797	0.2524
Mean years in country	9.6549	10.5599	8.8803
N	53,518	24,683	28,835

Table 1: Mean individual characteristics year 1996

*Note:* Earnings are the mean annual income from work with a zero income for those without income. Deflated to 1980 values. Mean earnings at the group local level is among individuals with a positive annual income. Distribution over countries of origin presented in table 9 in the appendix. The remaining municipal group level covariates are described in table 10 in the appendix.

 $(\theta_c)$  and municipality fixed effects  $(\psi_m)$  the variation used for identification is the variation within groups between municipalities and within municipalities between groups. I also control for year of immigration fixed effects  $(\rho_{t_0})$ , eliminating the bias from different entry cohorts meeting different labor markets at the time of entry, and year fixed effects  $(\pi_t)$ . A control for time-varying municipality size is also added. Standard errors are clustered at the level of the local group, as this is where the main identifying variation comes from.

$$pr(\text{Self-employed})_{it} = \alpha + \beta \text{Co-ethnic banker}_{it} + \gamma X_{it} + \mu V_{cmt} + \theta_c + \psi_m + \rho_{t_0} + \pi_t + \varepsilon_{it}$$
(1)

This analysis hinges on the assumption that local financial market knowledge and banks in the local area matters for credit decisions. The Agency for Economic and Regional Growth (2012) shows that entrepreneurs in remote and low population-density regions perceive capital access as a larger difficulty than entrepreneurs in larger cities, which implies that capital is to some extent supplied at the local level. This supports the claim that local banking is important for the capital acquisition for small firms.

In the baseline analysis I focus on the probability of being self-employed at a particular point in time. I will also explore the dynamics of the self-employment entry and exit in an extension.

#### 4.2 Municipal group level covariates

It is possible that *co-ethnic banker* will capture something more than the pure effect of improved access to banks and financial knowledge on self-employment probability. This would be the case if there is some omitted variable which is correlated both to the self-employment rate and the representation of the group in the banks, for example the groups' local labor market position. To rule out this possibility I control for an extensive set of covariates at the level of the local group  $(V_{cmt})$ .<sup>17</sup>

A number of variables are included to control for the labor market position of the group at the local level, such as the employment rate, mean earnings and share with income above the median. I also control for share with university education. But as these variables might be imperfect measures of how well the group is integrated at the local labor market, I also control for the mean income

as tied movers.

<sup>&</sup>lt;sup>17</sup>These variables are, just as the *co-ethnic banker* variable, defined in a sample consisting of all immigrants from non-western countries, regardless of their year of arrival or whether they arrived as tied movers. These variables are aggregated over each individual's co-ethnics who reside, or work, in the individual's municipality of residence. This means that even though these variables capture characteristics of the group at the local level, each individual has their

of, and share of university educated at, the workplaces where the co-ethnics of an individual work.

It should be noted here that even though ethnic businesses only constitute a small fraction of immigrant businesses in Sweden (Committee on Foreign Born Entrepreneurship, 1999) it is likely that ethnic businesses will start up as a response to a demand for ethnic goods. A high share of the co-ethnics working in the formal bank sector might imply high purchasing power of the group, which could possibly be correlated with a high demand for ethnic products. If this is the case, we could see that a high share of co-ethnics in the bank sector positively correlates with the probability of becoming self-employed due to the omitted demand variable. But this risk is reduced by the inclusion of group local level income variables, as discussed above. The demand for ethnic goods is also potentially larger among more recent immigrants, which means that another way to control for ethnic demand is to include controls for the mean time spent in Sweden in the local community.

As banks might employ clerks from a specific country in response to the credit demand from this specific group it is useful to control for the share of the individuals from ones' own country who are themselves self-employed. The own-group members defining this variable includes all immigrants from the same country in the municipality, except for the individuals in the main sample. The individuals in the main sample are excluded as this control variable otherwise will be a group municipality level aggregation of the outcome variable.<sup>18</sup> The group might also be represented in the local bank only because the group constitutes a large share of the population in the municipality. For this reason I control for the share of the population which the own group constitutes at the local level. For a full description of the covariates included, see variable list and table 10 in the appendix.

# 5 OLS results

Table 2 presents the baseline OLS estimates for the probability of being selfemployed in a given time period t. The estimate for the whole group is presented in column 1, and the remaining columns present results for different

own value for these variables.

<sup>&</sup>lt;sup>18</sup>The results are robust to both excluding this control variable, and to defining it by all co-ethnics (except for the own observation, as with the other group municipality level variables). As this control variable controls away for one possible mechanism through which co-ethnic bankers might affect an individuals' self-employment probability, namely through the groups' entrepreneurial activity, the least conservative estimates are attained if not including this control at all. The most conservative estimates are obtained when aggregating self-employment over all local co-ethnics. To avoid that the control variable captures the same thing as the outcome variable, my preferred choice is to include this control variable, but to define if by the individuals from the own group who are not themselves in the main sample.

subsamples. Column 1 shows that having a local co-ethnic working in a bank is associated with a 0.6 percentage point increase in the probability of being selfemployed which translates to about a 15% increase in the mean self-employment rate. In this study, both individuals who are unemployed and have a regular employment are coded as not being self-employed, why an increasing selfemployment rate should not be interpreted as an increased employment rate in the group.

We see that for those individuals who arrived to Sweden with at most high school education (column 2 and 3) co-ethnic bank employees seem to matter for the self-employment probability. For the least educated the size of the estimate indicates that the presence of at least one co-ethnic local bank employee is associated with an increase in the self-employment rate of about one fifth of the mean. Similar effects are found for males, while there are no indications of a positive relation for females. The mean self-employment rate among women is 1.7%, and thus much smaller than that of men (5.3%). The variables included in the model also explain a smaller share of the variation in female self-employment rates than it does for men. This indicates that the processes determining male and female self-employment differ and observed individual and group-municipality level factors are poorer determinants of female self-employment.

For individuals with more than high school education, the estimates are small and insignificant. The substantially smaller effect of local co-ethnic bankers on the self-employment probability for the highly educated can potentially be explained by the fact that this group might have better observable characteristics, and aim to start businesses in less over-crowded sectors. Both of these factors may function as signals of high productivity, and will thereby affect the banks' capital allocation decisions even in absence of the help from co-ethnic bank employees.

But the estimate for the highly educated might also be affected by the fact that bank employees do not only facilitate credit access. They may also offer labor market contacts which are beneficial for entry into the regular labor market, and this would counteract the positive effect on the likelihood of being self-employed. As individuals with higher education are more likely to share labor market with the bankers than less educated individuals are, this labor market referral effect is more likely to influence the estimates for this group.

	(1)	(1)	(2)	(3)	(4)	(5)	(9)	(2)
	All	Less than high school	High school	More than high school	Men	Women	Metro	Non-metro
Co-ethnic banker	0.00580***	$0.00840^{***}$	0.00554***	0.000160	$0.00912^{***}$	0.000937	$0.00361^{**}$	0.00700***
	(0.00148)	(0.00188)	(0.00185)	(0.00253)	(0.00203)	(0.00122)	(0.00161)	(0.00213)
Mean outcome	0.038	0.036	0.04	0.039	0.053	0.017	0.035	0.04
Observations R-squared	800,899 0.053	325,430 0.067	303,818 0.053	$171,651 \\ 0.049$	467,226 0.063	333,673 $0.024$	390,613 0.056	410,286 0.057
<i>Note:</i> The table rep errors are clustered municipality- and c	orts estimated $\not{E}$ at the level of ountry of origin	<sup>1</sup> parameters (standa) each municipality c are included as well	rd errors) from ec ountry of origin as controls for ser	pation (1). The out pair. *** p<0.01, * & age and education	come variable is a * p<0.05, * p<0 level. A full set o	l dummy for being 1.1. Fixed effects f municipal group	g self-employed. R for year-, year of level covariates are	obust standard immigration-, e also included.

Ś
Ľ
0
ī
- È-
Ē
.5
ਕ
÷
ŭ
р
зt
et
В
- 5-
10
đ
Ξ
မ
Ε
٦.
5
E
č
ē
-¥
ar
Ê,
U,
Ξ.
-q
et
Ļ
5
0
Ţ
9
3
SC
ă
X
Ľ
ö
$\mathbf{s}$
5
fe
ef
-
ĕ
at
Я
÷Ξ
, S
Ы
ä
ò
-7
al
Η

numerpanty- and country of origin are included as well as controls for sex, age and concation level. A full set of municipal group level as each see list in the appendix for overview of these variables and table 14 in the appendix for full regression output (for columns 1 and 2).

The last two columns show estimates for the individuals who reside in metropolitan and non-metropolitan regions separately.<sup>19</sup> The co-variation between co-ethnic bankers and self-employment rate is twice as large in the non-metropolitan region as in the metropolitan region.

## 6 Analysis based on initial exogenous variation

As discussed earlier the direction of causality of the OLS estimates is not clear, as banks and entrepreneurs act in response to one another. Individuals may also select into municipalities where the group's labor market position is favorable. As the presence of co-ethnic bankers is related to slightly better labor market outcomes of the group overall (see table 1), this selective migration may bias the estimates. But it is not clear which way the bias from this selective migration will affect the estimates, as own-group individuals in the bank sector may be associated with larger possibilities for employment in the regular labor market as well as facilitated credit access. The bias arising from the fact that banks might act in response to entrepreneurs' demands is most likely affecting the estimates positively.

To get closer at a causal understanding of the exposure to co-ethnic bankers and its' effect on self-employment probability I make use of a refugee dispersal policy, which provides arguably exogenous variation in the exposure to coethnic bankers the first year in the country. By studying initial year exposure we can rid the estimates of the potential bias arising from individuals' settlement patterns.

## 6.1 Placement of refugees: the policy

Between 1985 and 1994 refugees arriving to Sweden were subject to a dispersal policy through which the government was responsible for placing them in municipalities which met certain standards. At first, a smaller share of the municipalities were involved. But due to a substantially increased inflow of refugees the years following 1985, the number of municipalities contracted to receive refugees increased and by 1987 almost all Swedish municipalities were involved (Andersson, Musterd, and Robinson, 2003; Borevi and Myrberg, 2010). The placement was based on agreements made with the local municipalities and was designed to disperse the refugees from the metropolitan areas over a larger number of municipalities in Sweden (Borevi and Myrberg, 2010).

<sup>&</sup>lt;sup>19</sup>Actually, for facilitation of comparison with estimates using the initial year exogenous variation, these subsamples are defined by the municipality at time of arrival. Metropolitan regions are here defined as Stockholm county, Västra Götaland country (the region where the second largest city Gothenburg is located) and Scania county (the region where the third largest city Malmö is located).

The refugees themselves could ask for a residential location, but due to scarcity of available housing in many municipalities the wishes could most often not be fulfilled. During the years 1987 to 1991 the immigration numbers were large and thus the likelihood of being assigned to the preferred municipality was substantially reduced compared to earlier and later years. A reason for this was that priority was given to shortening the time span between receiving residence permit and moving to the placement municipality, rather than to accommodating to the individuals' preferences (The Immigration Board, 1997).

The officials in charge for placement did not meet with the refugees, but based the placement on information from the refugee centers (The Immigration Board, 1997). This means that placement was based primarily on observable factors, and thus we can arguably control for the factors which might have affected placement (Åslund and Fredriksson, 2009).

Many researchers have shown that the dispersal policy did indeed change the initial settlement patterns among the refugees (Andersson, 1998; The Committee on Immigration Policy, 1996). Prior to this reform about 7% of the inflow of non-OECD immigrants settled in the six northernmost counties of Sweden, and by the end of the policy period this fraction was 14 % (Åslund, Edin, and Fredriksson, 2003). During the same period the fraction of the refugees settling in Stockholm decreased.

After initial placement, there was nothing keeping the immigrants from onward migration to a municipality which they preferred better, and non-trivial mobility rates indicate that many did not settle with the municipality of placement (Andersson, Musterd, and Robinson, 2003). Åslund, Östh, and Zenou (2010) find that the immigrants who arrived during this period started out in localities which differed from those of previous immigrants, but over time the characteristics of the localities of residence became more similar. Similarly, Andersson (1998) shows that most secondary migration has been into cities with a larger concentration of the own ethnic group than the reception municipality. Despite these onward migrations rates around 56 percent of the immigrants remained in the municipality of initial placement after 5 years in the country.

The above description indicates that refugees arriving to Sweden between 1987 and 1991 were restricted in their first location of residence. Thus, conditional on observed covariates, the assignment of placement municipality can be treated as exogenous to the outcome.

## 6.2 Modeling the impact of placement

Due to the possibility of controlling for observable characteristics of the refugees, the policy described above can be used to provide exogenous settlement municipalities for the refugees placed during this period. This means that the exposure to co-ethnic bankers the first year in Sweden is arguably exogenous, conditional on observed characteristics, and we can model self-employment probabilities at later points in time as a function of this initial exposure. This rids the estimates of potential bias from residential sorting. Also, when modeling the self-employment probability as a function of earlier years' exposure to co-ethnic bankers, the potential bias due to reverse causality is reduced.

The estimate from this specification can be interpreted as the effect of initial exposure to own-group bank employees on the self-employment rate, and it should therefore be seen as a long term effect. Standard errors are here clustered at the arrival municipality group level, and the initial characteristics of the co-ethnics at the local level are controlled for, instead of the contemporary characteristics.<sup>20</sup> Also municipality of arrival fixed effects are used instead of fixed effects for the contemporary municipality.

Self-employed<sub>it</sub> = 
$$\beta$$
Co-ethnic banker<sub>it0</sub> +  $\gamma X_{it}$  +  $\mu V_{cmt0}$  +  $\theta_c$  +  $\psi_{mt0}$   
+  $\rho_{t0}$  +  $\pi_t$  +  $\varepsilon_{it}$  (2)

For this model to causally estimate the impact of initial year exposure on subsequent self-employment probability, the initial placement has to be randomly assigned, and independent of the potential outcome. As a test of this I predict the propensity to become self-employed based on observable individual level covariates and study the correlation between this propensity and *co-ethnic banker* variable in the arrival municipality the arrival year. This correlation is -0.0498, which implies that those with characteristics associated with a higher propensity to become self-employed are not to a larger degree placed in municipalities where co-ethnic bankers are present than where they are not present. This supports the claim that the settlement patterns is exogenous relative to the ethnic composition in the financial sector.

In contrast to the IV-version of equation (1) (described below), specification (2) can be seen as the direct effect of the instrument on the outcome. From here on I will therefore refer to specification (2) as the reduced form model.

