University of Sussex

Causes and Consequences of internal migration: Evidence from Brazil and Ghana

Eva-Maria Egger

Thesis submitted for the degree of Doctor of Philosophy

Department of Economics

University of Sussex

July 2017

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

I hereby also declare that chapter 3 is co-authored with Dr Julie Litchfield. My contribution to the paper comprises liaising with survey team to verify the second wave of the data was correct, data cleaning and preparation, developing the empirical strategy, conducting the data analysis, and a large part of the writing. My contribution can be quantified at 75%.

Signature:

Eva-Maria Egger

UNIVERSITY OF SUSSEX

Eva-Maria Egger
Doctor of Philosophy in Economics

Causes and Consequences of internal migration: Evidence from Brazil and Ghana

Summary

This thesis investigates socio-economic drivers and impacts of internal migration in two countries, Brazil and Ghana.

The first empirical chapter analyses the choice of Brazilian workers to move out of metropolitan cities. This direction of movement is substantial in the Brazilian context and leading against standard models of rural-to-urban migration. I estimate the role of living costs and local amenities in the determination of the destination choice of metropolitan out-migrants. Furthermore, I quantify the returns to migrating out of a metropolis by computing counterfactual wages applying matching techniques. The metropolitan out-migrants prefer to move to smaller towns where their real wage gain is positive. They minimize the physical and social costs of migration by moving to closer towns within their state of birth. Living costs in big cities appear to be a main driver for workers to leave these, especially if they are low-skilled.

In the second empirical chapter, I investigate the effect of internal migration on homicide rates in Brazil in the period from 2005 to 2010. I construct a retrospective panel of migration rates between municipalities and use local labour demand shocks in the manufacturing sector at the origins of migrants as instrument for immigration rates. An increase in immigration rates of 1% translates in an increase of 1.2% in crime rates at the local level. The effect is predominant in municipalities with historically higher homicide rates and there is no effect in locations with a large informal sector. While internal migration puts pressure on destination labour markets, these results suggest that it is the presence of a criminal or lack of a flexible sector that channel this pressure into negative outcomes.

The third empirical chapter explores dynamic patterns of internal migration from rural areas in Ghana. With a new household panel survey collected in 2013 and again in 2015, I document that many households have multiple migrants moving at different points in time and for various reasons. Conditional on having had a migrant in the past, I estimate the effect of having a new migrant on the asset welfare of origin households. The findings suggest that due to prior migration experience and consequently lower migration costs for new migrants, there is no decline in welfare from having a new migrant.

Acknowledgements

This thesis is a milestone on my journey to become an applied economist. I am indebted to my supervisors, Prof Andy McKay and Dr Chari Amalavoyal. Their patience, insights, and encouragement was essential to complete this thesis.

My special thanks go to Dr Julie Litchfield for her constant support. She gave me the opportunity to learn from and with her. I am grateful to have gained a co-author and friend in her.

Faculty in the Department of Economics at the University of Sussex has always been very supportive, especially Prof L. Alan Winters who encouraged my work.

This thesis benefited from the financial support of the Economic and Social Research Council (ESRC) as well as the Migrating out of Poverty (MooP) Research Programme Consortium funded by the Department for International Development (DFID). I thank the team of MooP for giving me the opportunity to learn together from our multi-disciplinary work. My trip to the project partners at the Center for Migration Studies (University of Ghana, Legon) was a truly enriching experience.

This journey was filled with wonderful as well as difficult moments. Neither would I have enjoyed nor gotten through them without my yoga teacher Gratia, my bicycle, and most importantly the beautiful collegiality and friendship I share with my fellow Ph.D. colleagues:

Cecilia Poggi, Elsa Valli, Samantha DeMartino, Antonia Schwarz, Rashaad Chowdhury, Hector Rufrancos, Francisco Cabrera, Ani Silwal, Mattia DiUbaldo, Amrita Saha, Mimi Xiao, Egidio Farina, Jorge Hombrados, Edgar Salgado, Tsegay Tekleselassi, Pedro Orraca, Nihar Shembavnekar, Andreas Eberhard-Ruiz, Rafael Parra-Peña, Nick Jacob, Eugenia Go, Marta Schoch, Sweta Gupta, Wiktoria Tafesse, Maika Schmidt and Daniele Guariso.

All of them shared their knowledge of economics and Stata with me as well as a great curiosity for each other's research - peer-education at its best.

There are many friends whose support and distraction I greatly appreciated during my time in Brighton: Javi, Samantha with my 'PhD-nephews' Shayon and Niraad, Marco, Mar, Jonas, Lorenz, Fatema, Stace, Tomás, Filippo, Sarah, Michela with Attilio, and Henry.

I would also like to thank my friends who were constantly and patiently supporting this process from a distance, especially Judith and Myriam, and Alexandra, my extended 'family' Nina, Brigitte and Matthias as well as Manfred and Brigitte.

Marcos receives a *Muito obrigada!* for inviting me to Rio de Janeiro and easing my access to Brazil, its people, data and language.

There is nothing comparable to the unconditional love and support of my family. I would have never chosen to pursue the Ph.D. if it had not been for my parents' encouragement. They made me believe in my strengths and curiosity. *Danke, Mama und Papa.* My brothers, Robert and Richard, I thank for sibling-silliness to gain perspective. I thank Oma Inge, who is not with us anymore, but always present in my heart.

To Michael goes my love and gratefulness for proof-reading, ensuring that I am always well fed and for his endless patience.

Contents

List of Tables

List of Figures

Introduction

Internal migration is a common phenomenon in most countries. An estimated 740 million people live outside their region of birth (Bell and Muhidin, 2009). Individuals move to locate where they can achieve a higher return to their skills or where they hope to find better opportunities than in their origin (Sjaastad, 1962). Households move in response to unfavourable environments or they might send members to diversify their income sources (Stark and Bloom, 1985). On the aggregate, migration out of the rural into the more productive urban sector eventually transforms a country's economy to a more advanced one (Harris and Todaro, 1970). The contribution to poverty reduction and economic development motivates research to better understand the socio-economic drivers and impacts of internal migration.

To broaden our understanding of internal migration it is important to study people's mobility with the latest data available. Using such data from Brazil and Ghana, this thesis documents recent migration patterns and applies econometric methods to analyse their causes and consequences at the individual, household and local level. Studying internal migration is inherently prone to endogeneity issues. The empirical chapters of this thesis address these with methods from the evaluation and the local labour market literature.

Models of internal migration in developing countries focus on rural to urban movements, because it is in line with the structural two-sector model of development (Harris and Todaro, 1970). However, data suggests that rural to rural migration is more common and even urban to rural migration exists. Recent studies started to move away from the rural-urban dichotomy and included smaller towns in their analysis (Christiaensen and Todo, 2014). In the first empirical chapter, I document that a substantial share of migrants in Brazil move out of the biggest cities to smaller towns. Why do workers move out of metropolitan cities? There are several possible explanations. On the side of pull-factors, smaller towns in Brazil have been catching up economically partially due to targeted public investment in lagging areas. The smaller cities might attract migrants with increasing job opportunities and fast growing wages. Furthermore, they often have lower crime rates than big cities in Brazil so that they offer a safer environment to live in. On the other hand, public service provision in smaller towns is often worse than in big cities in Brazil. It is hence possible that migrants might also be pushed out of big cities and not just pulled into the smaller destinations. Urbanization can be associated with overcrowding and scarcity of affordable housing so that workers might decide to leave the expensive city for a cheaper smaller town.

I use the detailed information of the Brazilian Population Census survey of 2010 to analyse the out-migration of workers from metropolitan cities. First, I estimate the role of wages, living costs and local amenities for the destination choice of metropolitan out-migrants. Secondly, I quantify the wage return to moving out of big cities in nominal and real terms by computing counterfactual wages with matching techniques. The metropolitan out-migrants prefer to move to towns where their real wage gain is positive and they minimize the physical and social costs of migration. Wage differences are not significant for the destination choice, whereas price differences are. The results suggest that high living costs in metropolises drive workers out into smaller towns whose economies are growing fast and thus offer an attractive alternative.

The second empirical chapter investigates the relationship of labour mobility and local crime. Crime rates in Brazil are among the highest in the world which is associated with high economic cost. The economic literature on crime finds that labour market factors, such as low income and unemployment, are important determinants for higher crime (Becker, 1968). One important aspect of labour market dynamics is internal migration because it can affect the employment rate and wages in migrants' destinations (Kleemans and Magruder, 2017). These effects can in consequence lead to higher crime. The existing evidence of the impacts of international immigration on crime in destination countries suggests, however, that there is no effect or, if one is found, it is associated with a specific group of immigrants (e.g. Bell et al. (2013)). Impacts of international immigration on labour markets differ to those found for internal migration. It is therefore not clear whether and how internal migration affects crime. The second empirical chapter provides evidence for a positive effect of internal migration on crime and explores possible channels that could explain the result.

