Essays on urban wage premium, returns to internal migration and migrants' selection in Norway

Darina Steskal

Dissertation submitted to the
Department of Economics,
NHH - Norwegian School of Economics,
in partial fulfilment of the requirements for the PhD degree.

February 2015

CONTENTS

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my supervisor, Kjell G. Salvanes, for his guidance and support throughout the course of my doctoral research. He introduced me to the exciting field of labor economics, helped me narrow down the research topic, came up with a lot of interesting ideas, and, among other things, encouraged me to spend time abroad, either for conferences, workshops, summer schools or a research visit, which all helped to shape my research interests.

Secondly, I would like to thank my co-supervisor, Erik Ø. Sørensen, from whose advice and comments I benefited a lot. I appreciate his time, constructive comments, productive consultations, and clever ideas.

Thirdly, I am grateful to my co-author and friend, Aline Bütikofer, for her support, time, readiness and willingness to help, often out of her own initiative. She helped me improve not only my research, but also presentation skills, for example. I learned a lot from her.

My special thanks go to Leah Platt Boustan, who contributed her time and inspiring thoughts. Her ideas helped to guide and shape my research interests to a large degree. I am very grateful for her unconditional help, productive meetings and hospitality. I would also like to thank UCLA for providing a very creative and inspiring research environment.

I would like to express my thankfulness to the Department of Economics at NHH for giving me the opportunity to set out on this journey and for providing a caring, supportive and stimulating environment. It has been my pleasure to be a part of this department and I will undoubtedly miss the positive atmosphere and good social environment. I am also grateful to the administrative staff at the department, who have always been very responsive and helpful to my needs.

I acknowledge the labor group at NHH for interesting seminars, as well as the opportunities to present my work in progress and get valuable feedback.

I further thank to all of my fellow Ph.D. students for their friendship, inspiring discussions, mutual understanding and the friendly and motivating environment they created.

My thanks goes also to the members of the doctoral committee; their time and efforts are highly appreciated.

Last but by no means least, I would like to thank my husband, Lubos, for his incredible help, patience, concern, moral support and discussions. He played an inevitable role in bringing this journey to an end.

Thanks also to my family, especially my parents and parents-in-law, for their support. A thank also goes to our daughter, Editka, for creating extra motivation.

I sincerely thank you all.

Darina Steskal

INTRODUCTION

"*Cities are our species' greatest invention.*"

Edward Glaeser

(Glaeser, 2011)

Even though there is plenty of space on this planet (the entire world population would take up approximately as much space as is the area of Norway - with the same population density as in Paris), we choose to live closer and closer together, in cities and metropolitan areas. About 60 thousand people worldwide become urban every day, and although urban areas currently occupy only 3 percent of the planet's surface, more than half of the world's population already lives there. Not even the declining travel costs, which basically changed our perception of what long distances mean, has been able to reverse or mitigate this trend in urbanization, and there is no evidence that anything is going to. But why do cities grow?

There are several possible drivers of urban growth. It might be amenities that attract people to cities, like a nice beach or a good selection of cinemas; it might be random shocks that increase the labor demand, for example technological innovations that lead to a production of a new product, or it might be the underlying productive advantages that draw more firms and workers to the cities, which become larger and larger as a result.

The productive advantages can be achieved in multiple ways. It can be static advantages that workers and firms experience immediately after arrival to a city. These include reduced transport costs, proximity to customers, access to public infrastructure, sharing facilities with large fixed costs, to name a few. Another way for cities to be more productive is through the matching mechanism, which utilizes the dense labor market. In a dense labor market the probability and/or quality of matches between firms and workers are higher, leading to increased productivity. Yet another way how cities can be more productive is through human capital spillovers. Higher population density gives the opportunity for more face-to-face interactions and fosters knowledge accumulation. As far back as in 1890 Alfred Marshall suggested that "The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus becomes the source of further new ideas" (Marshall, 1890, p. 271). Historically, cities have always been at the center of the process of spreading knowledge - from trader to trader, engineer to engineer, scientist to scientist or painter to painter. It was Athens that gave birth to ancient philosophy, it was the streets of Paris where modern democracy was fought for, and it was the streets of Manchester that played a pivotal role in the Industrial Revolution.

For estimating the empirical magnitude of the urban productivity advantages, one widely used strategy is to compare workers' wages in urban and rural areas. It has been well documented in literature that people earn higher wages in cities. From the point of view of a firm, if it is willing to pay higher wages to its workers in cities, there must be some productive advantages that offset these higher costs for the firm. Thus, by estimating the wage gap between urban and rural areas, which is referred to as the "urban wage premium" in the literature, we can estimate the urban productivity.

A key concern when interpreting the urban wage premium is that the observed spatial disparities may be a combination of treatment and sorting - not only places impact people and make them more productive but also more productive people may live in cities. City workers may differ from rural workers in many characteristics, both observable and unobservable, for example, they may be more educated, motivated, ambitious, skilled etc. Hence, the estimated urban wage premium may reflect these differences in addition to the true productivity effect of the city on its workers. While the observed differences can be easily controlled for, a standard way to tackle the concern with the unobservable differences is to use panel data for individual workers and use migrants to estimate wage effects with individual fixed effects. In this way any individual-specific attributes will be differenced out. Compared with a simple pooled OLS regression, a fixed effects regression has been found to reduce the estimates of urban wage effects in the literature. This drop is usually interpreted as evidence of more productive workers sorting into cities, and the remaining wage premium as evidence of the true productivity effect of cities on their workers. However, such an approach is not without caveats. It is based on migrants who are not necessarily a representative sample of the general population, but it nevertheless addresses the sorting problem in some way and tends to suggest that there exists a real treatment effect of working in a city.

The fact that city workers might differ from rural workers can partly be a consequence of migration being a self-selective process. Migrants are not randomly selected from the population, but they choose to work in the location that offers the best opportunities for them. The utility maximizing behavior was first addressed by Hicks in 1932 who argued that "differences in net economic advantages, chiefly differences in wages, are the main causes of migration" (Hicks, 1963, p. 76). Thus, it may happen that the more productive workers choose to live in cities.

These phenomena gave rise to a vast body of literature studying selection and returns to migration, both in internal and international contexts, in developed and developing countries, both empirically and theoretically, using micro and macro approaches. Questions have been raised about how migrants are selected from the sending population and whether and to what extent they benefit economically from moving.

Many models have been developed to study selection into migration. Roy (1951), for example, discussed self-selection in terms of hunting and fishing occupations, but the same argument can be made for region of residence as well, as later developed by Borjas (1987). According to Borjas's model, the composition of migrants will depend on the income distribution of the sending and receiving regions - migrants will move to a region that has the most favorable income distribution for their skill level. Thus, regions with higher returns to skills, reflected in larger income inequality, will mostly attract high-skilled workers (positive selection), whereas regions that reward skills less and thus have more narrow income distribution, will mostly attract low-skilled migrants (negative selection).

The involvement of the selective process makes the identification of the returns to migration challenging. The observed wage gains of migrants might reflect both treatment and selection, and thus identification strategies that account for this have to be employed. However, unlike in the case of the productivity advantages that are estimated from the *nominal* wage differences, the costs of living are now included in the migrants' utility maximization. Workers will relocate only if the *real* wages will be higher in their new destination. Thus, in any analysis of returns to migration the price level in cities should be taken into account in order to capture the real benefits from relocation. This is, however, not an easy task since there rarely are standard price indices available for spatial comparisons.

The quality of migration data has increased dramatically during the last 20 years and has allowed numerous important advances in migration research that otherwise would have been impossible. The availability of micro- and longitudinal data has been particularly noteworthy. Micro data allows us to focus on the individual's behavior and study the migration decision making process. The advantage of longitudinal data is that they offer econometric advantage of being able to eliminate any individual fixed effects that could potentially be correlated with unobserved factors that affect the individual's propensity to migrate. Such factors lead to selectivity problems in cross-sectional studies.

This thesis provides empirical evidence on a multitude of questions related to internal migration and urban productivity in one of the countries in the world - Norway. In particular, it consists of three self-contained chapters where the first studies the selection and returns to north-to-south migration, the second documents urban wage premium over time and studies urban-to-rural migration, and the last one analyzes the magnitude of urban productivity and the channels behind it. Norway is an interesting case study because it is among the few countries that currently have high-quality administrative data sets available that track individuals' full mobility histories and contain a large number of personal characteristics. All chapters utilize these data sets. Furthermore, Norway is an example of a country that has been late, in European and US terms, in urbanizing, where movement from rural areas to cities, or from agricultural north to industrial south, has dominated the post-war period and to a large extent coincides with the data period under study. Finally, analysis of urban areas of a much smaller scale, as Norwegian cities undoubtedly are in the international context, provides an interesting contrast to the literature on huge metropolitan areas. The access to rich data allows us to employ several different econometric identification strategies that account for sorting and selectivity bias, ranging from difference-in-difference and fixed-effects estimators through IQ scores as ability controls, to a comparison of migrants' pre-move wages or comparison of brothers' outcomes. The long data span enables us to document how the urban wage premium and selection into migration have evolved over time, which is a novel feature in the literature. In the thesis, we also empirically test the Borjas's predictions about the migrants' selectivity behavior and are able to construct regional price indices and thus evaluate the *real* returns to north-to-south migration.

We find that movers from northern Norway who settled in the south during the past 40 years are more educated, while they seem to be negatively selected on unobservable characteristics. Comparing two cohort groups, we find that the selection gets stronger over time. Attempting to indirectly control for selection, we estimate mean (real) migration gains of 7% and 15% for early and late cohorts, respectively. Comparing the estimated mean effects of migration using different specifications we confirm positive selection of migrants on unobservable characteristics in the early cohorts, but a pronounced negative selection among later cohort migrants. Attempting to control directly for ability using conscription IQ tests (on a subsample) gives mixed results. We further estimate the effects of migration on the earnings distribution of movers. The estimated effects differ across the two cohort groups, suggesting an even effect for the early cohort but a steadily increasing effect for the later cohort.

When analyzing the urban wage premium from 1967 until today, we find that there is a substantial wage premium for working in cities and that a significant part of this premium is due to positive sorting of workers on observable, as well as unobservable characteristics. Over time, the urban wage premium decreased, and sorting became less pronounced. Also the selection into migration changed during the observed time period; we find that the migrants are positively selected from the sending population on characteristics such as education or IQ score, the positive selection is however less pronounced in 2010 than it was in 1970. When comparing movers' wages to rural stayers' wages before they relocate or the wages of brothers where one brother moved to the city and one stayed in the rural area, we find some evidence that the selection of rural-to-urban movers changed from the positive selection in the 1970s to a more negative selection in later decades.

The insight into the mechanisms behind the urban wage premium suggests that spatial differences in unobserved skill composition accounts for roughly one-tenth of the observed urban wage premium. The remaining premium is attributed to the family of explanations called agglomeration economies, out of which the static mechanisms appear to be the most dominant source. The dynamic mechanisms, which include learning and matching, both play significant roles as well. Finally, learning advantages acquired in cities do not appear to be transferable outside of cities, implying their location-specific character. The results are found to be robust to different strategies handling endogeneity and an alternative measure of agglomeration.

The following sections present the three thesis chapters in greater detail.

CHAPTER 1: SWEET SUNNY SOUTH: ECONOMIC RETURNS AND SELF-SELECTION INTO NORTH-TO-SOUTH MIGRATION IN NORWAY (COAUTHORS KJELL G. SALVANES AND ERIK Ø. SØRENSEN)

In this paper we analyze the returns to moving as well as the selection of workers moving from the north to the south of Norway over almost the last half century. We use population wide register data to track the mobility at the individual level, and we use regional price indexes to calculate real returns. The population wide data makes it possible to construct the pre-move counterfactual distributions. We use DiD and fixed effects in order to help identify selection on unobservables. We also contribute to the literature on migration by analyzing changes in selection over a long time period, and also analyzing the heterogeneity in returns to mobility.

We find that movers are generally positively selected on observables or more educated, while we find a change in the selection on unobservables over time from positive to negative. We find substantial gains from migration; mean real returns purged of selection are 7% and 15% for early and late cohorts, respectively

Our results are comparable to a recent study by Bartolucci, Villosio, and Wagner (2014), who analyzed migration from the south of Italy to the north and find negative selection to northbound migration in Italy. Italy has the same difference as in Norway although the north is the most industrial region in Italy. When controlled for selection, the returns to migration to the north over the last 20 years are of a similar size as the returns for the recent cohorts in Norway. As found for Norway for the later cohorts, the selection on unobservables is negative. For Norway, due to the long panel series of data, we were able to estimate the change in selection over time, and it went from strongly positive to negative.

We also estimate the effects of migration on the entire earnings distribution of movers. The estimated effects differ across the two cohort groups, suggesting an even effect for the early cohort but heterogeneous, steadily increasing effect for the later cohort, or more precisely upper tail longer for the more recent cohorts. It is likely that this reflects the change in selection over time.

CHAPTER 2: WHAT IS THE CITY BUT THE PEOPLE? CHANGES IN URBAN WAGE PREMIUM AND MIGRANT SELECTION SINCE 1967 (COAUTHORS ALINE BÜTIKOFER AND KJELL G. SALVANES)

The analysis of spatial differences in wages and in particular the urban wage premium has been well documented (Roback, 1982; Henderson, 1982; Helsley and Strange, 1990). However, it is still an open question to what extend cities make a person more productive, how much of this premium is driven by positive sorting and who self-selects into moving from rural locations to cities. Since the urbanization process in Norway happened comparatively late and in a time period observable in the Norwegian population registry data, Norway offers a unique laboratory for exploring in particular the changes in the urban wage premium and selection into rural-to-urban migration over time. The novel feature of this paper is that the registry data in combination with the urbanization process provide a unique opportunity for studying the urban wage premium and the selection into rural-to-urban migration during an interesting time period and using different identification strategies.

Our population registry based panel dataset ranging from 1967 until 2010 allows us to trace individuals' municipality of residence in each year and link this information with characteristics such as education and earnings patterns. Based on this information, we first estimate the size of the urban wage premium and document its changes from 1967 until today using different identification strategies. In a second step, we focus on individuals who are attracted by cities, and analyze rural-to-urban migrants and especially how the characteristics and selection of these rural-to-urban movers changed over time.

We find evidence for a substantial urban wage premium in Norway: the estimated premium for living in cities when pooling data for all the years is 16 percent. Controlling for observed characteristics such as education and age reduces the estimated premium to 13 percent. This drop in the urban wage premium denotes that there is a positive sorting on observable characteristics of people living in cities. Adding individual fixed effects reduces the urban wage premium to about seven percent. Hence, there is still a substantial urban wage premium even after controlling for observed and unobserved, time-invariant characteristics. This indicates that the positive sorting cannot explain the entire premium and that there remains a positive productivity effect of living in a city. In addition, we show that the full productivity premium is not realized immediately after the relocation to the city, but that the wage grows gradually during the first few years after the arrival in the city. Alternatively, we use IQ test scores as a proxy for the unobserved ability. Here, the results confirm that sorting on cognitive ability measured by IQ tests scores is an important component of the urban wage premium, although less than the individual fixed effects that may also control for non-cognitive ability. Nevertheless, both approaches suggest that the skill composition of rural and urban residents is different and important in explaining the urban wage premium. Focusing on changes over time, we find that the urban wage premium has decreased steadily since the late 1960s. The raw premium dropped from more than 26 percent in the late 1960s to below 10 percent in 2010 (from 22 to 7 percent when controlling for education and age). Hence, the positive sorting on observables became less pronounced over time. In addition, the urban wage premium is substantially smaller after introducing individual fixed effects in each of the four decades. The productivity effect, however, increased slightly over time. That is, the role of sorting on individual-specific fixed effects has lost importance since 1967.

During the same time period the characteristics and selection of rural-to-urban movers also changed; movers are significantly younger and better educated than stayers and they also have higher IQ scores. The positive selection on education and IQ scores is strongest in the early period of our sample. When comparing the outcome of the brothers where one brother moved to the city and one stayed in the rural area, we find that the selection of migrant households is positive in the 1970s and changed into a negative selection in the following decades. These findings are confirmed by analyzing pre-migration incomes of movers to rural stayers. In the six years prior to migrating to the city, those moving into an urban area earn slightly more than those who remain in the rural area in the 1970s. In 2010, the pre-relocation income of rural-to-urban movers is significantly lower than the income of the rural stayers.

CHAPTER 3: HOW CITIES MAKE US MORE PRODUCTIVE: EMPIRICAL EVIDENCE FROM NORWAY

The aim of this study is to empirically examine the reasons why workers in urban areas earn more than outside urban areas. Using rich administrative panel data on all male workers in Norway from 1986-2010, I first analyze the time patterns of migrants between the two types of regions. I find that the urban wage premium, when identified of stayers and not incorporating sorting or the dynamic structure, is about 13%. I also find that movers to cities experience immediate wage gains, and similarly, movers away from cities experience immediate wage losses, which implies the presence of static agglomeration economies and the fact that sorting alone is not a sufficient factor to explain the whole wage differential. The earnings profiles of workers after coming to urban areas monotonously continue to rise. This finding is consistent with the presence of dynamic advantages, acquired either through learning or matching. The analysis further shows that after migration urban-to-rural migrants' earnings fall to match the average rural earnings level. This result makes learning an unlikely explanation, since the accumulated human capital is hypothesized to stay with a worker after relocation. I further find that workers experience a substantial negative transitory shock several years before migration, which might bias the estimates if not taken into account.

I next develop a framework that combines all three types of agglomeration economies in one model by including an urban dummy and allowing the returns to experience and each job change to differ by location. Moreover, this framework allows for a direct assessment of the importance of sorting by an introduction of worker FE or, alternatively, IQ score as a control variable. The introduction of worker FE leads to a substantial reduction in the observed UWP - dropping from 8.5% in the baseline OLS estimation to about 7.6%. The reduction is interpreted as the effect of spatial sorting and amounts to about 11% in this data. The remaining UWP of 7.6% serves as evidence for a large influence of the static mechanisms. The rest of the findings present empirical evidence of superior learning and job matching in urban labor markets. Specifically, the effect of learning is almost twice as high in cities, with 0.1% higher return to each year of urban experience. As for matching, with each job transition a worker experiences 0.8% higher premium, which is more than a two-fold increase in comparison with rural areas. In terms of relative importance of static, learning, and matching mechanisms in explaining the UWP, under the assumption that sample averages give reliable estimates of the mean number of jobs that people have in each type of location, I find that static agglomeration economies are responsible for the bulk of the wage disparities, while learning and matching explain in the short run about 5% or 17%, respectively, increasing their influence to 15% or 27%, respectively, in the long run. The numbers slightly change when I use the estimates from the specification including IQ control instead of FE, but lead to similar conclusions, perhaps only assigning a greater relative explanatory power to the learning mechanism. The results are robust to an alternative measure of agglomeration which includes only the six largest cities of Norway among urban areas.

The broad patterns found in both methodologies used are in accordance with each other - all sorting, static and dynamic advantages are important factors in explaining UWP. There is one potential discrepancy originating from the fact that the analysis of urban-to-rural movers suggests no effect of learning while the other specification finds evidence for significant learning effects. This pattern, however, can still emerge in case human capital accumulated in cities is not transferable to rural areas, which happens if it entails very distinct, city-specific knowledge.

The empirical evidence presented in this paper helps to shed more light on the potential explanations behind the increased productivity of workers in urban locations, and I hope it leads to a more thorough examination of the different mechanisms, especially the dynamic ones, in the future.

REFERENCES

BARTOLUCCI, C., C. VILLOSIO, AND M. WAGNER (2014): "Who migrates and why? Evidence from Italian administrative data," Mimeo, Boston College.

BORJAS, G. J. (1987): "Self-selection and the earnings of immigrants," *The American Economic Review*, 77(4), 531–553.

GLAESER, E. (2011): *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier and happier*. Pan Macmillan.

HELSLEY, R. W., AND W. C. STRANGE (1990): "Matching and agglomeration economies in a system of cities," *Regional Science and Urban Economics*, 20(2), 189–212.

HENDERSON, J. V. (1982): "Evaluating consumer amenities and interregional welfare differences," *Journal of Urban Economics*, 11(1), 32–59.

HICKS, J. R. (1963): *The theory of wages*, vol. 2. Macmillan London.

MARSHALL, A. (1890): *Principles of economics*. London: Mac-Millan.

ROBACK, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90(6), 1257–78.

ROY, A. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3(2), 135–146.

CHAPTER 1

SWEET SUNNY SOUTH: ECONOMIC RETURNS AND SELF-SELECTION INTO NORTH-TO-SOUTH MIGRATION IN NORWAY

Kjell G. Salvanes Darina Steskal Erik Ø. Sørensen[1]

Abstract: We study north-to-south migration in Norway over the past 40 years (1967-2010). In a slow but steady process, people from the less industrialized north moved south to access better economic conditions. We first investigate how the movers from north to south are selected from the north population, examining both the education and the residual earnings distributions. Movers are more educated. Comparing two cohort groups, we find that selection on unobservable characteristics changes from positive to negative over time. Attempting to indirectly control for selection, we estimate mean (real) migration gains of 7% and 15% for early and late cohorts, respectively. Comparing the estimated mean effects of migration using different specifications we confirm positive selection of migrants on unobservable characteristics in the early cohorts, but a pronounced negative selection among later cohort migrants. Attempting to control directly for ability using conscription IQ tests (on a subsample) gives mixed results. We estimate the effects of migration on the earnings distribution of movers. The estimated effects differ across the two cohort groups, suggesting an even effect for the early cohort but a steadily increasing effect for the later cohort.

[1]All authors: Department of Economics, Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. Emails: kjell.salvanes@nhh.no, darina.steskal@nhh.no, and erik.sorensen@nhh.no. We are grateful for excellent research assistance with the household expenditure surveys by Magnus Kvåle Helliesen.

1.1 INTRODUCTION

What are the returns to migration, and are the migrants a positively selected group as compared to those who stayed behind? These are the questions being analyzed for the recent inflow of cross-border migration to Europe from Asia and Africa (Kerr and Kerr, 2011), as well as the historic mass migration from Europe to The New World in the 18th and early 19th centuries (Abramitzky, Boustan, and Eriksson (2012) examine Norwegian mass migration to the US a hundred years ago). A recent parallel to these migrations waves is the large internal migration in Norway following WWII, from rural areas to regional centers, and especially from northern to southern regions of Norway.

This internal migration to better jobs in the southern part of Norway is not dissimilar to the internal migrations that feed the labor intensive export-oriented industries in rapidly growing developing economies today (Murphy, 2008), as well as the large internal black migration from the southern to northern US states where the new manufacturing jobs could be found (Boustan, 2009). As has been documented in this literature, migration led a strong convergence in income between regions. It seems reasonable to assume that the continuing Norwegian southbound migration has been an important ingredient in reducing inequality in Norway as well, but there is no systematic evidence of this.[2]

Southbound migration in Norway has probably been dampened by a strong political consensus to preserve population in rural areas, especially those in northern Norway. From Figure 1.1 we notice that the process is slow, as the population in northern Norway (as defined by the three northernmost counties) had 11.6 percent of the population in 1970, and this share had only been reduced to 9.6 percent of the population 40 years later. But if the decline is slow, it is also steady, as can be seen from the lower panel panel of Figure 1.1 where we present the annual gross southbound migration. The migration rate is quite stable across years, slightly above 2 percent per year.

The fact that migrants are not a random sample from their home country's population has been widely recognized and undisputed, guided by the early model of Sjaastad (1962). In this model migrants are assumed to be income-maximizers, so only those for whom moving is economically beneficial will relocate. Thus, migrants constitute a self-selected sample that can differ from the sending population along multiple dimensions. Several models have been pro-

[2]Sweden went through a similar (but probably more radical process) almost at the same time, but research on the economic consequences of migration covers a later time period (Axelsson and Westerlund, 1998; Nakosteen and Westerlund, 2004; Westerlund, 1997).

posed to predict the nature of the migrant self-selection. One important strand of literature goes back to Borjas (1987) (see Chiswick (1999) for alternative specifications of the migration selectivity models). Building on the Roy (1951) model of occupational choice, Borjas developed a selection model which predicts the selection patterns of movers compared to natives. According to this model, the composition of migrants will depend on the income distribution of the sending and receiving regions – migrants will move to a region that has the most favorable income distribution for their skill level. Thus, regions with higher returns to skills, reflected in larger income inequality, will mostly attract high-skilled workers. Regions that reward skills less, and thus have a narrower income distribution, will mostly attract low-skilled migrants.

While some studies confirm the predictions of the model (Abramitzky, 2009), others put the empirical relevance into question. Loosening the restriction of Borjas that the cost of migration is constant (in time units), it is not clear how many restrictions the Borjas model puts on the selection pattern observed in data. Chiquiar and Hanson (2005) develop a generalized version of Borjas' model, and show that the evidence they find against the Borjas-prediction of negative selection of Mexican migrants to the US can be made consistent with the generalized model.

In this paper we analyze the returns to migration from the north of Norway to the south over almost half a century. We also look at whether these returns are changing over time, and to what degree migration has been a selective process. Our contribution to the literature on migration is that we have population based data set for each individual's locations, earnings, and education per year. This means that we are able to construct pre-move earnings distributions to analyze selection, as well as a long panel of movers and non-movers, which will allow us to analyze changes in returns and selection over time. In addition, we are able to construct a *real* earnings measure that controls for local price variation, which is important in the analysis of migration gains.

We find that movers are positively selected on education. In the beginning of the period they are also selected positively on unobservable skills, but this turns into a negative selection on unobservables for the later cohorts. The negative selection on unobservables is reached by two different approaches – residual earnings analysis and the comparison of estimated mean migration effects using several different specifications (OLS, DiD, FE). The fixed effects estimates of the migration gains are sizable, about 7% for the early and 15% for the late cohorts (controlling for selection). The estimated effects on the whole earnings distribution differ between the early and the late cohorts, implying a homogeneous effect in the early period, but heterogeneous, steadily increasing effect on the distribution in the later period.

The paper is organized as follows: Section 3.3 describes the data and the procedure to construct the local price indices. Section 3.5 contains evidence on migrant selection and the mean and distributional effects of migration with some tests of robustness. Section 3.6 concludes.

1.2 INSTITUTIONAL BACKGROUND AND DATA

As noted in the introduction, the backdrop for southbound migration in the period after World War II is a strong political consensus to preserve population in rural areas especially in northern Norway. The northern-most county, Finnmark, was burned down by German forces as they withdrew towards the end of the war. Just after the war, an extensive program was established to rebuild northern Norway. This program was formalized in 1951 (Gerhardsen, 1951; Handelsdepartementet, 1951), although rebuilding had started earlier. The 'Northern-Norway plan' (*Nord-Norge-planen*), had a fund of about 200 million NOK for support for building houses, infrastructure etc. In addition there were tax breaks for firms in order to attract them to this region. Ten years later this regional support system was extended to include a support system in general for rural areas and tax breaks for firms in rural areas ("Distriktsutbyggingsfondet").[3]

There was a political worry that industries typical of the rural areas were in strong decline (also in Norway), and there was political pressure to preserve a certain level of population density in these areas. For instance the primary sectors (agriculture, fishery, and logging) had a share of 42 percent of the man-years of production in 1930, while it was only 15 percent in 1970. The manufacturing sector had grown from 22 percent of man-years in 1930 to 35 percent in 1972. Administration and the service sector had had a similar increase from 36 percent to 50 percent in 1970. Since large parts of the new and modern manufacturing sector was established in the southern urban areas (Oslo had the largest share of the manufacturing sector in 1970), and the primary sector was obviously concentrated in the rural areas (fishery was strong in the north), large resources were used to attempt reducing the speed of urbanization and depopulation of the north.

In the mid-1970s the rural support system took a new turn when the Parliament decided that the farmers should be guaranteed earnings at the same level as workers in the manufacturing

[3]This discussion draws on Norges offentlige utredninger (1974, 1984a,b, 1991b,a,c).

sector. This led to a dramatic increase in the support to rural areas and changed the trend in the urbanization process. In 1993, the policy changed again with the Parliament withdrawing the guaranty of earnings at the same level as manufacturing workers.

1.2.1 DESCRIPTION OF DATA

Data are compiled from a number of different sources. Our primary data source is the Norwegian Registry Data, a linked administrative data set that covers the population of Norwegians up to 2010 and is a collection of different administrative registers such as the education register, family register, and the tax and earnings register. These data are maintained by Statistics Norway and provides information about educational attainment, labor market status, earnings, and a set of demographic variables (age, gender) as well as information on families.[4] These data are merged with a data set unique to this paper where we have information from 1967 onwards on the exact municipality where people are located in each year. This is the data set we are using for constructing a mobility matrix, and since for every year we can match in earnings and location of residence, we know earnings both pre- and post-move.

Earnings are measured as total pension-qualifying earnings in NOK reported in the tax registry and are available on a calendar year basis. They include labor earnings, taxable sickness benefits, unemployment benefits, parental leave payments, and pensions. Records on earnings are available from 1967 onwards and so we are able to calculate the earnings profiles both for movers and non-movers for several years after moving. Earnings are corrected for inflation to the base year 2010 using the Consumer Price Index (CPI). However, for tracking the welfare of workers as they move from north to south, using the national CPI is not satisfactory since there is a relevant regional variation in price levels (Mogstad, Langørgen, and Aaberge, 2007). Subsection 1.2.3 describes the method used to construct regional price-levels based on the Engel curve for food. The earnings measure is then corrected for these local price variations. Before 1986 (with the exception of the year 1981) earnings are top- coded.[5] To overcome this problem we follow the approach of Bhuller, Mogstad, and Salvanes (2011) and simulate Pareto-distributed earnings above the top-coding.

The education variable is measured as the highest completed level of education during a person's lifetime. We create three education groups that we call low- (primary and lower secondary

[4]See Møen, Salvanes, and Sørensen (2003) for a description of these data.

[5]In years after 1970 less than 1.5 % of observations is top- coded; in years before 1970 less than 6%.

level), medium- (upper secondary school level), and high-educated (tertiary level). We are using the education registry which is based on school reports sent directly to Statistics Norway, thereby minimizing any measurement error due to misreporting.