#### 6.3 Instrumental variable analysis

Another way to use the exogenous first year variation in exposure to co-ethnic bankers is as an instrument for the endogenous regressor in the OLS regression of equation (1). This means using the *co-ethnic banker* variable in the arrival municipality the arrival year as instrument for the same variable in the municipality of residence the year of interest. Equation (3) shows the first stage regression

<sup>&</sup>lt;sup>20</sup>As a sensitivity test, I test whether the results are affected by defining these variables only by the individuals who were residing in the arrival municipality at the year of arrival. This means that the individuals who arrived the same year as yourself are not included in the aggregate local group level variables. The results are robust to this restriction.

for the potentially endogenous variable *co-ethnic banker*. The predicted contemporary co-ethnic banker variable is used as explanatory variable in the second stage.

Co-ethnic banker<sub>it</sub> = 
$$\beta$$
Co-ethnic banker<sub>it0</sub> +  $\gamma X_{it}$  +  $\mu V_{cmt0}$  +  $\theta_c$  +  $\psi_{mt0}$   
+  $\rho_{t0}$  +  $\pi_t$  +  $\varepsilon_{it}$  (3)

For the co-ethnic banker dummy in the arrival municipality to be a valid instrument, it must be correlated with the same variable in the municipality of residence the year of interest, conditional on the other covariates. This assumption is testable and results are presented in a first stage analysis. As the strength of the instrument decreases with time spent in the country, this assumption will lose its validity the longer the individual has been in the country.

Furthermore, for the instrumental variable analysis to be valid, the exclusion restriction has to be fulfilled. This means that the co-ethnic bank employees in the arrival municipality the arrival year cannot affect the probability of being self-employed in the year of interest in any other way than through the presence of a co-ethnic bank employee in year t. As co-ethnic bankers the arrival year might be related to own-group bank representation in years between the arrival year and year t, this assumption might not be fulfilled. This could affect self-employment probability in year t through the start-up of a business any of the earlier years.

The potential invalidity of the exclusion restriction makes it difficult to interpret the size of the instrumental variable estimate. Despite this, I will present these estimates alongside estimates from specification (2).

#### 6.4 Results based on initial exogenous variation

Table 3 presents results from estimations using the exogenous initial exposure to co-ethnic bankers as explanatory variable. Here the first panel shows the reduced form estimates of the effect of the co-ethnic bankers in the placement municipality on the contemporary self-employment rate. Panel B shows the IV estimate where the co-ethnic banker variable in the arrival municipality year  $t_0$ is used to instrument the same variable in the municipality of residence the observation year. Panel C presents the first stage estimates, which are large and highly significant throughout.

Here results for the whole sample and for the subsamples for which we saw significantly positive OLS estimates are presented. For individuals with less than high school education, males and those placed in a metropolitan region the estimates are in line with earlier OLS estimates. For these groups we see significantly positive point estimates for the reduced form and/or the IV.

For those with high school education and those placed in a non-metropolitan region, the OLS estimates were significantly positive, but IV and reduced form

estimates are insignificant. This can be seen as indications of that the OLS estimates are driven by residential sorting, and when taking this into account, there is no clear effect of co-ethnic bankers. The slightly smaller reduced form than OLS estimates for the low educated and for the males might be indications of that residential sorting influences the OLS estimates also for these groups. Results for the other subsamples are found in table 12 in the appendix (for these subsamples we see insignificant reduced form and IV estimates).

The placement of refugees over the Swedish municipalities has been used extensively for evaluating peer effects and effects of local community composition on later labor market outcomes. One general finding of these studies is that sorting between localities matters substantially for the conclusions of how the size of the ethnic community affects subsequent labor market and schooling outcomes (Åslund, Edin, and Fredriksson, 2003; Åslund et al., 2011; Åslund and Fredriksson, 2009). When not accounting for residential sorting, the estimates generally point towards worse schooling and labor market outcomes as a consequence of a high concentration of the own-group, while when taking this sorting into account, the opposite conclusion is reached.

The results of this study are not directly comparable to the above mentioned studies, as the exposure to a co-ethnic banker is not a direct measure of the quantity of the local group. The concentration of the local group thus seems to be a factor relevant in the residential mobility decision of immigrants, while it is less likely that residential decisions are made on the basis of the presence of an own-group banker. This might be a reason why residential sorting does not seem to affect these results as much as it does affect results from earlier studies using the same placement policy.

As discussed earlier, the size of the IV estimates may be affected by the fact that the exclusion restriction might not be fully valid. There is reason to believe that bank employees in years in between the arrival year and year t have been influential in the decision to become self-employed, something which could possibly influence the IV estimate. The reduced form estimates tell us that, for low educated, males and those placed in metropolitan regions (insignificant) there is a long term effect on self-employment probability of the exposure to local co-ethnic bankers at arrival to Sweden. The similar magnitudes of the OLS and reduced form estimates may be interpreted as an indication of that identification does not rely on individual variation over time, but rather on variation in the cross-section.

	(1) All	(1) Less than high school	(2) High school	(3) Men	(4) Metro	(5) Non-metro
<b>Panel A.</b> Reduced form	0.00207	0.00562*	1.88e-05	0.00576*	0.00283	0.000189
Co-ethnic banker , year $t_0$	(0.00186)	(0.00294)	(0.00291)	(0.00318)	(0.00244)	(0.00301)
Observations	800,899	325,430	303,818	467,226	390,613	410,286
R-squared	0.047	0.061	0.047	0.055	0.045	0.049
Panel B. IV Co-ethnic banker year t	0.00845 (0.00754)	$0.0219^{*}$ (0.0113)	8.27e-05 (0.0128)	0.0246* (0.0135)	0.0225**** (0.00842)	0.000755 (0.0120)
Observations	800,899	325,430	303,818	467,226	390,613	410,286
R-squared	0.046	0.057	0.047	0.052	0.041	0.049
<b>Panel C.</b> First stage	0.246***	0.256****	0.227***	$0.234^{***}$ $(0.0192)$	0.235***	0.250****
Co-ethnic banker, year $t_0$	(0.0193)	(0.0210)	(0.0212)		(0.0232)	(0.0302)
Observations	800,899	325,430	303,818	467,226	390,613	410,286
R-squared	0.315	0.344	0.297	0.297	0.376	0.186
<i>Note:</i> Panel A reports estimated $\beta$ paramete endogenous regressor co-ethnic banker <sub><i>ti</i></sub> is in variable is a dummy for being self-employed are clustered at the level of municipality of municipality of arrival- and country of origi covariates are also included, see list in the app	ers (standard errors) nstrumented by equa in panel A and B, w arrival and country in are included as w pendix for overview	from equation (2), 1 tion (3). Panel C re- thile it is a dummy fc of origin pairs.*** $p <$ ell as controls for se of these variables.	while Panel B repoi ports estimates froi or the exposure to a c0.01, ** p<0.05, * x, age and educatio	rts estimates from n the first stage re; co-ethnic banker p<0.1. Fixed eff on level. A full se	a version of equati gression, equation in year t in panel C ects for year-, year t of municipal of $\varepsilon$	on (1) where the (3). The outcome Standard errors of immigration-, urival group level

#### 6.5 The importance of local group quality

The literature using the dispersal of refugees over Sweden to study effects of community characteristics on labor market outcomes suggest that the characteristics of group in the placement municipality affects long run labor market outcomes for the placed immigrants (see for example Åslund and Fredriksson (2009) and Åslund et al. (2011)). Åslund, Edin, and Fredriksson (2003) show that the quality of the initial placement community, measured by either group annual earnings or group self-employment rates, is an important factor for determining labor market outcomes in the long run, independent of educational background. For this reason it is important to control for a range of municipal-group level characteristics of the initial placement to ensure that the co-ethnic banker dummy is not capturing other dimensions of the group's attachment to the local labor market.

Table 4 describes how the reduced form estimate changes with the inclusion of placement municipality group level covariates. Results are shown for the whole placement data as well as for only those with the lowest educational attainment, since this is one of the groups for which we see the strongest relation. The first column shows estimates when only controlling for individual level characteristics, as well as municipal level-, year-, year of immigration- and country of origin fixed effects. Column 2 adds controls for the municipality size and the share of the local population which the group constitutes, and this reduces the estimate by 0.1 percentage point for the full sample as well as for the low educated. For the full sample the estimate becomes insignificant.

When in the following columns adding controls for the share of the individuals from the own country who are themselves self-employed, and several other group municipality level covariates the estimates are relatively stable. This means that the co-ethnic banker dummy is not highly correlated with other group municipality level covariates related to the labor market position and the quality of the group at the local level. I therefore draw the conclusion that the estimated effect is not driven by characteristics of the group in the initial location.

	(1)	(2)	(3)	(4)	(5)
Panel A. All	0.00314*	0.00221	0.00208	0.00216	0.00207
Co-ethnic banker	(0.00186)	(0.00189)	(0.00187)	(0.00186)	(0.00186)
Observations	800,899	800,899	800,899	800,899	800,899
R-squared	0.046	0.046	0.047	0.047	0.047
<b>Panel B.</b> Less than high school	0.00684**	0.00590**	0.00556*	0.00558*	0.00562*
Co-ethnic banker	(0.00277)	(0.00297)	(0.00294)	(0.00293)	(0.00294)
Observations	325,430	325,430	325,430	325,430	325,430
R-squared	0.060	0.060	0.060	0.061	0.061
Individual controls, fixed effects Share of local population, municipality size Share self-employed of local co-ethnics Local*group labor market covariates Local*group workplace covariates	Yes	Yes Yes	Yes Yes Yes	Yes Yes Yes	Yes Yes Yes Yes
<i>Note:</i> The table reports $\beta$ parameters (standard errors) from at the level of municipality of arrival and country of ori, arrival- and country of origin are included in all columns,	m equation (2). The outc	ome variable is a du	ummy for being self-	-employed. Standar	rd errors are clustered
	gin pairs. <sup>***</sup> $p<0.01$ , <sup>**</sup> $p$	<0.05, * p<0.1 Fi	ced effects for year-	-, year of immigrat	ion-, municipality of
	, as well as controls for sex	, age and education	level. Column 2 ad	lds controls for the	municipality size and

Column 4 then adds further controls for the earnings and education level of the group at the local level, while column 5 also adds controls for characteristics of the workplaces where the own-group members locally are working. See list in the appendix for overview of these variables. the share of the local population which the group constitutes. Column 3 then adds a control for the share of the group at the local level who are self-employed.

1.0 -5 . 1 1 11. 15 ų ٤ Ľ, E

## 7 Extensions

#### 7.1 Own-group employees of a similar sector

As a test of whether the explanatory variable of interest captures more than the pure capital access effect we are interested in, I reestimate the baseline model adding a control for own-group employees in a similar sector. The sector that I study is the insurance sector. The two sectors are similar in the sense that they require the same amount and type of education and offer similar wages (see table 11 and figure 2 in the appendix). But the insurance market does not provide credit for entrepreneurs. Although we expect insurance company employees to have partial knowledge of the financial sector and potential contacts within it, and hence also potentially influence the capital access process, we expect the effect of co-ethnic bankers to exceed the effect of own-group insurance company employees. As exposure to own-group employees in these sectors is correlated, we include both variables in the model and study the effect of both simultaneously.<sup>21</sup>

Both banks and insurance companies are sectors in which non-western immigrants earn relatively high incomes and those who are employed have relatively high education.<sup>22</sup> Therefore, employment in any of these sectors can be considered a relatively high level of economic integration for a non-western immigrant. As discussed earlier, bank employees might affect the probability to enter self-employment not only through improved credit access but also through affecting the possibilities of employment in the regular labor market. Insurance sector employees from the own country should mainly capture the second of these effects.

Figure 1 in the appendix shows the total number of non-western immigrants employed in the bank and insurance sector during the time period studied. There are much fewer non-western employees in the insurance sector than in the bank sector, mainly because the insurance sector employs much fewer workers overall. The share of employees with a non-western background is similar in the two sectors. In the analysis I focus on a co-ethnic insurance dummy, defined analogously to the co-ethnic banker dummy. The share of individuals who are exposed to at least one own-group insurance employee is 0.31.

Table 5 presents the results from the comparison between the local co-ethnic bank and insurance sector employees, focusing on the whole group and the subsamples for which we found significant results in the previous analysis. I

<sup>&</sup>lt;sup>21</sup>Municipality-year observations where there is no workplace identified as insurance company are not included in this analysis.

<sup>&</sup>lt;sup>22</sup>Some non-western bank employees have relatively low income. These may be individuals employed at maintenance positions, such as janitors and cleaners. As we cannot distinguish these individuals, this may lead to an underestimation of the effect of bank employees who are involved in the credit access process.

also limit the attention to the OLS and reduced form estimates.

The results show that the patterns from the baseline analysis remain, with positive OLS and reduced form estimates of bank representation on the probability of being self-employed, particularly for low educated and for males. Reassuringly, we do not find corresponding OLS results for the effect of the coethnic insurance dummy. The reduced form estimates are significantly *negative*. For the groups not presented, I generally do not find any significant effects of the insurance sector employees.<sup>23</sup>

The negative estimates for the co-ethnic insurance dummy in the reduced form model may indicate that the insurance company employees capture the impact of a well (economically) integrated group. The exposure to such individuals may be positively related to possibilities of regular employment. The fact that the sign is the opposite to the estimate for co-ethnic bankers suggests that the two professions have very different relations to self-employment, despite the similarity in labor market status. In my view, this strengthens the argument that the co-ethnic banker dummy captures access to capital, and not the impact of having co-ethnics in prestigious labor market segments.

## 7.2 Dynamics of self-employment: entry and exit

The baseline analysis has been focusing on the stock of self-employed individuals at a particular point in time. This section replaces the outcome variable *Being self-employed* with the outcomes *Becoming self-employed* and *Leaving self-employment*. These are both defined from the same self-employment variable as before, and the variable for becoming self-employed takes on the value 1 if the individual was not self-employed in the previous year but is self-employed in year t, and 0 if she does not enter self-employment. The variable for leaving self-employment takes on the value 1 if the individual was self-employed the previous year but is not so this year. Similarly, this is set to 0 if she continues to be self-employed.<sup>24</sup>

If access to capital is important both in the start-up and expansion phase and for surviving downturns we could expect to see effects along the extensive margins of both entering and exiting self-employment. Andersson Joona (2010) shows that net wealth at business start-up is an important factor in determining the risk of exiting self-employment and entering unemployment. As is clear from table 6, the local co-ethnic bank employees seem to be related to both selfemployment entry and exit. The probability of starting a firm is much lower

<sup>&</sup>lt;sup>23</sup>For individuals placed outside the metropolitan region we see a significantly positive OLS estimate (significant on the 10 % level).

<sup>&</sup>lt;sup>24</sup>For all observations where an individual was self-employed the previous year, the variable for becoming self-employed is set to missing, as you cannot start up a firm if you are already running one. For observations where an individual was not self-employed the previous year the variable for leaving self-employment is set to missing.