Specifically, I estimate the effect of internal migration on homicide rates in Brazilian municipalities from 2005 to 2010. The Census survey data of 2010 is used to construct a retrospective panel of migration rates between municipalities. Local labour demand shocks in the manufacturing sector at the origins of migrants provide exogenous variation in immigration rates to overcome endogeneity issues. An increase in the immigration rate by 1% is related to an increase of 1.2% in the homicide rate. The effect persists in destinations where the informal sector is small and the criminal sector large. This suggests that labour market structures can channel the impact of internal migration on local labour markets in different ways.

Migration is also common in rural households in Ghana. Often more than one household member migrates at different points in time. Studies looking at the impact of migration on left-behind households most times do not differentiate whether the migrant was the first member to move or whether she or another household member moved in the past. This could imply different reasons for migration as well as different impacts on the household. For example, some migrants might be sent to support the household by earning an income and to diversify the income sources. Another migrant leaves to pursue higher education or to find better opportunities than those in his origin community without any obligations to send money back home. The third empirical chapter documents such repeated patterns of migration with a new household panel survey from five regions in Ghana.

This survey provides detailed information on the demographics of migrants and their households, reasons for migration and its support, the financing and costs of movement, as well as contacts, jobs and occupations at destination. This information is used in the third empirical chapter to document the nature of repeated migration within households. The econometric analysis then further estimates how having a new migrant is related to the asset welfare of households left behind. There is no impact of new migrants leaving a household on the asset index. This chapter focuses on households with prior engagement in migration. The data documents that such migration experience is related to lower moving costs for new migrants and a different relationship of the new migrants to the origin household compared to that of previous migrants. These observations help to explain the findings of no effect on the asset index.

In summary, this thesis offers new empirical evidence concerning drivers of internal migration and its consequences for receiving and sending communities. It is structured as follows: Chapter 1 describes the out-migration from metropolitan cities in Brazil and analyses the destination choice and wage return of out-migrants. Chapter 2 investigates the impact of internal migration on crime in Brazilian municipalities. In chapter 3, dynamic migration patterns of Ghanaian households are documented and their effect on household welfare is estimated. The conclusion then summarises the findings and discusses the limitations of this thesis providing an outlook for future research.

Chapter 1

Out-Migration from Metropolitan Cities in Brazil

1.1 Introduction

Urban areas attract workers with job opportunities, high wages and better services. Yet, with the urbanisation waves in developing countries, large cities face many problems associated with over-crowding, such as informal housing, congested infrastructure and unemployment. Furthermore, many metropolitan cities in these countries have been growing in population but not economically (Fay and Opal, 2000). City growth increases demand for housing and amenities, whose supply has a relatively inelastic. Higher living costs put pressure on workers whose budget share for these goods is relatively high (Giannetti, 2003).

With rising congestion externalities, one might expect significant out-migration from big cities as was observed in the 70s and 80s in the US and Europe as the so-called 'population turnaround' (Cochrane and Vining, 1988). This phenomenon was accompanied by the fact that previously lagging regions regained attractiveness as they were catching up economically. In the second half of the 2000's, the Brazilian government heavily invested in the development of lagging areas. Income has been converging between poorer and richer Brazilian cities (Mata et al., 2005). Previously small and remote towns can now offer an alternative to the congested metropolitan cities (Christiaensen et al., 2017; Ministério do Planejamento, 2010; Lall et al., 2009; de Oliveira and de Oliveira, 2011).

Retrospective migration data from Brazil shows that around 20 percent of internal migrants[1] moved out of metropolitan cities between 2009 and 2010, which equals the share of migrants moving into the metropolises in the same period. The majority of out-migrants (around 78 percent) move to live and work in medium-sized destinations,[2] not small and rural locations. It appears that both high and low educated out-migrants are equally likely to move which gives rise to the question what drives these workers out of the cities.

There is a vast literature on the spatial sorting of migrants by skills with a focus on agglomeration effects (Henderson, 1986; Mion and Naticchioni, 2009; Matano and Naticchioni, 2012; Cheng et al., 2012; Fu and Gabriel, 2012; Eeckhout et al., 2014). These studies document that high-skilled workers benefit from human capital concentration in big cities. For low-skilled workers, skill-complementarity determines whether they benefit from agglomeration. Other papers investigate the sorting decision from an individual choice perspective. These studies found that workers sort themselves to destinations by balancing the highest return to their skills and the chance to find employment against local living costs and the presence of local amenities according to their individual preferences (Borjas, 1987; Borjas et al., 1992; Dahl, 2002; Aroca Gonzalez and Maloney, 2005; Lokshin et al., 2007; Moretti, 2011; Aguayo-Téllez et al., 2010; Ham et al., 2011; Grogger and Hanson, 2011; Fafchamps and Shilpi, 2013).

Despite that metropolitan out-migration appears to be substantial in the Brazilian case, there is little known about the reasons motivating individuals to this move in the context of developing countries.[3] This chapter hence investigates which local

[1]These migrants leave their place of birth and are not return migrants.

[2]The median population size of the administrative unit, a *microregião*, is 173,453 inhabitants. I classify a medium-sized *microregião* as one that has between 170,000 and 1 million inhabitants and above 1 million as metropolitan city.

[3]One exception is McCormick and Wahba (2005) who analyse migration in and out of big cities in Egypt. However, their sample of migrants moving out of the big cities is only 82 observations and

characteristics are preferred by the metropolitan out-migrants and whether their destination choice is associated with an actual gain in nominal as well as real wages.

The analysis of the migrants' destination choice is motivated by the fact that not every destination yields the same returns for migrants, because economic development varies across a country. I estimate how various factors at the local level affect the individual destination choice conditional on migration with a conditional logit model. The focus lies on the established determinants of migration: wages, costs, and local amenities. I assume a cost-benefit model of migration in which benefits are proxied with the expected wage in the destination as a function of skills (Sjaastad, 1962; Dahl, 2002). Costs are modelled with the distance between origin and destination and the difference in living costs (Giannetti, 2003; Moretti, 2011; Kennan and Walker, 2011). Coarsened exact matching (CEM) is applied to control for selection bias in the prediction of expected wages. Following this analysis, I further investigate how the actual wages instead of expected wages reflect a gain or loss resulting from moving out of metropolitan cities. I use counterfactual wages of migrants to estimate the return to out-migration in nominal and real wages. A few studies analyse the counterfactual situation of households had their member not migrated, but not of the migrants themselves (Barham and Boucher, 1998; Rodriguez, 1998; Tunalı, 2000; Adams, 2006; Lokshin et al., 2007; Adams et al., 2008; Brown and Jimenez, 2008; Adams and Cuecuecha, 2013).

The results in this chapter demonstrate that the metropolitan out-migrants prefer smaller cities where cost of living are lower. The counterfactual analysis confirms that the return to metropolitan out-migration is positive in real wages. The difference in living costs between metropolitan origins and non-metropolitan destinations appears to exceed migration costs. This result is strongest for low-skilled workers, who would normally experience a decline in nominal wages from leaving metropolitan cities. Furthermore, the out-migrants prefer towns closer to their origin and within their own state of birth which reduces the economic and

their hypothesis focuses on the movement into compared to out of large cities and the concentration of specific skill and age groups in large cities.

social costs of moving. They seem to be willing to accept lower quality in health service provision, but prefer destinations where education quality is relatively better. These results suggest that high prices are pushing workers out of metropolitan cities.

With this chapter I contribute to the related literature in several ways: To the best of my knowledge, it is the first work to empirically document the economic determinants of out-migration from metropolitan cities at the individual level in a developing country. I exploit the detailed information on migration from a unique census survey. I account for migration between local labour markets which allows me to capture a large share of labour mobility within states that accounts for more than 50 percent of migration in Brazil. Even though migration is high within Brazilian states due to large inequality in economic performance (De Vreyer and Spielvogel, 2009), most studies of migration in Brazil look only at movements at a more aggregated level such as state or region (Yap, 1976; Santos and Ferreira, 2007; dos Santos Júnior et al., 2005; Lall et al., 2009; Aguayo-Téllez et al., 2010; Fally et al., 2010). The variation of wages, prices, service provision and other amenities across a country plays an important role in the location choice of workers (Moretti, 2011) and regional planning can influence these factors. Policy makers intending to relieve cities from congestion should thus understand patterns of migration into and out of metropolitan cities.