1.2.2 Sample selection

Our sample contains yearly observations of men who resided in the north of Norway the first year they appear in the database, which contains years 1967-2010. The age is restricted to 25–55, and individuals must have non-missing data on year of birth, gender, and education. In case of missing location data in at least one year we exclude all observations for these individuals in all years, thereby creating a panel data set keeping only persons with continuous time periods with available residence data.[6] The sample is further restricted to stayers and one-time movers; the return and multiple movers are excluded.[7] Stayers are defined as people who resided in the north every year of their spell; movers as those individuals who resided in the north at the beginning of their spell and who changed their location of residence to the south exactly once. The north is defined as municipalities that belong to one of the three northernmost counties (Nordland, Troms, or Finnmark).

Finally, we keep only persons with sufficiently high earnings to support themselves. We therefore exclude all observations with earnings lower than the basic amount, G, during a year. The basic amount (*grunnbeløpet*) is a social security base rate, which serves as a basis for calculating the Norwegian state pension.[8] The *final sample* consists of $3,620,141$ person-year observations on $214,195$ individuals, out of whom $174,988$ lived all their lives in the north (82%) and $39,207$ (18%) moved from north to south once.

The sample is restricted to men to minimize selection issues due to low female participation in the labor market. The reason for restricting the sample to 25 years old and above is to avoid migration induced by the schooling or family decision. Many young people moved to the south to get higher education and not in the pursuit of higher income opportunities. By the age of 25 most of the individuals have finished their education, so that both earnings and migration are not observed during the still-in-education period.

To be able to track the changes of the migration gains and selection over time, some analyses are made on the split sample of two cohort groups, early (born 1940-1950) and late (born 1955-

[6] By excluding all individuals with missing location data in one or more years we lose less than 4% of the data.
[7] Return and multiple movers make up approximately 30% of all movers from north to south.
[8] For example, for year 2010 the base rate was set to 74,721 NOK.

1965). The choice of cohort groups is based on several criteria. First, we want to ensure sufficient records on earnings for ages 25-55. Second, the span of 11 years within each cohort group provides us with a sufficient number of observations. Third, the cohort groups should be spaced from each other at least a couple of years in order to allow us to evaluate the trends over time, and, finally, such a cohort choice will be suitable when assessing the distributional effects in section 1.3.2, where our focus will be on 40-year old individuals, and both cohort groups have observable earnings at the age of 40.

For a subsample, we are able to use a direct measure of cognitive ability (IQ) that we obtain from military records. All male (and only a few female) citizens undergo a medical examination at age 17-18, intended to determine their fitness for military service. At this time, they also take a standardized test of cognitive ability, which is classified into Standard Nine (stanine) units with mean of 5 and a standard deviation of 2. This data is available for persons tested from 1968 onwards (see also Black, Devereux, and Salvanes, 2010).

1.2.3 Comparing income across space

There might be large local variations in price levels in north and south due to differential transportation costs, local labor costs, and varying degrees of local competition. However, official regional price indices at the national level are not common (Ferrari, Laureti, and Jiménez, 2010), and do not exist for Norway. Mogstad et al. (2007) constructed regional price-levels using the price of housing, an approach that is difficult to take to the historical data we are using. Instead we propose an approach based on Hamilton (2001), who showed that regional price indices can be constructed based on the Engel curve for food, a known stable relationship.[9]

Let ω be the food share of expenditures, and let the Engel curve for food for an individual household i in location r be

$$\omega_{ir} = \alpha + \beta \log \frac{y_{ir}}{P_r} + \varepsilon_{ir},$$

where y_{ir} is the nominal income and P_r is the price level of region r. Formulating the regression model

$$\omega_{ir} = \beta \log y_{ir} + \gamma z_{ir} + \delta_r + \varepsilon_{ir}, \tag{1.1}$$

[9]This method has also been applied by Beatty and Røed Larsen (2005) and adapted to international comparisons by Almås (2012).

where z_{ir} is a vector of household demographics, the price level, P_r, can now be recovered up to a normalization as $\widehat{P}_r = \exp(-\widehat{\delta}_r/\widehat{\beta})$.[10] We normalize the average correction from year-specific estimation of (1.1) to unity and use the official CPI to normalize purchasing power across time.

For estimation of the Engel curves using the expression in equation (1.1), we use data from the survey of consumer expenditures, starting with the survey in 1973 (Statistisk sentralbyrå, 1973). We estimate the CPI corrections using household level data, classifying observations into 4 geographical regions and 3 levels of centrality, for a total of 11 regions, since there is no large urban municipality in the northern-most geographic region (Statistisk sentralbyrå, 1989). The southern-most parts of Nordland are not part of "north" using the expenditure survey geographical definition ("handelsfelt").

The results from our estimation are plotted in Figure 1.2 for the years we have expenditure surveys. We see that the food share of the budget has been going steadily down in all regions over time, reflecting a growth in real income. We also see that there are price differences between regions, and the average price level is slightly higher in the regions we classify as "north" than in the southern regions.

The surveys we have do not cover the all years we are interested in. For the remaining years, we inter- and extrapolate a linear trend in CPI corrections. Mapping from the 11 regions we estimate price levels for to municipality is straight-forward for the most central places (Oslo, Bergen, and Trondheim), but not for the distinction between rural and densely populated area. For the distinction between a densely and sparsely populated area, we use official numbers on the share of population in each municipality that lives in a sparsely populated area; when that share is above 50%, we allocate the municipality the rural price index (within geographical region).[11]

1.2.4 Descriptive statistics

Figure 1.3 presents the income distributions in the north and the south in the beginning of the time period we look at, in 1970. The top graph shows estimated densities, which are not easily comparable in the tails. The bottom graph shows QQ-plots of the same data: the k-quantiles in each location are plotted against each other, and deviation from the 45-degree line corresponds

[10] Note that this is a behavioral approximation, since the indirect utility function corresponding to the demand system in equation (1.1) also includes a second price index that cannot be quantified without direct information on prices.

[11] http://www.ssb.no/emner/02/01/10/beftett/arkiv/tab-2000-12-18-04.html.

to differences in distribution. We see from the QQ-plots that it is not the case that the distribution in one location dominates that of the other. In the south, the low-income individuals have higher values, but in the middle of the distribution, the northeners actually have a slight edge over the southerners. According to the variance of log income, inequality is higher in the north (north: 0.56 vs south: 0.46); this difference seems to be much smaller using the Gini-coefficient (north: 0.2896 vs south: 0.2832). Taking the Borjas prediction at face value, this should indicate scope for negative selection of migrants from the north to the south in the beginning of the period.

Table 1.1 compares the education distributions of stayers and movers. As is evident from the table, the share of high-skilled individuals is markedly larger in the group of migrants compared to stayers. This gives evidence on the positive selection of migrants on education. We will return to the issue of selection in section 1.3.1.

To provide descriptive characteristics of our earnings data, we choose to plot the unconditional earnings distributions for movers and stayers at the age of 40 for two boundary cohorts (to allow for a large time difference between them) – cohorts born 1940 and 1965. At the same time we restrict the age at migration to be below 35, so that at the age of 40 movers are observed at least 5 years after the move.

Figure 1.4 shows that for the early cohort, the migrants' unconditional distribution completely dominates that of stayers; migrants earn more than stayers. We also see that the migrants' distribution is closer to that of the stayers in the middle of the distribution; it is the tails that are shifted most upwards among the migrants. These results might either be a reflection of different compositions of movers and stayers (in the case of a positive selection of migrants), or they might indicate a positive causal effect of migration that varies somewhat across the distribution.

Note that for the later cohort, the picture is not quite the same. Indeed, at the lower end of the distributions, it seems that stayers dominate movers (but slightly), and only in the upper two-thirds of the distribution do the earnings of movers dominate those of stayers, more than outweighing the slight advantage of the stayers at the bottom end. This difference between the cohorts might be the result of a change in the selection pattern over time or a change in the causal distributional migration gains.

1.3 RESULTS

First we look at how the migrants select themselves, both in terms of observable and unobservable (residual) characteristics. Then we examine the gains of migration to those that do migrate.

1.3.1 SELECTION INTO MIGRATION

SELECTION ON EDUCATION

We first examine how migrants are selected in terms of education, i.e., if and how the propensities to migrate differ with the education. If migrants were randomly drawn from the population, the propensity scores would be equal across the different skill groups. In the case of negative selection we would see increased propensities to migrate for the low-skilled, and similarly, positive selection would imply higher propensities for the high-skilled. In Table 1.1 we saw that the educational composition of movers differs substantially from stayers, with a much higher share of highly educated. In addition, Table 1.2 shows the propensities to migrate by education category. It splits the sample into early and late cohort groups and examines the patterns when the age of move is restricted to be below 35. The propensities monotonously increase with education level in both groups, with about 40% probability to migrate among the highest educated and only about 10% among the lowest educated. Thus, there is an evidence on the positive selection of migrants on education.

SELECTION ON UNOBSERVABLE CHARACTERISTICS

We now study how the composition of movers differs with respect to unobservable characteristics. Our data allows us to construct a proxy for unobservable skills pre-move, which is a rare opportunity; very few studies have information on migrants before they decide to move. We measure unobservable skills by residual earnings, which are the predicted standardized residuals from the following Mincer equation, estimated by OLS on the sample of stayers and movers *before* the move,

$$\log \mathrm{EARNINGS}_{it} = \alpha + \beta_1 \mathrm{YEAR}_t + \beta_2 \mathrm{EDU}_i + \beta_3 \mathrm{AGE}_{it} + \varepsilon_{it}, \tag{1.2}$$

where EARNINGS_{it} is our outcome variable, and YEAR_t, EDU_i, and AGE_{it} are dummy variables for year, education groups, and age, respectively. Since the residuals are standardized to have unit variance, they can be interpreted as the number of standard deviations that an observation is above or below the mean earnings in a given year for workers with the same age and education level.

Figure 1.5 plots the estimated residuals of movers against stayers for the early and late cohort groups. Among the early cohorts, we see that stayers dominate slightly at the lower end, indicating perhaps some negative selection into migration, outweighed by a much larger proportion in the upper part of the skill distribution, where movers are dominating the stayers. So the stayers have a more compressed residual earnings distribution and the selection is mostly from the tails, with the selection from the right tail dominating the selection from the left tail. Among the late cohorts we see some of the same picture, but the compression is much more pronounced among the lower skilled stayers than we see in the early cohort. In the upper tail, the movers dominate to about equal degrees among the early and the late cohorts. Thus, in this case the negative selection appears to be more prevalent.

1.3.2 GAINS FROM MIGRATION

MEAN EFFECTS

In this section we estimate mean impacts of migration on earnings for the movers from the north by using three different specifications: pooled OLS, DiD and an individual-fixed-effects estimator. By comparing the mean effects from these specifications we are also able to infer the average effect of selection of unobservables and verify its consistency with Figure 1.5. In all the following specifications we restrict age at move to 35 or below.

We first estimate the following regression model by OLS,

$$\log \text{EARNINGS}_{it} = \alpha + \theta_1 \text{SOUTH}_{it} + \beta_1 \text{YEAR}_t + \beta_2 \text{EDU}_i + \beta_3 \text{AGE}_{it} + \varepsilon_{it}, \quad (1.3)$$

where SOUTH_{it} is 1 if a person lives in the south (i.e., movers post-move) and other explanatory variables are the same as before in equation (1.2). The identification strategy in this specification relies on the assumption that movers and stayers are similar in unobserved individual characteristics. The parameter of interest is θ_1 and captures the average effect of migration on movers.

The time dummies are meant to capture the time variation and make the earnings comparable across years.

The results from specification (1.3) are reported in column 1 in Table 1.3. We estimate a large significant positive effect of 12% for both the early and the late cohorts. The second column restricts the sample only to observations at the age of 40, so now the comparison and the treatment groups are stayers and movers post-move, observed at the same age (controlling for time effects and education). Estimated effects are now smaller for the early cohorts (8%), but not much changed for the late cohorts.

Since the group that migrates might be selected in a way that directly impacts the estimated gains from moving, we want to allow for such permanent differences in unobserved skills or talent between groups. First we allow for a fixed effect for those individuals that eventually move, a migrant fixed effect M_i,

$$\log \mathrm{EARNINGS}_{it} = \alpha + \theta_1 \mathrm{SOUTH}_{it} + \theta_2 M_i + \beta_1 \mathrm{YEAR}_t + \beta_2 \mathrm{EDU}_i + \beta_3 \mathrm{AGE}_{it} + \varepsilon_{it}, \quad (1.4)$$

where M_i is 1 if a person is a migrant (i.e., movers both pre- and post-move). This specification can be thought of as a differences-in-differences identification strategy, where the treatment group is the movers, the comparison group is the stayers, and the θ_1 coefficient estimates the treatment effect on the movers.

The results from specification (1.4), reported in column 3 in Table 1.3, show that the mean migration gains have now further decreased to 7% in the early cohort, and increased to 16% for the late cohort. The parameter on migrant status M_i is positive 4.5% for the early cohorts, indicating that the group of movers is on average positively selected on unobserved characteristics. However, for the late cohorts, the coefficient on the dummy for moving is negative, and of about the same magnitude as for the early cohorts, at -4.1%. This indicates that the nature of selection has changed over time; with the migrants negatively selected in the later cohorts. This is consistent with our previous findings on the nature of selection in Section 1.3.1 based on the conditional residual distributions in Figure 1.5.

Finally, we allow for individual fixed effects and estimate the following regression,

$$\log \mathrm{EARNINGS}_{it} = \alpha + \theta_1 \mathrm{SOUTH}_{it} + \beta_1 \mathrm{YEAR}_t + \beta_3 \mathrm{AGE}_{it} + \nu_i + \varepsilon_{it}, \quad (1.5)$$

where v_i are individual fixed effects. The results, in column 4 of Table 1.3 yield very similar results to the difference-in- difference results in column 3, confirming that the OLS results have a positive selection bias in the early cohorts, and a negative selection bias in the late cohorts.

DISTRIBUTIONAL EFFECTS

In addition to the impact on the conditional mean of the earnings distribution, the decision to migrate might have affected the shape of the earnings distribution for migrants. To examine whether migration affected some parts of the earnings distribution differently than others, we will compare the factual and counterfactual earnings distributions of movers, represented by their complementary cumulative distribution function ($\tilde{F}(y|X) = \Pr(Y > y|X) = 1 - F(y|X)$, where $\tilde{F}(y|X)$ is a complementary cumulative distribution function CCDF and $F(y|X)$ is a cumulative distribution function CDF). Specifically, we estimate the change in probability $\tilde{F}_{\text{fac}}(y|X) - \tilde{F}_{\text{counterfac}}(y|X)$ for each threshold value y and study how these changes in probabilities (or population shares with earnings above y) look at different points of the outcome distribution.

The idea behind this method is illustrated in Figure 1.6. Note that we estimate the effects of migration on the earnings distribution of movers, not the distribution of gains to migration. To describe the change from the counterfactual distribution F_1 to the actual, observed distribution F_2 at a given point y^*, we calculate the vertical distance between F_2^* and F_1^* at that point (both distributions represented by the CCDF).

Another widely used method to evaluate distributional effects is the (conditional) quantile regression, which focuses on the change in quantiles conditional on a set of explanatory factors. Conditional quantile regressions can be weighted to provide causal effects on marginal distributions, but not without considerable computational cost (Mata and Machado, 2005). Since migration is a large and discrete change in the level of income, we examine the direct effects on the distribution function instead of the continuous approximations developed by Firpo, Fortin, and Lemieux (2009).

This method is straightforward to implement. For each earnings level y, we create a new outcome dummy variable $Y^y = \mathbb{1}\{Y \geq y\}$, indicating whether earnings are above the threshold level y. Then for each y, we run an OLS regression of the new dependent variable on the control variables, and the estimated effect gives us the change in the probability that earnings are above y. We can then get the impact on the whole distribution function by combining these effects for each y. The underlying identifying assumption in this analysis is that movers and stayers are similar in unobserved individual characteristics.

The regression model can be formulated as

$$\Pr(\log \text{EARNINGS}_{it} > y) = \alpha^y + \theta_1^y \text{SOUTH}_{it} + \beta_1^y \text{YEAR}_t + \beta_2^y \text{EDU}_i + \varepsilon_{it}^y, \tag{1.6}$$

where the variables are defined as before. In this type of analysis we want to be able to identify a post-move period for the movers. To achieve that we restrict the age of move to 35 or less and analyze the stayers' and movers' earnings distributions at the age of 40.[12] Thus, the sample here is based on the *final sample*, defined in the subsection 2.5.1, and contains only 40-year old stayers and movers, who moved no later than at the age of 35. We look at how the earnings distribution of movers, 5-15 years after the move, differs from the earnings distribution these migrants would have had if they had not moved.

The earnings levels y for the counterfactual earnings distribution of movers can be selected in multiple ways. Preferably, we would like to select those earnings levels that divide the distribution into equal population shares, i.e., that correspond to equally spaced percentiles. Since we cannot observe the counterfactual distribution, we approximate the earnings levels by those that correspond to $1, 2, \ldots, 100$ percentiles of the observed earnings distribution of stayers (the results are robust to the choice of this baseline distribution).

Figure 1.7 plots the estimated distributional effects and 95% confidence intervals. The effects differ across the two cohort groups substantially. For the early cohorts the effects are homogeneously positive and significant over most of the inner support of the earnings distribution, fluctuating around 6%. For the later cohort, on the other hand, the effects are steadily increasing from very small and barely significant in the left tail to almost 12% and strongly significant in the right tail. These results suggest that for the early cohorts migration had mostly a shift effect on the earnings distribution of movers and the distribution to a large extent retained its overall shape. For the later cohorts, the shape of the distribution changed, creating a less steep CCDF. Translated into densities, migration shifts the earnings density to the right and, for the more recent cohorts, makes the upper tail longer.

The results are in line with the conclusions based on the descriptive earnings distributions we saw in figure 1.4. Obviously, it is interesting primarily given the identifying assumption

[12] Approximately 82% of movers moved between ages 25 and 35.

of no systematic unobserved differences between stayers and movers. Otherwise this uneven effect might just reflect the change in the selectivity pattern of movers over time.

ROBUSTNESS ANALYSIS: CONTROLLING FOR COGNITIVE ABILITY

In Section 1.3.2 we examined whether the estimated mean effects are robust to including group- or individual fixed effects. Such fixed effects absorb group differences in inherent talent, but identification of mean effects becomes crucially dependent on the timing of migration. Since the dynamics around time of migration can be complicated (Bütikofer, Salvanes, and Steskal, 2015; Steskal, 2015), it could be beneficial to use more direct measures to control for unobserved skills and abilities. One such measure, available for a subset of the data, is the IQ test administered at the time of military conscription.

In Table 1.4 we report variant estimates of the mean gains from moving, based on equation (1.3), first for the unrestricted sample of late cohorts, then for the subset of about 80% that we have been able to match to military records. A striking feature finding in this table is that returns to moving, not controlling for IQ, are considerably higher in the conscription subset that can be matched to the conscription data (comparing column 1 with 2, and column 4 with 5). This indicates that the subset that cannot be matched is negatively selected. On the other hand, restricting attention to the sample that can be matched, estimated gains from migration are slightly lower when we control for IQ (comparing column 2 with 3, and column 5 with 6). This is in contrast to what we found for the late cohorts when introducing individual level fixed effects (in Table 1.3). If we use the difference in estimated coefficient as measure of the amount of selection, the magnitude is about half of that we estimated using fixed effects, indicating that much of the negative selection observed in Table 1.3 might be due to movers that we cannot match to the conscription IQ tests.

It is also still the case that the estimated mean gains are larger with the IQ controls than all the estimates for mean gains for the early cohorts found in Table 1.3. The patterns are the same whether one looks at the whole period or just at earnings at age 40.

1.4 CONCLUDING REMARKS

In this paper we analyze the returns to moving as well as the selection of workers moving from the north to the south of Norway over almost the last half century. We use population wide register data to track the mobility at the individual level, and we use regional price indexes to calculate real returns. The population wide data makes it possible to construct the pre-move counterfactual distributions. We use DiD and fixed effects in order to help identify selection on unobservables. We also contribute to the literature on migration by analyzing changes in selection over a long time period, and also analyzing the heterogeneity in returns to mobility.

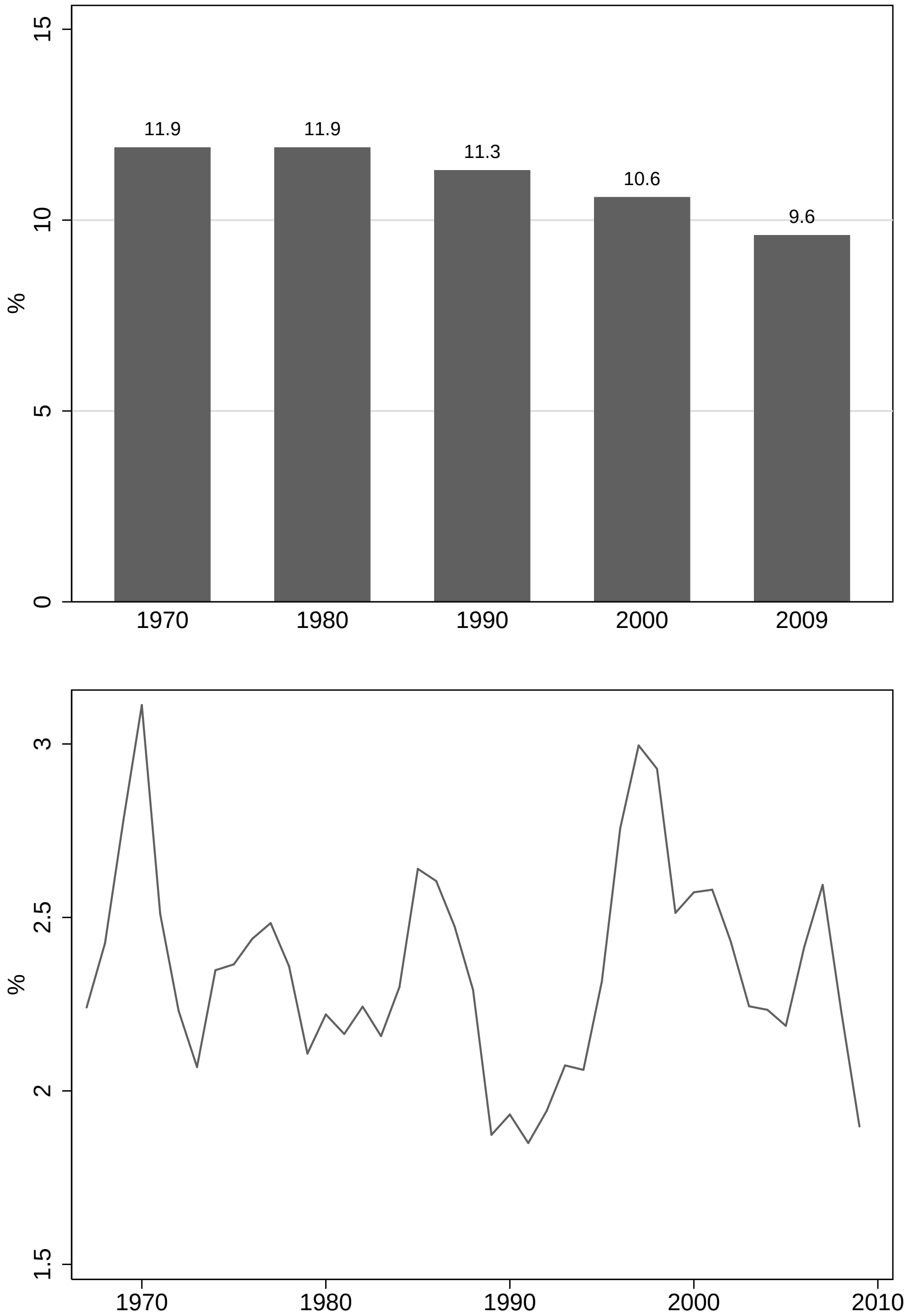
We find that movers are generally positively selected on observables or more educated, while we find a change in the selection on unobservables over time from positive to negative. We find substantial gains from migration; mean real returns purged of selection are 7% and 15% for early and late cohorts, respectively.

Our results are comparable to a recent study analyzing migration from the south of Italy to the north (Bartolucci et al., 2014), who find negative selection to northbound migration in Italy. Italy has the same difference as in Norway although the north is the most industrial region in Italy. When controlled for selection, the returns to migration to the north over the last 20 years are of a similar size as the returns for the recent cohorts in Norway. As found for Norway for the later cohorts, the selection on unobservables is negative. For Norway, due to the long panel series of data, we were able to estimate the change in selection over time, and it went from strongly positive to negative.

We also estimate the effects of migration on the entire earnings distribution of movers. The estimated effects differ across the two cohort groups, suggesting an even effect for the early cohort but heterogeneous, steadily increasing effect for the later cohort, or more precisely upper tail longer for the more recent cohorts. It is likely that this reflects the change in selection over time.

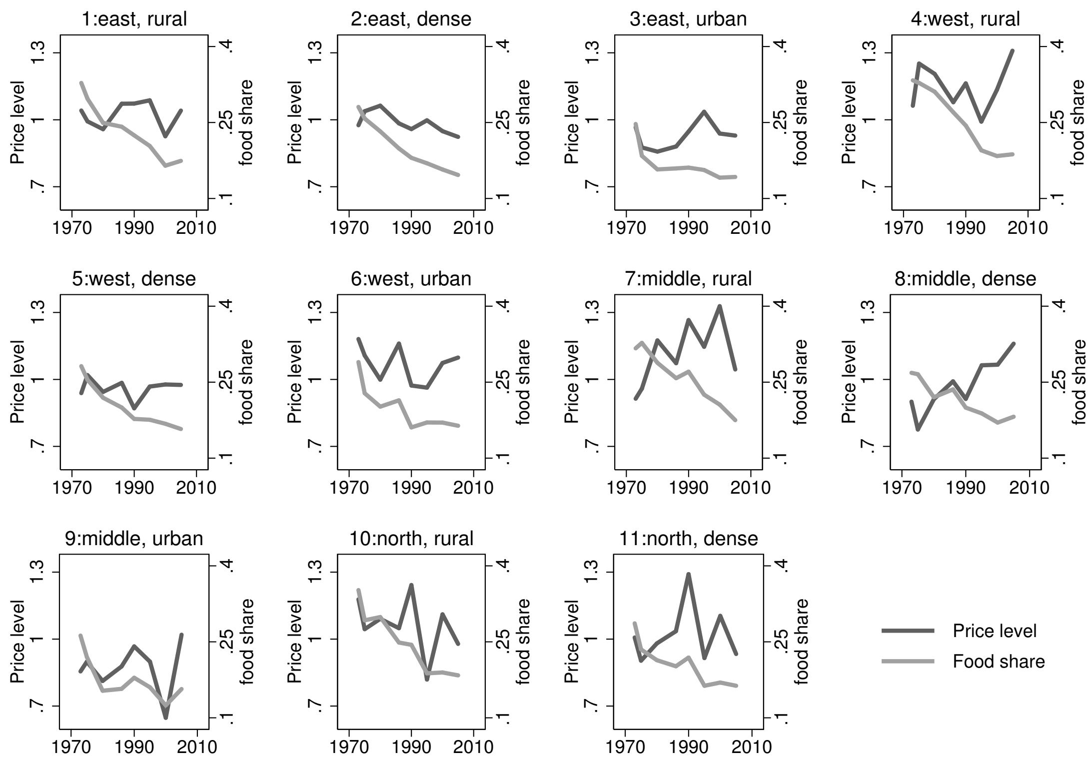
FIGURES AND TABLES

Figure 1.1: Population shares in the north and migration rates from north to south



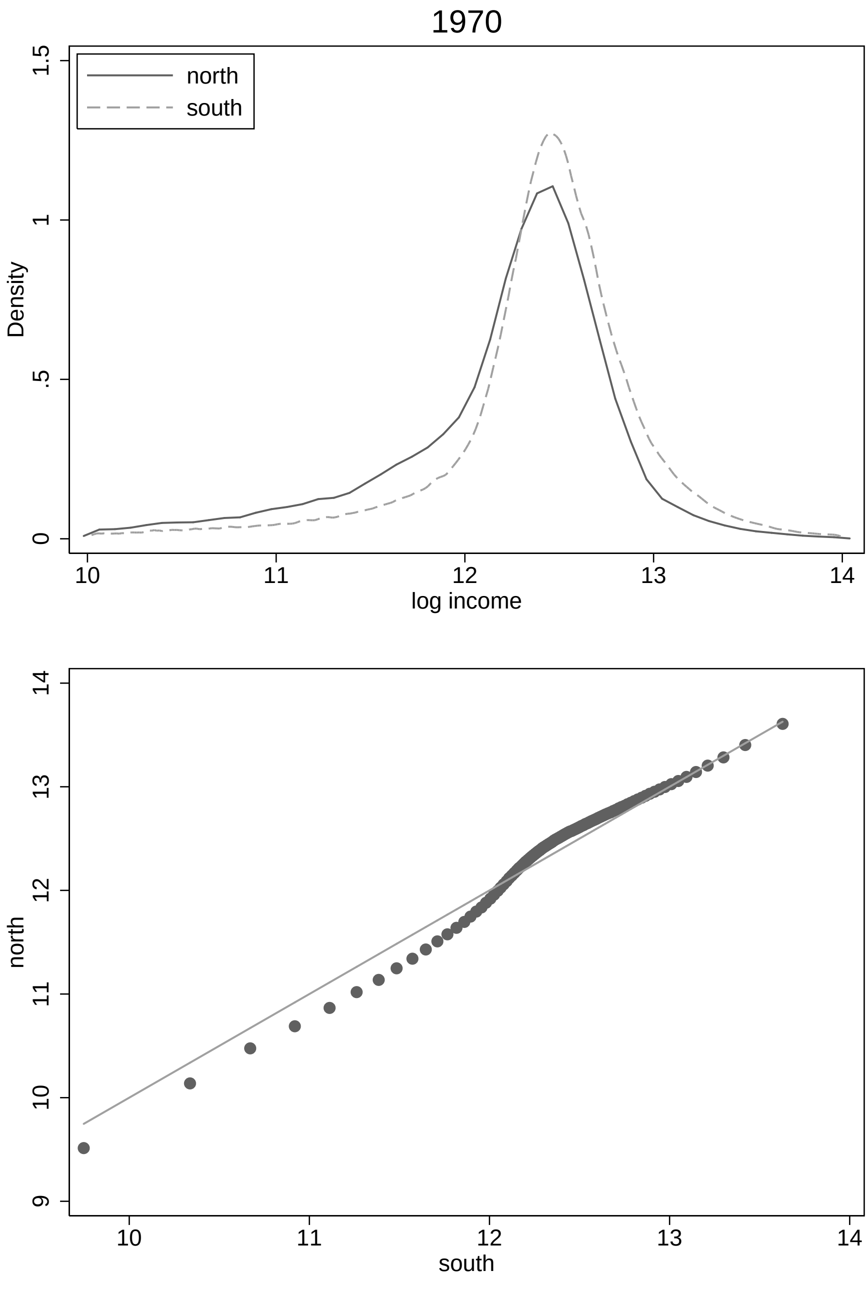
Notes: Migration rates in a given year are calculated as a share of northern population that resided in the south on January 1st the following year. The north in a given year contains those individuals who resided in either of the counties Nordland, Troms or Finnmark on January 1st that year, south the rest. The sample contains all men who were resident in Norway in at least one of the years 1967-2010 at the age between 25 and 55, and have non-missing data on year of birth, gender, and education and location in each year during their spell.