	(1)	(2)
Panel A. All	OLS	Reduced form
Co-ethnic banker	0.00570***	0.00201
	(0.00144)	(0.00186)
Co-ethnic insurance employee	0.00144	-0.00412
	(0.00172)	(0.00256)
Observations	800,899	800,899
R-squared	0.053	0.047
Panel B. Less than high school	OLS	Reduced form
Co-ethnic banker	0.00832***	0.00556*
	(0.00188)	(0.00294)
Co-ethnic insurance employee	0.00113	-0.00714**
	(0.00238)	(0.00328)
Observations	325,430	325,430
R-squared	0.067	0.061
Panel C. Male	OLS	Reduced form
Co. ethnia bankar	0.0000***	0.00557*
	(0.0088)	(0.00337)
Co-ethnic insurance employee	0.00308	-0.010/**
co-cumic insurance employee	(0.00244)	(0.00407)
Observations	467 226	467 226
R-squared	0.063	0.055
	01.0	
Panel D. Metropolitan region	OLS	Reduced form
Co-ethnic banker	0.00361**	0.00249
	(0.00158)	(0.00242)
Co-ethnic insurance employee	6.26e-05	-0.00697**
	(0.00192)	(0.00275)
Observations	390,613	390,613
R-squared	0.056	0.045

 Table 5: Estimated effects of exposure to co-ethnic bankers and insurance employees on self-employment probability: OLS and reduced form

*Note:* Column 1 (2) reports  $\beta$  parameters (standard errors) from a version of equation (1) ( (2)), where both Co-ethnic bankers and own-group insurance company employees are used as explanatory variables. The outcome variable used is a dummy for being self-employed. Standard errors for column 1 (2) are clustered at the level of municipality (of arrival) and country of origin pairs.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1 Fixed effects for year-, year of immigration-, municipality-and country of origin are included, as well as controls for sex, age and education level. In column 1 municipal group level covariates are contemporary, and in column 2 the initial year value is used. In column 2 also municipality of arrival fixed effects are used instead of contemporary municipality fixed effects. See list in appendix for overview over these variables.

than the probability of leaving self-employment once having started the firm. The large differences between the magnitude of the estimates for starting up a firm and leaving self-employment are related to the different mean probabilities. Both the effect on entry and the effect on exit is about one tenth of the mean probability, with slightly larger estimates for the low educated. This suggests that local co-ethnic bank employees are related to both facilitated entry into, and reduced exit out of, self-employment.<sup>25</sup>

There are, however, two reasons to be cautious when interpreting these results. As we saw earlier, the effect of initial exposure is almost the same size as the effect of contemporary exposure. This indicates that the effect may materialize with lags, and the transition estimates are identified by contemporary effects. Furthermore, the analysis on the dynamics of self-employment is also performed including insurance sector employees as an explanatory variable. Here we see similar estimates for the insurance sector employees as we do for the bank employees (see table 13 in the appendix).

	А	11	Less than h	igh school
	(1) Entry	(2) Exit	(3) Entry	(4) Exit
Co-ethnic banker	0.00146** (0.000642)	-0.0180* (0.00948)	0.00190** (0.000754)	-0.0335* (0.0189)
Mean outcome	.013	.228	.012	.234
Observations	774,251	26,495	315,157	10,192
R-squared	0.013	0.041	0.018	0.059

 Table 6: Estimated effect of co-ethnic banker on probability of entering and exiting self-employment, OLS

*Note:* The table reports  $\beta$  parameters (standard errors) from equation (1). The outcome variable used in column 1 and 3 (2 and 4) is a dummy for becoming self-employed (exiting self-employment). Standard errors are clustered at the level of municipality and country of origin pairs.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1 Fixed effects for year-, year of immigration-, municipality-and country of origin are included, as well as controls for sex, age and education level. Municipal group level covariates are contemporary, see list in appendix for overview over these variables.

# 8 Sensitivity

Table 7 shows OLS estimates for a wide range of sensitivity checks of the baseline results. Only results for the whole group and for those who arrived with

<sup>&</sup>lt;sup>25</sup>IV estimates are in line with these OLS estimates but standard errors are large so they are not

low education are presented.

As noted earlier the local co-ethnic bank employees can be defined based on their municipality of residence or municipality of work. The co-ethnic banker dummy variables based on either of these definitions are highly correlated. As a sensitivity analysis I define the local co-ethnic bank variable by banker's municipality of work. OLS estimates are presented in column 1 of table 7 and these estimates are very similar in magnitude to the estimates in the baseline analysis.<sup>26</sup>

The second column shows estimates using a variable defined as the *Share of local co-ethnics employed in bank*. The size of these estimates suggest that self-employment probability increases by 0.43 percentage points if the share of the local own-group members who work in a bank increases with one percentage point.

In column 3 I study the full sample of grown up non-western immigrants arriving after 1986. The positive OLS estimate is robust to using this data, which tells us that the individuals in the placement data and the individuals who arrived as tied movers do not differ substantially from each other in this regard. The estimates here are slightly smaller, something which could be due to the larger share of women in this data set.

Column 4 shows estimates when using logit instead of OLS and results are robust to this as well. The log odds of being self-employed, versus not being self-employed, for low educated, increase by a factor of 0.2 if exposed to a coethnic banker. This can be translated into that being exposed to a coethnic banker increases the odds of being self-employed by 22 %.

So far the analysis has relied on the education level at time of arrival to Sweden, but in column 5, I study how the results change when using contemporary education levels, in column 6 I use a self-employment definition only including sole traders, but not limited liability companies.<sup>27</sup> Lastly, column 7 checks to see if the results are robust to using firm level industry codes instead of the workplace level industry codes, which have been used in the baseline analysis. OLS estimates are robust to all these alterations.

significant.

<sup>&</sup>lt;sup>26</sup>Using this definition of local co-ethnic employees OLS estimates are more similar for bank employees and insurance company employees. But the estimates for the insurance company employees generally turn significantly negative in the IV and reduced form estimations, as in table 5. This definition is not used as baseline as it is not completely clear what it captures. Using this definition, exposure is mainly taking place in the larger cities, even though it might be the case that residents of commuting municipalities around the cities are also turning to these in-city bankers when applying for loans. This definition of the explanatory variable does not distinguish the effect of bank employees from the effect of employees in similar sectors, and thus it seems to capture the effect of the group's local labor market position rather than the effect of local exposure to co-ethnic bankers.

<sup>&</sup>lt;sup>27</sup>Figure 3 in the appendix show the development of the number of individuals in the data who are self-employed over time, using these two different definitions.

Panel A. All	(1)	(2)	(3)	(4)	(5) 	(9)	(2)
	Banker dummy by municipality of work	Share	Fùll data	Logit	Contemporary Education	Sole traders	Firm level code
Co-ethnic banker	$0.00527^{***}$ (0.00190)		$0.00212^{**}$ (0.000871)	0.0992*** (0.0384)	$0.00591^{***}$ (0.00149)	$0.00602^{***}$ (0.00148)	$0.00520^{***}$ (0.00145)
Share of own-group members in bank		0.439*** (0.111)					
Observations R-squared	800,899 0.053	800,899 0.049	2,536,065 0.042	798,654	791,278 0.053	801,126 0.050	800,899 0.053
Panel B. Less than high school							
Co-ethnic banker	0.00710*** (0.00201)		0.00329*** (0.00123)	0.196*** (0.0560)	$0.00914^{***}$ (0.00193)	0.00857*** (0.00193)	0.00756*** (0.00182)
Share of own-group members in bank		0.389** (0.175)					
Observations R-squared	325,430 0.067	325,430 0.067	825,282 0.058	320,677	287,731 0.072	325,550 0.065	325,430 0.067
<i>Note:</i> The table report: uses an explanatory var bank. Column 3 perfor: Column 5 uses the cont variables where banks ai *** p<0.01, ** p<0.05, *	s estimated $\beta$ parameter iable based on bankers' ms the baseline OLS est emporary education lew re defined by industry cc " $p<0.1$ . Fixed effects an	ts (standard error municipality of ' timation on a larg el, and column 6 ode at the firm le' od controls are the	s) from equation (1 work, and column 2 cer sample of non-we uses a self-employm vel. Robust standard e same as in earlier C	). The outcome v uses the share of stern immigrants. ent definition whi errors are clustere DLS estimations.	ariable is a dummy fr the own group at the Column 4 estimates ch only includes sole t ed at the level of each i	or being self-emp local level who equation (1) as a raders. Column municipality-cou	oloyed. Column 1 are employed in a logistic regression. 7 uses explanatory ntry of origin pair.

Overall, these robustness checks suggest that the baseline results for the full group and for the low educated are robust to a wide range of alterations of how the variables are defined and the way the model is specified.<sup>28</sup>

## 9 Concluding remarks

By studying how the self-employment rate of immigrants is related to exposure to co-ethnic bankers, we derive knowledge about whether shared background can help to overcome the information asymmetries which are associated with capital access for small firms. OLS estimates suggest that exposure to coethnic bank employees is positively related to the self-employment probability amongst immigrants, but also that this relationship differs between groups. We see a strong relation for low educated and for males, while we see no indications of positive effects for the highly educated and for women.

The differences between the groups become even clearer when employing a placement policy as a source of exogenous settlement to get closer at a causal understanding of the relation. By using this policy we rid the estimates of potential bias from selective residential sorting and reverse causality. The reduced form estimates for low educated, for males and for those placed in a metropolitan regions are in line with the earlier OLS estimates, indicating that there is a positive relation even after controlling for residential sorting. For high school educated and those placed in a non-metropolitan region, reduced form estimates are insignificant while the OLS estimates are significantly positive. This can be understood as signs of sorting into municipalities with own-group bank representation as well as beneficial business opportunities.

The low educated consistently show larger positive effects of bank representation than the more highly educated do. This indicates that with the lack of formal education, capital access is greatly facilitated by co-ethnic bank contacts. The size of the OLS estimate means that exposure to at least one local co-ethnic bank employee increases self-employment probability by around 20 %. This could be due to the improved knowledge, or relatively favorable behavior, at the local bank where immigrants apply for loans. But it could also be an effect of the increased knowledge of the local financial system. This study cannot distinguish between these different mechanisms.

A wide range of sensitivity analyses have been performed, on the full sample and on the low educated, to make sure that the baseline analysis captures the relation between self-employment rate and co-ethnic bankers. The results are

<sup>&</sup>lt;sup>28</sup>Some of these estimations are also performed in the reduced form model (results not presented here). For columns 1 and 5 reduced form estimates are significantly positive for the low educated, and for column 2 they are significant also for the full sample. For the remaining tests for the full sample, and for column 6 and 7 for the low educated, reduced form results are found to be positive but insignificant.

robust to changes in the definition of the explanatory variable and the outcome variable, as well as to changes in the model specification. This gives strong indications that the results are not driven by the specification and that the relation I estimate is stable. The model is also estimated studying both employees of the bank and the insurance sector, as these sectors are similar in many dimensions but differ in terms of supplying credit. The baseline results are robust to this, as they show much less of a relation between insurance company employees and self-employment than between bank employees and self-employment. Overall, this comparison suggests that the effect of co-ethnic bank employees differs from the generic effect on labor market outcomes of co-ethnics in well integrated sectors.

The results reported in this essay measure the effect of co-ethnic bankers on the self-employment probability. Without a measure of the actual capital accessed by the entrepreneurs, these measures serve as a proxy for the impact of shared background on capital access. Interesting avenues for future research would be to directly analyze the relationship between co-ethnic bankers and sources of finances. Due to data limitations, this has not been possible within the scope of this essay.

This essay has shown that the origin of bankers matters for the probability of immigrants from non-western countries to run small businesses. It does not address the question of whether the facilitated capital access for immigrants is improving efficiency in the credit market. But future studies on survival rates, and characteristics of firms started, could add to the understanding of the ethnic component in the credit market, highlighted by the analyses presented above.

## References

- Aaronson, D., R. Bostic, P. Huck, and R. Townsend. 2004. "Supplier relationships and small business use of trade credit". *Journal of Urban Economics* 55, 46–67.
- Abbasian, S. and D. Yazdanfar. 2012. "The Impact of Owner and Firm Characteristics on External Capital Acquisition at Start-up: Empirical Evidences from Swedish Data". *International Business Research* 5, 19–30.
- Agency for Economic and Regional Growth. 2007. "Finansieringssituationen vid företagande for utrikes födda kvinnor och män".
- ——— 2011. "Företagare med utändsk bakgrund".
  - 2012. "Företagande i Sveriges regioner- företagens villkor och verklighet 2011".
- Andersson Joona, P. 2010. "Exits from Self-Employment: Is There a Native-Immigrant Difference in Sweden?" *International Migration Review* 44, 539– 559.
- Andersson Joona, P. and E. Wadensjö. 2008. "A Note on Immigrant Representation in Temporary Agency Work and Self-employment in Sweden". *Labour* 22, 495–507.
- Andersson, R. 1998. "Socio-Spatial Dynamics: Ethnic Divisions of Mobility and Housing in post-Palme Sweden". Urban Studies 35, 397–428.
- Andersson, R., S. Musterd, and V. Robinson. 2003. Spreading the 'burden'? A review of policies to disperse asylum seekers and refugees. The Policy Press, Bristol, UK.
- Åslund, O., P-A. Edin, and P. Fredriksson. 2003. "Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment". *Quarterly Journal of Economics* 118, 329–357.
- Åslund, O., P-A. Edin, P. Fredriksson, and H. Grönqvist. 2011. "Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy". *American Economic Journal: Applied Economics* 3, 67–95.
- Åslund, O. and P. Fredriksson. 2009. "Peer Effects in Welfare Dependence: Quasi-Experimental Evidence". *Journal of Human Resources* 20, 798–825.
- Aslund, O., L. Hensvik, and O. Nordström Skans. Forthcoming. "Seeking similarity: How immigrants and natives manage at the labor market". *Journal of Labor Economics*.

- Åslund, O., J. Östh, and Y. Zenou. 2010. "How important is access to jobs? Old question—improved answer". *Journal of Economic Geography* 10, 389– 422.
- Bates, T. 1989. "The Changing Nature of Minority Business: A Comparative Analysis of Asian, Non-Minority and Black-Owned Businesses". *The Review* of Black Political Economy 18, 25–42.
- Blanchflower, D., B. Levine, and D. Zimmerman. 2003. "Discrimination in the Small Business Credit Market". *The Review of Economics and Statistics* 5, 930–943.
- Blanchflower, D. and A. Oswald. 1998. "What Makes an Entrepreneur?" *Journal of Labor Economics* 16, 26–60.
- Blume, K., M. Ejrnaes, H. Skyt Nielsen, and A. Wurtz. 2009. "Labor market transitions of immigrants with emphasis on marginalization and selfemployment". *Journal of Population Economics* 22, 881–908.
- Borevi, K. and G. Myrberg. 2010. "Välfardsstaten och de nyanlända: En flyktingplaceringspolitisk probleminventering". MIM Working Paper 3.
- Bradbury, M. and J. E. Kellough. 2011. "Representative Bureaucracy: Assessing the Evidence on Active Representation". *The American Review of Public Administration* 41, 157–167.
- Cavaluzzo, K. and L. Cavaluzzo. 1998. "Market Structure and Discrimination: The Case of Small Businessess". *Journal of Money, Credit and Banking* 30, 771–792.
- Cavaluzzo, K. and J. Wolken. 2005. "Small Business Loan Turndowns, Personal Wealth, and Discrimination". *The Journal of Business* 78, 2153–2178.
- Committee on Foreign Born Entrepreneurship. 1999. Invandrare som företagare- för lika möjligheter och ökad tillväxt- Betänkande av Utredningen om företagande för personer med utländsk bakgrund. SOU 1999:49. Stockholm fritzes.
- Evans, D. and B. Jovanovich. 1989. "An Estimated Model of Entrepreneurial Choice under Liquidity Constraint". *Journal of Political Economy* 97, 808– 827.
- Evans, D. and L. Leighton. 1989. "Some Empirical Aspects of Entrepreneurship". *The American Economic Review* 79, 519–535.
- Fisman, R., D. Paraviini, and V. Vig. 2012. "Cultural proximity and loan outcomes". NBER Working Paper 18096.