The paper is structured as follows. The data used for the empirical analysis is described in section 1.2. Descriptive maps, graphs and tables explore the nature of migration from metropolitan to non-metropolitan cities in detail in section 1.3. Thereafter, the conceptual framework of the destination choice model is discussed in section 1.4 and results are presented in the same section. The results of the counterfactual estimation are presented in section 1.5. Section 1.6 then concludes.

1.2 Data

1.2.1 Data source

Every ten years the Brazilian National Institute for Geography and Statistics (IBGE) conducts a 10 percent nationally representative household survey, the Census survey (*Censo Demogrfico* 2010, Instituto Brasileiro de Geografia e Estatística (IBGE) (2012)). The survey of 2010 comprises around 20 million individual observations in all municipalities of Brazil. It contains information on household composition, living conditions, labour market, education, geographic location, and on migration.

1.2.2 Definition of migration

The Census survey from 2010 allows to identify migrants in the sample using the questions "Were you born in this municipality?" to know whether people are living in their birthplace, "When did you move to this municipality?" provides the year of migration, and the question "In which municipality (in which state) did you live before you moved to this municipality that you are currently living in?" provides the exact origin of migrants. It further asks for the municipality of the current job as well as of the previous job. Migrants are individuals who used to live and work in a different location than the on they are living in at the time of the survey.

1.2.3 Sample

The sample comprises working-age migrants and non-migrant residents. The legal working age in Brazil is 16 years, and the retirement age for men 65 years. The age group for the sample has been restricted from 25 up to 65 years. This way it can be assumed that students are excluded. All individuals in the sample are currently not in school and are participating in the labour market which means that they are either employed or unemployed but looking for work. I restrict the sample of migrants to those who moved within the past year, between 2009 and 2010, in order to minimize recall bias.

1.2.4 Definition of origins and destinations

Migration is measured as change in living and working location at the level of a *microregião*. *Microregiões* are geographic and administrative agglomerations of municipalities sharing a labour market and economic activities, a bit larger than counties in the US. I define 22 of these *microregiões* as metropolitan based on their population size of 1 million and above.[4] There are 551 *microregiões*, 22 metropolitan and 529 non-metropolitan *microregiões*. Information on the local characteristics is aggregated to the *microregião* level using individual level data from the Census survey. I use survey weights to obtain local estimates of wages and housing prices measured with the amount of rent per room.

1.2.5 Other variables

Other information on local characteristics is obtained from *Ipeadata*. This is an online data pool provided by Instituto de Pesquisa Econômica Aplicada (Ipea), a Brazilian public research institute that collects data from several ministries and other public sources. It contains information on GDP, quality of education and health provision, and homicide rates as a measure of crime at the *microregião* level.

Quality of education and health are measured using an index constructed and annually updated by the Industrial Federation of the federal state of Rio de Janeiro (FIRJAN). The index for education provision combines information about subscription rate of pre-school children, dropout rate, rate of teachers with higher education, average daily teaching hours, as well as the results of a national education development score. The health provision quality index comprises the number of pre-natal consultations, deaths due to badly defined causes, and child-deaths due to evitable causes.

I also use data on formal sector wages from the national industrial census (RAIS) provided at aggregate level by the online portal Dataviva (DataViva, 2016). All

[4]This definition follows that of the United Nations' World Urbanization Prospects (UNWUP) (Christiaensen et al., 2013).

variables used and their source are specified in the appendix in table A.5 (page 183).

1.3 Descriptive Statistics

1.3.1 Patterns of internal migration in Brazil

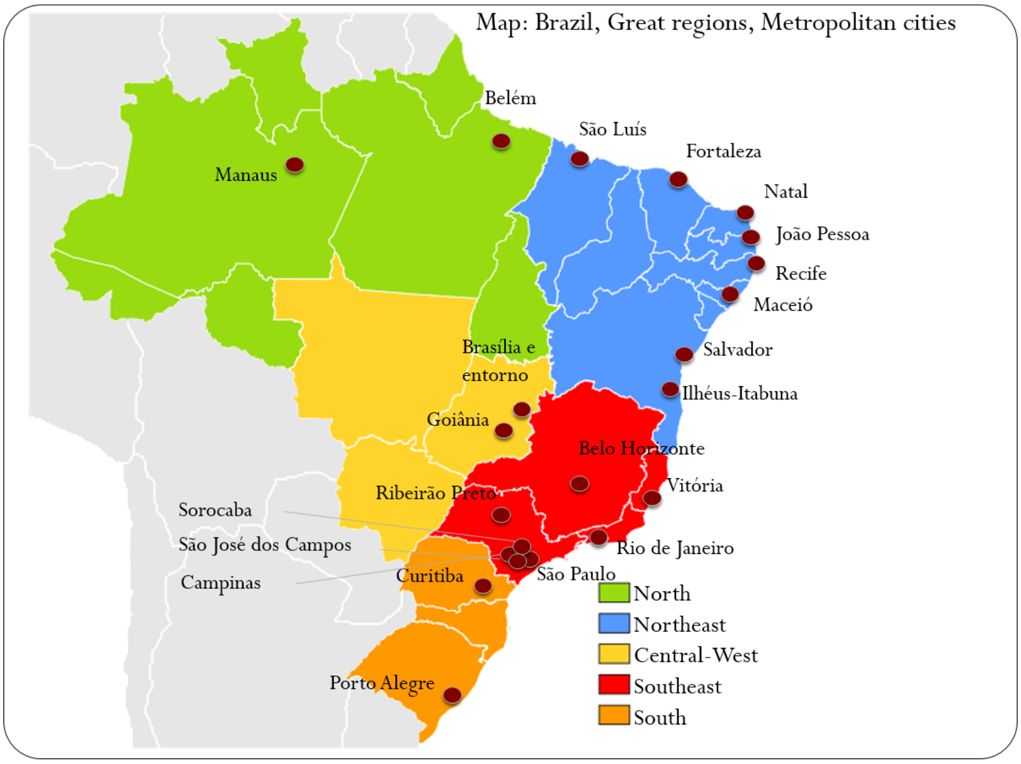

Figure 1.1: Map of greater regions and metropolitan cities of Brazil.

Figure 1.1 shows a map of Brazil, its five greater regions and the 22 metropolitan cities that are the focus of this analysis. The metropolitan cities are located mainly along the coast with the exception of the state capitals in the South-eastern region, Goiânia in the Central-West, Manaus in the Amazon as well as the national capital Brasília.

Labour migration within Brazil is historically very common and mainly attributed to socio-economic differences between regions and between the underdeveloped rural areas and several large urban centres (Yap, 1976). In recent years, migration patterns in Brazil have been changing. Of all Brazilian internal migrants in the year before the Census of 2010, 47 percent moved between non-metropolitan areas (Table 1.1). The second largest movement is into and out of metropolitan cities from and to non-metropolitan *microregiões* (Table 1.1) comprising around 20 percent each of all recent migrants, a substantial share of migration in the country. The remaining 12 percent of migrants move between the metropolises.

Table 1.1: Migrants between metropolitan and non-metropolitan *microregiões* between 2009 and 2010.

	Destination			
	Non-metropolitan		Metropolitan	
Origin	N	%	N	%
Non-metropolitan	380,627	46.9	167,781	20.7
Metropolitan	162,647	20.1	99,143	12.2

Total N=810,196, using survey weights

The graph in figure 1.2, plots the out-migration rate from cities with over one million inhabitants in Brazil from 2004 to 2009. The rate at which people leave big cities has been increasing.