Figure 1.2: Food shares and price levels



Note: The figure plots relative price levels (scaled to a unity national average) on the left axis and the food share of expenditures on the right axis for the 11 regions identified in the consumer expenditure surveys.

Figure 1.3: Income distributions in the north and in the south in 1970



Notes: "North" contains those individuals who resided in either of the counties Nordland, Troms or Finnmark on January 1st 1970, "south" the rest. The sample contains all men who were resident in Norway in 1970 at the age between 25 and 55, and have non-missing data on year of birth, gender, education, and location for each year during their spell (not only 1970). The top plot shows estimated density distributions, the lower shows QQ-plots of the earnings distribution: the figures plot the *k*-quantile in the (log) earnings distribution of movers against the *k*-quantile in the same distribution for stayers. The 45-degree line indicates equality of distribution.

Table 1.1: Education distribution for stayers and movers [in %]

	stayers	movers
low-educated	40	18
medium-educated	46	39
high-educated	14	43
	100	100
shares	82	18

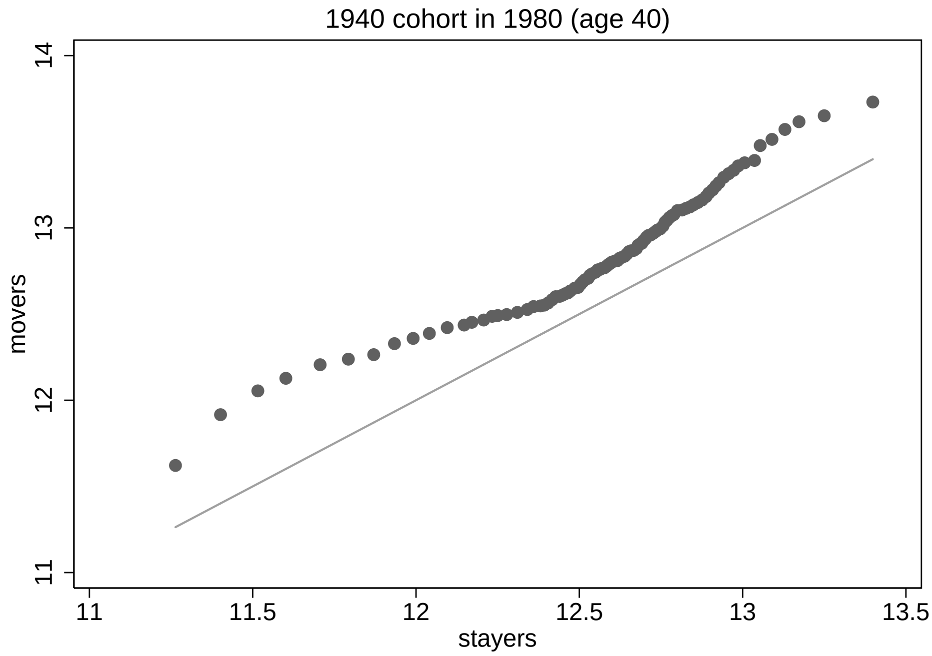
Notes: The sample used is the *final sample* of men aged 25-55 (see subsection 2.5.1 for definition). Stayers are defined as people who resided all their lives in the north; movers are those who originated in the north and moved to the south exactly once during their occurrence in the panel. Education is measured as the highest completed level of education during a person's lifetime and split into 3 education categories.

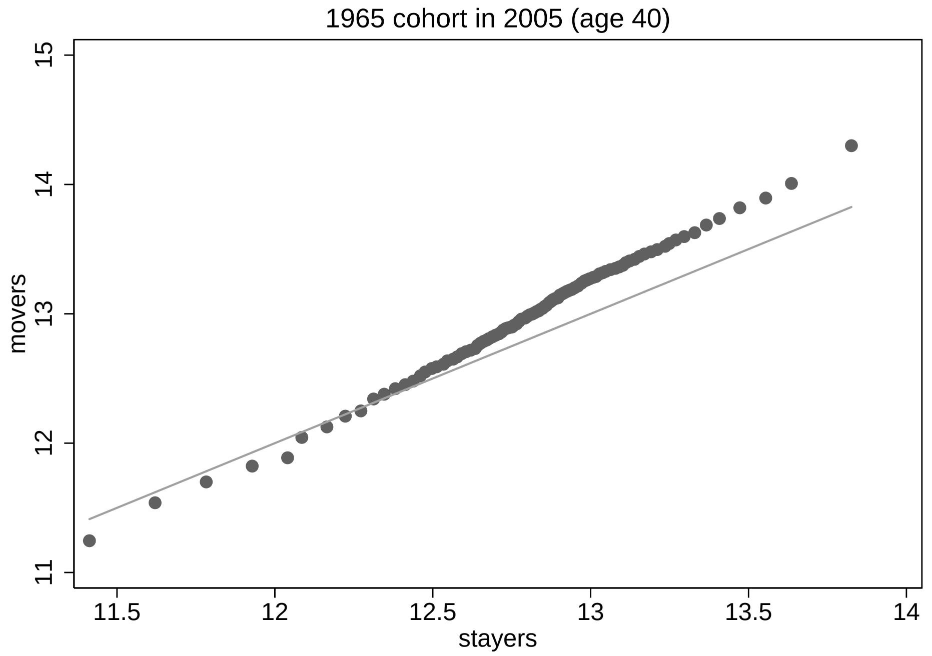
Table 1.2: Probability of move by education group [in %]

		moved at age	
Cohorts		any	35 or less
1940-1950			
	low-educated	11	8
	medium-educated	20	15
	high-educated	43	39
1955-1965			
	low-educated	17	11
	medium-educated	19	14
	high-educated	41	36

Notes:. Movers are defined as those who originated in the north and moved to the south exactly once during their occurrence in the panel. The sample used is the *final sample* of men aged 25-55 (see subsection 2.5.1 for definition).

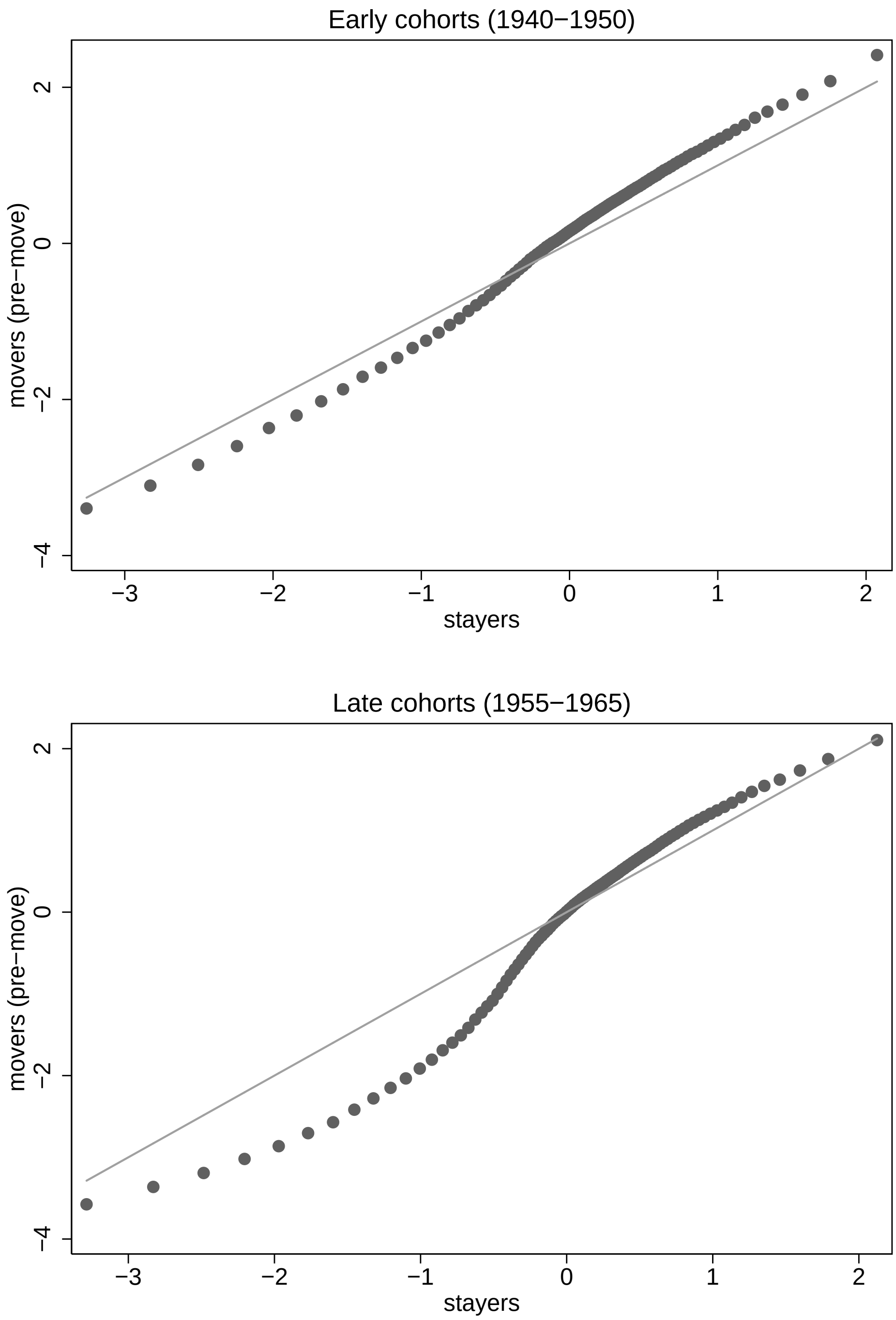
Figure 1.4: Unconditional earnings distributions





Notes: QQ-plots of earnings distribution: the figures plot the *k*-quantile in the (log) earnings distribution of movers against the *k*-quantile in the same distribution for stayers. The 45-degree line indicates equality of distribution. Stayers are defined as people who resided in the north every year of their occurrence in the panel; movers originated in the north and moved to the south exactly once during their occurrence in the panel. The sample used is the *final sample* of men (see subsection 2.5.1 for definition) restricted to be 40 years old in 1980 and 2005, respectively, with the age at move at most 35. Income is measured as total annual pension-qualifying earnings reported in the tax registry in NOK (1998) and is corrected for local price variations.

Figure 1.5: Conditional residual distributions



Notes: QQ-plots of predicted standardized residuals from a Mincer regression, equation (1.2), the figures plot the k-quantile in the (log) earnings distribution of movers against the k-quantile in the same distribution for stayers. The 45-degree line indicates equality of distribution. The *final sample* of men aged 25-55 (see subsection 2.5.1 for definition) is restricted to contain only stayers and movers *before the move*. The age at move is restricted to 35 or less. Stayers are defined as people who resided in the north every year of their occurrence in the panel; movers are those who originated in the north and moved to the south exactly once during their occurrence in the panel. Income is measured as total annual pension-qualifying earnings reported in the tax registry and is corrected for local price variations.

Table 1.3: Estimates of mean effects

	(1) OLS	(2) OLS age 40	(3) DiD	(4) FE
early cohorts 1940-1950				
South	0.116***	0.0827***	0.0714***	0.0698***
	(0.00129)	(0.00653)	(0.00341)	(0.00274)
Mover			0.0451***	
			(0.00323)	
N	875,014	29,237	875,014	875,014
R^2	0.164	0.108	0.165	0.149
late cohorts 1955-1965				
South	0.122***	0.116***	0.162***	0.150***
	(0.00140)	(0.00628)	(0.00326)	(0.00258)
Mover			-0.0411***	
			(0.00303)	
N	769,619	31,197	769,619	769,619
R^2	0.175	0.124	0.176	0.223

Notes: The estimated regressions are defined in subsection 1.3.2. The sample used is the *final sample* of men aged 25-55 (see subsection 2.5.1 for definition). Earnings are corrected for price levels. The second column restricts the sample to 40-year old men only. Movers are defined as those who originated in the north and moved to the south exactly once during their occurrence in the panel. The age at move is restricted to be 35 or less. Standard errors in parentheses. $^{*}: p < 0.05, ^{**}: p < 0.01, ^{***}: p < 0.001$.

Table 1.4: Controlling for IQ on the conscription subset

	All ages			At age 40		
	sample	IQ sample		sample	IQ sample	
South	0.122***	0.154***	0.134***	0.116***	0.148***	0.128***
	(0.00140)	(0.00157)	(0.00157)	(0.00628)	(0.00701)	(0.00700)
IQ control	No	No	Yes	No	No	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
N	769,619	611,026	611,026	31,197	24,870	24,870
R^2	0.175	0.187	0.199	0.124	0.131	0.149

Notes: This table examines estimate specification in equation (1.3) for the late cohorts (1955-65) for the sample of males, and for the sample of males that can be linked to conscription data on IQ scores. The other controls are indicators for the calendar year, education, and age (in the first three columns). Standard errors in parentheses. $^{*}: p < 0.05, ^{**}: p < 0.01, ^{***}: p < 0.001$.

Figure 1.6: Illustration of the method measuring the distributional effects

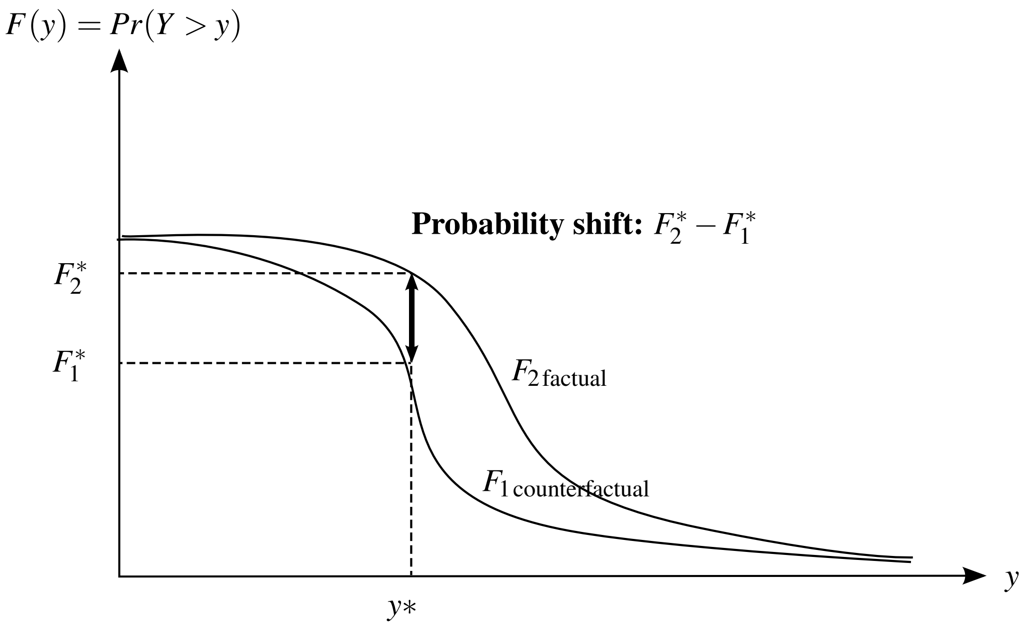
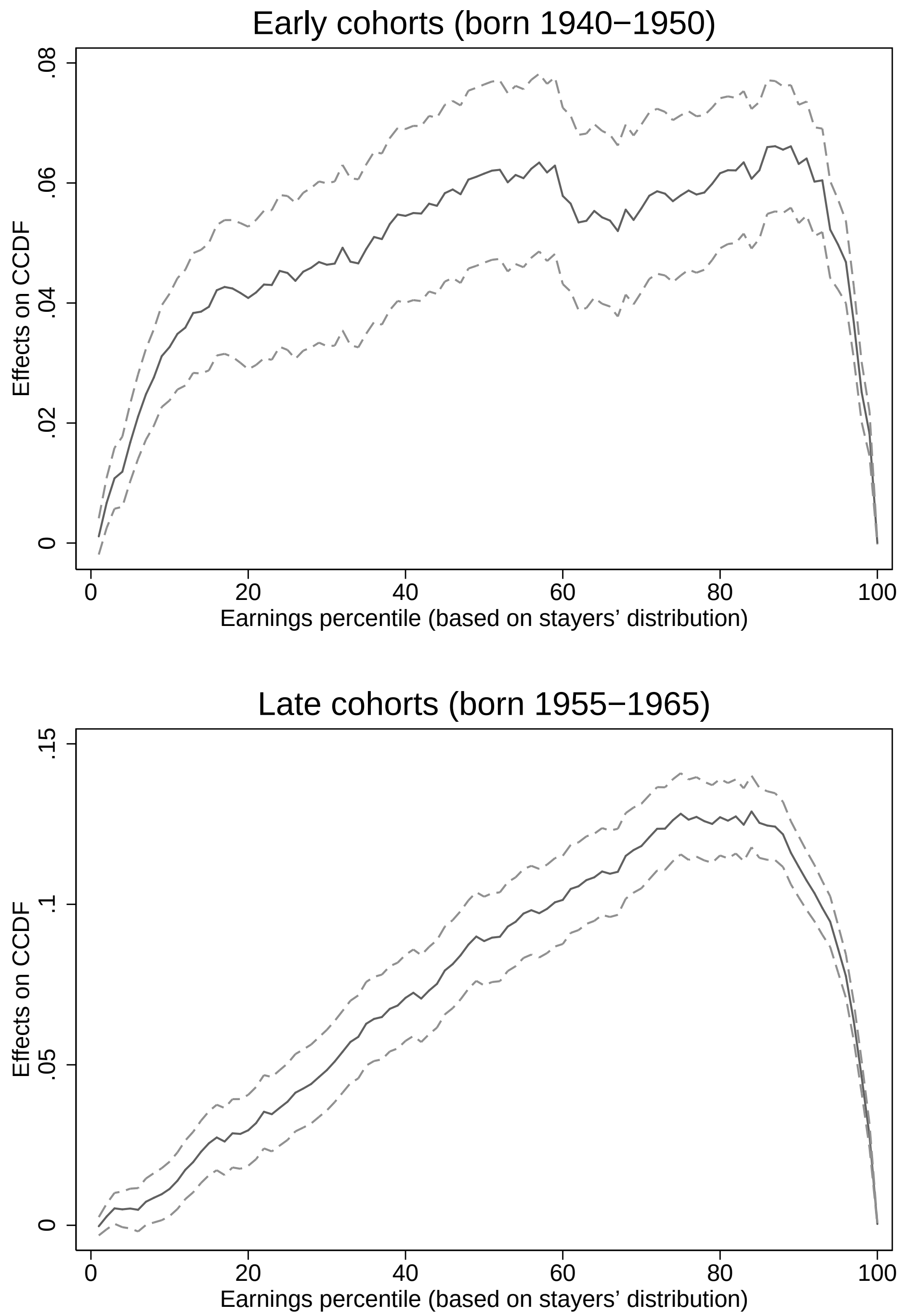


Figure 1.7: Estimated effects of migration on the distribution of earnings



Notes: The figures plot marginal effects on the probability of earning above a certain level represented by 1-100 percentiles of a chosen distribution (the distribution of stayers), together with 95% confidence intervals. The estimated regressions are defined in section 1.3.2. The *final* sample is restricted to 40-year old men. Stayers are defined as people who resided all their lives in the north; movers are those who originated in the north and moved to the south exactly once during their occurrence in the panel. The age at move is restricted to 35 or less.

REFERENCES

Abramitzky, R. (2009). "The effect of redistribution on migration: Evidence from the Israeli kibbutz," *Journal of Public Economics*, 93(3): 498–511.

Abramitzky, R., L. Boustan, and K. Eriksson (2012). "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration," *American Economic Review*, 102(5): 1832–1856.

Almås, I. (2012). "International income inequality: Measuring PPP bias by estimating Engel curves for food," *American Economic Review*, 102(2): 1093–1117.

Axelsson, R. and O. Westerlund (1998). "A panel study of migration, self-selection and household real income," *Journal of Population Economics*, 11(1): 113–126.

Bartolucci, C., C. Villosio, and M. Wagner (2014). "Who migrates and why? Evidence from Italian administrative data," Mimeo, Boston College.

Beatty, T. K. M. and E. Røed Larsen (2005). "Using Engel curves to estimate bias in the Canadian CPI as a cost of living index," *Canadian Journal of Economics*, 38(2): 482–499.

Bhuller, M., M. Mogstad, and K. G. Salvanes (2011). "Life-cycle bias and the returns to schooling in current and lifetime earnings," IZA Discussion Paper No. 5788.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2010). "Small family, smart family? Family size and the IQ scores of young men," *Journal of Human Resources*, 45(1): 33–58.

Borjas, G. (1987). "Self-selection and the earnings of immigrants," *The American Economic Review*.

Boustan, L. P. (2009). "Competition in the promised land: Black migration and racial wage convergence in the North, 1940–1970," *The Journal of Economic History*, 69(3): 756–783.

Bütikofer, A., K. G. Salvanes, and D. Steskal (2015). "What is the city but the people? Changes in urban wage premium and migrant selection since 1967," mimeo, Norwegian School of Economics, Bergen.

Chiquiar, D. and G. H. Hanson (2005). "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113(2): 239–281.

Chiswick, B. (1999). "Are immigrants favorably self-selected?" *The American Economic Review*, 89(2): 181–185.

Ferrari, G., T. Laureti, and J. M. Jiménez (2010). "Harmonized cross region and cross country CPI time-space integration in the Euro-zone: Three characters in search of an author," in L. Biggieri and G. Ferrari (eds.), "Price Indexes in Time and Space," Physica-Verlag, pp. 43–59.

Firpo, S., N. Fortin, and T. Lemieux (2009). "Unconditional quantile regressions," *Econometrica*, 77(3): 953–973.

Gerhardsen, E. (1951). "Innlegg i trontale- og finansdebatt," Stortingstidende, p. 634–639.

Hamilton, B. W. (2001). "Using Engel's law to estimate CPI bias." *American Economic Review*, 91(3): 619–630.

Handelsdepartementet (1951). "Utbyggingsprogrammet for Nord-Norge," Stortingsmelding nr. 85.

Kerr, S. P. and W. R. Kerr (2011). "Economic impacts of immigration: A survey," *Finnish Economic Papers*, 24(1): 1–32.

Mata, J. and J. A. F. Machado (2005). "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(4): 445–465.

Møen, J., K. G. Salvanes, and E. Ø. Sørensen (2003). "Documentation of the linked employer–employee data base at the Norwegian School of Economics, manuscript," *Norwegian School of Economics, Bergen*.

Mogstad, M., A. Langørgen, and R. Aaberge (2007). "Region-specific versus country-specific poverty lines in analysis of poverty," *Journal of Economic Inequality*, 5(1): 115–122.

Murphy, R. (2008). *Labour migration and social development in contemporary China*, volume 2, Routledge.

Nakosteen, R. and O. Westerlund (2004). "The effects of regional migration on gross income of labour in Sweden," *Papers in Regional Science*, 83(3): 581–595.

Norges offentlige utredninger (1974). “Støtteordninger i landbruket,” Norges offentlige utredninger 1974:26.

Norges offentlige utredninger (1984a). “Statlig næringsstøtte i distriktene (bygdeutvalget),” Norges offentlige utredninger 1984:21A.

Norges offentlige utredninger (1984b). “Vedlegg til utredning om statlig næringsstøtte i distriktene (bygdeutvalget),” Norges offentlige utredninger 1984:21B.

Norges offentlige utredninger (1991a). “Norsk landbrukspolitikk: utfordringer, mål og virkemidler – hovedinnstilling,” Norges offentlige utredninger 1991:2B.

Norges offentlige utredninger (1991b). “Norsk landbrukspolitikk: utfordringer, mål og virkemidler – sammendrag,” Norges offentlige utredninger 1991:2A.

Norges offentlige utredninger (1991c). “Norsk landbrukspolitikk: utfordringer, mål og virkemidler – vedlegg,” Norges offentlige utredninger 1991:2C.

Roy, A. (1951). “Some thoughts on the distribution of earnings,” *Oxford Economic Papers*, 3(2): 135–146.

Sjaastad, L. A. (1962). “The costs and returns of human migration,” *The Journal of Political Economy*, 70(5, Part 2): 80–93.

Statistisk sentralbyrå (1973). “Forbruksundersøkelse 1973,” Norges Offisielle Statistikk A705.

Statistisk sentralbyrå (1989). “Standard for handelsområder: korrigert 1989,” i serien Standarder for norsk statistikk.

Steskal, D. (2015). “How cities make us more productive: Empirical evidence from Norway,” mimeo, norwegian School of Economics, Bergen.

Westerlund, O. (1997). “Employment opportunities, wages and interregional migration in Sweden 1970-1989,” *Journal of Regional Science*, 37(1): 55–73.

CHAPTER 2

WHAT IS THE CITY BUT THE PEOPLE? CHANGES IN URBAN WAGE PREMIUM AND MIGRANT SELECTION SINCE 1967

Aline Bütikofer Kjell G. Salvanes Darina Steskal[1]

Abstract: We analyze the urban wage premium and selection into rural-to-urban migration in Norway from 1967 until today. As the urbanization process in Norway happened comparatively late and in a time period observable in the Norwegian population registry data, exploring changes in the urban wage premium and selection into rural-to-urban migration over time in Norway is of particular interest. We find that there is a substantial wage premium for working in cities and that a significant part of this premium is due to positive sorting of workers on observable, as well as unobservable characteristics. Over time, the urban wage premium decreased, and sorting became less pronounced. Also the selection into migration changed during the observed time period; we find that the migrants are positively selected from the sending population on characteristics such as education or IQ score, the positive selection is however less pronounced in 2010 than it was in 1970. When comparing movers' wages to rural stayers' wages before they relocate or the wages of brothers where one brother moved to the city and one stayed in the rural area, we find some evidence that the selection of rural-to-urban movers changed from the positive selection in the 1970s to a more negative selection in later decades.

[1] All authors: Department of Economics, Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. Emails: aline.butikofer@nhh.no, kjell.salvanes@nhh.no, and darina.steskal@nhh.no.

2.1 INTRODUCTION

Migration from rural to urban areas has offered an opportunity to economic advancement. Today, more than half the world's population lives in urban areas. In more developed regions, the share of urban population has already reached three quarters and there are no indications that this growth in the share of urban population will be reversed (UN, 2006). Cities are attractive as they offer better opportunities, and higher incomes. The income gap between rural and urban areas has been well documented for a long time (Roback, 1982; Henderson, 1982; Helsley and Strange, 1990) and also for Norway, a strong positive relationship between population size and average annual earnings for the urban areas is not debatable (see Figure 2.1). Although many empirical studies support that the wages are higher in urban areas, there is still the open question whether cities, for various reasons, make workers more productive, or whether there is positive sorting of inhabitants in cities.[2]

In this paper, we study the urban wage premium and the rural-to-urban migration from 1967 until today. First, we analyze how large the urban wage premium is and how population sorting contributes to the size of the urban wage premium. In a next step, we study whether cities attract a positive or a negative selection of workers from the rural population, and how this selection into cities evolved over time as the level of urbanization increased. Norway offers a unique laboratory for exploring the changes in the urban wage premium over time, and to analyze the selection into rural-to-urban migration over time. The urbanization process happened comparatively late and mostly from the late 1960s until 2010, which coincides with the time period we observe in our data. Whereas the urbanization rate in Norway was about 50 percent in the late 1960s, other Western countries including Sweden reached this level of urbanization in the 1920s (see Figure 2.2). After 1960, the Norwegian urbanization rate increased substantially until 1970. Since 1970, the urbanization rate increased at a slower pace and today the urbanization rate is about 80 percent as in most Western countries. As our register based panel data include annual information on the municipality of residence, the data allow us to trace individual mobility and link it to education and earnings. In addition, we are able to link individuals to their parents and siblings enabling us to compare the outcome of the brothers where one sibling moved to the city and one stayed in the rural area. Moreover, the data include IQ test scores

[2] Although nominal wages are higher in urban areas, this does not necessarily imply that people in urban areas are financially better off in real terms. The income correcting for local cost of living might even be lower in cities (see Glaeser and Gottlieb, 2009).

from Norwegian military records for cohorts born after 1949 which may serve as a control for cognitive ability.