- Fraser, S. S. 2009. "Is there Ethnic Discrimination in the UK Market for Small Business Credit?" *International Small Business Journal* 27, 583–607.
- Lindh, T and H. Ohlsson. 1996. "Self-employment and Windfall gains: Evidence from the Swedish Lottery\*". *The Economic Journal* 106, 1515–1526.
  - ------ 1998. "Self-employment and Wealth Inequality". *Review of Income and Wealth* 44, 25–41.
- Phelps, E. S. 1972. "The Statistical Theory of Racism and Sexism". *The Ameri*can Economic Review 62, 659–661.
- Raijman, R. and M. Tienda. 2010. "Ethnic foundations of economic transactions: Mexican and Korean immigrant entrepreneurs in Chicago". *Ethnic and Racial Studies* 26, 783–801.
- The Committee on Immigration Policy. 1996. *Sverige, framtiden och mångfalden*. SOU 1996:55. Stockholm fritzes.
- The Immigration Board. 1997. Individuell mångfald: Invandrarverkets utvärdering och analys av det samordnade flyktingmottagandet 1991–1996. Norrköping, Statens invandrarverk.

# Appendix A Variable description

- Bank Bank contains all firms defined as banks, but not other financial service companies. Industry coding is available from 1989 and onwards, so for the years 1987 and 1988 firms are given the industry coding they have 1989. This yields a possibility that firms that closed down before 1989 are not captured. Industry coding is also available at both firm and workplace level. Here I use workplace level industry coding in the baseline analysis but employ firm level industry coding in a sensitivity analysis.
- **Insurance sector** This includes insurance companies mainly focusing on injury insurances. The industry coding that it builds on is the same as the bank sector above.
- **Self-employed** Before 1993 only sole traders were counted as self-employed, and after 1993 also limited liability companies are included in this definition. For the baseline analysis I use a variable which uses the first definition the years before 1993 and after that the second definition. I furthermore only include self-employed with a positive annual income, and individuals for which self-employment income is the main income.
- **Countries/regions of origin** Some countries have a code of its own, as for example for Iran, while the whole of Latin America is categorized into 2 different codes. I use the 17 different countries/regions listed below.

Table 8: Countries of origin

-Bosnia and Herzegovina

-Former Yugoslavia, Croatia, Macedonia, Slovenia (Balkan States)

-Poland, Eastern Germany

-Estonia, Lithuania, Latvia (Baltic states)

-Romania, Former USSR, Bulgaria, Albania, Armenia, Azerbaijan, Georgia, Kazakhstan, Kirghizia, Moldova, Russia, Tadzhikistan's, Turkmenistan, Ukraine, Uzbekistan, Belarus (Eastern Europe)

-Former Czechoslovakia, Hungary

-Mexico, El Salvador, Other countries in the region (Central America)

-Chile

-Peru, Brazil, Columbia, Argentina, Uruguay, Other countries in the region (Latin America)

-Ethiopia, Eritrea, Somalia, Sudan, Djibouti (Eastern Africa)

-Lebanon, Syria, Morocco, Tunisia, Egypt, Algeria, Israel, Palestine, Jordan, Other countries in the region (Middle East)

-Gambia, Uganda, Zaire, Ghana, Other countries in the region (Africa South of Sahara)

-Iran

-Iraq

-Turkey

-Fillipines, Indonesia, Malaysia, Singapore, Thailand, Vietnam, Other countries in the region (Southeast Asia)

-Afghanistan, Bangladesh, Bhutan, India, Mongolia, Pakistan, Other countries in the region (East Asia)

	All	Own-group banker=1	Own-group banker=0	
Bosnia & Hercegovina	0.0007	0.000	0.001	
Balkan States	0.0409	0.055	0.029	
Poland	0.0338	0.042	0.027	
Baltic states	0.0031	0.005	0.001	
Eastern Europe	0.0681	0.068	0.069	
Former Czechoslovakia	0.0224	0.025	0.020	
Central America	0.0139	0.005	0.022	
Chile	0.1036	0.111	0.097	
Latin America	0.0237	0.032	0.017	
Eastern Africa	0.1011	0.128	0.078	
Middle East	0.1496	0.102	0.191	
Africa South of Sahara	0.0207	0.021	0.020	
Iran	0.2480	0.272	0.228	
Iraq	0.0644	0.026	0.097	
Turkey	0.0343	0.038	0.031	
Southeast Asia	0.0457	0.032	0.057	
East Asia	0.0259	0.037	0.016	
N	53,518	24,683	28,835	

Table 9: Country of origin distribution 1996

# Appendix B List of municipal group level covariates

Variables defined by those individuals from the own country who share the individual's municipality of residence

- Employment rate
- Share self-employed
- Share that group constitutes of local population
- Mean earnings
- Share with earnings above the median (defined by year)
- Share with university education
- Mean years since migration
- Share working in high income workplaces (defined as a workplace with median income above the yearly mean income)

Variables defined by mean values of workplaces where the local own-group members work (WP):

• Mean earnings at co-ethnics' workplaces

• Share with university education at co-ethnics' workplaces

Variables defined by those individuals from the own country who work in the individual's municipality of residence (W):

- Mean earnings
- Share with earnings above the median (defined by year)
- Share with university education
- Share working in high income workplaces (defined as a workplace with median income above the yearly mean income)
- Mean earnings at co-ethnics' workplaces
- Share with university education at co-ethnics' workplaces

	All	Co-ethnic banker=1	Co-ethnic banker=0
Share above median income (W)	0.2898	0.2982	0.2826
Mean earnings (1,000 SEK)(W)	41.4301	42.6855	40.3554
Share with university educ(W)	0.2858	0.2998	0.2737
Share working on high income workplaces	0.3364	0.3378	0.3351
Share working on high income workplaces(W)	0.3382	0.3390	0.3376
Mean income (1,000 SEK) (WP)	51.1261	52.8198	49.6762
Mean income (1,000 SEK)(W) (WP)	51.3051	52.9991	49.8550
Share with university educ(WP)	0.2665	0.2973	0.2401
Share with university educ(W) (WP)	0.2675	0.2991	0.2404
N	53,518	24,683	28,835

Table 10: Mean values of remaining municipality group level covariates, 1996

*Note:* (W) Variable defined by individuals who work in the municipality where one lives. (WP) Variable defined by the colleagues of an individuals' local co-ethnics.

# Appendix C Description; bank, insurance and self-employment



Figure 1: Number of non-western employees in bank sector- over time

Note: Industry codes before 1989 are imputed from 1989 years values.

Table 11: Educational distribution among non-western employees in the bank and insurance sector

Education level	Full labor market	Bank sector	Insurance sector
Less than 9 years schooling	15.50	2.77	3.52
9 years schooling	17.85	7.28	7.42
High school, at most 2 years	22.91	25.87	20.48
High school, 3 years	19.62	24.28	26.85
University, at most 2 years	11.31	18.55	19.28
University, at least 3 years	11.72	20.90	21.93
PhD	1.08	0.34	0.52
Total	100.00	100.00	100.00



Figure 2: Income distribution- non western employees

Note: Income in 1,000 SEK. Deflated to 1980 years values.



Figure 3: Number of self-employed non-western immigrants

Note: Before 1993 self-employment only includes sole traders, and from 1993 also limited liability companies are included



Figure 4: Share self-employed over years since migration

Appendix D Further regression results

	(1)	(2)
	More than high school	Female
Panel A. Reduced form		
Co-ethnic banker, year $t_0$	0.000617	-0.00181
	(0.00352)	(0.00154)
Observations	171,651	333,673
R-squared	0.044	0.022
Panel B. IV		
Co-ethnic banker, year t	0.00249	-0.00699
	(0.0142)	(0.00599)
Observations	171,651	333,673
R-squared	0.044	0.021
Panel C. First stage		
Co-ethnic banker, year $t_0$	0.248***	0.259***
	(0.0202)	(0.0203)
Observations	171,651	333,673
R-squared	0.318	0.344

 

 Table 12: Estimated effects of exposure to co-ethnic banker on self-employment probability: instrumental variable and reduced form. Further subsamples

*Note:* Panel A reports estimated  $\beta$  parameters (standard errors) from equation (2), while Panel B reports estimated  $\beta$  parameters (standard errors) from a version of equation (1) where the endogenous regressor Co-ethnic banker<sub>it</sub> is instrumented by equation (3). Panel C reports estimated  $\beta$  parameters (standard errors) from the first stage regression, equation (3). The outcome variable is a dummy for being self-employed in panel A and B, while it is a dummy for the exposure to a co-ethnic banker in year t in panel C. Standard errors are clustered at the level of municipality of arrival and country of origin pairs.<sup>\*\*\*</sup> p<0.01, <sup>\*\*</sup> p<0.05, <sup>\*</sup> p<0.1. Fixed effects for year-, year of immigration-, municipality of arrival- and country of origin are included as well as controls for sex, age and education level. A full set of municipality of arrival group level covariates are also included, see list in the appendix for overview of these variables.

	All		Less than high school	
	(1)	(2)	(3)	(4)
	Entry	Exit	Entry	Exit
Co-ethnic banker	0.00134*	-0.0159	0.00136*	-0.0306
	(0.000713)	(0.0106)	(0.000814)	(0.0207)
Co-ethnic insurance employee	0.00121*	-0.0175	0.00146*	-0.0225
	(0.000664)	(0.0114)	(0.000887)	(0.0198)
Observations	652,370	22,246	263,923	8,607
R-squared	0.013	0.039	0.018	0.054

 

 Table 13: Effects of co-ethnic bankers and insurance company employees on probability of start-up and self-employment exit, OLS

*Note:* The table reports  $\beta$  parameters (standard errors) from a version of equation (1), where both Co-ethnic bankers and own-group insurance company employees are used as explanatory variables. The outcome variable used in column 1 and 3 (2 and 4) is a dummy for becoming self-employed (exiting self-employment). Standard errors are clustered at the level of municipality and country of origin pairs.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1 Fixed effects for year-, year of immigration-, municipality- and country of origin are included, as well as controls for sex, age and education level. Municipal group level covariates are contemporary, see list in appendix for overview over these variables.

	(4)	
	(1)	(2)
	All	Less than high school
C. Ataisharka	0 00500***	0 000 10***
Co-etimic banker	0.00580	0.00840
F 1	(0.00148)	(0.00188)
Employment rate	$(0.0347)^{(0.00)}$	(0.0397
M	(0.00610)	(0.00809)
Wean earnings	1.44e-05	3.34e-05
61	(1.60e-05)	(2.22e-05)
Share with above median income	0.00849	-0.00837
	(0.0118)	(0.0166)
Share of local population	0.00962	-0.169
	(0.127)	(0.131)
Mean years in country	-0.000/41**	-0.000737
	(0.000330)	(0.000477)
Share with above median income (W)	0.00172	0.00440
()	(0.00869)	(0.0126)
Mean earnings (W)	2.76e-06	9.77e-06
	(1.22e-05)	(1.76e-05)
Share self-employed	0.353***	0.337***
	(0.0268)	(0.0336)
Municipality size	-8.48e-08	-1.27e-07
	(6.52e-08)	(7.98e-08)
Share with university education (W)	0.00231	0.00205
	(0.00626)	(0.00880)
Share with university education	0.00618	0.00697
	(0.0101)	(0.0139)
Share with university educ (W) (WP)	0.0416***	0.0496**
	(0.0153)	(0.0227)
Mean income (W) (WP)	-9.20e-05***	-9.36e-05***
	(1.84e-05)	(2.55e-05)
Share working in high income workplaces (W)	0.0109	0.00891
	(0.00726)	(0.00991)
Mean income (WP)	-9.50e-05***	-0.000102***
	(1.89e-05)	(2.67e-05)
Share with university educ(WP)	0.00102	-0.0330
	(0.0156)	(0.0211)
Share working in high income workplaces	0.0195**	0.0123
	(0.00830)	(0.0106)
Observations	800,899	325,430
R-squared	0.053	0.067

Table 14: Detailed regression results of column 1 and 2 in table 2, OLS

*Note:* The table reports estimated  $\beta$  parameters (standard errors) from equation (1). The outcome variable is a dummy for being self-employed. Robust standard errors are clustered at the level of each municipality country of origin pair. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Fixed effects for year-, year of immigration-, municipality- and country of origin are included as well as controls for sex, age and education level. (W) Variable defined by individuals who work in the municipality where one lives. (WP) Variable defined by the colleagues of an individuals' own-group members.
# Essay 3

# Negotiated wage increases and the labor market outcomes of low wage workers evidence from the Swedish public sector

(With Oskar Nordström Skans)

## 1 Introduction

Collective bargaining plays a central role for the setting of wages in most OECD countries (Freeman, 2007). In particular, this tends to be true for the low skilled segments of the labor market. However, wages in most OECD countries are set in a complex multi-stage process where sectoral collective agreements and local bargaining tend to coexist. As a consequence, the impact of collective wage agreements on actual outcomes, including wages, are therefore often unknown a priori, also to the unions and employer representatives that sign the agreements.<sup>1</sup> In this essay we provide empirical evidence on how wage increases as stipulated in a collective agreement for public-sector blue-collar workers affect key outcomes at the individual and establishment level.

It appears obvious that changes in workers' wage growth induced by collective agreements could have substantial effects on fundamental labor market outcomes such as labor turnover, working hours and the long run wage trajectories of the covered workers as well as the skill-composition of covered organizations.<sup>2</sup> Despite of this, we have found very few recent empirical studies on the topic. In particular, we have found very few recent studies for Sweden, from where we draw our data. The absence of studies is likely to reflect empirical challenges arising from the fact that the content of collective agreements should reflect expectations regarding the outcomes we are interested in.<sup>3</sup> In addition, it is often difficult to match the content of collective agreements to high-quality individual-level data (Forslund et al., 2012).