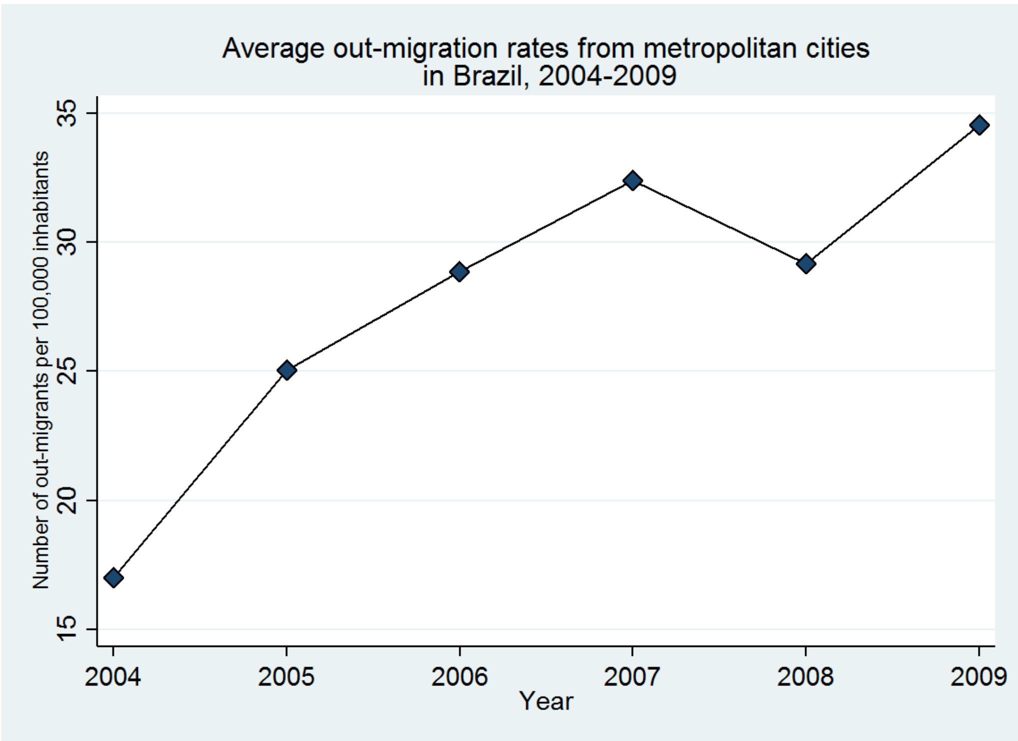

Figure 1.2: Out-migration rate from metropolitan cities in Brazil from 2004 to 2009.

Table 1.2 shows migration flows by region. The majority of migrants who leave metropolitan cities and move to non-metropolitan areas move within their own region. This is illustrated by the large diagonal elements in table 1.2. Only workers from the Central-West region cross the regional borders relatively more often.

Table 1.2: Cross-tabulation of origin and destination regions of migrants moving out of metropolitan *microregiões* between 2009 and 2010.

	Destination					
	North	Northeast	Southeast	South	Central-West	Total
Origin						
North	75 %	8 %	7 %	6 %	5 %	100 %
Northeast	5 %	75 %	13 %	4 %	4 %	100 %
Southeast	3 %	13 %	71 %	9 %	4 %	100 %
South	2 %	2 %	6 %	86 %	3 %	100 %
Central-West	18 %	15 %	16 %	6 %	44 %	100 %

Proportions by origin computed using survey weights.

Not presented in the table, at the level of the federal state, 43 percent of the metropolitan out-migrants leave their state of birth, the other 57 percent stay within the same state when they move. On average metropolitan out-migrants move 930km. For example, the distance from São Paulo to Rio de Janeiro is 440km whereas the distance from São Paulo to the capital Brasília is around 1,000 km. These observations confirm that the level of analysis at *microregião* level captures also intra-regional population dynamics, the largest movements in the country.

The map in figure 1.3 shows the destinations of the most recent metropolitan out-migrants. The majority of them moves to the Southeast and Central-West, but also to more remote *microregiões* in other parts of the country. Some of the non-metropolitan destinations are those neighbouring the metropolitan cities and thus reflect the agglomeration at work around large cities, others are located in areas far away from a metropolitan city.

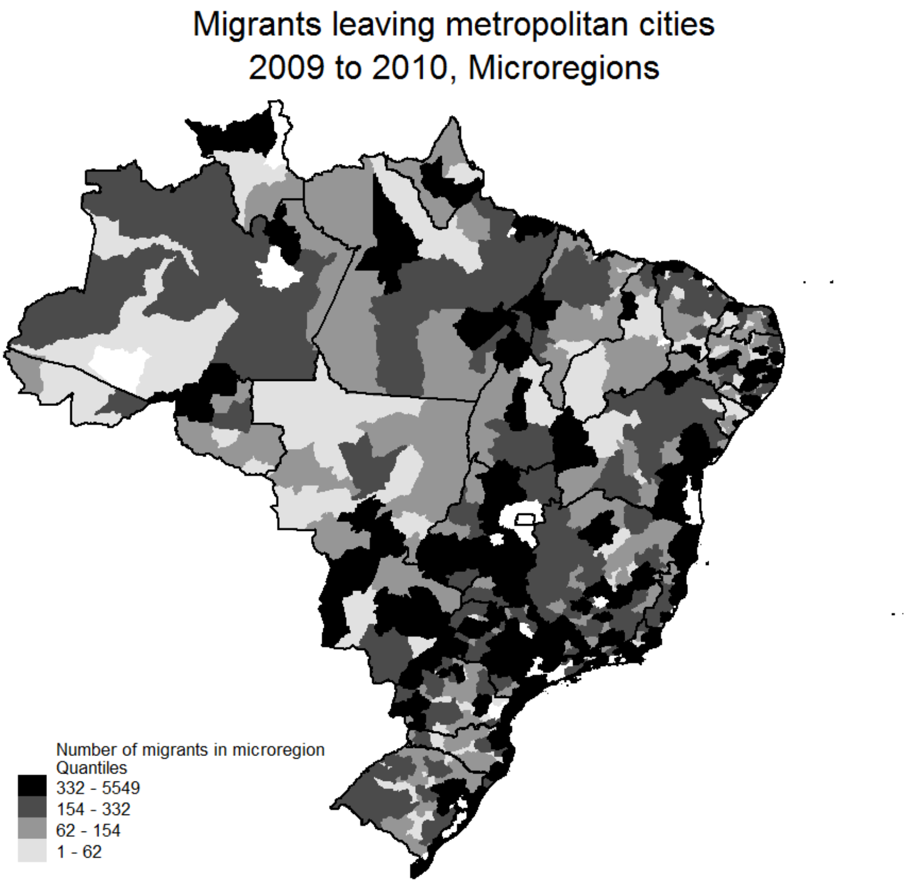

Figure 1.3: Map of destination *microregiões* of migrants leaving metropolitan cities in 2009.

1.3.2 Comparing origins and destinations

Table 1.3 compares metropolitan and non-metropolitan *microregiões* in terms of socio-economic characteristics. In the second and fourth column, I also include the coefficient of variation for the metropolitan and non-metropolitan *microregiões* to illustrate how diverse especially non-metropolitan areas are.

Table 1.3: Characteristics of metropolitan and non-metropolitan *microregiões* in 2010

	Metropolitan		Non-metropolitan	
	Mean	*Coeff. of Variation*	*Mean*	*Coeff. of Variation*
Population	2,679,687	1.11	213,680	0.79
Room rent (Brazilian Reais (R$), mean)	82.14	0.24	50.63	0.39
Room rent (R$, median)	72.47	0.23	45.22	0.42
Hourly wage (R$)	12.11	0.22	7.23	0.29
Share of				
Unskilled workers	0.37	0.09	0.37	0.14
Skilled workers	0.31	0.11	0.40	0.14
High skilled workers	0.24	0.17	0.16	0.23
Formally employed	0.58	0.11	0.40	0.36
Unemployed	0.06	0.29	0.05	0.41
Share of workers in				
Agriculture	0.09	0.36	0.30	0.39
Industry	0.21	0.23	0.18	0.37
Services	0.53	0.08	0.35	0.23
Public services	0.11	0.25	0.12	0.24
People living in				
Urban residence	0.97	0.04	0.73	0.20
Adequate living conditions	0.57	0.28	0.36	0.67
Other measures				
Log(GDP)	16.73	0.06	13.57	0.08
GDP growth 2005-2010	0.16	0.31	0.18	0.79
Health quality index (0 to 1)	0.82	0.09	0.79	0.11
Education quality index (0 to 1)	0.77	0.14	0.73	0.14
Homicide rate (per 100,000)	38.00	0.54	18.58	0.77

Skill level of workers is based on the occupation classification by the International Labor Organization (ILO).

Industries include extractive industry, processing industry, electricity/gas, sanitation/sewage, construction. Services include commerce, transport, housing/food, information/communication, financial services, real estate, professional consulting, science and technology, administrative services, arts/culture/sports, domestic services, and other services. Public services include public administration, security, education, health and social services, international organizations/foreign institutions. Six of the *microregiões* had missing values for homicide rates. I replaced them with 0 due to the way homicides are reported. For a more detailed discussion, see appendix section B.3 on page 188.