A major challenge when measuring the urban wage premium is to identify whether there is a real productivity premium from living in a city or whether there is positive sorting of people in cities. To evaluate the real productivity effect, we first exploit the panel structure of our data and estimate a fixed effects model which eliminates time-invariant unobserved characteristics. Alternatively, we use IQ test scores as a proxy for the unobserved ability of urban and rural residents. Whereas the urban wage premium documents the static productivity premium from living in a city, we also aim to document what selection of individuals relocates from rural to urban areas. As our data include all moves originated within Norway, we are able to analyze who self-selects into migration. The literature both on international and within-country migration finds support for positive as well as negative selection depending on the relative returns to skills in the regions of out- and in-migration (see, e.g., Borjas, 1987; Gould and Moav, 2010; Abramitzky, Boustan, and Eriksson, 2012). We first study the income inequality in the rural and urban areas to make predictions about the migrant selection and later examine whether the results are consistent with these predictions. We then compare characteristics such as education, age and IQ scores of rural-to-urban movers to rural stayers. Finally, we use migrants' brothers who remained in the rural area and compare their earnings to their brothers' earnings who relocated to the city. Assuming that ability is highly correlated among brothers, this identification strategy allows us to disentangle the unobserved abilities of migrants from the decision to move to the city. We document the urban wage premium and the selection into migration for different time periods and identify potential changes over time.

When pooling all years, we find that there is a substantial urban wage premium in Norway and that the urban population is positively sorted on observable and unobservable characteristics: in particular, conditioning on observable factors such as education and age reduces the mean urban wage premium from about 16 percent to about 13 percent; using a fixed effect estimator reduces the mean urban wage premium further to about 7 percent. The mean wage difference between rural and urban workers decreased steadily over time, where the raw premium dropped from more than 26 percent in the late 1960s to below 10 percent in 2010 (from 22 to 7 percent with controls). Comparing rural stayers with rural-to-urban migrants, we find that movers are on average more educated, significantly younger and they also have higher IQ scores than the rural stayers. In addition, we find support for a change in the selection of migrant households since the late 1960s. Whereas rural-to-urban migrants came from positively selected households in the 1970s, migrants appear to come from more negatively selected households today.

The reminder of this paper unfolds as follows. Section 2 discusses the related literature on the urban wage premium and migrant selection. Section 3 gives a short overview of the urbanization process in Norway. Section 4 describes our empirical strategies and Section 5 describes our data. Section 6 presents the results and Section 7 concludes.

2.2 Literature on Urban Wage Premium and Migrant Selection

2.2.1 Urban Wage Premium

A large body of literature shows that wages and productivity are higher in larger cities. However, estimates of the urban wage premium vary widely both across and within countries depending on the agglomeration variable and dataset: Combes, Duranton, and Gobillon (2008) find a wage elasticity of 4.9 percent in France with respect to employment density. Tabuchi and Yoshida (2000) find an elasticity of 10 percent with respect to the Japanese Standard Metropolitan Employment Area population (in real terms, the elasticity is negative). For the U.S., Glaeser and Mare (2001) find that in the MSAs containing at least one municipality with more than 500,000 inhabitants earnings are 24 to 28 percent higher than in rural areas. Furthermore, they find that the urban wage premium is reduced to about 4.5 percent after controlling for individual-specific fixed effects. Similarly, Yankow (2006) obtains a 19 percent wage premium in the MSAs with more than 1 million inhabitants falling to 6 percent after removing the time-invariant unobserved heterogeneity. Consequently, the empirical evidence shows that there is a remaining premium of living in a city even after controlling for individual fixed effects. There is however very little literature documenting the changes in the urban wage premium over time. An exception is found in Baum-Snow and Pavan (2012) who mention that the city size wage gap in the US became considerably more pronounced from 1980 until 2000: in 1980 large MSAs of over 1.5 million people had wages that were 24 percent higher than rural areas and small MSAs of less than 250 thousand people. In 2000, the wage gap was eight percent points larger.

But why do firms locate in cities where wages are higher and why do not all workers move to cities? Economic theory and empirical evidence present some explanations: Glaeser and Mare (2001) and Gould (2007) argue for example that the latter question can be explained by Roback's (1982) model. That is, nominal wage differentials do not seem to be the only reason why people move. Certain city-characteristics strongly influence a household's choice for migration. In particular, some people might have distaste for higher prices and certain non-market amenities of cities such as pollution, congestion, and crime rates, whereas others value positive aspects such as quality of life through the local schooling system as well as social and cultural activities. Thus, the workers' compensation includes a set of monetary and non-monetary benefits through the consumption of local amenities. If higher ability workers tend to value certain amenities of the city, such as cultural activities, more than low ability workers, better educated workers might be more prevalent in higher density areas (see, e.g., Gould, 2007). The question why firms locate in cities is more difficult to answer. The literature points out several possible reasons why firms might be willing to pay higher wages in cities. Transportation costs from nearby suppliers are lower, favorable geographical features such as a port or a bridge on a river might increase the marginal product in cities and better factors of production such as public or private capital, higher local demand, and local institutions might affect the marginal product positively (see, e.g., Krugman, 1991). Nevertheless, many firms located in cities and paying higher wages do not rely on local suppliers and local demand or geographical or natural advantages. Hence, the higher wages in cities might be explained by higher productivity. There are several different theories suggesting mechanisms why workers in cities are more productive. Higher productivity may occur if (i) workers in cities more easily find the most suitable job (Wheeler, 2001), (ii) human capital accumulation (i.e., 'learning') in cities is fostered through social interactions (Glaeser, 1999) and facilitated knowledge spillovers (Lucas, 1988), or (iii) workers in cities have higher abilities. Hence, the urban wage premium might partly reflect spatial differences in the skill composition of the workforce. It is therefore important to understand sorting of workers into cities and the selection of the rural-to-urban migrants.

2.2.2 Selection Into Migration

As noted above, the selection into rural-to-urban migration is an important issue in understanding the development of the urban wage premium. Following the early work of Sjaastad (1962), the economic literature assumes migrants to be income-maximizing and only those for whom moving is economically beneficial will relocate. Thus, the decision to migrate is based on a rational consideration of the relevant costs and benefits of a location change. Hence, migrants are self-selected and not a random sample from their sending region's population.

Several models have been proposed to predict the nature of the migrant's self-selection. Modifying the classic model of occupational choice by Roy (1951), Borjas (1987) develops a model in which migrant selection is determined by the relative return to skills in the sending and destination economies. That is, if the level of income inequality is smaller in the destination country than the sending country, the return to skill is lower in the destination country than the sending country. Hence, migrants will be drawn to a large extent from the lower end of the sending country's skill distribution. Migrants will however be drawn to a large extent from the upper tail of the source country's skill distribution if the level of income inequality is larger in the destination country. The model predictions are quite intuitive: individuals with high skills benefit from a more dispersed wage distribution as they are at the top of the skill distribution. On the other hand, lower skilled individuals benefit from a more compressed wage distribution.[3]

The model was tested empirically in many countries and several findings support the model's predictions. Gould and Moav (2010), for example, find that the returns to observable skills such as education are lower in Israel than in the United States and that this difference in returns to skills encourages higher ability Israelis to leave the country for the United States. Also Borjas (2008) finds support for the Roy model when studying migration from Puerto Rico to the United States and vice versa. He finds that Puerto Rico attracts high-skill in-migrants, whereas low-skill workers leave the county reflecting the higher income inequality in Puerto Rico compared to the United States. There is, however, evidence that the Roy model fails to predict the immigrant selection: Although Mexico has a higher level of income inequality than the United States, Chiquiar and Hanson (2005) find that Mexican immigrants in the United States are more educated than non-migrants in Mexico and that Mexican immigrants would be concentrated in the middle of Mexico's wage distribution. The authors suggest that migration costs systematically vary with the skill-level and wage of the workers potentially affecting the selection of migration. More specifically, they argue that Mexican migrants to the US are selected from the middle of the wage distribution since the migration costs are very high for low-skilled Mexicans. Similarly, Akee (2010) finds that highly educated workers have the highest likelihood of

[3] See Chiswick (1999) for alternative specifications of the migration selectivity models.

migrating from the Federated States of Micronesia to the United States, and McKenzie, Gibson, and Stillman (2010) find that a positive selection of Tongan migrants (in terms of observed skills) leaves for New Zealand. In addition, Docquier and Marfouk (2006) document that for the large majority of sending countries, individuals with more education are more likely to migrate abroad. Abramitzky, Boustan, and Eriksson (2012) argue that the current immigrant flow to the United States (and other developed countries) is not only a product of the migration decisions of immigrants, but also of entry rules. These restrictions which favor high skilled immigrants might determine whether immigrants are positively or negatively selected. They therefore study Norwegian migrants to the United States during the time of mass migration from 1850 to 1913 when the United States maintained an almost open border. They compare Norwegian men who emigrated to the United States to their brothers who stayed in Norway to disentangle the returns to migration from the migration decision and find that Norwegian movers from urban areas were negatively self-selected and came from the lower end of the occupational distribution whereas Norwegian movers from rural areas were positively self-selected.

Rural-to-urban migration is a special case of internal migration.[4] Borjas, Bronars, and Trejo (1992) apply Roy's self-selection model to internal migration. In their version of the Roy model, regions have different wage-generating characteristics. In some locations, average wages are relatively high and returns to skill relatively low. That is, these locations are characterized by wage distributions with comparably high means but relatively low returns to skills. On the other hand, there are locations which have relatively low means and relatively high returns to skills. Thus, wage distributions in different regions differ by means and in their returns to skills. They find that regional differences in skill returns will motivate young workers to migrate to areas where skill returns are suitable for their skill endowments. In other words, highly skilled workers will be attracted to areas which offer a higher return to skill and vice versa. Dahl (2002) develops a Roy model of within country mobility and earnings where individuals choose in which region (50 states) to reside and work. He finds that the OLS estimates of the returns to education in region-specific labor markets are biased upwards due to self-selection of better

[4] There are several attempts to disentangle the returns to migration within a country from the migration decision using natural experiments. Frank (2007), for example, analyzes the effect of migration from East to West Germany after the fall of the Berlin Wall on wages and employment in western Germany by instrumenting observed migration with characteristics of the sending region. By instrumenting for changes in black population, Boustan (2010) shows that whites strongly responded to the black influx into cities by leaving cities. Boustan, Fishback, and Kantor (2010) exploit a policy changes and natural disasters during the Great Depression to measure the effect of in-migration on the hourly earnings of existing residents. In particular, they use variation in the generosity of New Deal programs and extreme weather events to instrument for migrant flows to and from urban areas within the United States.

educated individuals to states with higher returns to education. Moreover, his results suggest that the relative state-to-state migration flows of college- versus high school-educated migrants are strongly affected by differences in the return to education and amenities across regions.

2.3 URBANIZATION IN NORWAY

In 18th century agricultural societies, only a tiny minority of the European population lived in cities. In modern societies, by contrast, the wide majority is concentrated in big urban centers. The 19th century marks the passage from a civilization based on agriculture to a civilization based on industry and services in many European countries. Urban centers grew during the industrialization as more and more people left the rural areas to work in cities. In Great Britain, for example, the rate of people living in urban areas was around 12 percent in 1700, 22 percent in 1800, 77 percent in 1901 and 79 percent in 1950 (see, e.g., Malanima and Volckart, 2008). In the US, the large increase in urban population happened in the 1950s when also the Southern United States became urban-majority. In Norway, however, urbanization started very late and was mostly pronounced in the 1960s. As noted above, Figure 2.2 plots the percentage of the population living in urban areas from 1910 to 2010. Less than half of Norway's population was living in urban areas until the 1950s, this number increased dramatically in the 1960s. The urbanization process leveled out in the 1970s, increased again from the mid-1980s and had reached almost 80 percent in 2010. The rapid growth of cities in the 1960s in Norway can be partly attributed to a larger population growth in the urban areas and rural-to-urban migration. Between 1968 and 2010, the out-migration from rural areas varies between 2.2 and 3.2 percent per year.

Also as compared to its neighbor Sweden, Norway's urbanization process happened late. Both countries had rather low urbanization rates during the 19th century. While the process of urbanization in Sweden increased rapidly after 1930, Norway stayed a mainly rural country until 1960. The large differences in the urbanization process in Norway and Sweden were mainly caused by different political views and efforts: whereas the Norwegian government aimed to slow down urbanization by strong regional development policies, Swedish politicians did not see a need for slowing down the urbanization process. In Sweden, urbanization was seen as 'modernization' and a process that positively affects living conditions. On the other hand, Norway invested a lot of resources building up the destroyed North after the end of World War II and the Norwegian government also established a fund granting loans for fishing and infrastructure projects in Northern Norway. Moreover, the government invested in the expansion of water power exploitation in rural Norway, placed large government funded industries in remote areas and even moved government institutions such as Statistics Norway out of Oslo. In addition, Norway implemented a generous system of agricultural subsidies in the 1960s (see SSB, 2005) which made rural life relatively attractive.

The decades when the urbanization process in Norway advanced at a high rate coincides with the time period observable in Norwegian Registry Data. This makes Norway an interesting case for exploring the changes in the urban wage premium and the selection into rural-to-urban migration over time. Figure 2.3 shows the share of individuals in our sample of males aged 25-55 who live in urban areas in our data period from 1967 to 2010. Note that trends in this share are similar to the development in the population plotted in Figure 2.2 during the same period.

2.4 EMPIRICAL STRATEGIES

We analyze (i) the size and composition of the urban wage premium and (ii) the rural-to-urban migration selection both from 1967 until today. As our data stretch over more than 40 years, we are able to apply all estimation strategies for four consecutive decades separately. This will allow us to get a sense of the importance and possible changes in the urban wage premium and selection into migration over time.

Similar to Glaeser and Mare (2001), we use individual level data to measure the urban wage premium and its development over time and estimate the regression:

$$ln(W_{it}) = X_{it}'\beta + U_{it}'\gamma + \phi_i + \varepsilon_{it}, \tag{2.1}$$

where $ln(W_{it})$ is the log of the yearly income for individual i at time t, X_{it}' is a vector of individual characteristics, and β is the price of those characteristics in the labor market. U_{it}' is a dummy variable describing whether the individual lives in an urban location. The vector γ indicates the productivity gain from living in an urban area, or in other words, the urban wage premium. The term ϕ_i represents individual ability.

Estimating an OLS regression omitting the individual ability ϕ_i, will bias γ if the individual ability is not randomly distributed across locations. We use two strategies to avoid this possible bias: first, we treat ϕ_i as an individual-specific time-invariant factor and estimate individual fixed effects regression. This strategy allows removing the part of the omitted ability bias that is individual specific and time invariant. In particular, the individual fixed effects estimators allow examining whether migrants to cities have wage gains. In the second identification strategy, we use the IQ test scores from Norwegian military records as a proxy measure for individual ability (see, e.g., Griliches, 1979) and estimate the following model:[5]

$$ln(W_{it}) = X_{it}'\beta + U_{it}'\gamma + A_i'\eta + \varepsilon_{it}, \tag{2.2}$$

where A_i' is the IQ score.

Similar to Glaeser and Mare (2001), we also analyze whether the urban wage premium is due to wage level or a wage growth effect in Norway. A wage level effect describes an immediate income gain for the workers right after the arrival in the city. On the other hand, if the urban wage premium is a wage growth effect, wages grow faster in cities and the urban wage premium emerges over time. Hence, this theory predicts a positive interaction between labor market experience and working in an urban area. We try to distinguish between these two effects by studying the wage patterns of migrants. We use the following regression:

$$ln(W_{it}) = X_{it}'\beta + U_{it}'\gamma + \sum_j \eta_j I_{t+j}^u + \sum_j \theta_j I_{t+j}^r + \varepsilon_{it}, \tag{2.3}$$

where I_{t+j}^u is an indicator variable that is equal to one if the individual will move from a rural area into an urban area at time $t+j$ and I_{t+j}^r is an indicator variable that is equal to one if the individual will move from an urban area into a rural area at time $t+j$. The extent to which earnings rise or decline immediately before or after the move is described by the parameters η_j and θ_j, whereas η_j denotes the earning changes for rural-to-urban movers and θ_j for urban-to-rural movers. The pattern in the parameters η_j right after the move will allow to determine whether the wage premium is a wage growth or a wage level effect. Moreover, the η_j's that describe the relative wages before the move, provide information about whether movers are positively or negatively selected from the rural population.

In a second step, we attempt to analyze the selection into migration more closely. Focusing in detail on the migrants is useful as rural-to-urban migrants give us identification in the above

[5]If selection into cities would be solely based on ability and IQ scores are a perfect measure of workers' ability, controlling for IQ test scores should result in an unbiased estimate of the urban wage premium.

described fixed effects estimation strategy. The estimated urban wage premium γ in Equation 2.1 is unbiased only if migration into urban areas is random conditional on the fixed difference across movers and non-movers over time. This assumption would be violated if an individual's decision is influenced by an unobserved factor that is not perfectly persistent over time and is correlated with wages. Thus, if there is selection into migration other than captured by the time invariant fixed effect, the fixed effect estimator is biased. To better understand the selection into migration, we first estimate the return to migration without accounting for selection for four different decades

$$ln(W_i) = X_i'\beta + M_i'\delta + \varepsilon_i, \tag{2.4}$$

where $ln(W_i)$ is the log of the yearly income for individual i, X_i' is a vector of individual characteristics including age, education, and birth order, and β is the price of those characteristics in the labor market. M_i' is a dummy variable equal to one if individual i migrated from a rural to an urban area within a given decade. The vector δ indicates the return to moving to an urban area. Note that the estimation of the selection into migration is based only on the sample of individuals residing in the rural area at the beginning of the decade.

There is most likely a substantial selection into migration also on unobservable factors. Hence, δ in Equation 2.4 might be biased if the individuals non-randomly self-select into migration after controlling for all observed variables. Our estimation strategy to measure the return to migration in the presence of selection is based on comparing rural-to-urban migrants with their brothers who stayed in the rural area. In particular, this strategy eliminates the across household component of migrant selection, which can for example occur if households are financially constrained and thereby have different propensities to migrate. In addition, this strategy removes the component of unobserved individual ability that is shared between brothers. Following a similar approach as Abramitzky, Boustan, and Eriksson (2012), we use the following regression equation in which the individual error term is decomposed into two components:

$$ln(W_{ij}) = X_{ij}'b + M_{ij}'d + \rho_j + \nu_{ij}, \tag{2.5}$$

where the component of the error that is shared between brothers in the same household j is given by ρ_j and ν_{ij} is the component that is idiosyncratic to individuals. Estimating a model with household-specific fixed effects will eliminate ρ_j and thereby the potential bias emerging due to aspects of family background that are correlated both with earnings and the probability of migration.

Comparing the return to migration δ in Equation 2.4 and d, the return to migration from estimating Equation 2.5, allows us to infer the direction of the selection across households. In particular, if $\delta < d$, δ is biased downward by negative selection of migrant households. On the other hand, if $\delta > d$, δ is biased upward by positive selection of migrant households. Thus, individuals from migrant-households would have had higher earnings even if they had stayed in rural areas.

2.5 DATA

Our primary data source is the Norwegian Registry Data maintained by Statistics Norway, a linked administrative data set that contains education, family, and earnings register. It allows us to construct a long panel data set with information about educational attainment, demographic variables as age and gender, and earnings for the Norwegian population from 1967-2010 on calendar year basis.[6] Thanks to the unique personal identifier given to all residents in Norway we are able to link these data to a central population register providing the municipality of residence in each year from 1967 onwards. The personal identifiers also enable us to identify siblings and parents.

Earnings are measured as log of total pension-qualifying earnings in Norwegian Kroner (NOK) reported in the tax registry. They include labor earnings, as well as taxable sickness benefits, unemployment benefits, parental leave payments, and pensions. Earnings are adjusted for inflation to the base year 2010 using the Consumer Price Index (CPI). Before 1986 (with the exception of the year 1981) the income is top-coded.[7] To overcome this problem we follow the approach of Bhuller, Mogstad, and Salvanes (2011) and simulate Pareto-distributed earnings instead of the censored earnings.

The education variable, measured as the highest completed level of education, is taken from the education register which is based on school reports sent directly to Statistics Norway, thereby minimizing any measurement error due to misreporting. We classify individuals into three education groups: low- (primary and lower secondary level), medium- (upper secondary school level), and high-educated (tertiary level).

[6] See Møen, Salvanes, and Sørensen (2003) for a more detailed description of the data.

[7] In years after 1970 less than 1.5 % of observations is top-coded; in years before 1970 less than 6%.

For cohorts born after 1949 we have access to IQ test scores from Norwegian military records (see Sundet, Barlaug, and Torjussen, 2004; Thrane, 1977, for details), which can be used as a control for personal ability. The military service in Norway is mandatory for men, and each male individual is tested for physical and psychological suitability during the examination around his 18th birthday. The IQ score is constructed as an unweighted mean of three tests–arithmetics, word similarities, and figures–and converted into a single digit number on a 0 to 9 scale, so that the scores are normally distributed with a mean of 5 and a standard deviation of 2.

Statistics Norway divides municipalities in 4 different levels (on a scale from 0-3) in terms of centrality (see, e.g., SSB, 1994). Centrality is defined as a municipality's geographic location in relation to a center where there are service functions such as a post office and bank. Urban areas are divided into three levels according to population and available service functions. Municipalities on level 3 are regional centers (municipalities that have a population of at least 50 000 inhabitants) or municipalities that are located within 75 minutes (90 minutes for Oslo) travel time from an urban settlement of level 3. Level 2 includes settlements with a population between 15 000 and 50 000 or municipalities that are located within 60 minutes travel time from an urban settlement of level 2. Level 1 includes municipalities with a population between 5 000 and 15 000 or municipalities that are located within 45 minutes travel time from settlement of level 1. Level 0 covers remote municipalities. Figure 2.11 in the Appendix displays the centrality level for each municipality within Norway. Following this definition of centrality level, we define municipalities as urban areas if their level of centrality is 3 or 2 in accordance with the international standard for comparing urbanization level across countries (see, e.g., Figure 2.2). According to this definition 42.7 percent of the municipalities are defined as urban locations.

2.5.1 Sample Selection

Our sample contains yearly observations of men aged 25 and 55, who were resident in Norway in at least one of the years from 1967 to 2010 and have non-missing data on year of birth, gender, and education. In case of missing location data in at least one year we exclude all observations for these individuals in each year, thereby creating a panel data set keeping only persons with continuous time periods with available residence data.[8]

The sample is restricted to men in order to minimize selection issues due to low female participation in the labor market in the 1960s and 1970s. The reason for restricting the sample

[8]By excluding all individuals with missing location data in one or more years we lose less than 4% of the data.

to individuals which are at least 25 years old is to avoid observations that have no earnings as they are still in education. By the age of 25, most men have completed their education.

We keep only individuals with sufficiently high income to support themselves. We therefore exclude all observations with income lower than is a social security base rate (one G or "grunnbeløpet", which serves as a basis for calculating the Norwegian state pension.)[9]

Our final sample consists of $33,714,448$ person-year observations of $1,992,703$ individuals. Summary statistics for the key variables we use are presented in Table 3.1. The incomes for individuals in urban areas on average are 20% higher than in rural areas. Individuals in urban areas are also better educated: the share of people with tertiary education is 14 percentage points higher in cities. In addition, Table 3.1 shows that rural-to-urban migrants have higher incomes and higher education than rural stayers.

2.6 Results

2.6.1 Urban Wage Premium

Table 2.2 presents the basic regression results for the whole sample period (1967-2010) using Equation 2.1. Column 1 shows that the estimated raw wage premium for living in cities, that is the estimated γ when only controlling for calendar year dummies, is 16 percent. Controlling for education and age in column 2 reduces the estimated premium to 13 percent. Thus, controlling for observable characteristics does not eliminate the urban wage premium. The significant difference of the urban wage premium with and without control confirms that there is a positive sorting (on observables) of people living in cities.

In addition, it is possible that unobservable ability is much higher in urban areas, that is, the urban population is positively sorted on unobserved characteristics. Column 3, therefore, estimates a specification that includes individual fixed effects, which provides a possibility to eliminate time-invariant unobservable ability. Controlling for individual-specific fixed effects further reduces the urban wage premium to 7 percent. This indicates that positive sorting on unobserved ability is an important component of the urban wage premium as well, and that the real productivity effect of cities is overestimated when workers' unobserved characteristics are not taken into account.[10] The results show that there is still a substantial urban wage premium

[9]For example, for year 2010 the base rate was set to 74,721 NOK or about 11,000 US

[10]Note however, that the fixed effects result should be interpreted with caution. As discussed in Section 2.4, the estimate might be biased if there is selection into migration other than captured by the time invariant fixed effect.

even after controlling for individual fixed effects. Hence, cities have a productivity advantage and the observed urban wage premium is not simply a result of sorting. Our findings tie in well with the literature (see, e.g., Glaeser and Mare, 2001; Yankow, 2006).

As discussed in Section 4, we attempt in a next step to determine whether individuals moving to the city receive an immediate wage increase or whether the urban wage premium is caused by a slow wage increase over time. Figure 2.4 displays the results from Equation 2.3. The earnings increase right after the relocation, and they continue to rise in each year. One year after the relocation, the wage premium is 5.9 percent; five years after the move the earnings difference to those remaining in the rural area is 11.2 percent. Overall, there is some support for both the wage level effect and wage growth effect. The evidence suggests that there are gains to moving into the city in the first year after the move and also that the full wage premium is not realized immediately but that the wage grows substantially during the first few years in the city. Hence, there is an immediate level effect and there is also a positive interaction between labor market experience and working in cities. These results also correspond well with earlier findings on the distinction of the wage level and wage growth effect (see, e.g., Glaeser and Mare, 2001).

URBAN WAGE PREMIUM OVER TIME

A major aspect of this paper is to analyze the development of the urban wage premium over time. Using the basic earnings equation with and without controlling for education and age, Figure 2.5 displays the urban wage premium in each year from 1967 to 2010. The urban wage premium decreased substantially from the 1960s until today - the raw premium dropped from more than 26 percent in the late 1960s to below 10 percent in 2010 (from 22 percent to 7 percent when controlling for observables). The largest decrease happened in the 1960s and 1970s. The urban wage premium was fairly stable in the 1980s and 1990s and decreased again during the 2000s. A possible explanation for the large decrease in the urban wage premium in the 1970s and 2000s is that mobility was higher in those years causing a supply shift and thereby lower returns.

Figure 2.5 also shows that there is positive sorting on observed characteristics; in the specifications controlling for education and age, the estimates are smaller than the raw differentials. In addition, the gap between the raw differential and the urban wage premium controlling for observed characteristics decreased. This indicates that the positive sorting on observables becomes less pronounced over time.

To control again for individual-specific time-invariant fixed effects, we divide the sample period into four decades and apply fixed effects estimation for each decade. Table 2.3 contains estimation results for all four decades. Focusing on the basic model with and without controls confirms the decrease in the urban wage premium over time, as discussed above, and the presence of positive sorting. When including the individual fixed effects, the urban wage premium is substantially smaller in each decade than in the basic regressions. This shows that under the assumption that the estimates including fixed effects are unbiased; the positive sorting on individual-specific fixed effects is a significant component of the observed urban wage premium. The fixed effects estimates increase from the 1960s to the 2000s, i.e., the gap between the OLS and the fixed effects estimates become smaller. This finding suggests that positive sorting on individual-specific fixed effects becomes a less important component of the urban wage premium over time. Moreover, it also indicates that even though the observed raw urban wage premium steadily decreases over the four decades, the productivity premium, net of sorting, in fact increases over time.[11]

Additionally, we analyze whether there are any changes in wage patters of rural-to-urban migrants after the move in each decade. This allows us to determine whether the wage growth or wage level effect became more important after 1967. As above, we use Equation 2.3 to determine how fast wages of movers grow once they arrived in a city. The results are provided in Figure 2.6. In all four decades, we find that in addition to an immediate level effect, there is a positive interaction between labor market experience and working in cities. Thus, the wage growth is faster in cities relative to outside of cities.

A Proxy For Cognitive Ability

As discussed in Section 2.4, another possible identification strategy that allows us to control for the unobserved cognitive ability but at the same time does not rely on identification from movers, is to use IQ scores as proxy for ability. As noted above, for cohorts born after 1949

[11] As mentioned earlier, the fixed effects estimates of the urban wage premium γ are identified by rural-to-urban migrants. Since movers do not necessarily constitute a representative sample, the fixed effect estimates might over- or understate the true productivity effect of living in a city. We are interested in the average wage premium a random person receives from living in a city. Biased fixed effects estimates might for example arise when the reason for people to move would be the fact that they become unemployed. The fixed effects estimates would be overstated because the average movers' wage before the move is lower due to unemployment. In addition, if the probability of moving to urban areas after a lay-off has changes in time, the time pattern we observe in the fixed effects estimates might also be affected by the biased results. Thus, it is not clear whether the increase in our fixed effects estimates over time is due to the fact that the true productivity premium in cities is indeed increasing or whether it is a mere consequence of the changes in the individuals' incentives to relocate.

the data provide IQ score for each man in the sample. As the military service is mandatory for men, each male individual born in 1950 or later is tested with a standardized test.[12] This allows us to re-examine the importance of unobserved cognitive ability. We estimate Equation 2.2 using IQ scores at the age of 18 as a proxy measure for individual ability ϕ_i (Griliches, 1979). Although IQ scores are arguably not a perfect measure of workers' ability as they only picture the cognitive ability and not the non-cognitive characteristics such as motivation, willingness to compete, or self-discipline, this estimation strategy does not rely solely on rural-to-urban movers.

We use all individuals with non-missing IQ score data and estimate Equation 2.2 as well as the basic model including controls for education and age. The results for the estimated urban wage premium in the 1990s and 2000s when controlling for education and age are displayed in Column 1 in Table 2.4. Column 2 shows the resulting urban wage premium when adding IQ scores as a control variable. The results confirm the positive sorting of urban residents, although the drop in the estimates after controlling for IQ scores is not as substantial as it was in the fixed effects case. Here, sorting explains a relatively smaller part of the urban wage premium and there remains a substantial productivity effect of living in a city. Unlike in the case of the fixed effects model, the estimated urban wage premium in the later decade is lower than in the first decade. This may indicate that the fixed effects estimates in later decades might be overstated or that non-cognitive skills which are not pictured by the IQ scores become a more important part of the individual-specific fixed effects.