The intervention we study in this essay (see below for details) was motivated by a union initiative with the aim of increasing wages among low-paid women. To achieve this end, a collective agreement stipulated additional, establishmentlevel, wage increases in proportion to the number of female workers earning below a pre-specified wage threshold. We use panel data over the universe of covered workers and document that actual establishment-level wages increased as stipulated by the agreement, and analyze the extent to which this increase affected the long run outcomes of the employees.

The previous literature contains a host of empirical studies of how wages and other outcomes are affected by wage setting institutions, such as, e.g., studies of the union wage premium, or the wage impact of works councils. In terms of more precisely documenting the impact of wage changes on economic outcomes within a stable institutional setting, the previous literature is however almost exclusively focused on the impact of minimum wages. This literature has, on the other hand, resulted in hundreds of empirical studies using a multi-

 <sup>&</sup>lt;sup>1</sup>A good example is Böckerman and Uusitalo (2009) who finds that a sectoral exception from minimum wages in Finland had no effects since wages at the local level failed to respond.
 <sup>2</sup>See e.g. Neumark and Wascher (2008) and Salop and Salop (1976).

<sup>&</sup>lt;sup>3</sup>See for example the literature on "wage drift", (Hibbs and Locking, 1996; Holden, 1998)

and the literature on "wage cushions", e.g. Cardoso and Portugal (2005).

tude of outcomes. The minimum wage literature is dominated by studies from the US; Schmitt (2013) provides a recent overview. But as noted by Neumark and Wascher (2008), the effects of collectively agreed minimum wages may differ from legally imposed minimum wages since the bargaining process may take into account adverse effects of minimum wage changes. Rare examples of studies in a collective agreement context include Skedinger (2006, forthcoming) and Böckerman and Uusitalo (2009).

The empirical focus on minimum wages in the broader literature is a natural response to the fierce political debate regarding the appropriate level of minimum wages in the US (in particular). However, to the extent that we are interested in how wage-setting practices affect the labor market outcomes of low-skilled workers in a European industrial-relations context, it is important to note that a one-sided focus on the minimum wage will provide us with a very partial answer. In general, a very small share (mostly very young) of the low-skilled workers tend to be covered by the minimum wages (Garnero, Kampelmann, and Rycx, 2013), whereas the coverage of collective agreements is near universal within sizable segments of many European labor markets. As a consequence, bargained wage increases are likely to be relevant for significant fractions of the workforce. In a Swedish context, Forslund et al. (2012) documented that actual wages for young entrants are far above the sectoral minimum wages within many bargaining areas. Yet, the wage costs of these workers are still affected by collective bargaining through local agreements on entry wages and (potentially) through both sectoral and local agreements on wage increases. In contrast to archetypichal state-level minimum wages, the decentralized bargaining outcomes also allow workers to move across sectors in response to changes in collective agreements. In this paper, we try to derive first-order evidence of the effects of collectively agreed wage hikes within such a setting.<sup>4</sup>

Our empirical case is based on a very unusual collective agreement. The agreement stipulated that average wage increases among incumbent workers should be a function of the interaction between gender and the previous wage of incumbent employees within the establishment. This agreement was the result of a very turbulent discussion within the Swedish Confederation of Labor Unions (LO) which resulted in a general agreement to focus the attention towards wage increases for low paid women in the 2006-2007 round of wage nego-

<sup>&</sup>lt;sup>4</sup>Our analysis is also related to other strands of the economic literature. The literature on the effects of changes in labor costs, induced by subsidized labor or changes in payroll taxes, provides another way of studying labor costs changes similar to the collectively agreed wage growth we study here. Examples include Bennmarker, Mellander, and Öckert (2009), Huttunen, Pirttilä, and Uusitalo (2012), and Kramarz and Philippon (2001). A large literature focus on the inter temporal dependence of wages by either analyzing real or nominal wage rigidities, or by documenting the impact of past economic conditions on current wages (following Beaudry and DiNardo (1991); see Grant (2002) for a more recent study).

tiations across the Swedish economy. The exact implementation of this target differed greatly across sectors however. The most literal interpretation, which we study in this paper, was implemented by the Swedish Municipal Workers' Union in their agreement with the Swedish Association of Local Authorities and Regions (National Mediation Office, 2007).<sup>5</sup> This agreement (HÖK) stipulated an extra scope for wage increases amounting to 400 SEK (200 SEK) for each woman earning below 20 000 SEK during 2007 (2008). Since the general principle for wage determination within the bargaining was (and is) that negotiated (aggregate) wage increases should be distributed locally in accordance with individual performances, the agreement did not stipulate that the extra wage increase necessarily needed to be paid to the women that earned below this threshold, instead it was stipulated that average (local) wage increases should reflect the number of low-paid women.

The design of the agreement allows us to study its impact on actual wage growth and subsequent economic outcomes while separately controlling for wages and gender since the eligibility is determined by the interaction of these variables. Within this essay, we focus on the impact on wage trajectories, hours worked and exit probabilities. We also explore changes in hiring practices at the establishment level. Throughout, we rely on population-wide micro-level panel data capturing all workers covered by the agreement.

Our results show that the agreement had an effect on actual wage growth along the lines suggested by the agreement, inducing additional wage growth at establishments which employed a high fraction of women earning below the threshold. In other words, the agreements did induce a substantial change in the relationship between wage growth and the interaction between initial wages and gender composition at the workplace. The impact on wage increases among the covered incumbents had effects which lasted at least four years after the initial impact, with no signs of reversion within our four year follow-up period. This result implies that the impact of this one-time shock to individual wages had persistent effects on individual wages.<sup>6</sup>

These results suggest that one-time interventions through collective agreements can serve as a tool to achieve long-lasting changes in the wage distribution. This is a question of direct importance for wage setting unions (and policy-makers) who may perceive observable wage differentials between, e.g., men and women as a reflection of historical discrimination. Notably, the rationale for such interventions (and the result of the evaluation) will crucially hinge on whether wages are set according to contemporary market forces or if they remain a function of past wages, and hence also of past events.

<sup>&</sup>lt;sup>5</sup>The agreement between the Swedish Paper Worker Union and the the Swedish Forest Industries Federation was designed similarly.

<sup>&</sup>lt;sup>6</sup>Path dependence of wages have previously been documented in other settings, see e.g. Beaudry and DiNardo (1991).

We also study the extent to which the agreement affected the hours worked and the probability to remain within the establishment, which serve as channels of possible adjustment for the employer as labor costs change, as has been noted in the literature on minimum wages (Hirsch, Kaufman, and Zelenska, 2011). The results show that, as wage growth increased, hours worked appear to have increased less than within other workplaces at the same time. We also find that the overall separation rate remained unaffected, while for young (more mobile) workers with grades above the median the separation probability declined. This final result suggests that higher wages allowed the employers to retain mobile workers with relatively good outside options. Further results suggest that the number of new hires at the workplace level decreased as a consequence of the rising wages.

The essay is structured as follows. Section 2 presents the institutional background and, in particular, the 2007 agreement that we study empirically. Section 3 describes the data. Section 4 present the first stage results on actual wages of covered workers. Sections 5 documents the long-run impacts. Section 6 concludes.

## 2 Institutional background

Since 1997, Sweden has a system of pattern bargaining where the "industrial agreement" for blue and white collar workers within the manufacturing sector serves as the focal point for sectoral agreements that are to follow. The (formal) relationship between sectoral agreements and final wages differs greatly between sectors. In contrast to many other countries, public sector wages are (again, at least formally) set with an element of individual bargaining within all major agreements. National wage scales were abandoned throughout the Swedish public sector already in the early 1990s (OECD, 1996).

Our analysis focus on the impact of the 2007 sectoral agreement between the Swedish Municipal Workers' Union (Kommunal) and the Swedish Association of Local Authorities and Regions (SKL). The agreement covers more than 400,000 workers, which is more than any other agreement on the Swedish labor market. The agreement covers some of the lowest paid jobs in the Swedish economy and most employees are female (about 16 percent male employees, varying slightly over the years). The main occupations are within basic services such as child care and health care.<sup>7</sup>

In normal years the (wage part of the) agreement has three main components. First, it stipulates a minimum wage. This minimum wage is so low that it is only paid to about 10 percent of young first-time employees and it has no direct impact on the wages of incumbent workers or older entrants (see Forslund et

<sup>&</sup>lt;sup>7</sup>See a description of the occupations covered by the agreement in table 10 the appendix.

al. (2012)). Second, the agreement (sometimes) stipulates a lowest possible wage increase which is payable to all incumbent employees. This minimum amount is typically far below the average wage increase. Finally, it stipulates a required (average) rate of wage increases which should be calculated according to the number of full-time equivalent employees on permanent contracts. This average should be distributed over all employees according to local, in principle individualized, bargaining. Wages are not allowed to discriminate against those on temporary contracts. Table 1 shows the central elements of the agreements and the outcomes between 2004 and 2011.

#### 2.1 The 2007 agreement

The different unions within the blue collar confederation (LO) have, to a varying degree, made attempts to coordinate their bargaining strategies before the negotiations concerning the industrial agreement. The bargaining round resulting in the 2007 municipal agreement was preceded by heated discussions between different unions within the LO confederation where unions dominated by low-wage females, including the Swedish Municipal Workers' Union, were pushing for a general emphasis on lifting the wages of low-wage females during the upcoming negotiations. As a result, many of the 2007 agreements across the Swedish economy resulted in larger increases in minimum wages than in average wages and larger wage increases in sectors with large shares of low-paid female workers (National Institute of Economics Research, 2007).

The municipal agreement covering 2007 to 2009 was signed in April 2007. The agreement implied that workers, instead of wage increases, should receive two lump sum payments (5400 SEK in total) during 2007. The motivation for this odd solution was a perceived problem with lags in the local negotiations. By skipping one round of local negotiations, the partners hoped that the 2008 round of local negotiations would be completed already in January 2008.

		)	•		
Year	Required average wage increase	Guaranteed individual wage increase	Change in minimum wage - workers 19 or older	Change in minimum wage - workers with at least 1 year experience	Year specific details
2004	2 %	0	0	7.7 % *	5 % average wage increases
2005	510 SEK	175 SEK ¤	3.1 %	0	
2006	500 SEK	175 SEK $\alpha$	3 %	7.1 %	
2007	0	0	0	0	5400 SEK lump sum
2008	1400 SEK	0	0	0	400 SEK per female earning below 20,000 SEK
2009	720 SEK	175 SEK	10.9 %	13.3 %	200 SEK per female earning
2010	490 SEK	100 ¤	2.5 %	2.26 %	DEION ZU,UUU SEN
2011	480 SEK	100 ¤	2.45 %	2.21 %	
<i>Note:</i> *C: Negotiate	alculated as increase from th d centrally, and can be overr	ne level of minimum wage for v uled by local negotiations.	vorker 19 or older 2003. The	latter minimum wage definitio	on did not exist until 2004. $^{\mathrm{a}}$

Table 1: Details of wage agreement for municipal and county blue collar workers 2004-2011

The second unusual part of the agreement, which we study in this paper, was a specific focus on low paid females *within* the agreement. On top of the normal required average wage increases (on average 1400 SEK/month and full-time employee), it was stipulated that an additional 400 SEK/month should be added to the required average increases for each *female* earning below a full-time equivalent of 20,000 SEK/month. Thus, during 2008, the share of low-wage females mattered for wage increases according to collective agreements. The third year of the agreement (2009), had a similar construction but instead added 200 SEK/month. Minimum wages were kept unchanged during 2007 and 2008, and there were no guarantees for individual wage increases for these years. For 2009, the minimum wage increased by 10.9 percent<sup>8</sup> and all workers were guaranteed a wage increase of at least 175 SEK/month.

Notably, while 20,000 SEK was a very low wage in many sectors at the time, this was not the case within this particular part of the labor market. Therefore, a large share of the covered workers within this agreement are considered low-wage earners when using this cut-off (see wage distribution in section 3).

## 3 Data and descriptive statistics

The paper primarily relies on wage data from the Wage and Salary Structure Data (WSSD) originally collected from the municipalities and counties by the National Mediation Office. These data allow us to separate between workers covered by blue and white collar agreements and hence identify our population of interest. The data set contains wages, occupations and working hours in November during 2003-2011 for all employees within the municipalities and counties covered by the agreement. Wages are reported as monthly "full-time equivalents" which means that they are adjusted to take into account differences in working hours (i.e. they measure hourly wages multiplied by normal monthly full-time hours within the agreement). The wages we use in this paper are "base wages", which include all persistent wage components (including, e.g. supplements for managerial work), but not time-varying supplements such as overtime compensation.

Individuals can be followed over time through an anonymized indicator variable. Place of work (or "establishment") is added by Statistics Sweden using data from Swedish tax authorities that are available for all years except 2011. To these data we add basic demographic characteristics and final grades from compulsory schools (age 16) which are available for all individuals graduating from 1988 onwards.

<sup>&</sup>lt;sup>8</sup>The minimum wage for workers with at least one year of experience increased with 13.3 percent in 2009.

Panel A. Wage Dispersion					
	Mean	Variance	90/10	50/10	Wage gap ¤
	(Ln SEK	) (Ln SEK)	)		(ln SEK)
Total SEK	9.7932	0.0091	1.2608	1.1579	-0.0269
Organization		0.0085	1.2484	1.1516	-0.0243
Workplace		0.0067	1.2202	1.1286	-0.0022
Local Occupation		0.0064	1.2106	1.1252	-0.0028
Job		0.0048	1.1715	1.0977	0.0069
Within large occupations					
Child-care workers	9.7814	0.0078	1.2556	1.1605	-0.0109
Assistans nurses, hospital assistants	9.8210	0.0051	1.1859	1.1151	0.0032
Home based personal care workers	9.7435	0.0093	1.2767	1.1639	0.0099
Attendants, psychiatric care	9.8073	0.0085	1.2544	1.1611	0.0011
Panel B. Wage Growth Dispersion					
0. I I I I I I I I I I I I I I I I I I I	Mean	Variance	90-10	50-10	Wage gap
Total SEK	0.0334	0.0006	0.0410	0.0146	0.0003
Organization		0.0005	0.0305	0.0096	0.0003
Workplace		0.0005	0.0323	0.0103	0.0002
Local Occupation		0.0005	0.0312	0.0096	-0.0002
Job		0.0004	0.0289	0.0091	-0.0003
Within large occupations					
Child-care workers	0.0325	0.0005	0.0381	0.0137	-0.0005
Assistans nurses, hospital assistants	0.0323	0.0004	0.0374	0.0140	-0.0010
Home based personal care workers	0.0364	0.0008	0.0455	0.0138	-0.0002
Attendants, psychiatric care	0.0338	0.0007	0.0463	0.0191	0.0008

Table 2: Wage dispersion 2006 and wage growth dispersion 200.	5-2006
---	--------

*Note:* P90/P10 an P50/P10 are calculated within each organization/Workplace/Local Occupation/Job. Mean of ratios presented here. In Panel B P90-P10 and P50-P10 is used instead as some P10 equals 0. ¤ Wage gap is defined as the estimate in a simple wage regression controlling only for female dummy and Age. Summary statistics presented for the unrestricted data, which means the whole population of covered workers excluding only observations with unidentified workplaces. We place two initial restrictions on the data: First, we use one observation per worker and year. Here we give precedence to the observation with the highest annual earnings according to Statistics Sweden (if there are multiple entries) and the observation with the highest wage if there are multiple observations for the same employer. Second, we remove workers with missing establishment codes. These are all employees without a physical place of work, including many within the home care sector.