Metropolitan cities have on average around ten times more inhabitants than non-metropolitan *microregiões*. In terms of prices, metropolitan residents face room rents that are more than 50 percent higher than in non-metropolitan areas. At the same time they earn similarly higher wages. As expected, high skilled occupations are concentrated in the metropolitan areas and labour markets are much more formalized in these big cities. The employment share of various sectors is highest for services with 53 percent in the metropolitan areas and 35 percent in non-metropolitan *microregiões*. Yet, agriculture in the non-metropolitan areas employs around 30 percent compared to only 9 percent in metropolitan cities. While GDP is higher in metropolitan regions, it is growing faster in the non-metropolitan regions. In terms of living standards, almost 60 percent of metropolitan residents live with adequate sewage, water and electricity provision compared to only 36 percent outside of these cities.[5] This illustrates the stark spatial inequality not only in economic but also in social aspects. Similarly lower are the indices for the quality of health and education provision in non-metropolitan areas in contrast to higher standards in the big cities. In contrast, crime is concentrated in big cities with a homicide rate of 38 homicides per 100,000 inhabitants compared to around 19 in non-metropolitan areas.

The variation in these characteristics among non-metropolitan *microregiões* is large. The second and fourth columns show the coefficients of variation for metropolitan and non-metropolitan *microregiões* respectively. It is relatively larger for non-metropolitan areas in almost all categories but population, public service worker share and education quality. This motivates the analysis of the metropolitan outmigrants' destination choice. Labour mobility is expected to respond to this spatial variation of real wages and other socio-economic characteristics.

Furthermore, smaller *microregiões* have been catching up economically. GDP and wages increase more in these locations than in metropolitan cities over the 2000's

[5]The definition which type of sewage, water and electricity provision is adequate comes from the report on subnormal agglomerations in Brazil (Instituto Brasileiro de Geografia e Estatística (IBGE), 2010)

as shown in figures 1.4 and 1.5. The non-metropolitan areas can offer opportunities as alternative to expensive and congested metropolitan cities.

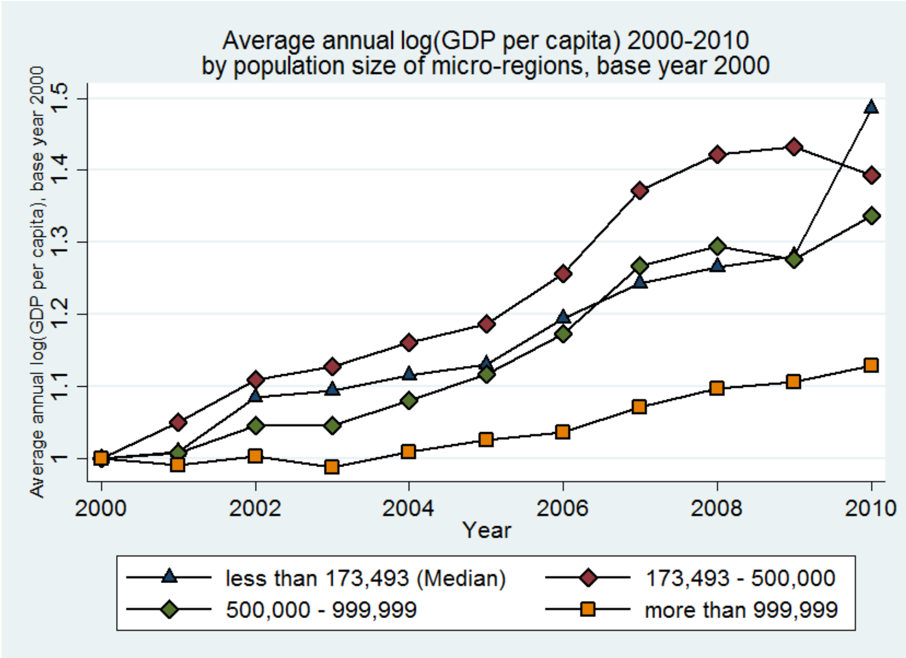

Figure 1.4: Average annual GDP per capita in *microregiões* 2000-2010 by population size of *microregião*.

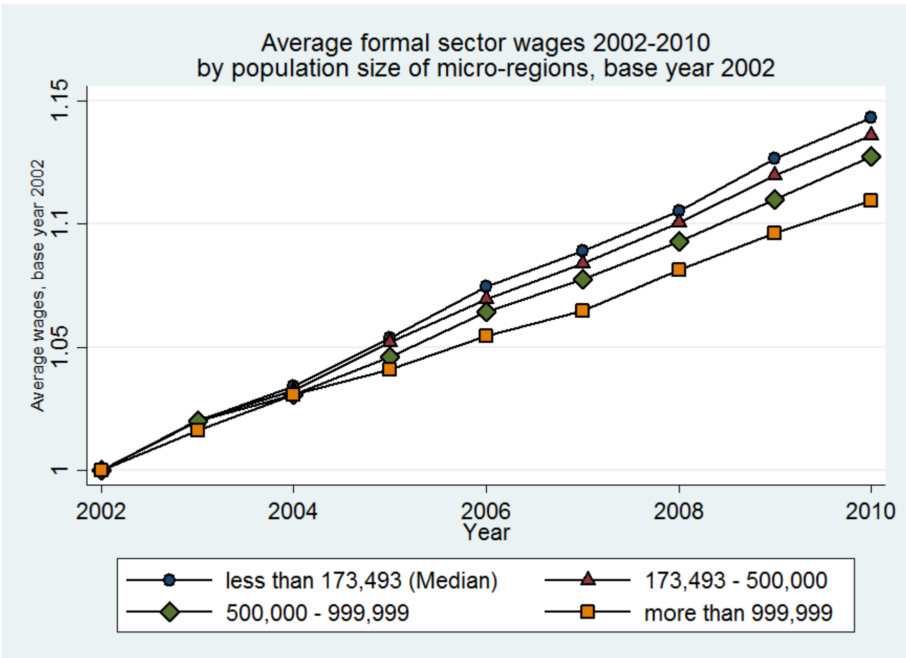

Figure 1.5: Average annual formal sector wages in *microregiões* 2002-2010 by population size of *microregião*.

1.3.3 Comparing migrants and residents

Metropolitan out-migrants are unlikely to be representative of the population of metropolitan cities. In tables 1.4 and 1.5, I compare the characteristics of non-metropolitan residents, of metropolitan out-migrants and of metropolitan residents. Non-metropolitan residents are typically workers who live in non-metropolitan *microregiões* and who have never migrated. Metropolitan residents are also defined as non-migrants, but they live in metropolitan *microregiões*. The comparison of these residents with metropolitan out-migrants allows to shed light on the differences between migrants and residents at the origin and destination. For simplification, in the descriptive analysis, this comparison does not account for the fact that migrants could all be concentrated in a specific sub-set of non-metropolitan *microregiões* so that residents in all non-metropolitan *microregiões* might not represent the exact comparison group of that specific subset.

Migrants are on average slightly younger than residents and relatively more of them are male. Overall, they are much better educated than the average resident at the non-metropolitan destination and their education is very similar to that of metropolitan residents. A slightly larger share of migrants has a tertiary education compared to metropolitan residents. From this comparison, it does not seem that low or high educated workers are more likely to leave metropolitan cities than the respective other group.

Compared to non-metropolitan residents, out-migrants live in more expensive housing, but they also are on average richer with a per capita household income of R$1,350 in contrast to only R$770 of non-metropolitan resident workers. Yet, around 21 percent of these out-migrants live below the national poverty line compared to only 16 percent of metropolitan residents.

Table 1.4: Characteristics of migrants and non-migrants 2010

	Non-metropolitan residents	Metropolitan out-migrants	Metropolitan residents
Number of observations	4,184,904	19,318	1,598,869
Population size	31,054,034	155,288	24,508,716
Age	40.25	36.85	40.22
Female	0.41	0.37	0.45
White	0.51	0.51	0.51
Education level			
None, primary incomplete	0.47	0.29	0.29
Primary, secondary incomplete	0.16	0.16	0.17
Secondary, higher incomplete	0.26	0.33	0.36
Higher complete	0.11	0.21	0.19
Marital status			
Separated, divorced, widowed	0.09	0.11	0.10
Married	0.52	0.43	0.47
Single	0.39	0.46	0.42
Household characteristics			
Partner works	0.57	0.45	0.53
Share of children in household	0.19	0.2	0.17
Urban area	0.79	0.88	0.98
Rent per room	57.12	80.39	91.66
Household income p.c. (R$)	767.4	1,348.3	1,316.1
Poor (National poverty line)	0.28	0.21	0.16
Adequacy of living situation			
Adequate	0.46	0.51	0.63
Semi-adequate	0.51	0.48	0.37
Inadequate	0.03	0.01	0

Proportions and means computed using survey weights.