2.6.2 Migrant Selection

So far, we have studied the size of the urban wage premium and the role of sorting on both observable and unobservable characteristics. In addition, we analyzed the evolution of the urban wage premium over time and also whether the urban wage premium accrues immediately or gradually after the worker's arrival in the city. Next, we examine who self-selects into moving from rural locations to cities, and how the characteristics of rural-to-urban movers, as well as urban and rural stayers, changed over time.

[12] For a more detailed description of the IQ scores see Section 3.3.

MIGRANT SELECTION

As a first step, we study income inequality in the rural and urban areas to make predictions about the rural-to-urban migrant selection based on the Roy model (see Borjas, 1987). The rural areas have a more equal income distribution than the urban areas in the 1970s (see Figure 2.7). To be consistent with the Roy model, we would expect positive selection of migrants moving from rural to urban areas in the 1970s. Since then, the income distributions in the urban and the rural area have only shifted to minor degree.

Figure 2.8 shows the different average education levels for urban stayers, rural stayers and rural-to-urban movers in each year. In all years, rural stayers have the lowest average level of formal education. Urban stayers are on average better educated than rural stayers. The rural-to-urban movers are on average the best educated group in all years from 1967 until 2010. Since the 1960s, the average education level increased for all groups. The increase was, however, slowest for the rural-to-urban movers indicating that the positive self-selection was strongest in the early period of our sample. The finding that the migrants are on average more educated than the sending population is in line with the prediction based on the Roy model (see Figure 2.7). Moreover, the positive selection became less pronounced over time as the relative skills of the movers, compared to long-term rural residents, decreased. The pattern of the average age for the three groups is displayed in Figure 2.9. The rural-to-urban movers in our sample are very young. They are on average around 32 to 34 years old, whereas the rural stayers and rural movers are around 40 years. The differences over time are rather small.

Migrant selection might be driven by unobserved characteristics as ability or motivation. As a proxy for cognitive ability we again use IQ test scores from Norwegian military records (measured around age 18). Figure 2.10 shows the average IQ scores for rural stayers, urban stayers and rural-to-urban movers. Rural-to-urban movers score higher on average in the IQ test scores than both rural and urban stayers. Rural stayers have the lowerst average IQ score, whereas the average IQ score of the urban stayers stays stable over time and the average IQ score of urban-to-rural movers decreases after 1990. Hence, migrants are positively selected on IQ score over the whole time period, the positive selection however becomes less pronounced.

A further possibility to analyze migrants' characteristics is to compare their wages to rural stayers. The graphs in Figure 2.6 are based on Equation 2.3 and picture the migrants' wages prior to the move, compared to rural stayers (controlling for education and age). The graphs suggest a negative time pattern in migrant selection. In the first two decades, rural-to-urban migrants earn slightly more than those who remain in the rural area in the six years prior to relocation. In the two later decades, however, the rural-to-urban migrants earn substantially less than rural stayers prior to moving. These patterns might indicate that the selection of people moving into the city changed over time from individuals earning on average more than rural stayers to individuals earning on average less than rural stayers (taking the differences in education and age into account).

SELECTION OF MIGRANT HOUSEHOLDS

Next, we examine whether the migrant flow from the rural areas to the cities was drawn from households from a lower or higher stratum. We apply a method similar to Abramitzky, Boustan, and Eriksson (2012) and compare brothers where one sibling is moving from a rural to an urban area and one sibling stays in the rural area. The sample is limited to households with at least two brothers who are in the sample in the beginning (e.g., 1971, 1981, 1991, 2001) and the end of the decade (e.g., 1980, 1990, 2000, 2010).[13] As described in Section 2.4, the return to migration δ is estimated using Equation 2.4. δ is however only the true return if migrants were selected randomly from the rural population. δ will be biased if migrants are self-selected. We therefore compare the earnings of migrants and their non-migrant brothers to eliminate selection across households and estimate Equation 2.5 including household fixed effects.

The first panel of Table 2.5 shows the return to migration estimated using Equation 2.4 for the sample of brothers. The return to migration δ is identified by all rural-to-urban migrants that moved to the city in the past decade. The second panel displays the results of the within-household estimation. Here, the return to migration d is only identified by brothers where one sibling is moving from a rural to an urban area and one sibling stays in the rural area in the past decade. In the 1970s, the return to migration, δ, is 14 percent. Adding family fixed effects and thereby eliminating the component of migrant selection that takes place across households reduces the return to migration d to 10 percent. This change indicates a positive selection of migrant households and is consistent with our predictions based on the Roy model. In the subsequent decades, the estimated return to migration is, however, increased by adding

[13]Since the family identifiers only allow us to identify parents for cohorts born in 1950 or later, the sample of men is restricted to ages 40 or less. Moreover, we restrict our sample to individuals aged 25 or older at the beginning of a decade. To estimate the return to migration from 1971 to 1980 we use family information from the 1960 census. In the census, children can be linked to their parents if they live in the same household with at least one parent, are non-married and non-divorced. For children aged 20 or less, we are able to link 95.5 percent to their parents. Thus most men aged 34-40 in 1980 can be linked to their brothers.

family fixed effects. This implies a change to negative selection of the migrant households. A bit surprising is the large increase in the estimated return to migration in the 1980s, where the increase from 11 to 16 percent indicates a strong negative selection of migrant households. This might be linked to the recession in the 1980s and the large deindustrialization in the remote areas in this decade.[14]

As one might suspect that the sample of brothers is a much selected sample, we also estimate Equation 2.4 using all men aged 25 to 40 who are in the sample in the beginning and the end of the decade. The third panel in Table 2.5 shows the return to migration in each decade for the whole sample of men. The estimated returns to migration are generally similar to those using the brothers-sample, with the estimates being slightly higher in the first and the third decades.

To sum up, we find that migrants who decide to move to urban areas are younger, more educated and have a higher IQ score, and that the positive selection on education and IQ score is strongest in the beginning of our period. When comparing the migrants' wages prior to the move to rural stayers in an individual fixed effects model, our results suggest a transition from a positive selection in 1970 to a negative selection in 2010 based on unobserved characteristics. Comparing brothers where one sibling is moving from a rural to an urban area and one sibling stays in the rural area indicates that the rural-to-urban migrant flow was drawn from positively selected households in 1970 and from negatively selected households in 2010.

2.7 CONCLUSION

The analysis of spatial differences in wages and in particular the urban wage premium has been well documented (Roback, 1982; Henderson, 1982; Helsley and Strange, 1990). However, it is still an open question to what extend cities make a person more productive, how much of this premium is driven by positive sorting and who self-selects into moving from rural locations to cities. Since the urbanization process in Norway happened comparatively late and in a time period observable in the Norwegian population registry data, Norway offers a unique laboratory for exploring in particular the changes in the urban wage premium and selection into rural-to-urban migration over time. The novel features of this paper is that the registry data in combination with the urbanization process provide an unique opportunity for studying the

[14] As we discussed above, the estimated premium might be overstated if there are unemployed workers among brothers who will move to the city. In that case, the migrant households might appear to be more negatively selected, even though they only experience a transitory negative earnings shock.

urban wage premium and the selection into rural-to-urban migration during an interesting time period and using different identification strategies.

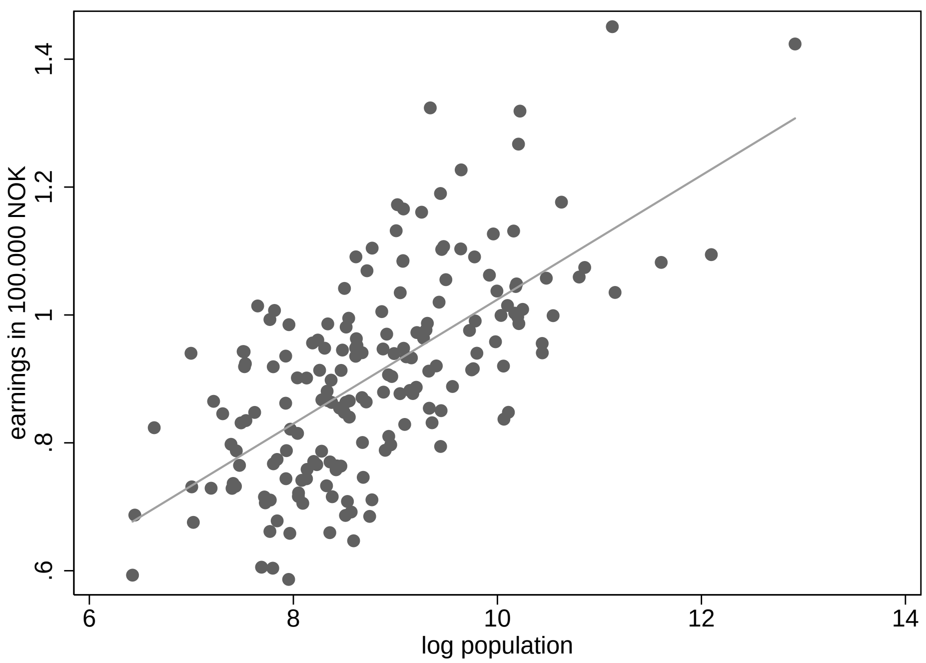
Our population registry based panel dataset ranging from 1967 until 2010 allows us to trace individuals' municipality of residence in each year and link this information with characteristics such as education and earnings patterns. Based on this information, we first estimate the size of the urban wage premium and document its changes from 1967 until today using different identification strategies. In a second step, we focus on individuals who are attracted by cities, and analyze rural-to-urban migrants and especially how the characteristics and selection of these rural-to-urban movers changed over time.

We find evidence for a substantial urban wage premium in Norway: the estimated premium for living in cities when pooling data for all the years is 16 percent. Controlling for observed characteristics such as education and age reduces the estimated premium to 13 percent. This drop in the urban wage premium denotes that there is a positive sorting on observable characteristics of people living in cities. Adding individual fixed effects reduces the urban wage premium to about seven percent. Hence, there is still a substantial urban wage premium even after controlling for observed and unobserved, time-invariant characteristics. This indicates that the positive sorting cannot explain the entire premium and that there remains a positive productivity effect of living in a city. In addition, we show that the full productivity premium is not realized immediately after the relocation to the city, but that the wage grows gradually during the first few years after the arrival in the city. Alternatively, we use IQ test scores as a proxy for the unobserved ability. Here, the results confirm that sorting on cognitive ability measured by IQ tests scores is an important component of the urban wage premium, although less than the individual fixed effects that may also control for non-cognitive ability. Nevertheless, both approaches suggest that the skill composition of rural and urban residents is different and important in explaining the urban wage premium. Focusing on changes over time, we find that the urban wage premium has decreased steadily since the late 1960s. The raw premium dropped from more than 26 percent in the late 1960s to below 10 percent in 2010 (from 22 to 7 percent when controlling for education and age). Hence, the positive sorting on observables became less pronounced over time. In addition, the urban wage premium is substantially smaller after introducing individual fixed effects in each of the four decades. The productivity effect, however, increased slightly over time. That is, the role of sorting on individual-specific fixed effects has lost importance since 1967.

During the same time period also the characteristics and selection of rural-to-urban movers also changed; movers are significantly younger and better educated than stayers and they also have higher IQ scores. The positive selection on education and IQ scores is strongest in the early period of our sample. When comparing the outcome of the brothers where one brother moved to the city and one stayed in the rural area, we find that the selection of migrant households is positive in the 1970s and changed into a negative selection in the following decades. These findings are confirmed by analyzing pre-migration incomes of movers to rural stayers. In the six years prior to migrating to the city, those moving into an urban area earn slightly more than those who remain in the rural area in the 1970s. In 2010, the pre-relocation income of rural-to-urban movers is significantly lower than the income of the rural stayers.

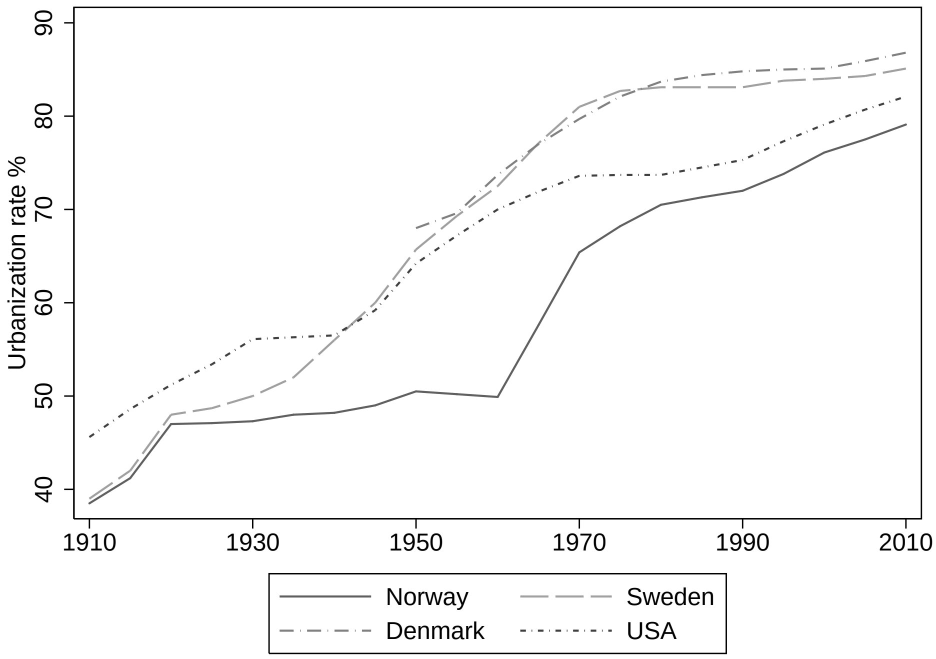
FIGURES AND TABLES

Figure 2.1: Municipality size and average annual earnings in 1970



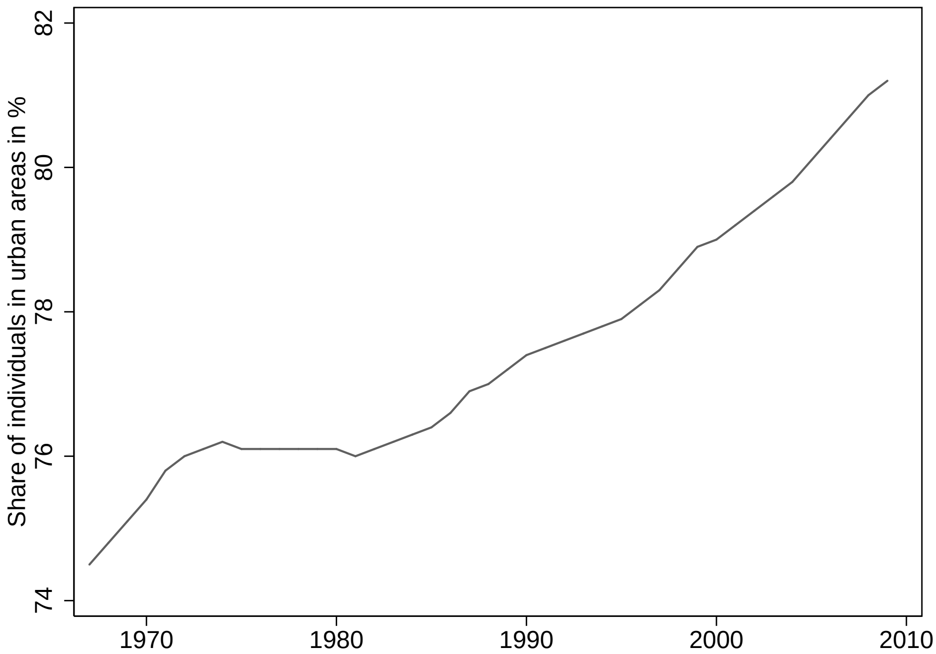
Note: Wages and population in municipality with centrality level 2 and 3 according to the centrality levels provided by Statistics Norway (see SSB, 1994). Earnings = 0.097 + 0.053 log(population); R-squared = 0.43, N= 181. The sample includes all men and women aged 16-75 with positive income in 1970. Earning are expressed in 2010 NOK.

Figure 2.2: Rate of population living in urban areas in Sweden, Denmark and Norway from 1910-2010



Data source: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, US Census Bureau

Figure 2.3: Share of the urban population



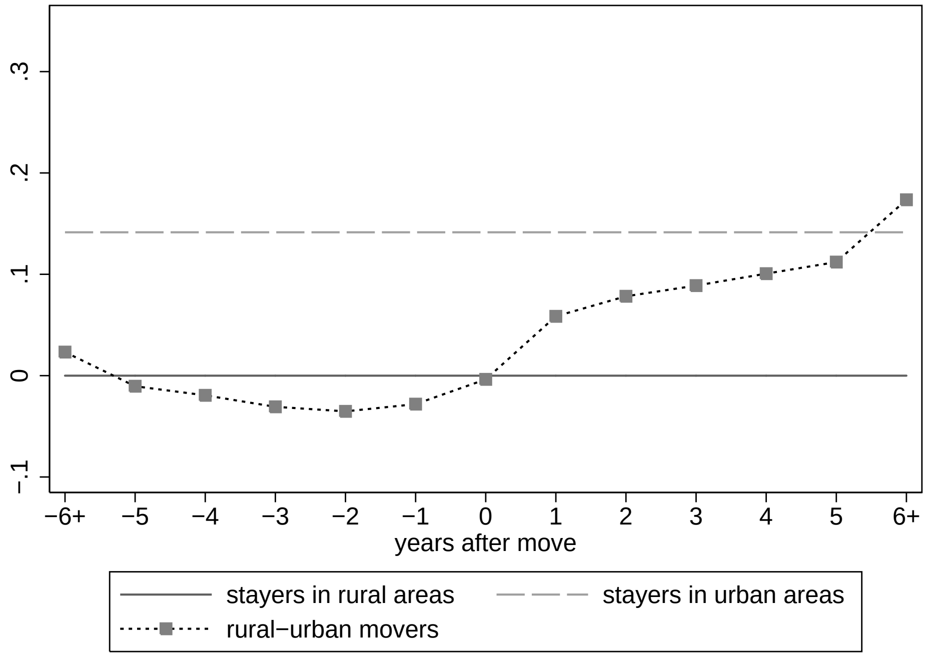
Note: Individuals are defined to reside in an urban area if their municipality of residence belongs to centrality level 2 or 3 according to the centrality levels provided by Statistics Norway (see, e.g., SSB, 1994).

Table 2.1: Summary statistics

	Total	Rural Stayers	Urban Stayers	Rural-to-Urban Movers
Log earnings	12.6	12.4	12.6	12.6
s.d.	(0.78)	(0.78)	(0.76)	(0.83)
Age	39.3	40.2	39.3	38.0
s.d.	(8.9)	(9.0)	(9.0)	(8.5)
Tertiary educ.	0.24	0.10	0.24	0.40
Secondary educ.	0.49	0.54	0.49	0.42
N (mil)	36.75	5.96	25.49	2.77
% sample	100	16.2	69.4	7.5

Note: Rural (urban) stayers are defined as those who lived in rural (urban) municipalities each year during their spell. Rural-to-urban movers started in rural municipality and moved to urban municipality at least once during their spell.

Figure 2.4: Wages of rural-to-urban movers pre- and post-move compared to rural and urban stayers



Note: Wage premia are calculated relative to stayers in rural areas. Regressions include a full set of year, education and age dummies. All estimates are significant at 1% level.

Table 2.2: Urban wage premium 1967-2010

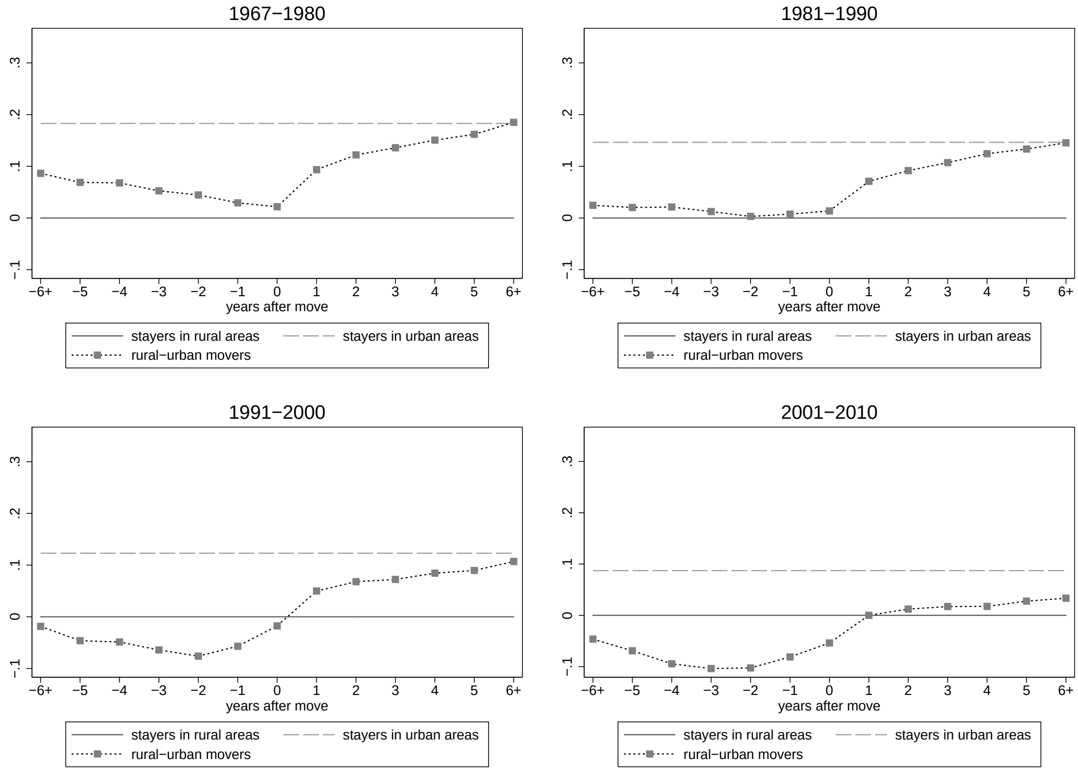
	(1) No Controls	(2) Basic Wage Equation	(3) Fixed Effects Estimator
Urban premium	0.158***	0.133***	0.0719***
	(0.000193)	(0.000182)	(0.000876)
Dummies:			
Education	no	yes	-
Age	no	yes	yes
Year	yes	yes	yes
N	33,714,448	33,714,448	33,714,448
adj. R-sq	0.132	0.233	0.246

Note: Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

Figure 2.5: Urban wage premium by year

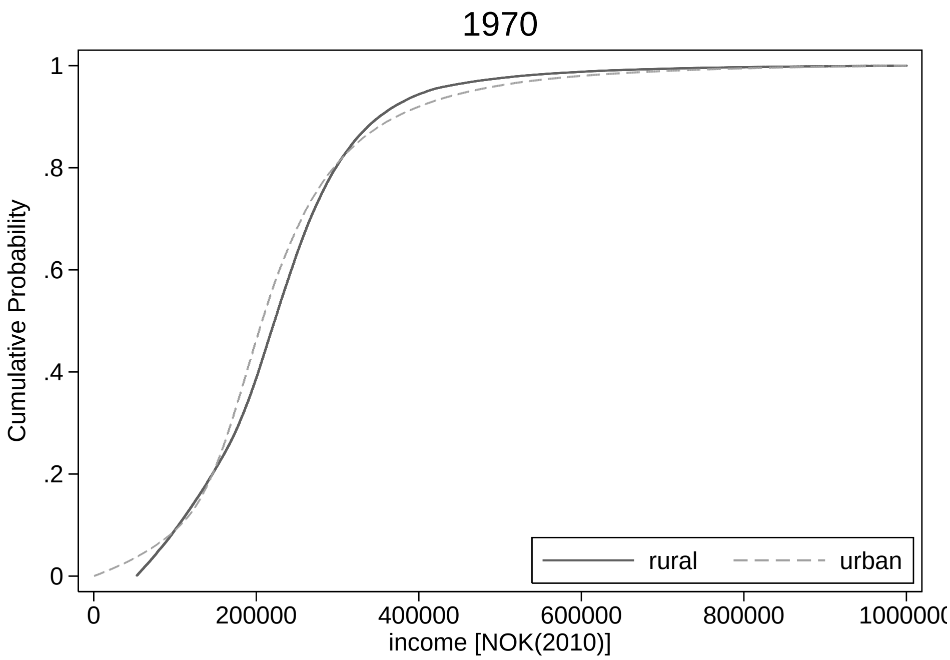
Note: Regressions include a full set of year dummy variables. All estimates are significant at 1% level.

Figure 2.6: Wages of rural-to-urban movers pre- and post-move compared to rural and urban stayers



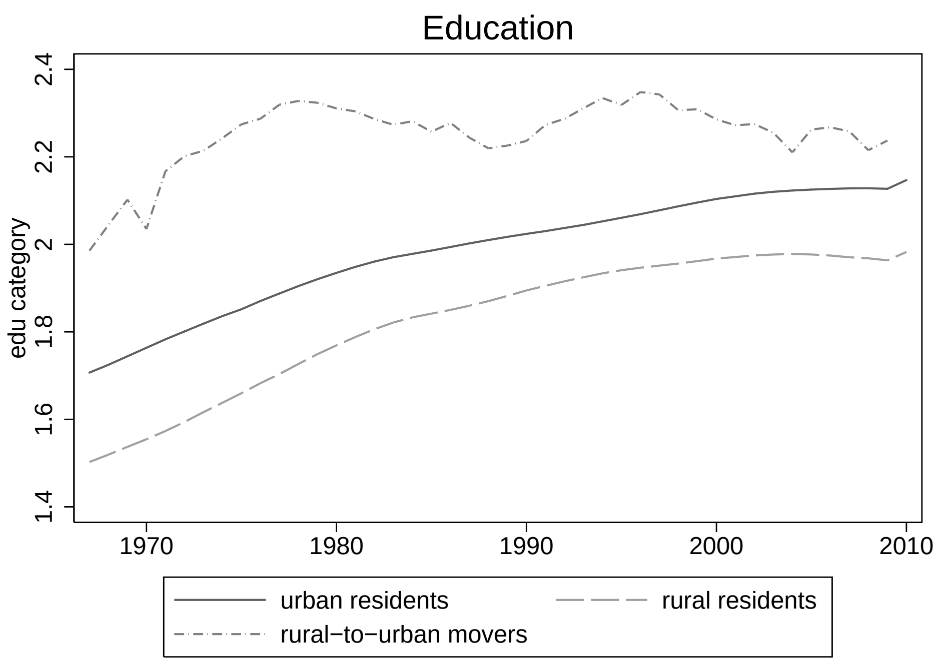
Note: Wage premia are calculated relative to stayers in rural areas. Regressions include a full set of year, education and age dummies. All estimates are significant at 1% level.

Figure 2.7: Cumulative income distribution function in urban and rural areas



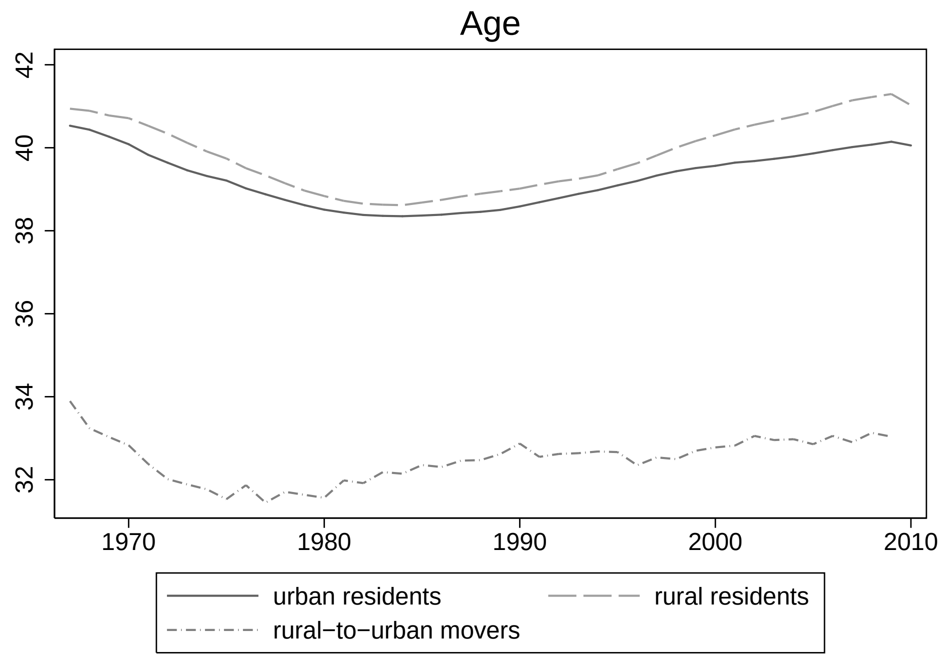
Note: Income in rural areas is scaled to have the same mean as in urban areas.

Figure 2.8: Education of movers versus stayers



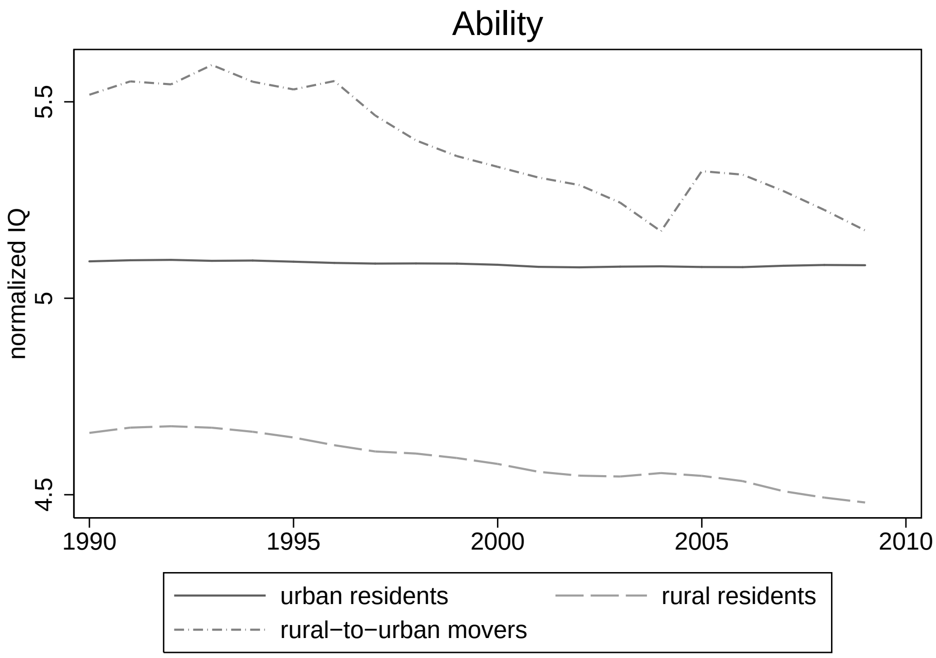
Note: Rural-to-urban migrants are defined as those who moved from a rural to urban area in a given year. Those who did not move that year are defined as either urban or rural stayers. Education categories: 1=primary, 2=secondary, and 3=tertiary.