As could be seen from table 1 there are clear differences in the agreement for 2007 compared to the previous and following years. In 2007 all wage growth was paid as a lump sum and therefore the base wage which we use for studying wage changes remained at the exact same level for 80 % of workers between 2006 and 2007.<sup>9</sup> For this reason we will consistently use wage changes between 2005 or 2006 (or before that) whenever we study the evolution before the agreement of interest (covering changes between 2007 and 2008).

In table 2 we describe the dispersion of wages and wage increases as well as wage differences between men and women before the agreement was struck. As is evident from the table, wages within these jobs are fairly compressed and there are very small wage differences between men and women: On average men earn 2.7 percent more than women, which reflects the fact that the municipal sector is the sector with the smallest gender wage gap in the Swedish economy (National Institute of Economics Research, 2007). The ratio between the 90th and the 10th percentile is 26 percent. We also see that most of the overall wage dispersion remains within municipalities, occupations, establishments and even jobs (occupation-establishment combinations), whereas gender differences appear to primarily originate from differences across occupations.

A similar pattern pertains to wage increases. The average wage increase was 3.3 percent with a 90-10 ratio of 4.1 percentage points. Most of this dispersion remains within workplaces and within the occupations. Differences in wage growth between men and women are virtually non-existent.

In figure 1, we show the wage distribution in November 2007. Since the targeted increase we are interested in was supposed to take effect at the turn of the year, this is a reasonable proxy for the target distribution. The figure shows that most of the wage distribution falls below the "low wage" threshold of 20,000 SEK, emerging at the 91st percentile of the distribution of base wages. The average base wage within our sample is 18,050 SEK.

Figure 2b shows wage increases by gender and (half) wage percentiles during 2007-2008 for an interval including wages 1,000 SEK above or below the low-wage cut-off of 20,000 SEK in 2007. The percentiles of the wage distribution which this interval translates to are also plotted for the wage growth between 2005-2006 for comparison (figure 2a). As is evident, wage increases

<sup>&</sup>lt;sup>9</sup>Figure 4 in the appendix shows the share of all workers who remained at their workplace but who did not change their base wage between year t and year t+1 for years 2003 to 2009.





Note: Plotted for unrestricted data excluding only workers with unidentified workplaces. The vertical line shows the cut-off for 20,000 SEK.

for females within this interval close to the low-wage cut-off point were on average much higher relative to men during the second period than during the first period. It is also evident that the shift in gender differences is particularly pronounced for those earning below 20,000 SEK.<sup>10</sup> Notably, however, there is a strong relationship between individual wages and the fraction of low-wage (below 20,000) women at the establishment (see Figure 5 in the appendix). For this reason we will study the separate effects of the own treatment status, as well as of the composition of the workplace.

## 4 The first stage impact: wage growth during 2007-2008

The agreement specifies an extra wage increase for each female earning below 20,000 SEK. According to the agreement, actual wage increases should be determined at a level where performance can be monitored, but local partners have large degrees of freedom concerning how this should be implemented in practice. Our initial conjecture was that wage increases either were allocated by individual treatment status (being a low paid female), or by the treatment

<sup>&</sup>lt;sup>10</sup>The higher wage growth overall between 2007 and 2008 than between 2005 and 2006 is due to that wage increases 2007 were paid as a lump sum and thus base wages does not change between 2006 and 2007.



Figure 2: Wage increases by wage percentile

(b) 2007-2008

Note: Plotted for unrestricted data excluding only observations with unidentified workplaces. The vertical line shows the cut-off for 20,000 SEK in 2007.

intensity (share of low paid females) at the establishment level. In order to study the impact on other outcomes, we first need to separate between these two hypotheses since if, e.g., the impact depended on the individual treatment status, we expect the wage of low paid women to respond more to the intervention than others in other dimensions as well. A third alternative hypothesis is that the wages were smoothed across the entire organization (municipality or county) which would make it difficult to study the impact on other outcomes.

#### 4.1 Restrictions

For the empirical analysis we add four restrictions on our data.<sup>11</sup> Firstly, we exclude all individuals with very few contracted working hours (less than 25 % of full time), as the wage growth of these individuals are likely to be subject to a different wage setting scheme.<sup>12</sup> We furthermore remove outliers in terms of wage growth (1 percent in each tail). We thereafter exclude newly employed individuals (defined as individuals who have not been employed in the organization any of the last two years). We do this as a proxy for workers on permanent contracts, but we also present results from a robustness check where new entrants are added to the data and the results are completely unchanged. Lastly, we focus our baseline wage growth analysis on workers remaining within the original establishment since these are the ones that are directly treated by the agreement. For long-run estimates we require that workers remain within the organization (i.e. Municipality or County). Again, robustness exercises suggest that the restriction to establishment-level stayers is unimportant for the conclusions.

#### 4.2 Effect of individual treatment status

To study the impact of the agreement we first model the wage increase as an individual process with an individual-level treatment T taking the value one for females earning below 20,000 SEK and 0 for all males and high-wage women. This allows us to identify the effect of the agreement using a cross-sectional differences-in-differences strategy controlling (separately) for the direct impacts of both gender and initial wages.

Our outcome of interest is the individual log wage growth  $(\Delta w_i)$ . Since this outcome is derived by taking first differences at the individual level, we will, throughout the analysis, account for persistent wage differences across workers (i.e. the model corresponds to traditional "individual fixed effects" models specified in log levels). We control for gender by a dummy  $(D_i^{Female})$ . Our controls for initial wages are defined by a set of dummies for each percentile

<sup>&</sup>lt;sup>11</sup>For a description of how sample sizes changes with each restriction, see table 11 in the appendix.

<sup>&</sup>lt;sup>12</sup>The estimated effects on wages are robust to relaxing this restriction.

(denoted by  $\pi^p$ ) in the (initial) wage distribution. Using percentile-dummies allows us to contrast the estimated outcomes of the agreement with patterns during preceding years.<sup>13</sup> Formally, we estimate:

$$\Delta w_i = \sum_{p=75}^{96} \phi^p \pi_i^p + \delta D_i^{Female} + \gamma T_i + \beta X_i + \varepsilon_i, \tag{1}$$

where  $T_i = I[\pi_i^p < 91; D_i^{Female} = 1].$ 

The vector  $X_i$  include controls for education level, a dummy for being foreign born, a set of controls for working hours measured in percent of full-time work, <sup>14</sup> as well as dummies for different tenure levels<sup>15</sup> and a set of fixed effects defined either at the organization (Municipality/County) or establishment level. We estimate the model on the sample range between 19,000 and 21,000 SEK in 2007 which translates to the 75th to 96th percentile of the 2007 wage distribution.

The results are presented in table 3. Columns 1-4 show an individual-level effect of being a low paid woman. Column 1 controls for demographics and wage percentile, while column 2 adds organization (i.e. for the employing municipality or county) fixed effects. Importantly, the fact that the effect remains if we control for organization fixed effects implies that the wages are indeed not smoothed across the entire organization. The estimated effect of being a low paid female, compared to being a highly paid female or a low paid male, is about 0.26 % (approx. 50 SEK) when including organization fixed effects.

As shown in Figure 5 in the appendix, individual treatment status is very highly correlated with the treatment intensity at the establishment level. To identify the individual level impact, net of the impact of the establishment-level treatment intensity, we re-estimate the model after including establishment fixed effects. The results of this exercise, presented in column 3, shows that the estimate decreases radically and becomes insignificant once we control for these establishment fixed effects. This implies that the wage effect of being a low-wage woman is to a large degree driven by employment in workplaces with a large share of low-wage women. Notably, this is fully in line with the spirit of the agreement which stipulated that *average* wage increases should be a function

<sup>&</sup>lt;sup>13</sup>The wage percentiles are calculated on the wage distribution in the unrestricted data. We focus this model on the sample range between 19,000 and 21,000 SEK in 2007 which translates to the 75th to 96th percentile of the 2007 wage distribution. Thus, when studying pretreatment years, we focus on individuals at the same place in the wage distribution, i.e; between the 75th and 96th wage percentile of that year.

<sup>&</sup>lt;sup>14</sup>The distribution of individually contracted hours as a fraction of full-time work is displayed in figure 6 in the appendix. There are three noticeable spikes are at 50, 75 and 100 percent of full time and we therefore control for three dummies corresponding to these spikes, as well as for a linear function of individually contracted hours.

<sup>&</sup>lt;sup>15</sup>Tenure is defined as number of years at the workplace. Dummies for each year of tenure,

	(1) $lnw_{t+1} - \ln w_t$	$t=2007$ (2) $lnw_{t+1} - \ln w_t$	$(3)$ $lnw_{t+1} - \ln w_t$	$ \begin{array}{c}     t=2005 \\     (4) \\     lnw_{t+1} - \ln w_t \end{array} $
Female*Below	0.00307***	0.00262**	0.00124	-0.000910**
wage percentile 91	(0.00118)	(0.00109)	(0.000986)	(0.000444)
Observations	61,471	61,471	61,471	60,260
R-squared	0.070	0.123	0.431	0.712
Fixed effects	Organization	Organization	Establishment	Establishment
Clusters		Organization	Organization	Organization

Table 3: Test of individual level treatment

*Note:* Robust standard errors in parentheses. \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1. Estimation of equation (1). Individual level controls include dummies for wage percentile, gender, age, age squared, education level dummies, a dummy for being foreign born, tenure level dummies, three dummies for spikes in, and a linear control for, contracted hours of work as fraction of full time.

of the fraction of low-wage females.<sup>16</sup> We find no corresponding effects when reestimating the model for the 2005-2006 period (column 4) using the same percentile range and same controls as in column 3. Instead we here find a very small (significant) negative effect of being a low-wage woman. This indicates that before this special agreement low-wage women had slightly lower wage increases than both low-wage men and high-wage females.

### 4.3 Effect of the establishment level treatment intensity

Next, we turn to estimating models using the establishment level treatment intensity (ETI) as the main variable of interest. We calculate this fraction (by establishment) as the share of all employees that are female and earned less than 20,000 SEK in 2007. The mean share of female low-wage workers in 2007 is 0.78 with a standard deviation of 0.24.<sup>17</sup> In addition, we calculate the share of females (*FS*) and the share of low-wage workers overall (*LWS*).<sup>18</sup> We then

truncated at 5 years.

<sup>&</sup>lt;sup>16</sup>Controlling for establishment level treatment (see section 4.3) instead of using establishment fixed effects in the model corresponding to column 3 of Table 3 show similar results.

<sup>&</sup>lt;sup>17</sup>We find similar means and standard deviations of "treatment intensities" for all years in our data if we using a low-wage cut-off defined by the percentile corresponding to 20,000 SEK in 2007 for the other years.

<sup>&</sup>lt;sup>18</sup>All establishment level variables are defined by individuals who have working hours more than 25 percent of full time, and who are not recent entrants in the organization. As the wage setting scheme is based on full time equivalents, ETI, FS and LWS are calculated from the total number of full time employments among the workers. This means that two half-time workers are counted as one full time employed worker.

estimate the impact on the individual wage growth of the establishment level treatment intensity, controlling for these shares. Formally,

$$\Delta w_i = \sum_{p=1}^{100} \phi^p \pi_i^p + \delta D_i^{Female} + \beta X_i + \mu \overline{w}_j + \gamma^e ETI_j + \Psi[FS_j^E, LWS_j^E] + \varepsilon_i.$$
<sup>(2)</sup>

The model controls for the same individual level controls as equation (1) through  $X_i$  as well as the establishment mean log wage  $(\overline{w}_j)$  in 2007. When estimating this model, we do not restrict the analysis to individuals within the narrow wage range defined above, but include all covered workers, regardless of initial wage. We do restrict the analysis, though, by focusing on establishments with common support in terms of both gender and wages, implying that all establishments should have shares of low-wage workers and males strictly between zero and one.<sup>19</sup>

	Full data	a, t=2007	Narrow data, t=2007
	(1)	(2)	(3)
	$lnw_{t+1} - \ln w_t$	$lnw_{t+1} - \ln w_t$	$lnw_{t+1} - \ln w_t$
ETI	0.0172***	0.0160***	0.0174***
	(0.00375)	(0.00345)	(0.00382)
Observations	116,765	116,765	39,341
R-squared	0.136	0.155	0.132
Fixed effects		Organization	Organization
Clustered	Organization	Organization	Organization

Table 4: Test of establishment level treatment

*Note:* Robust standard errors in parentheses. \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1. Estimation of equation (2). Individual level controls include dummies for wage percentile, gender, age, age squared, education level dummies, a dummy for being foreign born, tenure level dummies, three dummies for spikes in, and a linear control for, contracted hours of work as fraction of full time. Workplace level controls include share low paid, share female and mean log wage level.

The results presented in table 4 suggest that the agreement had a significant impact on actual wages. Extrapolating from the share-estimates suggest that wages increased by an amount which is fairly close to the stipulated 400 SEK if the share of low-wage females increased from 0 to 1, keeping constant the direct impact of the share of females and the share of low-wage workers.<sup>20</sup> Column 1 estimates the model without organization fixed effects. Adding organization dummies in column 2 only has a marginal effect on the estimate of interest.

<sup>&</sup>lt;sup>19</sup>In general, the results are robust to relaxing this restriction, as seen in section 5.5.

<sup>&</sup>lt;sup>20</sup>When using change in Wage (SEK) as the regression outcome we get an estimate of 346 SEK.



Figure 3: Establishment treatment effect 2004-2009

Note: Plotted estimates of equation (2) for years 2004-2009.

In column 3, we estimate the same model, but for the narrow range of individuals with pre-period wages between 19,000 and 21,000 to compare to the earlier estimates of the individual level effect. The effect is noticeably stable considering that we remove two thirds of the data (standard errors remain largely unaffected since we cluster on organization).

To ensure that our models are not capturing generic differences in wage growth between establishments with a high share of low-wage females and other establishments, we have also re-estimated the model using data from several years (compare with column 2 in table 4 above). Here, we define workers with wages below wage percentile 91 as earning low wages when calculating low-wage shares and treatment intensities for other years. The results for two leading and lagged years respectively are presented in figure 3.<sup>21</sup> As is evident, the results suggest that 2007-2008 is a true outlier in terms of the estimates. The fact that we fail to find any effects in the succeeding years suggests that the ambition to target these particular establishments disappeared once new agreements where struck.