Table 1.5 documents the labour market characteristics of out-migrants and residents. Around 12 percent of workers who left the metropolitan cities for non-metropolitan destinations are unemployed in contrast to an unemployment rate of only 5 percent among non-metropolitan residents. This indicates, that while metropolitan out-migrants on average have a higher income than residents at their destinations, they are a heterogeneous group and some clearly lose out at their new destination. However, the high unemployment rate might just capture a period of adjustment for very recent migrants who have not found a job yet at their new destination.

In terms of wages, migrants earn on average more than their non-migrant counterparts at non-metropolitan destinations and they earn almost as much as residents in metropolitan areas. This might only reflect differences in the productivity of locations where migrants live as well as different observable and unobservable characteristics of migrants in contrast to residents. The regression analysis in this chapter aims to disentangle these factors.

More than 60 percent of non-migrants in the metropolitan cities are employed either in the public formal or private formal sector, whereas only around 46 percent of non-migrants in non-metropolitan towns work in the formal sector. Migrants appear to find relatively more formal employment at the destinations outside of the big cities compared to the residents there.

Most migrants work in service sectors. Only few work in agriculture at the non-metropolitan destinations even though the main sector of activity there is agriculture. In the metropolitan cities, services are the main sectors of employment. This suggests that most migrants are unlikely to change their sector of activity when they move out of metropolitan areas.

Table 1.5: Labour market characteristics of migrants and non-migrants 2010

	Non-metropolitan residents	Metropolitan out-migrants	Metropolitan residents
Unemployed	0.05	0.12	0.06
Log(monthly wages)	6.59	6.95	6.98
Sector			
Formal private	0.40	0.43	0.56
Formal public	0.06	0.08	0.06
Informal	0.26	0.23	0.21
Self-employed	0.02	0.02	0.01
Small business	0.26	0.24	0.15
Industry, ISIC			
Agriculture	0.26	0.14	0.09
Industry	0.22	0.27	0.21
Services	0.38	0.44	0.54
Public services	0.15	0.17	0.17

Proportions and means computed using survey weights. Industries include extractive industry processing industry, electricity/gas, sanitation/sewage, construction. Services include commerce, transport, housing/food, information/communication, financial services, real estate, professional consulting, science and technology, administrative services, Arts/culture/sports, domestic services, and other services. Public services include public administration, security, education, health and social services, international organizations/foreign institutions.

These observations highlight three findings: First, there is a significant difference in economic and social characteristics between metropolitan and non-metropolitan *microregiões* that are likely to determine migration between these. Prices of housing, non-tradable living costs, are much higher in the metropolitan cities and the non-metropolitan areas are catching up economically. Secondly, there is a large spatial variation in the characteristics of non-metropolitan *microregiões* across the country. Hence, metropolitan out-migrants are unlikely to be indifferent between destinations in their choice where to move. Thirdly, migrants are not a random draw of the population and they are a heterogenous group. It is important to account for underlying selection in the econometric analysis of this chapter.

1.4 Destination choice of migrants

1.4.1 Empirical methodology

The empirical analysis now focuses on the estimation of the effect of various local attributes on the destination choice of migrants. The analysis is based on a multiple choice setting such as presented by McFadden (1974). The empirical application is restricted to those who migrated.[6] As in Fafchamps and Shilpi (2013), I model destination choice conditional on the individual being a migrant.

Migrants are assumed to choose their location in order to maximize their utility. Motivated by a random utility model, a migrant i residing in current location o chooses among all possible destinations J. Let $\mathbf{z}_{ij}$ be a vector of destination attributes that vary across alternatives and can vary by migrant i and let c_j be the cost of moving to destination j from the current location o. Therefore, I define $c_j = 0$ if $j = o$. The utility of moving to destination j is assumed to have the following form:

$$U_{ij} = \boldsymbol{\beta}'\mathbf{z}_{ij} - c_j + \epsilon_{ij} \tag{1.1}$$

The utility of migrant i from moving to destination j depends on the destination attributes, moving costs and an idiosyncratic random component ϵ_{ij}. The observed choice by the migrant is assumed to reflect the maximum utility of all J utilities. The probability that migrant i chooses destination j is therefore

$$\text{Prob}(U_{ij} > U_{ik}) \qquad \text{for all other} \qquad k \neq j \tag{1.2}$$

It is assumed that the error terms are distributed independently and identically with Weibull distribution as in McFadden (1974):

[6] The model allows to include also residents in the analysis and assume that they chose not to move. In the empirical application, this would result in a sample so large that it is not feasible to handle. The decision to migrate itself yields a selection bias distinct from the location choice. Costs of moving are heterogeneous for workers so that some of those who did not move might have done so due to high costs or risk which gives rise to a selection bias in the decision to migrate. By excluding the choice to stay at one's origin, and estimate the destination choice model with migrants only, this specific selection bias does not arise.

$$F(\epsilon_{ij}) = exp(-e^{-\epsilon_{ij}}) \tag{1.3}$$

The probability of moving to destination j is now modelled conditional on migration (i.e. leaving location o). If Y_i represents a random variable indicating the destination choice of migrant i, the probability that this choice is destination j conditional on migration can then be expressed as:

$$\text{Prob}(Y_i = j | Y_i \neq o) = \frac{e^{\beta' z_{ij} - c_j}}{[\ \sum_{j=1}^{J} e^{\beta' z_{ij} - c_j}] - e^{\beta' z_{io} - c_o}} \tag{1.4}$$

This is equal to:

$$\text{Prob}(Y_i = j | Y_i \neq o) = \frac{e^{\beta' z_{ij} - c_j}}{\sum_{j \neq o}^{J} e^{\beta' z_{ij} - c_j}} \tag{1.5}$$

Equation 1.5 represents a conditional logit model. The vector z_{ij} may comprise individual-specific but destination-invariant characteristics w_i and the attributes of each destination x_{ij}, which vary across destinations and can also vary across individuals:

$$z_{ij} = g(w_i, x_{ij}) \tag{1.6}$$

In this analysis, the interest lies on the attributes of destinations and not on migrants' characteristics. Greene (2000) shows how w_i drops out of the probability in equation 1.5 so that this model automatically controls for any individual-specific factors in the destination choice.[7] However, this also implies that I cannot estimate the effect of such factors as age of the migrant etc. Hence, the alternative specific conditional logit model takes the following form:

$$\text{Prob}(Y_i = j | Y_i \neq o) = \frac{e^{\beta' x_{ij} - c_j}}{\sum_{j \neq o}^{J} e^{\beta' x_{ij} - c_j}} \tag{1.7}$$

[7] In some applications, e.g. Fafchamps and Shilpi (2013), this is called individual fixed effect alternative specific conditional logit. It is however not to be confused with the inclusion of fixed effects like in a panel model.

This model can be estimated by the method of maximum likelihood. Let $d_{ij} = 1$ if $Y_i = j$ and 0 otherwise. Then the log-Likelihood function is:

$$log\ L = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log \mathrm{Prob}(Y_i = j | Y_i \neq o) \tag{1.8}$$

For the estimation, there are N observations and regressors for each of the 17,501 metropolitan out-migrants.[8] They choose from $J = 529$ possible non-metropolitan *microregiões* as destinations. Only one of the destinations will have a positive outcome as chosen destination, i.e. the one observed in the data. This results in 9,258,029 individual-destination observations for the multivariate analysis. I cluster standard errors at the metropolitan *microregião* of origin o because it is likely that migrants from the same origin share a pattern in their destination choice.

Based on the human capital migration model (Sjaastad, 1962), the destination attributes of interest in this analysis are wages and prices. One measure for prices would be a Consumer Price Index, which is not available at the level of *microregiões*. Deaton and Dupriez (2011) showed that food trade is strongly integrated across Brazil, so that prices for these consumption goods do not vary much across space. Price differences in non-tradable goods, such as housing, therefore appear more suitable to capture differences in cost of living between metropolitan origins and non-metropolitan destinations. Living costs are proxied with the amount of rent per room reported in the survey. If individuals do not report rent, the average rent in the *microregião* of residency is imputed (see Morten and Oliveira (2016) who use the same approach for Brazil). This is the most disaggregated available measure of prices for non-tradable goods in Brazil. In order to measure the impact of local development, I additionally include population and GDP, as well as homicide rates as destination attributes. I also look at public service provision. This is an index for the quality of education and one of health provision at local level.