Figure 2.9: Age of movers versus stayers



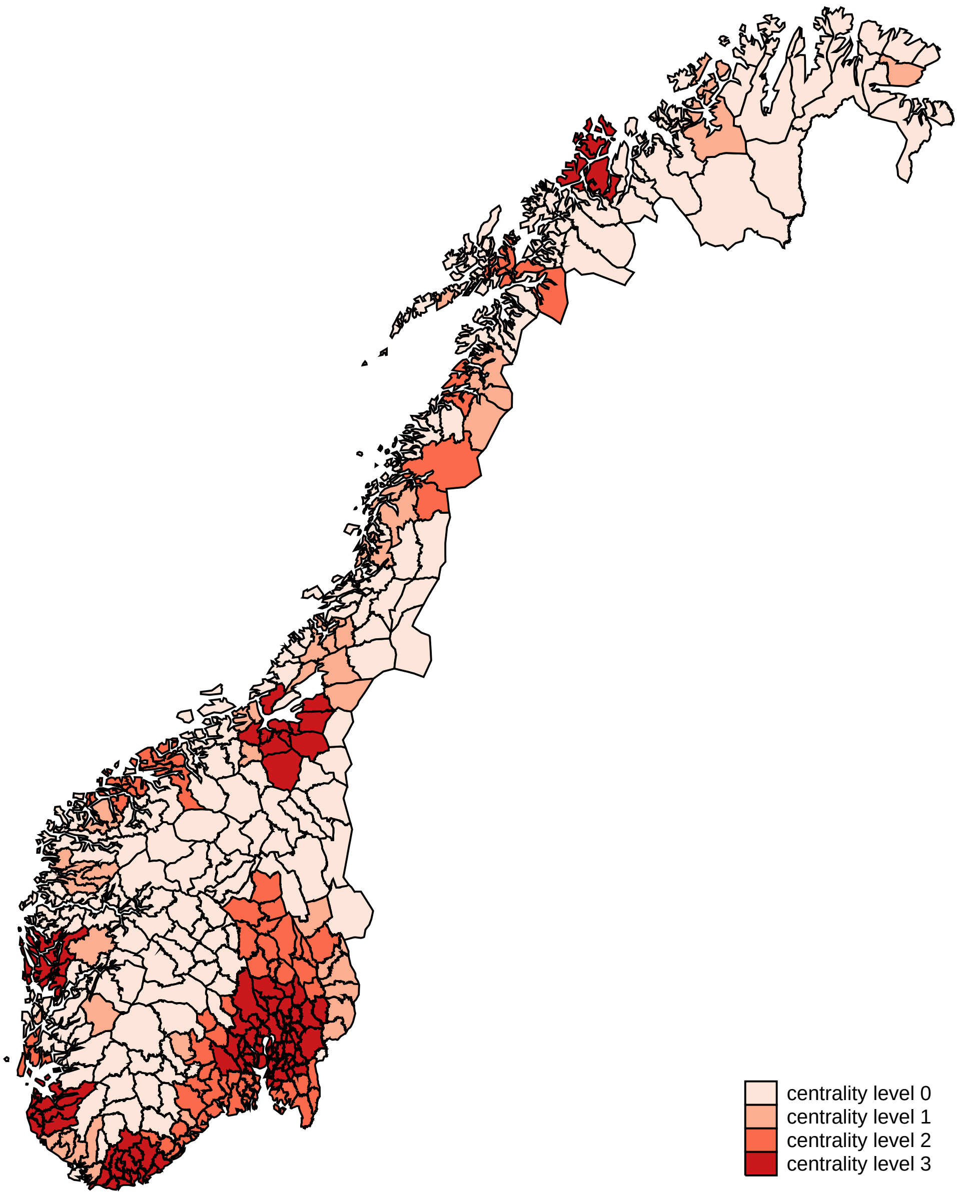
Note: Rural-to-urban migrants are defined as those who moved from a rural to urban area in a given year. Those who did not move that year are defined as either urban or rural stayers.

Figure 2.10: IQ score of movers versus stayers



Note: Rural-to-urban migrants are defined as those who moved from a rural to urban area in a given year. Those who did not move that year are defined as either urban or rural stayers. IQ scores are measured on the scale of 0-9 (for a more detailed description see Section 3.3).

Figure 2.11: Centrality level of municipalities



Data source: Statistics Norway

Table 2.3: Urban wage premium in four sub-periods

period	1967-1980			1981-1990			1991-2000			2001-2010		
	No Controls (1)	Basic (2)	FE (3)	No Controls (1)	Basic (2)	FE (3)	No Controls (1)	Basic (2)	FE (3)	No Controls (1)	Basic (2)	FE (3)
	0.206			0.162			0.143			0.109		
	(0.0003)			(0.0004)			(0.0004)			(0.0004)		
Urban		0.171			0.140			0.123			0.091	
		(0.0003)			(0.0004)			(0.0004)			(0.0004)	
			0.038			0.030			0.056			0.063
			(0.0014)			(0.0017)			(0.0017)			(0.0017)
Dummies:												
Educ.	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
Age	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	9,368,241			7,366,253			8,308,584			8,671,370		
adj. R-sq	0.101	0.217	0.195	0.030	0.138	0.050	0.034	0.144	0.152	0.032	0.152	0.191

Note: Standard errors in parentheses. All estimates are significant at 1% level.

Table 2.4: Urban wage premium in two sub-periods including IQ scores

period	1991-2000		2001-2010	
	Basic (1)	Including IQ (2)	Basic (1)	Including IQ (2)
Urban	0.126***		0.106***	
	(0.0005)		(0.0004)	
		0.118***		0.098***
		(0.0005)		(0.0004)
Dummies:				
Educ.	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
IQ	No	Yes	No	Yes
N	4,962,901	4,962,901	6,798,388	6,798,388
adj. R-sq	0.158	0.170	0.163	0.175

Note: The sample contains all men with non-missing IQ score. Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

Table 2.5: Return to migration per decade, analyzing brothers

Period	1971-1980	1981-1990	1991-2000	2001-2010
OLS	0.140***	0.108***	0.091***	0.062**
brothers sample	(0.033)	(0.026)	(0.024)	(0.021)
FE	0.097*	0.158***	0.098**	0.077**
brothers sample	(0.046)	(0.034)	(0.032)	(0.027)
OLS	0.163***	0.099***	0.128***	0.062***
whole sample	(0.012)	(0.011)	(0.011)	(0.009)
N brothers sample	4,896	9,816	11,473	7,567
N whole sample	26,613	40,921	41,790	36,833

Note: The brothers sample contains all brothers aged 25 to 40 who are rural in the beginning of the decade and urban or rural at the end of the decade. The whole sample contains all men (i.e., including men without brothers). Regressions include a full set of education and age dummies. Standard errors in parentheses. * p<0.05, ** p<0.01, and *** p<0.001.

REFERENCES

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102(5), 1832–56.

AKEE, R. (2010): "Who Leaves? Deciphering Immigrant Self-Selection from a Developing Country," *Economic Development and Cultural Change*, 58(2), 323–344.

BAUM-SNOW, N., AND R. PAVAN (2012): "Understanding the City Size Wage Gap," *The Review of Economic Studies*, 79(1), 88–127.

BHULLER, M., M. MOGSTAD, AND K. G. SALVANES (2011): "Life-cycle Bias and the Returns to Schooling in Current and Lifetime Earnings," Discussion Paper 5788, IZA Discussion Paper.

BORJAS, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 77(4), 531–553.

——— (2008): "Labor Outflows and Labor Inflows in Puerto Rico," *Journal of Human Capital*, 2(1), 32–68.

BORJAS, G. J., S. G. BRONARS, AND S. J. TREJO (1992): "Self-selection and Internal Migration in the United States," *Journal of Urban Economics*, 32, 159–85.

BOUSTAN, L. P. (2010): "Was Postwar Suburbanization 'White Flight'? Evidence from the Black Migration," *The Quarterly Journal of Economics*, 125(1), 417–443.

BOUSTAN, L. P., P. V. FISHBACK, AND S. KANTOR (2010): "The Effect of Internal Migration on Local Labor Markets:American Cities during the Great Depression," *Journal of Labor Economics*, 28(4), 719–746.

CHIQUIAR, D., AND G. H. HANSON (2005): "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113(2), 239–281.

CHISWICK, B. R. (1999): "Are Immigrants Favorably Self-Selected?," *The American Economic Review*, 89(2), 181–185.

COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2008): “Spatial Wage Disparities: Sorting Matters!,” *Journal of Urban Economics*, 63(2), 723–742.

DAHL, G. B. (2002): “Mobility and the Return to Education: Testing a Roy Model with Multiple Markets,” *Econometrica*, 70, 2367–2420.

DOCQUIER, F., AND A. MARFOUK (2006): “International Migration by Educational Attainment, 1990-2000,” in *International Migration, Remittances, and the Brain Drain*, ed. by C. Ozden, and M. Schiff, pp. 151–200. Washington, DC: The World Bank and Palgrave McMillan.

FRANK, D. (2007): “The Effect of Migration on Natives’ Employment Outcomes: Evidence from the Fall of the Berlin Wall,” Discussion Paper 54, INSEAD Working Papers Collection.

GLAESER, E. L. (1999): “Learning in Cities,” *Journal of Urban Economics*, 46(2), 254 – 277.

GLAESER, E. L., AND J. D. GOTTLIEB (2009): “The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States,” *Journal of Economic Literature*, 47(4), 983–1028.

GLAESER, E. L., AND D. C. MARE (2001): “Cities and Skills,” *Journal of Labor Economics*, 19(2), 316–42.

GOULD, E., AND O. MOAV (2010): “When is ’Too Much’ Inequality Not Enough? The Selection of Israeli Emigrants,” CReAM Discussion Paper Series 1014, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

GOULD, E. D. (2007): “Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium,” *The Review of Economic Studies*, 74(2), 477–506.

GRILICHES, Z. (1979): “Sibling Models and Data in Economics: Beginnings of a Survey,” *Journal of Political Economy*, 87(5), 37–64.

HELSLEY, R. W., AND W. C. STRANGE (1990): “Matching and Agglomeration Economies in a System of Cities,” *Regional Science and Urban Economics*, 20(2), 189–212.

HENDERSON, J. V. (1982): “Evaluating Consumer Amenities and Interregional Welfare Differences,” *Journal of Urban Economics*, 11(1), 32–59.

KRUGMAN, P. (1991): *Geography and Trade*. MIT Press, Cambridge, MA.

LUCAS, R. J. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3–42.

MALANIMA, P., AND O. VOLCKART (2008): "Urbanisation 1700-1870," in *Unifying the European Experience: An Economic History of Modern Europe, Volume 1: 1700-1870*, ed. by S. Broadberry, and K. O'Rourke.

MCKENZIE, D., J. GIBSON, AND S. STILLMAN (2010): "How Important is Selection? Experimental vs Non-experimental Measures of Income Gains from Migration," *Journal of the European Economic Association*, 8(4), 913–945.

MØEN, J., K. G. SALVANES, AND E. SØRENSEN (2003): "Documentation of the Linked Employer-employee Data Base at the Norwegian School of Economics," Discussion paper, Norwegian School of Economics and Business Administration, Bergen.

ROBACK, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90(6), 1257–78.

ROY, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

SJAASTAD, L. A. (1962): "The Costs and Returns of Human Migration," *The Journal of Political Economy*, 70(5), 80–93.

SSB (1994): "Standard for Kommuneklassifisering 1994 (Standard Classification of Municipalities 1994)," *Norges offisielle statistikk (Official Statistics of Norway): Oslo-Kongsvinger*.

——— (2005): *Hundre års ensomhet? Norge og Sverige 1905-2005*. Statistics Norway.

SUNDET, J., D. BARLAUG, AND T. TORJUSSEN (2004): "The End of the Flynn Effect? A Study of Secular Trends in Mean Intelligence Test Scores of Norwegian Conscripts During Half a Century," *Intelligence*, XXXII, 349–362.

TABUCHI, T., AND A. YOSHIDA (2000): "Separating Urban Agglomeration Economies in Consumption and Production," *Journal of Urban Economics*, 48(1), 70–84.

THRANE, V. (1977): "Evneprøving av Utskrivingspliktige i Norge 1950-53," Discussion Paper nr. 26, INAS.

UN (2006): *World Urbanization Prospects: The 2005 Revision*. United Nations.

WHEELER, C. H. (2001): "Search, Sorting, and Urban Agglomeration," *Journal of Labor Economics*, 19(4), 879–99.

YANKOW, J. J. (2006): "Why do Cities Pay More? An Empirical Examination of Some Competing Theories of the Urban Wage Premium," *Journal of Urban Economics*, 60(2), 139–161.

CHAPTER 3

HOW CITIES MAKE US MORE PRODUCTIVE: EMPIRICAL EVIDENCE FROM NORWAY

Darina Steskal[1]

Abstract: The aim of this paper is to contribute to the understanding of the mechanisms that lie behind the wage differentials between urban and rural areas. Using rich administrative panel data on all male workers in Norway from 1986-2010, I first analyze the wage patterns of migrants year by year and then develop a framework that allows for a direct assessment of the relative importance of several possible explanations of rural-urban wage differentials. I find that spatial differences in unobserved skill composition account for roughly one-tenth of the observed urban wage premium. The remaining premium is attributed to the family of explanations called agglomeration economies, of which the static mechanisms appear to be the most dominant source. The dynamic mechanisms, which include learning and matching, both play significant roles as well. Finally, learning advantages acquired in cities do not appear to be transferable outside of cities, implying their location-specific character. The results are found to be robust to different strategies for addressing endogeneity and an alternative measure of agglomeration.

[1]Department of Economics, Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. Email: darina.steskal@nhh.no.

3.1 Introduction

It has been well documented that workers earn more in urban areas. This phenomenon is generally referred to as urban wage premium (UWP). The spatial nominal wage differences are indicative of a higher urban productivity because for firms to be willing to pay higher wages in cities, there must be productive advantages that offset these higher costs. Otherwise, firms would relocate to localities outside urban areas where wages are lower.[2]

While the presence of UWP has been well established, empirical evidence on its sources is rather limited. Different theories have been proposed in order to explain the sources of the productivity differential between urban and rural areas. Recent research has focused on two main lines of explanations. The first relates to a higher urban productivity as a result of non-random spatial sorting of workers, e.g. that more skilled workers choose to live in cities. This explanation does not involve any causal effect of urban location on worker productivity. The second line of explanations is based on the very act of bringing many workers and firms together, which in itself generates productivity benefits. These productivity benefits belong to the family of explanations generally referred to as agglomeration economies. If agglomeration economies are the source of the increased workers' productivity (as opposed to sorting), the causality of urban location on worker productivity is implied, i.e., urban location makes workers earn more.

Agglomeration economies can be generated by a multitude of underlying mechanisms. Some of the potential mechanisms include sharing indivisible facilities[3], sharing risks, reduced transport costs, proximity to customers, larger capital stocks, access to general public infrastructure, to name a few. All of these underlying theoretical mechanisms are observationally equivalent, with the empirical predictions implying urban wage *level* effects and *immediate* wage gains after moving to a city. I will refer to this type of agglomeration economies as *static*.

Agglomeration economies can also have a *dynamic* character, if the benefits accrue to workers only with time spent in cities. These dynamic agglomeration economies include *learning* and *matching* mechanisms. *Learning* fosters human capital accumulation in cities through knowledge spill-overs (for example because face-to-face interactions are more frequent in urban

[2] I am not concerned with obtaining *real* wages, corrected for the local cost of living, because I am interested in the firms' willingness to pay higher wages, reflected in nominal values, and not in workers' incentives to relocate which are based on real wages.

[3] Duranton and Puga (2004) illustrate the benefits from sharing indivisible facilities using an example of an ice hockey rink. An ice hockey rink is an expensive facility with large fixed costs, and there needs to be a community of a certain size that is willing to share the rink in order for it to be a productive facility. Other examples include opera houses, airports and power plants.

areas), while *matching* facilitates a higher quality or probability of matches between the firm and a worker due to a denser labor market. The joint empirical prediction for these dynamic effects is a faster urban wage *growth* and *gradual* wage gains after moving to a city.

A widely used way to address the fact that workers with some sort of unobserved comparative advantage choose to live in cities is to employ a fixed-effect (FE) estimator. This approach, pioneered by Glaeser and Mare (2001), takes away any time-invariant individual heterogeneity. However, such an identification of UWP relies on movers who have changed their location over time, and this group might not be representative of the population as a whole. These caveats of the FE estimator notwithstanding, in the absence of random allocation it is often the only way to tackle this issue. In this study, in addition to the above mentioned approach of including FE, I use IQ score as a proxy for unobserved heterogenous ability among the control variables and find that the results are robust with respect to this alternative method of handling workers' heterogeneity.

The remaining UWP after accounting for sorting generally serves as evidence for the presence of agglomeration economies. The empirical methods to distinguish among the different types of agglomeration economies are more diverse. Some studies attempt to distinguish between the static and dynamic mechanisms by finding support for either wage level or wage growth effects. An attempt to further differentiate between the learning and matching mechanisms is to analyze the within- and between-job growth. Faster learning in cities is hypothesized to be reflected in a faster within-job wage growth, while the matching mechanism would imply a faster between-job wage growth. By analyzing movers migrating from urban to rural areas, the support for the learning hypothesis can be found in case these movers do not lose their accumulated human capital from urban areas, since human capital is hypothesized to stay with a worker. More job transitions in cities may lead to conclusions about the importance of the matching mechanism. Various proxies for job quality have also been used in the evaluation of the job match quality in urban areas.

In this study, I use several different approaches to discriminate between the potential theories. I start with analyzing the wage paths of movers year by year, both before and after they move, to shed some light on the dynamics behind the migration process between urban and non-urban locations. I then develop a specification which allows me to quantify the relative importance of the particular sources by simultaneously considering them in one model, net of observed and unobserved sorting bias.

I find that some of the observed UWP is an artifact of high ability individuals sorting into cities: about one-tenth of the spatial wage disparities can be traced back to differences in the unobserved skill composition of the workforce. The remaining wage differential between urban and rural regions can be explained by agglomeration economies. In particular, the static effects appear to be the most dominant source, accounting for roughly three-quarters of the causal effect of cities on wages in the short run, and about one-half in the long run. The rest of the increased productivity effect is attributed to the learning and matching mechanisms, with matching roughly accounting for one-fifth in the short run and one-third in the long run. Finally, learning benefits acquired in cities do not appear to be transferable outside of cities, implying their location-specific character. The results are found to be robust to an alternative, more narrow measure of agglomeration.

This study includes several features that contribute to the prior literature. I use a long panel data set of every working man in Norway from 1986 to 2010. This rich data set allows me to track workers' earning histories over a long time period, as well as to identify when they switch jobs. I am able to define urban status based on municipality of work instead of residency, which reduces the inaccuracies caused by commuting patterns. Additionally, for cohorts born after 1950 the data provide IQ score for each man in the sample, which is utilized when dealing with spatial sorting bias. Furthermore, as argued by Andersson, Klaesson, and Larsson (2013) in the context of Sweden, most of the prior literature has been studying large metropolitan areas, for instance in the US, France or Germany. Analysis of urban areas of a much smaller scale, as Norwegian cities undoubtedly are in the international context, provides an interesting contrast to this literature on huge metropolitan areas. Finally, I attempt to evaluate the relative importance of the potential explanations behind UWP and offer a broad overview over the roles they play in the short, medium and long run.

The paper is organized as follows. Section 3.2 offers a review of the existing literature on urban wage premium and its sources. Section 3.3 describes the data set, defines variables and sample restrictions. Section 3.4 provides a methodological discussion of the two econometrical approaches used, including the discussion about potential biases that might be present. Section 3.5 presents the results and section 3.6 concludes.

3.2 Related literature

Many studies confirm that workers in urban areas earn on average more than workers in rural areas.[4] For example, in a seminal contribution, Glaeser and Mare (2001) show that inhabitants in US dense metropolitan areas earn on average 26-34 percent more. Using French data, Combes, Duranton, and Gobillon (2008) find that average wages in Paris are 15 percent higher than in other large French cities and 60 percent higher compared to rural areas. Di Addario and Patacchini (2008) find that the mean wages in the largest Italian labor markets are five percent higher than elsewhere. For Norway, Bütikofer et al. (2015) document UWP of 16 percent from the mid-1960s till today. Other studies to confirm the existence of a sizeable UWP include Lehmer and Möller (2010) on German data and Andersson, Klaesson, and Larsson (2013) for Sweden.

The core objective of the UWP literature is to disentangle the causal effect cities have on workers' productivity from the fact that workers with more favorable characteristics sort themselves into living in cities. The main difficulty with the existence of such endogenous location decisions is that workers can base their decision to migrate not only with respect to their observable, but also unobservable characteristics. If these sorting effects are not taken into account, the observed wage differences might be falsely attributed to the true agglomeration effect.[5] Glaeser and Mare (2001) try to eliminate the effect of ability sorting by including individual fixed effects in their estimation. After controlling for unobserved heterogeneity, the UWP drops to about 11 or 4.5 percent, depending on the sample used. Using a similar approach, Yankow (2006) finds that higher unmeasured skills and ability can explain between one-half to two-thirds of the US raw urban wage premium. Similarly, Bütikofer et al. (2015) find that UWP reduces to about one-half in Norway. Combes, Duranton, and Gobillon (2008) further confirm that individual skills account for a large fraction of existing spatial wage disparities, by estimating a model of wage determination across local labor markets, using a very large panel of French workers. Gould (2007) develops a structural model that accounts for the self-selection of workers moving into urban areas. He shows that for blue-collar workers the whole observed UWP is due to self-selection, while for white-collar workers a true UWP of 11 percent remains. On the contrary, using a rich panel data set for workers in Spain, a recent study by De la Roca and Puga (2012)

[4]Melo, Graham, and Noland (2009) propose a meta-analysis of 729 elasticities of urban agglomeration taken from 34 different studies in order to make sense of the range of values for agglomeration economies found in the literature.

[5]See Combes, Duranton, and Gobillon (2011) for a detailed discussion of the main sources of bias in the proper identification of agglomeration effects.

finds that sorting of initially more able workers into bigger cities plays at best a minor role in explaining wage differentials.

Some of the recent studies have shifted their focus to the sources of the productive urban wage premium.[6] The first analysis to study the wage growth in addition to wage level effects by Glaeser and Mare (2001) examines the migrants' wage paths before and after the move. They find that in addition to the immediate effect, the migrants experience a continuing wage growth over time that roughly accounts for one-third of the UWP. Additionally, they find evidence that workers who leave cities do not experience wage declines, which supports the explanation in terms of learning. However, Glaeser and Mare (2001) do not address the mechanisms by which the wage growth occurs. Using US data, Yankow (2006) confirms that the wage premium consists of both level and growth elements. He further examines the role of job mobility in the wage growth process. He finds no statistical difference in the wage gain from a single job change experienced by urban and non-urban workers as predicted by the matching hypothesis, but he finds that workers are more likely to change jobs in urban areas, leading to a potentially higher cumulative wage growth. The study by Wheeler (2006) again offers consistent evidence that the wages of workers located in large markets tend to grow faster over time. However, in contrast to Yankow (2006), he finds that much of this association is driven through job changes (as measured by between-job wage growth). Interpreting faster within-job wage growth as a necessary implication of a learning mechanism and faster between-job wage growth as a necessary implication of a matching mechanism, Wheeler (2006) provides a greater support for the latter explanation. Similarly, Andersson, Burgess, and Lane (2007) utilize a uniquely suited data set from the US to show that the quality of firm and worker match does contribute to the urban premium. On the other hand, using U.S. Census microdata, Bleakley and Lin (2012) show that older workers change occupation less often in more densely populated areas. Baum-Snow and Pavan (2012) estimate a structural model that incorporates all the relevant components of the UWP (unobserved ability, search frictions, job match quality, human capital accumulation and endogenous migration). In contrast to Wheeler (2006), this decomposition reveals that within-job wage growth generates more of the city size wage gap than between-job wage growth. They further find that wage level effects are more important for generating wage gaps between medium and small cities, while human capital accumulation is more important in

[6] For a survey on the theoretical microfoundations of urban agglomeration economies see Duranton and Puga (2004). Rosenthal and Strange (2004) review the empirical literature on the nature and sources of agglomeration economies. For more recent surveys, see Puga (2010) and Heuermann, Halfdanarson, and Suedekum (2010).

explaining large-small city size wage gaps. The recent study by D'Costa and Overman (2013) addresses two central issues using data from Britain: sorting and whether the wage premium is received immediately, or through faster wage growth over time. Controlling for sorting they find no evidence of an urban wage growth premium. They find, however, that having worked in a city at some point does have an impact on wage growth of rural workers compared to rural workers who have never worked in a city. Analyzing Spanish workers, De la Roca and Puga (2012) find that about one-half of the gains from working in bigger cities is static advantages, while another half accrues to workers over time gaining urban experience. The further evidence that these workers are able to take their dynamic gains with them when they relocate is interpreted as the importance of learning. A Swedish study by Andersson, Klaesson, and Larsson (2013) find that after accounting for sorting, agglomeration economies are in general small. They analyze two groups of skills - routine and non-routine - and find that agglomeration economies are larger for workers with skills associated with non-routine job tasks. They also appear to involve human capital accumulation, as evidenced by the fact that workers retain their premium after moving away from denser regions. Krupka and Noonan (2013) examine the relative importance of the different productivity advantages and find that learning efficiencies are most important in medium-sized cities, while both learning and matching are important in the largest and smallest cities. Finally, Carlsen, Rattsø, and Stokke (2013) follow the approach of De la Roca and Puga (2012) and calculate dynamic wage premium for Norwegian workers. They analyze learning across education groups and separate between large-, medium-, and small-size regions. They find that the initial premium does not vary across education groups, but the learning advantages differ with respect to education and are concentrated to the most educated individuals.

3.3 DATA AND SAMPLE SELECTION

The data source is administrative register data maintained by Statistics Norway, which allows me to construct a long panel data set for the Norwegian population from 1986 to 2010 on calendar year basis.[7] It includes earnings, municipality of work and a battery of other personal and job characteristics such as age, gender, immigrant and marital status, hours worked, type of employment and sector.[8]

[7]See Møen, Salvanes, and Sørensen (2003) for a description of the data.

[8]Some studies (e.g. Andersson, Klaesson, and Larsson (2013)) analyzing UWP include sector dummies among control variables. Even though the data contains this information, I do not include sector identifiers among the

The dependent variable is log of gross earnings in 2010 Norwegian Kroner (NOK) reported in the tax registry. Earnings include labor earnings, as well as taxable sickness benefits, unemployment benefits, parental leave payments, and pensions.

Education is measured as the highest completed level of education an individual has achieved. The data is obtained from the education register which is based on school reports sent directly to Statistics Norway, thereby minimizing any measurement error due to misreporting. I create three education groups that I call low- (completed primary and lower secondary level), medium- (completed upper secondary school level), and high-educated (completed at least tertiary level).

Worker's value of experience is generated as number of years worked since I start observing them in the data. The urban experience is the cumulative work experience acquired in urban locations. I further generate a job index variable that indexes the number of jobs a worker has had so far (including the current one). This variable is computed from the information about job start date and job stop date for each job spell of a worker.[9] The urban job index variable counts only the number of jobs an individual has had in urban location.

I allocate workers to urban or rural location according to their municipality of work. Statistics Norway divides municipalities into four different levels (on a scale from 0-3) according to population and centrality (SSB, 1994). Centrality is a measure of a municipality's provision of central functions and services or its geographical position in relation to a centre with such functions, with the centrality criterion having greater weight than the population criterion (i.e., municipalities that fulfill the population criterion can belong to a lower level if functional requirements are not satisfied). Municipalities on level 3 are regional centres and have a population of at least 50 000 inhabitants. Level 3 further includes municipalities that are located within 75 minutes (90 minutes for Oslo) travel time from these regional centres. Level 2 includes settlements with a population between 15 000 and 50 000 or municipalities that are above 50 000 but are not regional centres, plus municipalities located within 60 minutes travel time from an urban settlement of level 2. Level 1 includes municipalities with a population between 5 000 and 15 000 and municipalities that are located within 45 minutes travel time from a settlement of level 1. Level 0 covers all other municipalities.[10] Figure A.1 in the Appendix displays the centrality level for each municipality in Norway. Following this definition of centrality level, I

regressors since I tend to think of sector-change effect as being part of the agglomeration effect I try to measure. I utilize the information about the sector only to exclude public sector workers.

[9]Since by the construction of the data the workers are observed in their job establishment only once a year, I am only able to identify a maximum of one job change per year.

[10]If the municipality fulfils the requirements for centrality on more than one level, the highest of these levels applies.

define municipalities as urban if their level of centrality is 3 or 2 (i.e., population above 15 000 or within commuting distance from such a municipality). According to this definition 183 out of 430 Norwegian municipalities[11] are classified as urban. As a robustness check, I also reproduce the main results using an alternative definition of urban based only on centrality level 3 (see the Appendix).[12] According to this alternative urban classification, 104 out of 430 Norwegian municipalities are classified as urban.