Overall, the results of this section suggest that the agreement of interest had an impact on the wage growth of the individuals employed in establishments with a larger share of low-wage females. This additional wage growth does not

<sup>&</sup>lt;sup>21</sup>As explained above, we cannot use data for 2006 since very few workers had their wages revised during 2007.

differ significantly between low-wage females and other females or low-wage workers within the same establishments. Therefore, apart from learning about the actual outcome of the agreed wage change, we can use this property of the agreement to examine further outcomes for the covered workers.

## 5 Post-agreement outcomes for the covered individuals

#### 5.1 Wage trajectories

To document the impact of the agreement beyond the direct effects on the workers employed by the covered establishments we first follow the career trajectories of the covered workers. Clearly, the implications of a one-time change in the wage growth distribution differs depending on whether the effects survives over a number of years. The extent to which a one time wage-push persists over time is a first-order diagnostic over how sensitive the contemporary wage distribution is to historical wage agreements.

In order to study the impact on wage trajectories, we replace the 1 year differences of equation (2) above by 2, 3 and 4 year differences which capture the total accumulated wage growth from 2007 to 2009, 2010 and 2011 respectively. These estimates should be stable if the total gain from the treatment intensity in 2007 remained during the years to follow. If, on the other hand, other workers are catching up, estimates should go towards zero and if wage growth begets future wage growth, the estimates should grow over time.

Here it should also be noted that interpretation of, in particular, the 2-year difference (i.e. from 2007 to 2009) will reflect the impact of the treatment intensity in 2008, but also the fact that each female that remained below the 20,000 SEK threshold in 2008 provided an extra wage increase of 200 SEK during 2009 (thus, treatment intensities may be correlated).<sup>22</sup>

We estimate the model for the year-by-year sample of workers remaining within the establishment the first year, and workers remaining within the organization 2, 3, and 4 years after 2007. We also estimate the model for a balanced sample of workers remaining within the organization during all four years. The results presented in table 5 shows that the agreement had lasting effects on the wages of covered workers. Indeed we find little evidence of a declining effect. The fact that the wage growth is very similar for those who stayed all the four years (see panel A) as for those in the year-by-year sample (panel B) suggests that the results are not driven by exit of workers with lower wage growth. The

<sup>&</sup>lt;sup>22</sup>Although Figure 3 shows that the effect of low-wage females in the workplace during 2008 had an insignificant effect of wage growth between 2008 and 2009, it should be noted that the analysis of the figure used a threshold defined from percentiles and not from nominal 20,000 SEK as in the agreement.

Panel A. Balance	d data (4 years staye	rs in organization), 1	t=2007	
	$(1) \\ lnw_{t+1} - \ln w_t$	$(2) \\ lnw_{t+2} - \ln w_t$	$(3) \\ lnw_{t+3} - \ln w_t$	$(4) \\ lnw_{t+4} - \ln w_t$
ETI	0.0165***	0.0185***	0.0177***	0.0203***
	(0.00403)	(0.00469)	(0.00555)	(0.00587)
Mean outcome	.09	.13	.15	.18
Observations	82,942	82,942	82,942	82,942
R-squared	0.154	0.205	0.250	0.224
Panel B. Unbalan	nced data, t=2007			
ETI	0.0160***	0.0200***	0.0181***	0.0217***
	(0.00345)	(0.00442)	(0.00544)	(0.00553)
Mean outcome	.09	.13	.15	.18
Observations	116,765	115,680	107,263	102,179
R-squared	0.155	0.201	0.232	0.210
Fixed effects	Organization	Organization	Organization	Organization
Clusters	Organization	Organization	Organization	Organization

#### Table 5: Long term effects of establishment level treatment

Fixed effectsOrganizationOrganizationOrganizationOrganizationClustersOrganizationOrganizationOrganizationOrganizationNote:Robust standard errors in parentheses.\*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1. We only conditionon staying in the establishment the first year.Subsequent years we condition on staying in theorganization.Estimation of equation (2).Individual level controls include dummies for wagepercentile, gender, age, age squared, education level dummies, a dummy for being foreign born,tenure level dummies, three dummies for spikes in, and a linear control for, contracted hours ofwork as fraction of full time.Workplace level controls include share of low paid, share female

and mean log wage level.

119

table also presents the mean wage growth over 1,2,3 and 4 years respectively.<sup>23</sup>

In principle, the model could also be estimated using an IV strategy, where the initial wage growth was instrumented by the treatment intensity in 2007. This would require the additional assumption that the agreement only affected the long run evolution of wages through its short-run wage impact, an assumption which seems fairly uncontroversial (apart from the complication arising from the 2009-component). Since the short run (first stage) and long run (reduced form) estimates are of a very similar magnitude, it is however straightforward to infer that the corresponding IV-estimates would be close to unity.

#### 5.2 Hours of work

Next we turn to the impact of the wage trajectories on the number of hours worked by the covered employees. The agreement under study provides a very unusual example of an exogenous shock to workers' wages which in principle could allow us to present evidence on the hours elasticity to changes in wages.<sup>24</sup> However, although we are able to document the causal impact of wage changes on changes in hours worked, our estimates must be interpreted as a (partial) equilibrium outcome. We are unable to isolate whether the changes in hours worked depend on the behavior of workers or employers. In particular it should be noted that the Municipal workers' union considers the reduction of involuntary part-time work within the sector as one of its key objectives, which indicates that part-time workers may be constrained in their hours of work.

As our measure of hours of work we use individually contracted hours measured as a fraction of full-time work (usually 40 hrs per week). We use the model of equation (2), with the change in percentage of full time, between 2007 and subsequent years as, the dependent variable.<sup>25</sup> Note that the model includes very detailed controls for initial working hours.

The results presented in table 6 show that the effects on contracted hours of work are negative, but also that they are estimated with fairly poor precision.<sup>26</sup>

<sup>&</sup>lt;sup>23</sup>The high wage growth the first year is a result of an extra scope for wage increases during that year, this phenomena is related to the zero wage growth during the preceding year (2006 to 2007).

<sup>&</sup>lt;sup>24</sup>From the minimum wage literature it is not clear what to expect in terms of effects on hours of work, see Neumark and Wascher (2008) for a general discussion. For Sweden, Skedinger (forthcoming) show a modest average reduction in hours of work from an increased minimum wage in the retail sector, although the decline is larger amongst the young. Skedinger concludes that the modest average decline in hours in part is due to selective separation, as many marginal workers (i.e. employed very few hours) exited their jobs due to the minimum wage increase. As our specifications are estimated in first-differenced form with detailed controls for initial hours, we are less exposed to these types of selection effects.

<sup>&</sup>lt;sup>25</sup>Outliers (1 percent in each tail) in change in hours worked are dropped.

<sup>&</sup>lt;sup>26</sup>These results are, in contrast to other main results in the study, sensitive to the inclusion of extreme values. If these are included these estimates are no longer significant.

The negative effect could either be due to income effects dominating substitution effects as in the traditional labor supply setting, or due to cost saving reductions in hours from the demand side. Notably, the mean contracted fraction of full time is increasing within the full study population, which suggests that the negative estimates can be interpreted as smaller increases in contracted hours of work for the covered workers. Comparing the estimates for the change in fraction of full time to the wage growth estimates, we see that the elasticity between wages and hours is near -1. As wages are full time equivalents, this elasticity can be interpreted as a net effect on gross disposable income close to 0.

	(1) $h_{t+1}^C - h_t^C$	(2) $h_{t+2}^C - h_t^C$	(3) $h_{t+3}^C - h_t^C$	$(4) \\ h_{t+4}^C - h_t^C$
ETI	-0.958** (0.414)	-0.102 (0.510)	-0.928* (0.528)	-0.367 (0.673)
Mean outcome	.9	1.14	1.5	1.4

 Table 6: Effect of treatment intensity in 2007 on changes in (individually contracted)

 hours since 2007

Mean individually contracted hours as a fraction of full-time (2007): 86.6 %

Observations	115,249	113,981	105,636	101,057
R-squared	0.126	0.167	0.203	0.212
Fixed effects	Organization	Organization	Organization	Organization
Clusters	Organization	Organization	Organization	Organization
	8	0.0	8.	8

*Note:* Robust standard errors in parentheses. \*\*\*\* p<0.01, \*\* p<0.05, \* p<0.1. We only condition on staying in the establishment the first year. Subsequent years we condition on staying in the organization. Estimation of equation (2). Outcome variable is the change in contracted hours worked as fraction of full time between year t and subsequent years. Individual level controls include dummies for wage percentile, gender, age, age squared, education level dummies, a dummy for being foreign born, tenure level dummies, three dummies for spikes in, and a linear control for, contracted hours of work as fraction of full time year t. Workplace level controls include share of low paid, share female and mean log wage level.

Around 50 percent of the remaining workers do not change their hours of work at all. We have therefore also estimated models where we let the outcome be the probability of increasing or decreasing hours of work. As outcomes we use dummies taking on the value 1 if the employee has increased (decreased, in corresponding models) her hours of work between 2007 and 2008 and the value zero if hours remained unchanged or decreased (increased). Results, presented in table 12 in the appendix, are well in line with the results of table 6, in particular considering that average hours were increasing within the study

population: A higher treatment intensity is associated with a lower probability of increasing hours of work, whereas the probability of decreasing contracted hours is unchanged.<sup>27</sup>

#### 5.3 Separations

Next we turn to the impact on the separation rates of covered workers. Here, we follow in the tracks of the minimum wage literature which to a large extent has focused on the impact of changes in the minimum wage on the rate of separations. Examples include Kramarz and Philippon (2001) and Abowd et al. (1997) who find disemployment effects of minimum wage increases for France and US. For Sweden, Skedinger (forthcoming) uses data from the the retail sector, finding evidence of substitution (in terms of separations) between workers affected by the minimum wage and those unaffected. A main reason for focusing on separations rather than entrants is that the target population (workers at risk) is well-defined when analyzing separations, but not when analyzing recruitments. This is particularly true when analyzing changes in sectoral agreements since workers may change their sectoral allocation of search efforts in response to sector-specific changes in wage levels.

Here we, again, estimate a straightforward version of equation (2) where we use a dummy taking the value one for individuals who leave the establishment as the outcome.<sup>28</sup> The results, presented in Table 7, column 1 suggest that the average exit rate remained unaffected.

Since mobility among older workers tends to be fairly low, we have also explored the exit rates among younger workers. Here, we focus on workers aged 36 and younger which implies that we study the cohorts for which we have data on compulsory school grades. Results, presented in column 2, suggest that the average exit rate for this group also remained unaffected.

Finally, we turn to the selectivity of workers who remain within their jobs.<sup>29</sup> Here we focus on the young workers and characterize the workers by their percentile ranked grades from the end of compulsory school (age 16). This is

<sup>&</sup>lt;sup>27</sup>Data also contains information on actual hours of work during the survey month. This variable deviates from contracted hours due to overtime and absence and thus provides a noisier measure of hours. Using this variable we find results that are sensitive to the choice of empirical model: We find positive effects of the establishment treatment intensity on changes in actual hours worked, but also an increased probability of decreasing actual hours of work. Our impression is that contracted hours is a less noisy and hence more suitable variable for studying the response of hours worked.

<sup>&</sup>lt;sup>28</sup>This implies that separations are defined as individuals working in an establishment during November during year t (2007) but not in November during t+1. The dummy variable is set to missing (i.e. the observation is removed) if the entire establishment is no longer in the data.

<sup>&</sup>lt;sup>29</sup>We have also estimated models analyzing separation probabilities separately for females and males, as well as for native born and foreign born. We found no significant differences in

an indicator which previous research has shown to be strongly related to the incidence and duration of unemployment among young workers in Sweden (OECD, 2008), and we therefore use it as an indicator of the workers' market value. In column 3, we interact the estimate of interest with an indicator for having grades below the median at the end of compulsory school. The results suggest large differences in the impact of separation probabilities depending on position in the grade distribution. Workers with lower than median grades have significantly larger probability of exiting the establishment due to a higher treatment intensity.<sup>30</sup>

	Without new entrants		With nev	v entrants	
	(1)	(2)	(3)	(4)	(5)
	Pr(exit)	Pr(exit)	Pr(exit)	Pr(exit)	Pr(exit)
ETI	0.0159	-0.00477	-0.145	-0.0334	-0.171**
	(0.0562)	(0.0751)	(0.0995)	(0.0693)	(0.0849)
Lower grades		-0.0259***	-0.0257***	-0.0336***	-0.0332***
		(0.00831)	(0.00776)	(0.00724)	(0.00695)
Lower grades*ETI			0.204*		0.222**
0			(0.116)		(0.0858)
Mean separa- tion rate	.21	.35	.35	.41	.41
Observations	147,835	24,394	24,394	34,346	34,346
R-squared	0.114	0.113	0.114	0.141	0.141
Fixed effects	Yes	Yes	Yes	Yes	Yes
Clusters	Yes	Yes	Yes	Yes	Yes

Table 7: Effects on separations 2007-2008

*Note:* Robust standard errors in parentheses. Clustered at the level of the organization. \*\*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Estimation of equation (2). Outcome variable is a variable taking on the value 1 if the workers leaves the establishment between year t and t+1, and 0 otherwise. In the specifications where ETI is interacted with Lower grades (column 3) also the variables Share of Female and Share of Low paid are interacted. Individual level controls include dummies for wage percentile, gender, age, age squared, education level dummies, a dummy for being foreign born, tenure level dummies, three dummies for spikes in, and a linear control for, contracted hours of work as fraction of full time. Workplace level controls include share of low paid, share female and mean log wage level. Organization fixed effects are also included.

As many young workers are recent entrants (i.e. starting their job between

estimated effects.

<sup>&</sup>lt;sup>30</sup>The results are very similar if we instead define separations at the organization level. The differences between the high and low skilled remain, with the workers with grades above the median reducing their separation probability. Mean separations rates are about 7 percentage point lower when defined by organization.

2005 and 2007), who are excluded in our base sample, we have re-estimated the model on an extended sample where these workers are included. Results presented in column 4 and 5 of table 7, are very similar to the models for the base sample, but the estimates are more precise.

Overall, these results suggest that the young stock of employees become more selected when wages go up as those young workers who have better outside options find stronger incentives to stay within their jobs, or as employers become more selective in who, among the young, they allow to stay. This result is well in line with recent research on the effects of minimum wages hikes on separations, which indicates that employers substitute low skilled workers with slightly more skilled labor in response to minimum wage increases (Neumark and Wascher, 2008).