Migration costs are measured by the Euclidean distance between origin and des-

[8] The sample of metropolitan out-migrants is slightly reduced as I only include those in the analysis who are matched so that the results are comparable. Those dropped were not matched.

tination in kilometres and additionally by a dummy whether a migrant moved out of her state of birth to a different state. This captures social proximity of a destination to the migrant's origin as in Brazil people have a strong identity with their birth state. Both of these variables imply also the social cost of being farther away from one's family and friends. I further include an interaction term of distance and the 'other-state' dummy in order to disentangle the impact of these two factors.

Average wages in a location need not reflect the wages a migrant can expect to earn. I therefore predict expected wages for migrants based on their characteristics and the coefficients from a wage estimation of residents at each location.

First, I estimate a wage regression separately for all 6.9 million resident observations in each *microregião*. The wage regression takes the following form:

$$W_i^j = \alpha_j(a_i^j - \bar{a}_j) + \beta_j(E_i^j - \bar{E}_j) + \gamma_j(S_i^j - \bar{S}_j) + \chi H_i^j + \delta_j + v_i^j \tag{1.9}$$

Log hourly wages of individual i in location j are determined by the individual characteristics a_i^j, E_i^j, S_i^j, household characteristics H_i^j, a dummy for the *microregião* δ_j and an idiosyncratic error term v_i^j. The variable a_i^j summarises age and age-squared, E_i^j the education level, and S_i^j measures gender and race (white vs. non-white). Each of these variables is demeaned at the level of the *microregião*, so that the coefficients α_j, β_j and γ_j capture the return to these characteristics specific to each location. Additionally, this implies that δ_j measures the unconditional *microregião*-specific average wages. Household characteristics, H_i^j, include the proportion of children and a dummy for whether the partner works, as these might vary by region, e.g. in more rural areas households tend to be larger and female labour force participation lower, so that wages would be overestimated in these areas if this was not controlled for. I use the survey weights in these regressions in order to make the estimates representative of the population.

For each *microregião*, the coefficients $\widehat{\alpha}_j, \widehat{\beta}_j, \widehat{\gamma}_j$ from this regression are then used to predict a measure of expected wages for each migrant. This predicted wage reflects what each migrant can expect to earn in each *microregião* conditional on her characteristics a_i, E_i and S_i, and the unconditional local wage level $\widehat{\delta_j}$:

$$\widehat{E[W_i^j|X_i]} = \widehat{\delta_j} + \widehat{\alpha_j}(a_i - \bar{a}_j) + \widehat{\beta_j}(E_i - \bar{E}_j) + \widehat{\gamma_j}(S_i - \bar{S}_j) \tag{1.10}$$

In appendix table A.1 (page 177), I present the coefficients of the wage predictions corresponding to equation 1.10. The results confirm the relationships documented in the literature: age has a positive, but diminishing effect on expected wages, women earn less, white Brazilians more, and wages increase with the level of education.

This approach assumes that migrants are a random draw from the resident population so that the returns to individual characteristics should be the same for migrants and residents. In the descriptive statistics, I showed that migrants differ from the resident population in a number of observable characteristics. This implies that the expected wage measures used in the analysis so far could be biased by unobservable characteristics. I thus estimate another measure for expected wages that should reduce the selection bias. I predict expected wages in non-metropolitan destinations from a sample of previous migrants from the same origin as the migrant. These migrants have moved more than a year ago to the destinations. They are assumed to be more comparable to migrants than residents in terms of unobservable characteristics specific to migrants, for example risk-taking preferences.

For the metropolitan origins, I predict expected wages based on a matched sample of residents at origin. I apply coarsened exact matching (CEM) to use only those residents that look most similar to the migrants. In the appendix section A.1 (page 172), I discuss the advantages of CEM over Propensity Score Matching in this analytical setting and I explain in detail the matching procedure. In brief, CEM bounds the degree of model dependence in the main analysis and the data is automatically restricted to common support. The large dataset of the Census at hand allows for this exact matching method, without facing the trade-off of conventional matching methods between bias and variance.

1.4.2 Results

In the specific application of this chapter, metropolitan out-migrants choose their destination not only based on destination attributes, but on these attributes relative to the attributes of migrants' metropolitan origins. For each location attribute, I thus compute the difference between the destination and origin, e.g. the cost of living in destination j minus the cost of living in origin o. Table 1.6 gives an overview over the differences between destinations and origins of all variables of interest and how these differences vary between the destinations that migrants chose to those that they did not choose.

Table 1.6: Difference between non-metropolitan destination and metropolitan origin comparing chosen destination to alternative destinations

Difference between destination and origin in:	Chosen destination	Alternative destinations	t-statistic, difference in mean
Expected hourly wages (log)	-0.52	-0.58	-21.16
Rent per room (log)	-0.43	-0.63	-57.24
Other state than origin	0.57	0.92	174.73
Distance to origin (km)	1020.31	1123.94	28.83
Population (log)	-2.32	-2.81	-58.14
GDP (log)	-2.82	-3.59	-68.47
Homicide rate	-9.52	-12.52	-16.77
Education provision quality index (0 to 1)	-0.03	-0.07	-33.04
Health provision quality index (0 to 1)	-0.02	-0.05	-32.03

This table already indicates some patterns of destination choice. Similar to what we observed earlier in the descriptive part, nominal wages are on average always lower in non-metropolitan areas. Migrants tend to choose locations, where this gap is relatively smaller, -0.52 compared to -0.58. Similarly, prices for housing, measured in rent per room, are on average higher in the big cities. Migrants settle in locations where this price gap is not as big as in other possible destinations. This could indicate a trade-off between higher wages and lower prices at destinations. The difference of prices between chosen destination and alternative locations is larger than that of wages. This could indicate that prices matter more than wages for the location decision.

A clearer pattern is revealed with regards to the geographic and social distance of chosen destinations. 57 percent of chosen destinations are in a different state than the origin, that is also the state of birth of the migrants, contrasting 92 percent of the destination alternatives and reflected also in a lower average distance of chosen destinations to the migrants' origin.

The difference in GDP is also statistically significant. Migrants choose locations with a higher level of GDP compared to alternative destinations, but GDP is always lower in non-metropolitan destinations on average than in metropolitan origins. The same is the case for population size. Chosen destinations are on average larger than their alternatives. Based on their average size, they are however not the smallest locations, but still medium-sized *microregiões*. In terms of amenities, chosen destinations have relatively higher levels of homicide rates in contrast to alternative destinations. Education and health service provision is on average higher in the chosen *microregião* than in alternative destinations.

These averages are all statistically different between chosen and alternative destinations. Many of these factors are highly correlated with each other which makes it necessary to apply a multivariate analysis to disentangle their influence on the metropolitan out-migrants' destination choice.

Table 1.7 reports the results of the alternative specific conditional logit model that estimates the probability for destination choice conditional on migration as specified in equation 1.8. The interpretation of coefficients in the alternative specific conditional logit model is not straightforward. It is not possible to compare the coefficient size directly, but only in relative terms which I will do later in section 1.4.4. To interpret the the direction of the effect, I draw on the formula of the elasticity as presented in Greene (2000). Let the probability of choosing destination j be P_j, then the elasticity of P_j with respect to an attribute x_{ij} evaluated at the mean $\bar{x_{ij}}$ can be written as:

$$\frac{\delta\log(P_j)}{\delta\log x_{ij}} = \bar{x_{ij}}(1 - P_j)\beta_x \tag{1.11}$$

where β_x is the coefficient of the destination attribute from the conditional logit estimation. Because the attributes x_{ij} are defined in terms of differences between destination and origin, it is important to keep in mind for each variable whether it was on average higher or lower at the destinations compared to the origins in order to know whether $\bar{x_{ij}}$ is positive or negative.

Table 1.7 reports the results. The specification in column 1 is that of only wage and price differences. The first coefficient in column 1 is that of wage differences. Its value is -0.132. Wage differences are on average negative as documented above in table 1.6 implying that the elasticity itself is positive. An increase in the wage difference in destination j compared to an alternative implies an increase in the likelihood for migrants to choose this destination. In the current setting, an increase in wage differences means that wages get closer to those in the metropolitan origin of migrants. Metropolitan out-migrants hence prefer to move to non-metropolitan *microregiões* that have a relatively smaller wage difference to the origin than alternative destinations.