For cohorts born after 1949 I have access to IQ test scores from Norwegian military records (measured around age 18), which can be used as a control for personal ability (see Sundet, Barlaug, and Torjussen (2004) and Thrane (1977) for details). The military service in Norway is mandatory for men, and each male individual is tested for physical and psychological suitability during the examination. The IQ score is constructed as an unweighted mean of three tests - arithmetics, word similarities, and figures - and converted into a single digit number on a 0-9 scale, so that the scores are normally distributed with a mean of 5 and a standard deviation of 2.

The sample used contains yearly observations of working men aged 16 - 65, who were resident in Norway in at least one of the years from 1986-2010.[13] I keep only full-time workers (who worked at least 30 hours per week), exclude self-employed (since I am interested in labor earnings) and public sector workers (whose wages are not determined by market outcomes but national regulations). I further exclude observation years in which a worker switched jobs or worked less than a full twelve months. Such years might indicate that a worker is searching in the labor market, and these years might introduce a bias associated with lower earnings (for example due to unemployment). I also exclude return and multiple movers (i.e., those who changed their municipality of work from rural to urban or vice versa more than once) in order to avoid the bias introduced by migrants' selection into return migration, or to avoid the period in between two moves to count as both the pre- and post-move period.[14]

The final panel consists of about 12.4 million person-year observations on 1.33 million individuals. Table 3.1 presents summary statistics for the key variables that are included in the analysis. It also demonstrates how urban and non-urban workers differ in regards to their

[11] Even though municipalities are undergoing continuous consolidation, only minor changes occurred between 1986 and 2010. Of even smaller importance is the consolidation with respect to changes in centrality level. I use the 2010 municipality and centrality classification for all years.

[12] Level 3 is restricted to the following 6 largest urban settlements in Norway: Oslo, Kristiansand, Stavanger, Bergen, Trondheim and Tromso.

[13] Following the previous studies, I concentrate on male workers due to concerns connected with the labor force participation and different mobility patterns of female workers.

[14] I also estimate my regressions on a sample that does not exclude years of job change or return and multiple movers. The results (not reported) are robust to this alternative sample choice.

personal and job-related characteristics. Urban wage premium is apparent in the group averages: urban workers make 19 log points more than rural workers. Urban workers also tend to have on average 0.4 more jobs than those in rural areas and somewhat higher IQ score. Urban regions also host twice as large a share of tertiary educated and display only slightly higher shares of married and immigrants. Overall, the sample includes 10.5 million observations contributed by urban workers and 1.9 million observations contributed by rural workers.

3.4 EMPIRICAL SPECIFICATIONS

3.4.1 ANALYSIS OF MOVERS

The first empirical model follows Glaeser and Mare (2001) and identifies workers that move to and from urban areas. It allows for separate dummies for each calendar year in the labor market before and after the move, which capture the complete movers' wage path over time. Specifically, I estimate

$$\ln \mathrm{W_{it}} = \alpha \mathrm{X_{it}} + \beta \mathrm{USTAYER_i} + \sum_{j\in J} \gamma_j^{RU} \mathrm{M_{jit}^{RU}} + \sum_{j\in J} \gamma_j^{UR} \mathrm{M_{jit}^{UR}} + \phi_i + \varepsilon_{it}, \tag{3.1}$$

where i refers to individuals, t to years, and j indexes the number of years after the move and ranges from -10 to 10, with negative values for years before the move and 0 for the calendar year in which the move occurs. There are two additional border categories in the set J that are called "less than -10" and "more than 10", and contain all observations that are less than -10 or more than 10 years after the move.[15] The variables of interest are $\mathrm{M_{jit}^{RU}}$ and $\mathrm{M_{jit}^{UR}}$ indicators, which take on a value of 1 if an individual i in year t is j years after the rural-to-urban (RU) or urban-to-rural (UR) move, respectively. $\mathrm{USTAYER_i}$ is an indicator variable equal to 1 for people who work in urban municipalities all the time. The omitted category is stayers in rural areas. Thus, the estimates of γ_j^{RU} and γ_j^{UR} give us the percentage wage difference for movers j years after the move relative to the comparison group that stays in the rural area all the time. $\mathrm{X_{it}}$ is a vector of control variables including year, age, continent of origin and marital status dummies. The individual fixed effects ϕ_i are included in the fixed-effects estimation. The dependent variable is annual log earnings.

[15]These border categories are added for the dummies to be comprehensive.

The parameter estimate of the variable USTAYER_i, if significantly positive, documents the existence of an urban wage premium that remains after controlling for basic observable characteristics.[16] The method further sheds some light on the dynamics behind the migration process between urban and non-urban locations. It leads to a better understanding of how well the movers are doing in terms of earnings compared to stayers, both before and after they move, whether and how their earning paths change with the approaching time of the migration, whether and how much they gain or lose immediately upon relocation, and how their earnings evolve in the destination location.

The analysis of movers, while being interesting on its own, further provides an empirical test of the predictions about movers' wages that depend on the source of UWP. If spatial sorting is the main source of the observed urban wage differential, i.e., the whole wage discrepancy originates from the fact that people living in cities are inherently different from people living outside of cities, then migrants moving between these two types of regions should not experience any wage gain (or loss) after relocation.[17] Thus, if I observe that migrants gain (or lose) some wage premium (either immediately or gradually) after relocation, the spatial sorting is not likely to be the only source of UWP and agglomeration economies play a role as well.[18]

If agglomeration economies are the main source of the UWP, then empirical implications are as follows. Static premia should be gained (lost) immediately after moving to (away from) an urban area. In the case of learning and matching, the empirical implication of both effects predicts urban wage gains over time. On the other hand, upon moving away from a city the advantages of accumulated human capital through learning should stay with a worker, while the benefits from higher average quality or frequency of firm-worker job matches should clearly not follow workers when they arrive in a rural area since they are not subject to a dense labor market any more. However, one has to be cautious in drawing strong conclusions from the empirical predictions of wage loss upon out-migration from urban locations since the human

[16]This UWP is identified off of rural and urban stayers and is a mixture of sorting on unobservable characteristics as well as both static and dynamic urban effects.

[17]The term spatial sorting throughout the text refers to sorting on unobservable time-invariant characteristics as I always account for the observable individual heterogeneity.

[18]While the empirical prediction of sorting being an exclusive source of UWP is rather straightforward (no wage changes after relocation), the determination of the importance of sorting is less so. The common approach in the literature (Glaeser and Mare (2001), Yankow (2006), Combes, Duranton, and Gobillon (2008) etc.) to quantify the extent of sorting relies on the comparison of the baseline OLS estimate of UWP that incorporates both sorting and agglomeration economies, to the FE estimate of UWP that is net of sorting (though only identified off of movers). The extent to which the estimates change after accounting for sorting then reflects the importance of sorting. The set-up I am using here, however, does not provide a single FE estimate that could be directly compared to the baseline OLS estimate identified by stayers. This is because the FE estimates differ with respect to the number of years after migration.

capital accumulated in cities might be city-specific. This would be the case if learning entailed a very distinct, city-specific knowledge that is not fully transferable to non-urban areas. In that case workers might well experience wage loss upon moving away from an urban location and learning and matching would be observationally equivalent.

The changes in earnings for migrants after relocation needed to test these empirical predictions could, in fact, be inferred from comparing the pre-move to the post-move estimates of wage premia (calculated relative to rural stayers) from the OLS estimates of equation (3.1). However, since the sample is not balanced and movers move at different stages of their lives and are heterogeneous in unobserved characteristics, more credible estimates of the wage developments over time within individuals can be obtained by employing a fixed-effect (FE) estimation, which is based on the within-person transformation and accounts for individual permanent heterogeneity.[19]

Finally, the method provides some sense of the relative importance of static and dynamic effects by comparing the size of the immediate and gradual wage gains after migration.

3.4.2 THE SOURCES OF UWP

The primary focus of the next specification is to differentiate between particular mechanisms behind the UWP. This specification simultaneously considers both static and dynamic urban advantages while controlling for the unobserved worker characteristics. The model implies that wages of a worker *i* at time *t* are given by:

$$\ln \mathrm{W_{it}} = \alpha \mathrm{X_{it}} + \beta \mathrm{U_{it}} + \gamma_1 \mathrm{EXPR_{it}} + \gamma_2 \mathrm{JOBS_{it}} + \gamma_3 \mathrm{UEXPR_{it}} + \gamma_4 \mathrm{UJOBS_{it}} + \phi_i + \varepsilon_{it}, \quad (3.2)$$

where $\mathrm{U_{it}}$ is equal to 1 if an individual i worked in an urban municipality in year t, $\mathrm{EXPR_{it}}$ is worker's experience in years accumulated up to time t, $\mathrm{JOBS_{it}}$ indexes how many jobs a worker has had up to time t (including the current one), $\mathrm{UEXPR_{it}}$ is the number of years of experience a worker has accumulated up to time t in an urban area, and, similarly, $\mathrm{UJOBS_{it}}$ indexes how many jobs a worker has had up to time t in an urban area. $\mathrm{X_{it}}$ includes age, year, education, continent of origin and marital status, and ϕ_i are individual fixed effects that enter the equation in the case of the fixed-effects estimation.

[19] The omitted category in the fixed-effect estimation is 1 year prior to move.

I start with estimating equation (3.2) by pooled OLS. The variable U_{it} allows for the level wage gap between urban and rural areas, and the importance of static mechanisms should be reflected in the positive estimate of this coefficient. The coefficient on $EXPR_{it}$ represents the returns to working experience, and in line with general findings in the literature I expect that increasing experience is positively associated with earnings. Since the likely outcome for workers searching in labor markets is greater wage growth when changing employers, $JOBS_{it}$ is as well assumed to be positively associated with earnings. The variables $UEXPR_{it}$ and $UJOBS_{it}$ allow the returns to experience and job switches to differ for urban and rural areas. If learning is a relevant mechanism making workers in cities more productive, it should be reflected in the positive estimate of the coefficient on $UEXPR_{it}$. If the improved quality of matches in urban regions drives the urban-rural wage differential, it should lead to positive estimates of the coefficient on $UJOBS_{it}$. The dependent variable is annual log earnings.

In order to obtain a causal interpretation of the estimated urban coefficient on U_{it}, I have to account for spatial sorting to ensure that the coefficient does not only reflect the population heterogeneity. The result from the OLS estimation is biased in case urban and rural populations differ in innate characteristics that are correlated with earnings. Generally, workers' heterogeneity represents an important challenge in the estimation of causal effects of cities. If the heterogeneity is based on time-invariant characteristics, however, there is a way to account for it. A common approach used in the literature is to include worker fixed-effects in the estimation. Such estimates are based on the changes in location over time within workers, hence conditioning out any permanent individual characteristics. The importance of spatial sorting can subsequently be quantified by the magnitude of the drop of the FE estimates compared to the baseline OLS estimates.[20] The estimate of the level premium that remains after controlling for spatial sorting of workers should then capture the causal effect of static agglomeration economies.

Nontheless, such a method for accounting for spatial differences in the composition of the workforce is not flawless. It works well as long as workers choose their location based on their time-invariant characteristics. As highlighted by Baum-Snow and Pavan (2012), the source of concern is that the identification comes from movers who possibly constitute a highly selected

[20]This approach is often applied in specifications that ignore the dynamic component of the UWP. Since the FE estimates can reduce the bias associated with the omitted dynamics, the drop in the earnings premium that occurs after incorporating worker FE can reflect both the importance of sorting as well as the importance of learning and matching. As a consequence, the importance of sorting might be overrated. In my specification (3.2) I simultaneously consider both static and dynamic mechanisms, hence I know that the drop in FE estimates addresses sorting only.

set of workers.[21] For instance, mostly workers with a particularly attractive wage offer might move, or the decision to migrate might depend on the negative earnings shock a worker experiences in the original location. In both cases the movers to cities will bias the static earnings premium upwards. However, if the endogeneity behavior is similar for moving in the opposite direction, the movers leaving cities will to some extent reduce this bias since they will understate the urban wage premium. In order to reduce the potential bias from the negative earnings shock experienced right before the migration (Ashenfelter dip), I exclude the years immediately preceding the migration in some specifications. This decision is based on the results from the analysis of movers in the first model (3.1), which indeed indicate the presence of a substantial negative pre-migration earnings shock.[22]

To alleviate the concern with a non-random subsample, one would ideally like to instrument for the migration decision. While a number of instruments have been used in the literature, e.g. distance, natural shocks, political or historical factors (Card (1990), Combes, Duranton, Gobillon, and Roux (2010), McKenzie, Stillman, and Gibson (2010), Zaiceva (2006) etc.), these instruments rely on time- and/or location-specific events or economic conditions and cannot be applied in a different context or for a different time period. Bartik (1991) suggests to instrument local labor demand growth using a shift-share index that is created by interacting the initial employment share in each industry in each geographical region with national employment growth in those industries. I tried using this instrument to estimate UWP, but unfortunately the instrument turned out to be a weak predictor of migration at the municipality level.

To challenge the robustness of the FE results, I re-examine the importance of sorting and static UWP and employ an alternative approach to account for workers' heterogeneity - IQ score as a proxy for workers' ability (for an alternative application of this identification strategy in the Norwegian context, see, for example, Bhuller, Mogstad, and Salvanes (2011)). Although IQ score is arguably not a perfect measure of workers' heterogeneity, unlike the FE estimator this identification strategy does not rely exclusively on movers.

[21] Note that in specification (3.2) only the identification of the static UWP comes from movers. In case of the dynamic premia all workers contribute to the estimation as urban experience or urban job switching vary over time for urban stayers as well.

[22] Another possible way for movers to constitute a non-representative subsample is when those with higher potential for learning in cities, or those who are less prone to switch jobs unless a substantially better wage is offered, disproportionately choose to move to cities. Estimations based on such a subsample will, however, leave the static premium estimate unbiased, but will bias the estimates of the dynamic mechanisms. Since the dynamic premia are estimated both from movers and stayers, this bias should be of minor importance given that migrants represent a relatively small fraction of the urban population.

Another important feature of the model given by equation (3.2) is that in addition to the qualitative analysis of whether a particular mechanism plays a role, it also allows for the possibility to quantify the relative importance of the different sources of agglomeration economies. I therefore calculate the short- (5 years) , medium- (10 years), and long-term (20 years) earnings premia based on the results from equation (3.2) for workers who spend an equal amount of time (5, 10, or 20 years) in urban areas as their rural counterparts with similar characteristics. I can then evaluate the contributions of particular sources into these short-, medium-, or long-term wage gaps.

It is worth noting that learning and matching are not necessarily mutually exclusive hypotheses. First, as argued by Krupka and Noonan (2013), learning can, for instance, mean that in cities it is easier to learn about better jobs. Second, matching can be interpreted as a match to a firm whose employers one would benefit most from interacting with. Third, it can be argued that there is another channel through which the matching mechanism can operate. Specifically, a faster wage growth within the same job in an urban area can be experienced even without the presence of faster learning, but due to an increased negotiation power based on better outside options in the denser labor market. In my setting, in the first case the learning effect will be falsely attributed to matching, while in the two latter cases matching will be incorporated as part of the learning mechanism, and the effect of learning will thus be overestimated. In this study, I concentrate on the basic, mutually exclusive interpretation of learning and matching and assume the aforementioned biases to play a minor role.

3.5 RESULTS

3.5.1 ANALYSIS OF MOVERS

The results from both OLS and FE estimation of equation (3.1) are plotted in figure 3.1, which shows the estimates of coefficients on migration dummies and the urban stayer dummy (if relevant).[23] The estimated coefficients can be interpreted as the earnings premium in log points (which is interpreted as percentages) in relation to the comparison group (rural stayers or the year -1 for the OLS or FE estimates, respectively). For the detailed regression results, see the corresponding tables A.1 and A.2 in the Appendix.

[23] The estimate for the year of move $(j = 0)$ drops out since the sample excludes the years when the worker switched jobs.

Following the OLS results, urban stayers earn a significant premium of about 13% compared to rural stayers. Note that this estimate accounts for observable characteristics, but still comprises the sorting on unobserved characteristics together with the effects of both static and dynamic agglomeration economies and is identified off of stayers only since all movers are represented by one of the migration dummies. Such a magnitude is similar to what have previously been found in the literature - in a similar set-up for US workers Glaeser and Mare (2001) obtain the estimate of 16% - 20% depending on the sample choice. The documentation of the existence of an earnings premium in urban regions is an important premise for my subsequent analysis of the sources of this earnings differential.

The figure 3.1 further sheds some light on the characteristics of the migration process as such. Migrants' earnings begin to deteriorate several years prior to the move, with the largest dip in years -1 and -2. This phenomenon has been observed and described in the literature as the "Ashenfelter dip". Ashenfelter (1978) observed that the wages of future participants in a training program had a tendency to fall in the period before they entered the program. Many subsequent studies have confirmed this observation in various applications.[24] If Ashenfelter's dip describes a transitory phenomenon, then the within-estimator overestimates the effect of the treatment. Given that only years when workers worked a full 12 months are included in the sample, and thus separation into unemployment is not a likely explanation, this dip is strikingly large; the future rural-to-urban migrants drop from an earnings premium of about 4% in year -3 to a wage penalty of about 7% in year -1 (relative to average rural earnings level). One plausible explanation of this phenomenon, in addition to decreased earnings with the same employer prior to the move, is that workers have been laid off, but are still affiliated with the company. Not an uncommon practice in Norway, this can in fact cause unemployed workers to appear as employed in the dataset during this period. In order for this negative transitory shock not to bias my estimates of UWP, I drop two years prior to the move in my next specification.

Both estimation procedures, OLS and FE, offer the same qualitative conclusions in relation to underlying sources of UWP. First, after the move to (away from) the city movers' earnings immediately rise (fall), which suggests that sorting alone can not explain the whole observed UWP and that the agglomeration economies are present as well. In particular, it is the evidence for the presence of static agglomeration economies. Second, the wages of migrants to cities

[24] Huttunen, Møen, and Salvanes (2011), for example, analyze the average effect of displacement on earnings for Norwegian workers who remain in the labor force. They find that the earnings of displaced workers begin to deteriorate four years prior to the displacement incident.

continue to monotonously rise after the arrival to the city, which provides evidence for an urban wage growth that is faster than in rural areas (due to a flat earnings profile of these movers prior to the move). Thus, movers do not acquire their whole premium at once but earn it gradually by gaining urban experience. This observation is consistent with findings of Glaeser and Mare (2001), and is a common prediction for both hypotheses - learning and matching. Finally, concluding from the OLS estimates, urban-to-rural migrants' earnings reach the average wage level in rural areas after migration. This suggests that having been in an urban location does not provide the migrants with any significant advantage compared to long-term rural workers. Under the assumption that the accumulated human capital stays with a worker, this provides evidence against learning. However, if the accumulated human capital is city-specific and thus non-transferable outside cities, learning can still be an underlying mechanism.

The overall picture from this analysis is that there is a substantial wage differential between urban and non-urban regions that cannot be solely explained by spatial sorting. Agglomeration economies seem to play an important role as well; in particular, deriving the results from the FE estimates, static advantages appear to be the dominant source, but the dynamic advantages are present as well. If learning is one of the dynamic mechanisms, then it is not likely to be transferable outside urban areas.

Figure A.2 and corresponding tables A.3 and A.4 in the Appendix provide evidence that very similar time patterns emerge for the alternative definition of urban municipalities, based on the centrality level 3 only.

3.5.2 The sources of UWP

The results from estimating equation (3.2) are reported in table 3.2. Column 1 shows the OLS results from the full sample. The positive coefficient of the urban dummy documents the existence of a large static urban wage premium of 8.5%, accounting for basic observable characteristics and the dynamic components. The estimates of the experience and number of previous jobs variables have an expected positive sign - each year of rural working experience yields 0.9% wage growth, and each new job in a rural area brings along 0.5% earnings premium. Significantly positive urban experience and number of urban jobs indicate the presence of the dynamic effects in explaining the existing UWP; each year of urban working experience yields 0.2% faster wage growth, and each new job in an urban area brings along 0.9% higher earnings premium than each new job in rural areas.

In view of previous research, I expect a reduction of the static UWP when accounting for workers' unobservable permanent heterogeneity. The FE estimate in column 2 confirms this expectation; the earnings premium drops to 7.6%, suggesting that sorting matters. Sorting can explain about 10% of the observed wage gap.[25] The remaining UWP of 7.6% after accounting for sorting reflects the magnitude of static advantages. As for learning, each year of urban experience yields 0.1% higher wage growth compared to rural experience. The effect of experience is thus almost doubled in cities (since the rural estimate for experience is 0.1% as well). As for matching, with each job change in a city a worker gains 0.8% higher premium than what he would gain outside a city. Given that the wage growth arising from a single job change outside cities is 0.5%, it is more than a two-fold increase.

Results from the sample with non-missing IQ data reveal very similar patterns to the full sample estimates. After accounting for individual IQ score, the static urban wage premium drops from 6.7% to 6.2%, suggesting that the sorting explains 7% of the premium. This is somewhat less than for FE estimates, which is reasonable since the IQ control captures only the IQ-based permanent workers' heterogeneity. The extra effect of learning in cities is estimated to be 0.3%, and the the extra effect from changing a job in a city 0.9%.

Overall, depending on the sample choice and specification, the static effects range from about 6.2% to 7.6%, the effect of learning is reflected in 0.1-0.3% higher return to urban experience, and the effect of matching translates into 0.8-0.9% higher return to each new job in urban regions, relative to rural areas. While the statistical significance of the learning and matching mechanisms has been confirmed, the estimates may appear rather low in magnitude. I will next illustrate the economic significance of learning and matching mechanisms in explaining short-, medium-, and long-term UWP.

To evaluate how the three different agglomeration economies are interrelated, I calculate the urban wage premia for three different period lengths (5, 10, and 20 years) using the FE estimates from column 2 in table 3.2. The premium from the static agglomeration economies is common for all three period lengths and amounts to 7.6%. Learning and matching advantages, on the other hand, accrue to workers with time and thus depend on the length of period spent in cities. Learning premium can be calculated as years of urban experience multiplied by the estimated coefficient on UEXPR (0.001). The premium that originates from the matching

[25] A lot of studies find sorting to explain a much bigger part of the UWP. It is worth noting that these studies, while not allowing for the dynamic mechanisms in their specifications, interpret the drop between the FE and OLS estimates as evidence of the importance of sorting, but part of this drop can, in fact, be attributed to reduced bias from omitting the dynamics, thus overrating the importance of sorting (see De la Roca and Puga (2012)).

mechanism, however, is more complex because it can be gained through two different channels. First, workers find more productive job matches in cities over time, and second, workers are more likely to engage in search for new employment in urban labor markets and have a higher frequency of job changes, which can both lead to a greater cumulative wage growth.

To account for both of these channels, I need to know the average number of jobs workers have had in their location during a given period. The total effect from matching will then be calculated as the average number of jobs in urban areas times the coefficient estimate of the UJOBS variable (0.00795), plus the difference between the average number of urban and rural jobs, multiplied by the estimated coefficient on JOBS (0.00454).

The average number of jobs in each type of location is computed from the sample averages for stayers for a given number of years of working experience.[26] The results are reported in the first two columns of table 3.3. Interestingly, in line with the descriptive patterns in table 3.1, workers tend to have a higher rate of occupational transitions in urban areas. For example, during the time period of 20 years urban workers have had on average 1 employer more than rural workers.

Plugging in the average number of jobs reported in table 3.3 into the calculation of the, for example, short-term effect from matching, the total effect is computed as 2 multiplied by 0.00795, plus (2-1.8) multiplied by 0.00454. The calculation yields a premium of 1.7%. All together, the estimated average earnings differential between a person with 5 years of urban and a person with 5 years of rural working experience consists of the following effects: 7.6 % premium arises from static advantages, 0.5% premium comes from urban experience (interpreted as learning advantages, computed as 5 times urban experience coefficient, 0.001) and the impact of matching is 1.7% (calculated as explained above). Expressed in relative terms, out of the total 7.6 + 0.5+ 1.7 = 9.8% short-term premium, 78% is static premia, learning makes up for 5%, and 17% is due to a higher incidence of job changes together with their increased quality.[27]

Table 3.3 further summarizes the results of the analogous computations for the medium- and long-term premia. For a comparison, table 3.4 reports analogous results based on parameter estimates from the IQ sample in column 4 in table 3.2. In summary, in the short run static premia are responsible for the bulk (64-78%) of the wage disparities between urban and rural locations

[26]Obtaining the average number of jobs from the sample means might yield biased estimates if, for example, workers that are more prone to change jobs sort themselves into living in cities. I rely on these estimates and trade off the lack in appropriateness for simplicity and transparency.

[27] It is important to note that even though the calculations are done exactly, the whole exercise should rather serve as a rough measure of the importance of the different factors in explaining the urban wage differential for an average worker, providing us with a general idea about their economic significance rather than exact numbers.

(after controlling for sorting). Learning and matching mechanisms together contribute with a relatively smaller share in the short run (22-36%), but roughly double their overall importance in the long run (42-61%). In the short run, it is the matching that is the more important contributor, while in the long run it is less clear since the results are sensitive to the sample choice. Nevertheless, the results suggest that both learning and matching mechanisms are of economic significance.

As a robustness check, I repeat the empirical analysis described above for an alternative definition of urban municipalities, based on centrality level 3 only. The estimates of equation (3.2) and the relative contributions of the single factors into urban wage premia are reported in tables A.5 and A.6-A.7 in the Appendix. The tables provide a compelling evidence that the results are robust to an alternative urban definition.

3.6 Conclusion

The aim of this study is to empirically examine the reasons why workers in urban areas earn more than workers outside urban areas. Using rich administrative panel data on all male workers in Norway from 1986-2010, I first analyze the time patterns of migrants between the two types of regions. I find that the raw urban wage premium, when identified of stayers and not accounting for sorting or for the dynamic structure, is about 13%. This magnitude is in line with the estimates from a similar set-up by Glaeser and Mare (2001) for U.S. data. I also find that movers to cities experience immediate wage gains, and similarly, movers away from cities experience immediate wage losses, which implies the presence of static agglomeration economies and the fact that sorting alone is not a sufficient factor to explain the whole wage differential. The earnings profiles of workers after coming to urban areas monotonously continue to rise. This finding is consistent with the presence of dynamic advantages, acquired through either learning or matching. The analysis further shows that after migration urban-to-rural migrants' earnings fall to match the average rural earnings level. This result makes learning a likely explanation only in case the human capital accumulated in cities does not stay with a worker after relocation. I further find that workers experience a substantial negative transitory shock several years before migration, which might bias the estimates if not taken into account.

I next develop a framework that combines all three types of agglomeration economies (static, learning, and matching) in one model by including an urban dummy and allowing the returns to experience and each job change to differ by location. This framework allows for a direct assessment of the importance of sorting by an introduction of worker FE or, alternatively, IQ score as a control variable. The introduction of worker FE leads to a substantial reduction in the observed UWP - dropping from 8.5% in the baseline OLS estimation to about 7.6%. The reduction is interpreted as the effect of spatial sorting and amounts to about 10% in this data. The remaining UWP of 7.6% serves as evidence for a large influence of the static mechanisms. The rest of the findings present empirical evidence of superior learning and job matching in urban labor markets. Specifically, the effect of learning is almost twice as high in cities, with 0.1% higher return to each year of urban experience. As for matching, with each job transition a worker in urban areas experiences 0.8% higher premium, which is more than a two-fold increase in comparison with rural areas. In terms of relative importance of static, learning, and matching mechanisms in explaining the UWP, under the assumption that sample averages give reliable estimates of the mean number of jobs of workers in each type of location, I find that static agglomeration economies are responsible for the bulk of the wage disparities, and learning and matching explain in the short run about 5% or 17%, respectively, increasing their influence in the long run to 15% or 27%, respectively. The numbers slightly change when I use the estimates from the specification including IQ control instead of FE, but lead to similar conclusions, perhaps only assigning a greater relative explanatory power to the learning mechanism. The results are robust to an alternative measure of agglomeration which includes only the six largest cities of Norway among urban areas.

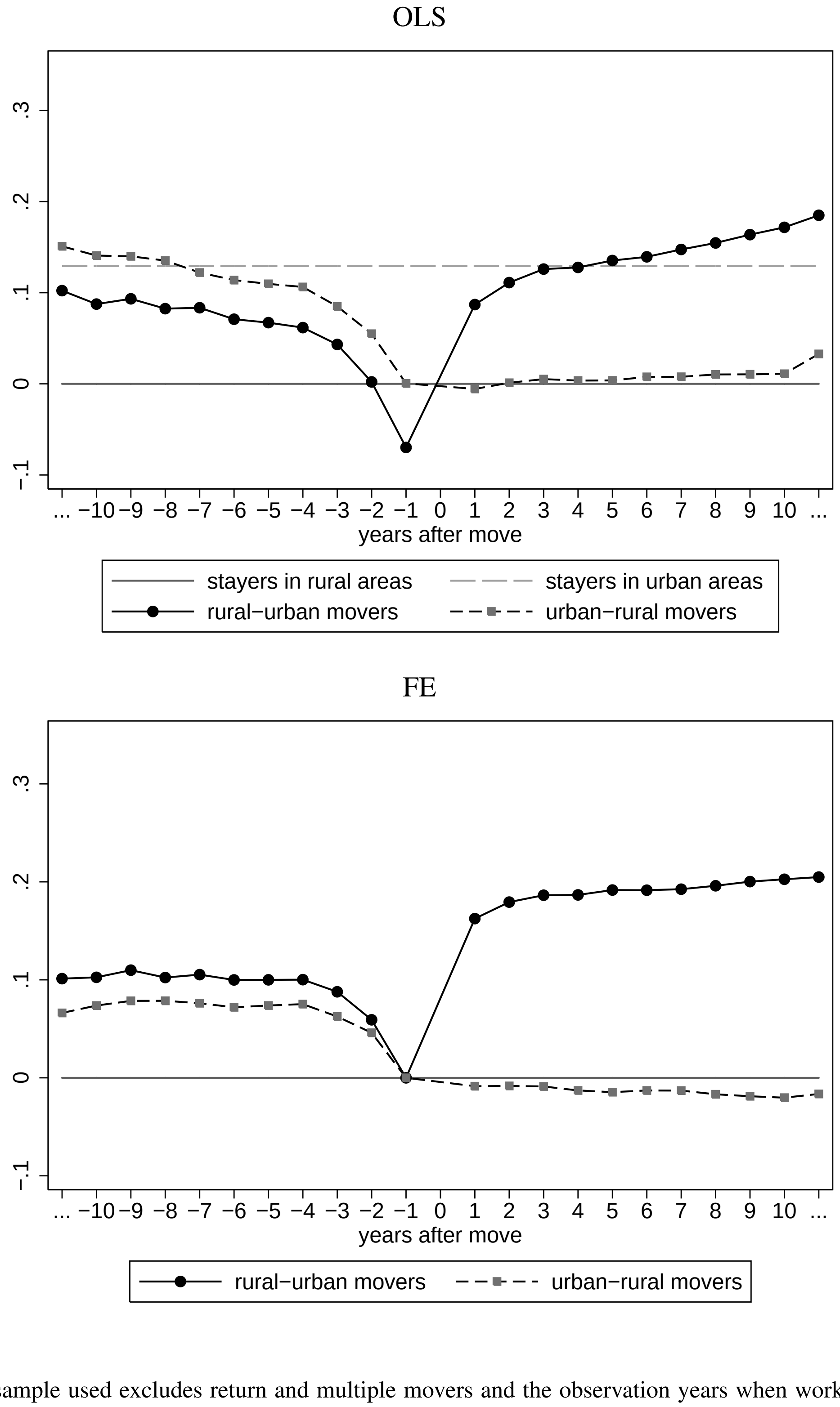
The broad patterns found in both methodologies are in accordance with each other - all sorting, static and dynamic advantages are important factors in explaining UWP. The empirical evidence presented in this paper helps to shed more light on the potential explanations behind the increased productivity of workers in urban locations, and I hope it will lead to a more thorough examination of the different mechanisms, especially the dynamic ones, in the future.