#### 5.4 Establishment-level responses

In order to be able to, at least tentatively, analyze the impact on hirings, we have explored empirical models defined at the workplace level. These models control for individual level variables aggregated at the workplace level year 2007, as well as industry dummies (see equation (3)).<sup>31</sup> As outcome variables we use the normalized (by the stock of employees during 2007) number of hires in 2008, normalized separations between 2007 and 2008, as well as the normalized change in workplace size between 2007 and 2008. Estimates are weighted by workplace size the initial year to reflect the population distribution. Formally we estimate:

$$Y_{j} = \gamma ETI_{j} + \Psi[FS_{j}, LWS_{j}, \overline{X}_{j}, \overline{w}_{j}, \text{Industry}_{j}] + \varepsilon_{j}.$$
(3)

The results presented in table 8 suggest small positive but far from statistically significant effects on average separations which is well in line with the average effect of individual exits presented above.<sup>32</sup> More importantly, we find negative effects on establishment level hirings (see table 8). As a consequence, the final column shows that the net effect on employment is negative. This suggests that the higher wage growth induced by the agreement had negative effects on employment, but through reduced hires and not through increased separations.

#### 5.5 Robustness

All our models presented above control for permanent differences in wage levels (and hours) between different workers through individual fixed effects which

<sup>&</sup>lt;sup>31</sup>As with earlier workplace level variables, these variables are aggregated over all permanent employees at the workplace.

<sup>&</sup>lt;sup>32</sup>Here we include recent entrants, as well as individuals with low hours of work. The estimates are therefore not directly comparable to the estimate in column 1 of table 7.

	(1) Separations, 2007-2008	(2) Hires, 2008	(3) Change in size, 2007-2008
ETI	-0.0132	-0.144**	-0.131*
	(0.0541)	(0.0678)	(0.0746)
Observations	2,486	2,486	2,486
R-squared	0.480	0.402	0.189
Fixed effect	Organization	Organization	Organization
Clustered	Yes	Yes	Yes

Table 8: Effects on establishment level - separations, hires and employment

*Note:* Robust standard error in parenthesis. \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1 Standard errors clustered at the level of organization\*Industry. Estimations of equation (3). The outcome variables are normalized by the workplace size year t, why they should be understood as relative changes. Controls here include individual level controls aggregated at the workplace level, as well as dummies for industry affiliation. Estimates are weighted by workplace size year t. Outliers in normalized total hirings, normalized total separations and normalized change in workplace (1 percent in each tail) are excluded.

are implicitly accounted for by taking individual-level first differences. However, the estimates may be biased by the selective separations rates for the young (documented above) if these are related to differences in individual-specific wage or hours *trajectories* (i.e. trends). For this reason we have re-estimated the key models while excluding the group of young workers (aged 36 of less) for which we found selective separation rates.<sup>33</sup> These results are presented in Panel A of table 9. Here we see that the positive wage growth effect from being in an establishment with a higher share of low-wage females in year 2007 remains largely unaffected when the young workers are excluded (columns 1 and 2). The same is true for the negative effect on hours worked (column 3). This supports the notion that the wage trajectories as well as the reduction in hours worked is not caused by selective separation rates among the young.

Panel B of the same table instead shows how the wage and hours estimates are influenced by including recent entrants in the estimation sample. Overall, earlier estimated effects are unaffected (somewhat, but insignificantly, larger for hours) when including this group.

Finally, our main data set excludes workplaces which either do not have both men and women, or have only low-wage or high-wage employees in order to isolate the impact without letting other key controls be affected by outliers. In panel C, we explore the sensitivity of our estimates related to wage growth,

<sup>&</sup>lt;sup>33</sup>We also worry less about selection amongst the older workers since they are considerably less mobile overall.

changes in hours and separation rates to this restriction. The results imply that the wage and separation estimates remain robust to relaxing this sample restriction. A possible concern is that we find no significant effect on hours worked when relaxing this restriction. A potential explanation may be that the composition of part-time workers, and the evolution of working hours, is very different within all-male or all-female establishments.

As further robustness tests, we have also estimated the effect on individual wage growth during the first year, as presented in table 4, within a broader sample which also include those that change establishments within the organization (Municipality or County). The estimates are completely robust.<sup>34</sup> Overall, the effect of share of low paid females at the workplace on individual wage growth is robust to inclusion of further workplace level controls and more flexible controls for the workplace wage level defined by share of female and male workers within each decile of the overall wage distribution.<sup>35</sup>

### 6 Conclusions

This essay has focused on a documentation of the impact of negotiated wage increases on actual wages, worker career trajectories and separation rates as well as establishment-level separation and hiring responses. To this end, we have exploited the impact of a Swedish collective agreement for blue-collar workers employed by municipalities and counties. Our results show that the agreement had a substantial first-order effect on the wages of workers within the covered establishments. The wage increases appear to be lasting over at least four years for covered workers.

When studying labor input responses, we first find evidence of reduced hours of work (relative to other workers) which could imply that employers adapt to the wage hike by reducing labor input along the intensive margin. In a context where unions consider involuntary part-time work as a major concern, this is likely to be a negative side-effect from the workers' perspective. The negative impact on hours is large enough to greatly reduce the positive impact on workers' gross disposable income, but the estimates are fairly imprecise and sensitive to some of our robustness checks. We also see an increased selectivity amongst the young workers who remain within their jobs as well as indications of a reduced number of new hires in workplaces with a higher treatment intensity. Jointly, the estimates imply that employers reduce their use of labor inputs, and benefit from a relative skill upgrading, when wages are increased.

Although the particular agreement we are analyzing appear quite unique in

<sup>&</sup>lt;sup>34</sup>The estimate is 0.0152 with a standard error of 0.00345.

<sup>&</sup>lt;sup>35</sup>Added workplace level controls are aggregates of individual level controls among the employees who are not new entrants; share with at least high school education, share with low tenure, share foreign born and mean age.

	(1)	(2)	(3)	(4)
	$lnw_{t+1} - \ln w_t$	$lnw_{t+3} - \ln w_t$	$h_{t+1}^C - h_t^C$	Pr(Exit)
Panel A. With	nout Young Workers	3		
ETI	0.0142***	0.0141***	-0.920**	
	(0.00323)	(0.00523)	(0.450)	
Mean values	.089	.152	.798	
Observations	101,149	92,501	100,008	
R-squared	0.150	0.214	0.114	
Panel B. With	n New Entrants			
ETI	0.0155***	0.0188***	-1.381***	
	(0.00343)	(0.00564)	(0.438)	
Mean values	.09	.154	1.057	
Observations	126,483	115,867	124,545	
R-squared	0.149	0.234	0.142	
Panel C. With	n workplaces withou	t common support of	females and low wag	e employees
ETI	0.0138***	0.0162***	-0.0538	-0.0496
	(0.00222)	(0.00303)	(0.254)	(0.0321)
Mean values	.092	.157	.889	.212
Observations	224,226	209,059	221,513	285,846
R-squared	0.150	0.240	0.120	0.107
Fixed effects	Organization	Organization	Organization	Organization
Clusters	Organization	Organization	Organization	Organization

Table 9:	Sensitivity	analysis,	t=2007
----------	-------------	-----------	--------

*Note:* Robust standard errors in parentheses. \*\*\*\* p<0.01, \*\*\* p<0.05, \* p<0.1. Estimations of equation (2). We condition on staying in the establishment the first year. For column 2 we condition on staying in the organization. Outcome variable in column 1 (2) is the wage growth between year t and t+1 (t+3). Outcome variable in column 3 is the change in contracted hours of work between year t and year t+1. In column 4 the outcome variable is a dummy taking on the value 1 if the worker leaves the workplace and 0 otherwise. Individual level controls include dummies for wage percentile, gender, age, age squared, education level dummies, a dummy for being foreign born, tenure level dummies, three dummies for spikes in, and a linear control for, contracted hours of work as fraction of full time year t. Workplace level controls include share of low paid, share female and mean log wage level.

its seemingly arbitrary allocation of wage increases, our conjecture is that there is considerable room for more empirical work on how collective agreements affect the outcomes of covered workers and establishments. In order to reduce transaction costs, social partners (just as legislators) often resort to fairly arbitrary cut-offs when signing collective agreements and these cut-offs can provide very useful foundations for econometric evaluations. Yet, in particular since the wages of low skilled workers across Europe (at least formally) are determined through collective agreements, it is striking how little we know about the real impact of these agreements on the outcomes of the covered groups. We thus believe that a further investigation of the consequences of negotiated wage increases is a promising avenue for future research on key topics within the core of labor economics, such as documenting the nature and extent of wage rigidities, the allocative power of establishment-level wages and spill-over effects of wages between groups.

## References

- Abowd, J. M, F. Kramarz, T. Lemieux, and D. N. Margolis. 1997. "Minimum wages and youth employment in France and the United States". In *Youth Employment and Joblessness in Advanced Countries*. Ed. by D. G. Blanchflower and R. B. Freeman. University of Chicago Press, 427–472.
- Beaudry, P. and J. DiNardo. 1991. "The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data". *Journal of Political Economy* 99, 665–688.
- Bennmarker, H., E. Mellander, and B. Öckert. 2009. "Do regional payroll tax reductions boost employment?" *Labour Economics* 16, 480–489.
- Böckerman, P. and R. Uusitalo. 2009. "Minimum wages and youth employment: Evidence from the Finnish retail trade sector". *British Journal of Industrial Relations* 47, 388–405.
- Cardoso, A. Rute and P. Portugal. 2005. "Contractual wages and the wage cushion under different bargaining settings". *Journal of Labor Economics* 23, 875–902.
- Forslund, A., L. Hensvik, O. Nordström Skans, and A. Westerberg. 2012. "Kollektivavtalen och ungdomarnas faktiska begynnelselöner". IFAU Report.
- Freeman, R. B. 2007. "Labor market institutions around the world". NBER Working Paper 13242.
- Garnero, A., S. Kampelmann, and F. Rycx. 2013. "Sharp Teeth Or Empty Mouths?: Revisiting the Minimum Wage Bite with Sectoral Data". IZA Working Paper.
- Grant, D. 2002. "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from the National Longitudinal Surveys". *Industrial and Labor Relations Review* 56, 393.
- Hibbs, D. A. and H. Locking. 1996. "Wage compression, wage drift and wage inflation in Sweden". *Labour Economics* 3, 109–141.
- Hirsch, B., B. Kaufman, and T. Zelenska. 2011. "Minimum wage channels of adjustment". IZA Working Paper.
- Holden, S. 1998. "Wage drift and the relevance of centralised wage setting". *The Scandinavian Journal of Economics* 100, 711–731.
- Huttunen, K., J. Pirttilä, and R. Uusitalo. 2012. "The employment effects of low-wage subsidies". *Journal of Public Economics* 97, 49–60.

- Kramarz, F. and T. Philippon. 2001. "The impact of differential payroll tax subsidies on minimum wage employment". *Journal of Public Economics* 82, 115–146.
- National Institute of Economics Research. 2007. "Lönebildningsrapporten 2007".

National Mediation Office. 2007. "Avtalsrörelsen och lönebildning 2007".

Neumark, D. and W. Wascher. 2008. Minimum wages. The MIT Press.

OECD. 1996. "Pay Reform in the Public Service".

— 2008. "Country Survey Sweden".

- Salop, J. and S. Salop. 1976. "Self-selection and turnover in the labor market". *The Quarterly Journal of Economics* 90, 619–627.
- Schmitt, J. 2013. "Why does the minimum wage have no discernible effect on employment?" Report, Center for Economic and Policy Research.
- Skedinger, P. 2006. "Minimum wages and employment in Swedish hotels and restaurants". *Labour Economics* 13, 259–290.

------ Forthcoming. "Effects of increasing minimum wages on employment and hours: evidence from Sweden's retail sector". *International Journal of Manpower*.

# Appendix A Descriptives

	Unrestricted data		Restricted data	
	Mean	Share	Mean	Share
	values	males	values	males
Legislators, senior officials and managers	0.001	0.345	0.002	0.365
Professionals	0.004	0.419	0.003	0.478
Technicians and associate professionals	0.018	0.450	0.015	0.498
Clerks	0.003	0.640	0.003	0.636
Service workers and shop sales workers (not the below)	0.023	0.458	0.022	0.374
Cooks	0.029	0.098	0.028	0.130
Child-care workers	0.162	0.097	0.080	0.162
Assistant nurses and hospital ward assistants	0.309	0.066	0.420	0.072
Home-based personal care and related workers	0.187	0.105	0.166	0.116
Attendants, psychiatric care	0.124	0.237	0.113	0.324
Skilled agricultural and fishery workers	0.006	0.770	0.007	0.796
Craft and related trades workers (not the below)	0.011	0.968	0.013	0.965
Building caretakers	0.023	0.910	0.023	0.932
Plant and machine operators and assemblers	0.009	0.871	0.010	0.790
Elementary occupations (not the below)	0.009	0.822	0.014	0.823
Helpers and cleaners in offices, hotels etc.	0.035	0.069	0.042	0.076
Helpers in restaurants	0.047	0.047	0.041	0.047

#### Table 10: Occupational Distribution 2006

*Note:* Unrestricted data refers to the raw data over the covered workers, excluding only observations without identified workplace. The restricted data refers to the data used from section 4.3 and onwards, which is the most restricted data used.

	# Obs 2007
Before restrictions	433,022
After imposed restriction:	
Drop unidentified workplaces	359,532
Drop if hours worked $< 25$ percent	336,227
Drop 1 percent outlier in wage growth in each tail	330,751
Drop new entrants	289,620
Drop workplaces without common support in share males and share low paid	149,248

Table 11: Sample restrictions

*Note:* Numbers of observations for the year 2007 remaining in sample after restrictions imposed. The numbers are similar for other years.



Figure 4: Share of remaining workers with zero nominal wage growth

Note: Plotted for unrestricted data excluding only observations with unidentified workplaces.



Figure 5: Establishment treatment intensity by wage percentile

Note: Plotted for unrestricted data excluding only observations with unidentified workplaces. The vertical line shows the cut-off for 20,000 SEK in 2007.



Figure 6: Distribution over contracted hours of work as fraction of full time

Note: 46 percent are employed at 100 percent. The spikes in the distribution here are at 50 percent of full time (5 percent of observations), and at 75 percent of full time (12 percent of observations). 50, 75, and 100 are the spikes controlled for in the regressions.

## Appendix B Further regression result

VARIABLES	$(1) \ \Pr(h_{t+1}^C > h_t^C)$	(2) $\Pr(h_{t+1}^C < h_t^C)$	
ETI	-0.0335* (0.0172)	0.0111 (0.0140)	
Mean outcome	.113	0.067	
Observations R-squared	115,249 0.185	115,249 0.039	
Fixed effects Clusters	Organization Organization	Organization Organization	

 Table 12: Effects on probability of increasing and decreasing hours worked as fraction of full time, t=2007

*Note:* Robust standard errors in parentheses. \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1 Estimation of equation (2). Outcome variable is dummy for increasing contracted hours of work (column 1), and dummy for decreasing contracted hours of work (column 2). Individual level controls include dummies for wage percentile, gender, age, age squared, education level dummies, a dummy for being foreign born, tenure level dummies, three dummies for spikes in, as well as linear control for, contracted hours as fraction of full time. Workplace level controls include share of low paid, share female and mean log wage level.