FIGURES AND TABLES

Table 3.1: Sample statistics

Variable	Total		Urban area		Rural area	
	mean	s.d.	mean	s.d.	mean	s.d.
Log earnings	12.86	(0.5)	12.89	(0.5)	12.7	(0.4)
Age	41.4	(11.5)	41.3	(11.5)	41.4	(11.8)
Experience	9.3	(6.2)	9.3	(6.2)	9	(6.2)
Number of jobs	2.4	(1.6)	2.4	(1.7)	2	(1.4)
IQ score	5	(2)	5.01	(2)	4.5	(1.9)
Share of primary educated	0.21		0.21		0.24	
Share of secondary educated	0.56		0.55		0.64	
Share of tertiary educated	0.23		0.24		0.12	
Share of married	0.58		0.58		0.56	
Share of immigrants	0.03		0.03		0.02	
Observations (mil)	12.4		10.5		1.9	

Figure 3.1: Analysis of movers' wages by years after migration



Notes: The sample used excludes return and multiple movers and the observation years when workers switched jobs or did not work a full 12 months. OLS estimates in the upper panel are relative to wages of stayers in rural areas. FE estimates in the bottom panel are relative to wages one year before migration. Most estimates, except for the "after urban-to-rural migration" dummies, are highly significant (see tables A.1 and A.2 in the Appendix for standard errors, parameter significance and list of controls included).

Table 3.2: Estimation results

	Full sample		IQ sample	
	(1) OLS	(2) FE	(3) OLS	(4) IQ-control
Urban	0.0852***	0.0762***	0.0674***	0.0619***
	(0.000606)	(0.00212)	(0.000762)	(0.000756)
Experience	0.00909***	0.00127***	0.00628***	0.00658***
	(0.0000573)	(0.000218)	(0.0000708)	(0.0000703)
Number of jobs	0.00477***	0.00454***	0.00359***	0.00509***
	(0.000223)	(0.000544)	(0.000253)	(0.000251)
Urban experience	0.00166***	0.00101***	0.00275***	0.00276***
	(0.0000573)	(0.0000905)	(0.0000690)	(0.0000684)
Number of urban jobs	0.00920***	0.00795***	0.0102***	0.00899***
	(0.000239)	(0.000578)	(0.000271)	(0.000269)
IQ score	-	-	no	yes
Education	yes	-	yes	yes
Age	yes	yes	yes	yes
Year	yes	yes	yes	yes
Continent of origin	yes	-	yes	yes
Marital status	yes	yes	yes	yes
N	12,388,859	12,388,859	7,325,995	7,325,995
R-sq	0.314	0.312	0.374	0.383

Notes: The samples exclude 2 years prior to move for the migrants. The IQ-sample in columns 3 and 4 is restricted to individuals with non-missing observations on IQ test scores. The regressions include full set of dummies for the control variables listed in the table. Standard errors in parentheses are clustered on individual level. *p<0.05, **p<0.01, ***p<0.001.

Table 3.3: Relative importance of static effects, learning, and matching

time period	average # of jobs		% wage premium due to:		
	rural	urban	static	learning	matching
5 years	1.8	2	7.6 (78%)	0.5 (5%)	1.7 (17%)
10 years	2.3	2.9	7.6 (68%)	1 (9%)	2.6 (23%)
20 years	2.8	3.8	7.6 (58%)	2 (15%)	3.5 (27%)

Notes: Relative shares in parentheses. Results based on FE estimates on the full sample reported in table 3.2.

Table 3.4: Relative importance of static effects, learning, and matching, IQ sample

time period	average # of jobs		% wage premium due to:		
	rural	urban	static	learning	matching
5 years	1.9	2.2	6.2 (64%)	1.4 (14%)	2.1 (22%)
10 years	2.5	3.2	6.2 (51%)	2.8 (23%)	3.2 (26%)
20 years	2.9	4.1	6.2 (39%)	5.5 (34%)	4.3 (27%)

Notes: Relative shares in parentheses. Results based on estimates including IQ-control in the IQ-sample reported in table 3.2.

APPENDIX

Figure A.1: Map of municipalities by centrality levels

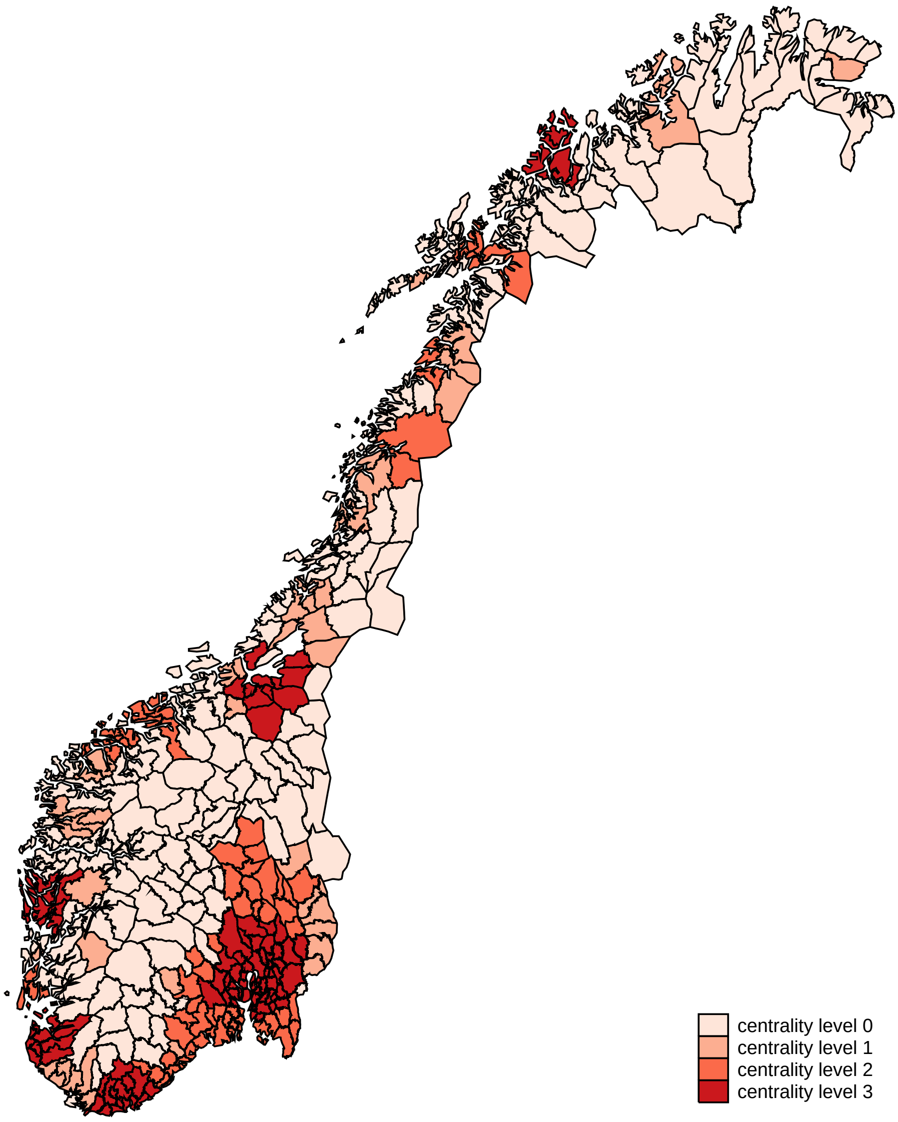


Table A.1: Analysis of movers' wages by years after migration (OLS)

	OLS			
	rural-to-urban	s.e.	urban-to-rural	s.e.
urban stayer		0.129***		
		(0.000359)		
years after the move:				
less than -10	0.102***	(0.00266)	0.151***	(0.00279)
-10	0.0875***	(0.00550)	0.141***	(0.00589)
-9	0.0933***	(0.00508)	0.140***	(0.00543)
-8	0.0825***	(0.00471)	0.135***	(0.00507)
-7	0.0835***	(0.00435)	0.122***	(0.00468)
-6	0.0709***	(0.00400)	0.114***	(0.00433)
-5	0.0672***	(0.00367)	0.110***	(0.00401)
-4	0.0617***	(0.00334)	0.106***	(0.00370)
-3	0.0432***	(0.00301)	0.0850***	(0.00339)
-2	0.00213	(0.00264)	0.0550***	(0.00307)
-1	-0.0698***	(0.00267)	0.000492	(0.00315)
1	0.0870***	(0.00230)	-0.00559*	(0.00276)
2	0.111***	(0.00238)	0.00120	(0.00295)
3	0.126***	(0.00250)	0.00518	(0.00320)
4	0.128***	(0.00265)	0.00360	(0.00345)
5	0.135***	(0.00279)	0.00377	(0.00370)
6	0.139***	(0.00294)	0.00765	(0.00391)
7	0.147***	(0.00307)	0.00774	(0.00414)
8	0.155***	(0.00324)	0.0103*	(0.00439)
9	0.164***	(0.00346)	0.0105*	(0.00467)
10	0.172***	(0.00370)	0.0111*	(0.00498)
more than 10	0.185***	(0.00181)	0.0328***	(0.00229)
education		yes		
age		yes		
year		yes		
continent of origin		yes		
marital status		yes		
N		12,422,118		
R-sq		0.305		

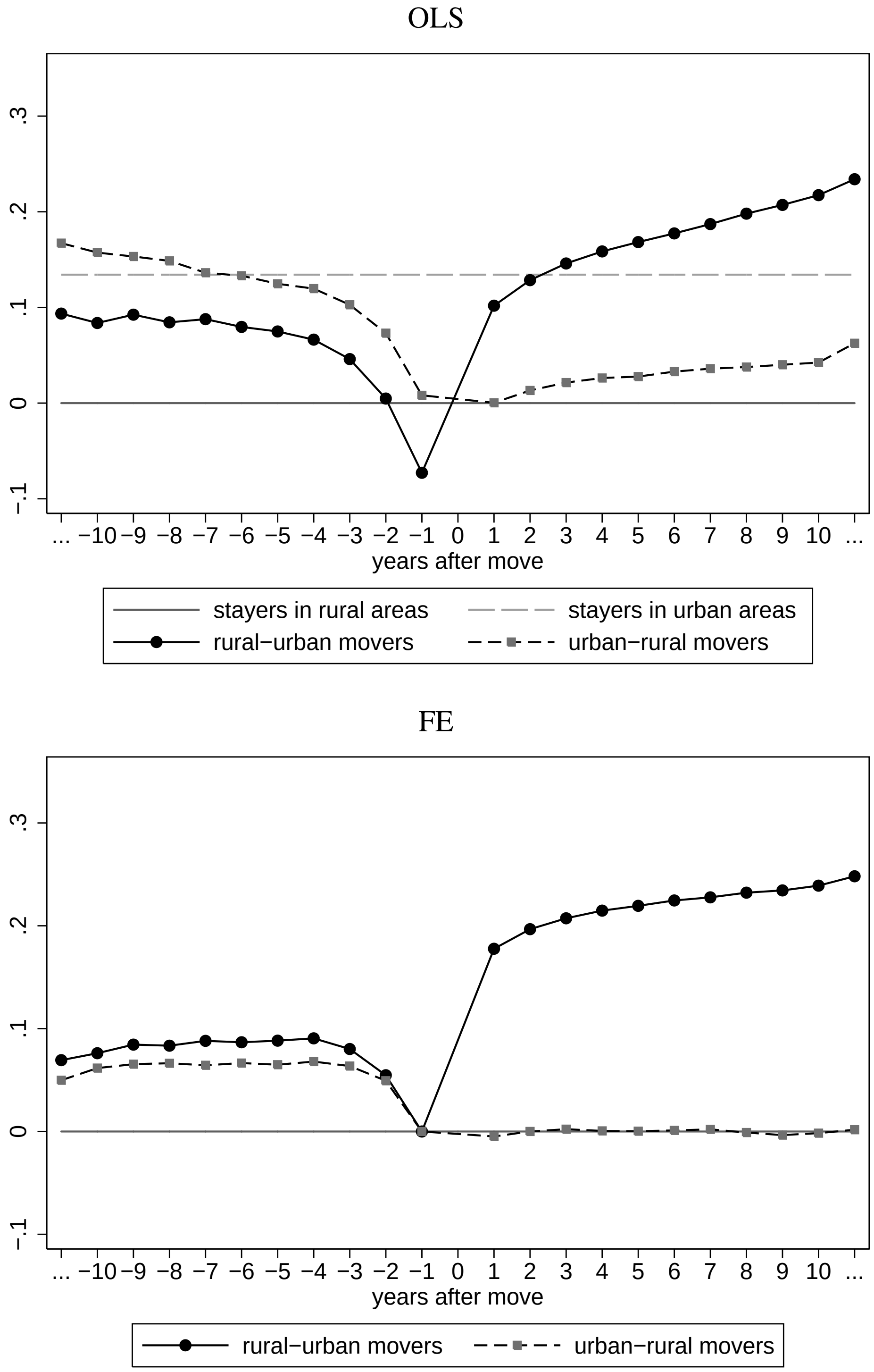
Notes: The table corresponds to the upper panel in figure 3.1. The reference group is stayers in rural areas. The regressions include full set of dummies for the control variables listed in the table. Standard errors in parentheses. *p<0.05, **p<0.01, ***p<0.001.

Table A.2: Analysis of movers' wages by years after migration (FE)

	FE			
	rural-to-urban	s.e.	urban-to-rural	s.e.
years after the move:				
less than -10	0.101***	(0.00486)	0.0662***	(0.00568)
-10	0.103***	(0.00504)	0.0738***	(0.00574)
-9	0.110***	(0.00487)	0.0785***	(0.00533)
-8	0.102***	(0.00485)	0.0786***	(0.00528)
-7	0.105***	(0.00452)	0.0761***	(0.00514)
-6	0.0999***	(0.00439)	0.0720***	(0.00480)
-5	0.100***	(0.00424)	0.0738***	(0.00460)
-4	0.100***	(0.00402)	0.0752***	(0.00438)
-3	0.0878***	(0.00390)	0.0626***	(0.00415)
-2	0.0592***	(0.00355)	0.0459***	(0.00383)
-1	-		-	
1	0.162***	(0.00394)	-0.00856*	(0.00417)
2	0.179***	(0.00398)	-0.00828	(0.00427)
3	0.186***	(0.00405)	-0.00882*	(0.00438)
4	0.187***	(0.00415)	-0.0129**	(0.00457)
5	0.192***	(0.00415)	-0.0146**	(0.00461)
6	0.191***	(0.00429)	-0.0129**	(0.00462)
7	0.192***	(0.00435)	-0.0130**	(0.00480)
8	0.196***	(0.00439)	-0.0168***	(0.00489)
9	0.200***	(0.00445)	-0.0188***	(0.00499)
10	0.203***	(0.00454)	-0.0203***	(0.00510)
more than 10	0.205***	(0.00463)	-0.0165***	(0.00494)
age		yes		
year		yes		
marital status		yes		
N		12,422,118		
R-sq		0.311		

Notes: The table corresponds to the lower panel in figure 3.1. The reference group is earnings one year prior to the move. The regressions include full set of dummies for the control variables listed in the table. Standard errors in parentheses are clustered on individual level. *p<0.05, **p<0.01, ***p<0.001.

Figure A.2: Analysis of movers' wages by years after migration, alternative urban definition



Notes: Municipality of work is defined as urban if it has centrality level 3 (see section 3.3 for details). The sample used excludes the return and multiple movers and the observation years in which workers switched jobs or did not work a full 12 months. OLS estimates in the upper panel are relative to wages of stayers in rural areas. FE estimates in the bottom panel are relative to wages one year before migration. Most estimates, except for "after urban-to-rural migration" dummies, are highly significant (see tables A.3 and A.4 in the Appendix for standard errors, parameter significance and list of controls included).

Table A.3: Analysis of movers' wages by years after migration, alternative urban definition (OLS)

	OLS			
	rural-to-urban	s.e.	urban-to-rural	s.e.
urban stayer	0.134*** (0.000264)			
years after the move:				
less than -10	0.0936***	(0.00190)	0.167***	(0.00218)
-10	0.0838***	(0.00407)	0.157***	(0.00453)
-9	0.0924***	(0.00378)	0.153***	(0.00418)
-8	0.0844***	(0.00353)	0.149***	(0.00386)
-7	0.0878***	(0.00327)	0.136***	(0.00357)
-6	0.0796***	(0.00303)	0.133***	(0.00329)
-5	0.0749***	(0.00280)	0.125***	(0.00304)
-4	0.0663***	(0.00258)	0.120***	(0.00281)
-3	0.0460***	(0.00235)	0.103***	(0.00257)
-2	0.00477*	(0.00209)	0.0731***	(0.00232)
-1	-0.0728***	(0.00212)	0.00813***	(0.00237)
1	0.102***	(0.00183)	0.00035	(0.00206)
2	0.129***	(0.00193)	0.0132***	(0.00219)
3	0.146***	(0.00205)	0.0214***	(0.00234)
4	0.159***	(0.00219)	0.0263***	(0.00249)
5	0.168***	(0.00233)	0.0278***	(0.00265)
6	0.177***	(0.00246)	0.0329***	(0.00280)
7	0.187***	(0.00259)	0.0360***	(0.00295)
8	0.198***	(0.00275)	0.0377***	(0.00310)
9	0.207***	(0.00293)	0.0401***	(0.00330)
10	0.217***	(0.00314)	0.0424***	(0.00352)
more than 10	0.234***	(0.00155)	0.0626***	(0.00161)
education	yes			
age	yes			
year	yes			
continent of origin	yes			
marital status	yes			
N	11,893,206			
R-sq	0.311			

Notes: Municipality of work is defined as urban if it has centrality level 3 (see section 3.3 for details). The table corresponds to the upper panel in figure A.2. The reference group is stayers in rural areas. The regressions include full set of dummies for the control variables listed in the table. Standard errors in parentheses. *p<0.05, **p<0.01, ***p<0.001.

Table A.4: Analysis of movers' wages by years after migration, alternative urban definition (FE)

	FE			
	rural-to-urban	s.e.	urban-to-rural	s.e.
years after the move:				
less than -10	0.0694***	(0.00383)	0.0498***	(0.00441)
-10	0.0761***	(0.00403)	0.0616***	(0.00435)
-9	0.0845***	(0.00392)	0.0655***	(0.00423)
-8	0.0834***	(0.00377)	0.0664***	(0.00408)
-7	0.0881***	(0.00361)	0.0644***	(0.00394)
-6	0.0868***	(0.00353)	0.0665***	(0.00366)
-5	0.0883***	(0.00344)	0.0649***	(0.00356)
-4	0.0906***	(0.00326)	0.0680***	(0.00338)
-3	0.0802***	(0.00316)	0.0636***	(0.00320)
-2	0.0547***	(0.00290)	0.0493***	(0.00291)
-1	-		-	
1	0.178***	(0.00324)	-0.00480	(0.00324)
2	0.197***	(0.00329)	0.000004	(0.00332)
3	0.207***	(0.00338)	0.00225	(0.00340)
4	0.215***	(0.00345)	0.000556	(0.00345)
5	0.219***	(0.00350)	0.000296	(0.00353)
6	0.225***	(0.00360)	0.00105	(0.00358)
7	0.228***	(0.00366)	0.00210	(0.00361)
8	0.232***	(0.00374)	-0.00101	(0.00373)
9	0.234***	(0.00393)	-0.00352	(0.00388)
10	0.239***	(0.00395)	-0.00161	(0.00386)
more than 10	0.248***	(0.00406)	0.00171	(0.00378)
age		yes		
year		yes		
marital status		yes		
N		11,893,206		
R-sq		0.308		

Notes: Municipality of work is defined as urban if it has centrality level 3 (see section 3.3 for details). The table corresponds to the lower panel in figure A.2. The reference group is earnings one year prior to the move. The regressions include full set of dummies for the control variables listed in the table. Standard errors in parentheses are clustered on individual level. *p<0.05, **p<0.01, ***p<0.001.

Table A.5: Estimation results, alternative urban definition

	Full sample		IQ sample	
	(1)	(2)	(3)	(4)
	OLS	FE	OLS	IQ-control
Urban	0.0945***	0.0847***	0.0763***	0.0697***
	(0.000460)	(0.00172)	(0.000579)	(0.000575)
Experience	0.00988***	0.00142***	0.00726***	0.00751***
	(0.0000413)	(0.000215)	(0.0000526)	(0.0000523)
Number of jobs	0.00486***	0.00633***	0.00383***	0.00495***
	(0.000141)	(0.000350)	(0.000161)	(0.000160)
Urban experience	0.00164***	0.00124***	0.00298***	0.00299***
	(0.0000442)	(0.0000751)	(0.0000537)	(0.0000533)
Number of urban jobs	0.00922***	0.00722***	0.0104***	0.00950***
	(0.000178)	(0.000443)	(0.000204)	(0.000203)
IQ score	-	-	no	yes
Education	yes	-	yes	yes
Age	yes	yes	yes	yes
Year	yes	yes	yes	yes
Continent of origin	yes	-	yes	yes
Marital status	yes	yes	yes	yes
N	11,837,131	11,837,131	6,922,995	6,922,995
R-sq	0.319	0.308	0.381	0.389

Notes: Municipality of work is defined as urban if it has centrality level 3 (see section 3.3 for details). The samples exclude 2 years prior to move for the migrants. The IQ-sample in columns 3 and 4 is restricted to individuals with non-missing observations on IQ test scores. The regressions include full set of dummies for the control variables listed in the table. Standard errors in parentheses are clustered on individual level. *p<0.05, **p<0.01, ***p<0.001.

Table A.6: Relative importance of static effects, learning, and matching, alternative urban definition

time period	average # of jobs		% wage premium due to:		
	rural	urban	static	learning	matching
5 years	1.8	1.9	8.5 (81%)	0.6 (6%)	1.4 (13%)
10 years	2.4	2.7	8.5 (72%)	1.2 (10%)	2.1 (18%)
20 years	2.9	3.6	8.5 (61%)	2.4 (17%)	3 (22%)

Notes: Municipality of work is defined as urban if it has centrality level 3 (see section 3.3 for details). Relative shares in parentheses. Results based on FE estimates on the full sample reported in table A.5.

Table A.7: Relative importance of static effects, learning, and matching, IQ sample, alternative urban definition

time period	average # of jobs		% wage premium due to:		
	rural	urban	static	learning	matching
5 years	2	2.1	7 (67%)	1.5 (14%)	2 (19%)
10 years	2.7	3	7 (54%)	3 (23%)	3 (23%)
20 years	3.1	3.8	7 (41%)	6 (35%)	4 (24%)

Notes: Municipality of work is defined as urban if it has centrality level 3 (see section 3.3 for details). Relative shares in parentheses. Results based on estimates including IQ-control in the IQ-sample reported in table A.5.

REFERENCES

ANDERSSON, F., S. BURGESS, AND J. I. LANE (2007): “Cities, matching and the productivity gains of agglomeration,” *Journal of Urban Economics*, 61(1), 112–128.

ANDERSSON, M., J. KLAESSON, AND J. P. LARSSON (2013): “The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies?,” *Papers in Regional Science*.

ASHENFELTER, O. (1978): “Estimating the effect of training programs on earnings,” *The Review of Economics and Statistics*, 60(1), 47–57.

BARTIK, T. J. (1991): “Who benefits from state and local economic development policies?,” *Books from Upjohn Press*.

BAUM-SNOW, N., AND R. PAVAN (2012): “Understanding the city size wage gap,” *The Review of Economic Studies*, 79(1), 88–127.

BHULLER, M., M. MOGSTAD, AND K. G. SALVANES (2011): “Life-cycle bias and the returns to schooling in current and lifetime earnings,” NHH Dept. of Economics Discussion Paper 2011/4.

BLEAKLEY, H., AND J. LIN (2012): “Thick-market effects and churning in the labor market: Evidence from US cities,” *Journal of Urban Economics*, 72(2), 87–103.

BÜTIKOFER, A., D. POLOVKOVA, AND K. G. SALVANES (2013): “Cities and wages: Measuring the urban wage premium in Norway,” mimeo, Norwegian School of Economics, Bergen.

CARD, D. (1990): “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, 43(2), 245–257.

CARLSEN, F., J. RATTSØ, AND H. E. STOKKE (2013): “Education, experience and dynamic urban wage premium,” mimeo, Norwegian University of Science and Technology, Trondheim.

COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2008): “Spatial wage disparities: Sorting matters!,” *Journal of Urban Economics*, 63(2), 723–742.

——— (2011): "The identification of agglomeration economies," *Journal of Economic Geography*, 11(2), 253–266.

COMBES, P.-P., G. DURANTON, L. GOBILLON, AND S. ROUX (2010): "Estimating agglomeration economies with history, geology, and worker effects," in *Agglomeration Economics*, pp. 15–66. University of Chicago Press.

D'COSTA, S., AND H. G. OVERMAN (2013): "The urban wage growth premium: sorting or learning?," Discussion paper, Spatial Economics Research Centre, LSE.

DE LA ROCA, J., AND D. PUGA (2012): "Learning by working in big cities," CEPR Discussion Paper No. DP9243.

DI ADDARIO, S., AND E. PATACCHINI (2008): "Wages and the city. Evidence from Italy," *Labour Economics*, 15(5), 1040–1061.

DURANTON, G., AND D. PUGA (2004): "Micro-foundations of urban agglomeration economies," *Handbook of regional and urban economics*, 4, 2063–2117.

GLAESER, E. L., AND D. C. MARE (2001): "Cities and skills," *Journal of Labor Economics*, 19(2), 316–42.

GOULD, E. D. (2007): "Cities, workers, and wages: A structural analysis of the urban wage premium," *The Review of Economic Studies*, 74(2), 477–506.

HEUERMANN, D., B. HALFDANARSON, AND J. SUEDEKUM (2010): "Human capital externalities and the urban wage premium: Two literatures and their interrelations," *Urban Studies*, 47(4), 749–767.

HUTTUNEN, K., J. MØEN, AND K. G. SALVANES (2011): "How destructive is creative destruction? Effects of job loss on job mobility, withdrawal and income," *Journal of the European Economic Association*, 9(5), 840–870.

KRUPKA, D. J., AND D. S. NOONAN (2013): "City air and city markets worker productivity gains across city sizes," *International Regional Science Review*, 36(2), 183–206.

LEHMER, F., AND J. MÖLLER (2010): "Interrelations between the urban wage premium and firm-size wage differentials: a microdata cohort analysis for Germany," *The Annals of Regional Science*, 45(1), 31–53.

McKenzie, D., S. Stillman, and J. Gibson (2010): “How important is selection? Experimental vs. non-experimental measures of the income gains from migration,” *Journal of the European Economic Association*, 8(4), 913–945.

Melo, P. C., D. J. Graham, and R. B. Noland (2009): “A meta-analysis of estimates of urban agglomeration economies,” *Regional Science and Urban Economics*, 39(3), 332–342.

Møen, J., K. G. Salvanes, and E. Sørensen (2003): “Documentation of the linked employer-employee data base at the Norwegian school of economics,” manuscript, Norwegian School of Economics, Bergen.

Puga, D. (2010): “The magnitude and causes of agglomeration economies*,” *Journal of Regional Science*, 50(1), 203–219.

Rosenthal, S. S., and W. C. Strange (2004): “Evidence on the nature and sources of agglomeration economies,” *Handbook of regional and urban economics*, 4, 2119–2171.

SSB (1994): “Standard for kommuneklassifisering 1994 (Standard Classification of Municipalities 1994),” *Norges offisielle statistikk (Official Statistics of Norway): Oslo-Kongsvinger*.

Sundet, J., D. Barlaug, and T. Torjussen (2004): “The end of the flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century,” *Intelligence*, XXXII, 349–362.

Thrane, V. (1977): “Evneprøving av utskrivingspliktige i Norge 1950-53,” Discussion Paper nr. 26, INAS.

Wheeler, C. H. (2006): “Cities and the growth of wages among young workers: Evidence from the NLSY,” *Journal of Urban Economics*, 60(2), 162–184.

Yankow, J. J. (2006): “Why do cities pay more? An empirical examination of some competing theories of the urban wage premium,” *Journal of Urban Economics*, 60(2), 139–161.

Zaiceva, A. (2006): “Self-Selection and the returns to geographic mobility: What can be learned from the German reunification ’experiment’,” IZA Discussion paper.