Table 1.7: Destination choice conditional on migration, alternative specific logit

	Likelihood to select chosen destination					
	(1)	(2)	(3)	(4)	(5)	(6)
Expected wages (log)	-0.132	-0.683***	-0.526***	-0.389*	-0.216	-0.272
	(0.286)	(0.213)	(0.192)	(0.208)	(0.185)	(0.183)
Rent per room (log)	1.323***	1.163***	0.647**	0.962***	0.876***	0.854***
	(0.190)	(0.197)	(0.264)	(0.262)	(0.263)	(0.295)
Population (log)			0.500***	0.676***	0.673***	0.684***
			(0.170)	(0.176)	(0.179)	(0.179)
GDP (log)			0.137	-0.055	-0.032	-0.048
			(0.151)	(0.128)	(0.124)	(0.122)
Homicide rate			0.010**	0.009*	0.008	0.009*
			(0.005)	(0.005)	(0.005)	(0.005)
Education quality index				-3.766**	-4.044**	-4.281***
				(1.714)	(1.681)	(1.637)
Health provision quality index				4.042**	4.054**	4.171**
				(1.782)	(1.836)	(1.828)
Destination specific:						
Distance to origin (log)		1.381*	1.548	1.284	1.299	-0.195
		(0.838)	(1.096)	(0.997)	(1.038)	(0.426)
Other state		9.353	10.175	9.412	9.694	-2.302***
		(6.141)	(7.491)	(7.035)	(7.258)	(0.306)
Distance * Other state		-1.696*	-1.812*	-1.720*	-1.760*	
		(0.883)	(1.081)	(1.021)	(1.054)	
Observations	9,258,029	9,258,029	9,258,029	9,258,029	9,258,029	9,258,029
Chi-squared	56.6	184	1,112	1,326	1,360	1,558
Number of cases	17,501	17,501	17,501	17,501	17,501	17,501
Number of alternatives	529	529	529	529	529	529

Significance levels * 10%, ** 5%, *** 1%. Standard errors are clustered at the metropolitan *microregião* of origin. Estimator is alternative specific conditional logit. In each column, the first set of regressors are the difference between destination and origin for each destination alternative. The second set, indicated as *Destination specific*, are measured at destination relative to origin. Column 5 and 6 use expected wage differences based on past migrants at destination and matched residents at origin as explained in section 1.4.1.

For living costs, it is the opposite is the case. The second coefficient in column 1 is positive. Due to much higher prices in metropolitan cities the difference in the cost of living is on average negative, but migrants prefer to keep this difference as large as possible. Metropolitan out-migrants prefer destinations that are cheaper than their origin and relatively cheaper than alternative destinations. This coefficient is strongly significant. The coefficient for wages is, however, insignificant.

In the second column, I extend the specification by the costs of migration, the distance to move and the social distance. The latter is the dummy whether a destination is in a different state than a migrant's origin. I also interact these two variables. The distance between destination and origin and the interaction term appear weakly significant. There is a negative relationship between the interaction term and the likelihood to choose a specific destination. Migrants are less likely to move to locations that are farther away and at the same time outside of their state of origin. The coefficients of distance to origin and of leaving one's own state are both positive which means that migrants who stay within their state of birth move longer distances.

Wages and prices remain with their expected signs and are now both significant predictors of destination choice once I add other control variables in columns 2, 3 and 4, but their coefficients become a bit smaller.

In column 3, I add population, GDP and homicide rates. By construction metropolitan out-migrants move to locations with fewer inhabitants than their metropolitan origin. They also seem to prefer relatively smaller locations among their destination choices. The level of economic activity measured in the log of GDP does not appear to significantly predict the location choice. Metropolitan out-migrants significantly prefer destinations where the homicide rate is relatively lower than in their metropolitan origins.

Then I add local amenities in column 4. The contribution of wages, prices, population and homicide rates to the destination choice remain significant. Additionally, it appears that migrants show preferences for locations with relatively better education quality, but they accept relatively lower health provision quality.

Lastly, in column 5, I use expected wages that are corrected for selection as explained in the previous section. The selection-corrected wage differences between non-metropolitan destinations and metropolitan origins are thus based on wage predictions using the previous migrants at destinations and matched residents at origins. The coefficient of expected wage differences becomes smaller and insignificant now that I control for selection. It is plausible that wage differences are less stark once observable and some unobserved characteristics are accounted for. This indicates that such characteristics of migrants led to an upward bias in expected wage differences. The coefficients of the other variables remain the same in terms of sign. The coefficient of housing prices becomes a bit smaller, but it remains positive and strongly significant. Homicide rates are no longer significantly correlated with the destination choice. The coefficient of education quality becomes larger. The other results are unchanged.

Due to concerns about collinearity between the variables of distance, moving to another state and their interaction, I also present results of a regression that excludes the interaction term in column 6. This reveals that only the indicator for moving to another state remains significant and its coefficient becomes negative. This result, however, is in line with the initial interpretation that migrants prefer locations that reduce the distance to their origin. The exclusion of the interaction term minimally changes the sizes of the coefficients of other variables, but it does not affect their direction or significance.

In summary, the results of the multivariate analysis confirm that prices matter significantly for the destination choice of metropolitan out-migrants. Wages are weakly significant and become insignificant when I account for potential selection bias. Furthermore, the interaction of physical and social moving costs appear to matter. Lastly, I document that also amenities in terms of public service provision are significantly correlated with the destination choice.

1.4.3 Robustness: Hedonic housing prices

The second main variable of interest in the destination choice analysis are the prices or living costs, proxied so far by the average rent per room in a *microregião*. In the Census data, households are asked to state the monthly rent they pay if they live in a rented apartment or house and the number of rooms of the unit. I used this data to aggregate the average room rent at the *microregião* level. This measure ignores the possibility that the price differences might just reflect differences in housing quality. Similar to Li and Gibson (2014), I construct a hedonic housing index that measures the differences in housing costs based on location-specific amenities rather than housing specific characteristics.

Households are asked to provide information on the quality of walls, floors, and the presence of toilets. Additional questions inform about the quality of sewage, waste water and electricity access. With these variables I can estimate a hedonic housing price for each location. I regress the rent per room on these characteristics weighted by the household survey weight and I include a dummy for each *microregião*. These regression results are presented in the appendix table A.4 (page 179). The coefficients of the *microregião* dummies capture any location specific amenities that contribute to spatial price differences. I extract these estimates to construct a location specific hedonic living cost measure. This variable is independent of differences in housing quality.

Using this measure provides no change in the signs of the estimates of the influence of price differences on destination choice compared to the initial results (see table 1.8). The coefficients of the hedonic price index are positive and strongly significant. They become slightly smaller when I use selection corrected wage differences (column 2 and 3). These results suggests that migrants prefer destinations where living costs are relatively cheaper than at their metropolitan origins. Other coefficients of the analysis do not change significantly.

Table 1.8: Destination choice of metropolitan out-migrants. Changes in wage and price measures; Inclusion of unemployment rate

	Likelihood to select chosen destination		
	(1)	(2)	(3)
Difference in:			
Expected wages (log)	-0.430**	-0.232	-0.241
	(0.185)	(0.175)	(0.159)
Housing prices (log)	0.916***	0.799***	0.761***
	(0.230)	(0.241)	(0.226)
Population (log)	0.692***	0.676***	0.657***
	(0.142)	(0.146)	(0.152)
GDP (log)	-0.026	0.008	0.030
	(0.087)	(0.087)	(0.100)
Homicide rate	0.009*	0.008	0.008*
	(0.005)	(0.005)	(0.005)
Education quality index	-3.529**	-3.835**	-3.894**
	(1.765)	(1.745)	(1.839)
Health provision quality index	4.167**	4.118**	3.925**
	(1.878)	(1.934)	(1.607)
Unemployment rate			-1.870
			(5.087)
Destination specific:			
Distance to origin (log)	1.225	1.248	1.250
	(0.977)	(1.023)	(1.020)
Other state	9.276	9.578	9.610
	(6.941)	(7.209)	(7.194)
Distance * Other state	-1.702*	-1.744*	-1.751*
	(1.007)	(1.046)	(1.045)
Price measure:	Hedonic price index (instead of rent per room)		
Samples for expected wages:	Unmatched residents	Past migrants at destination and matched residents at origin	
Observations	9,258,029	9,258,029	9,258,029
Wald-test	1,189	1,162	1,153
Number of migrants	17,501	17,501	17,501
Number of alternatives	529	529	529

Significance levels * 10%, ** 5%, *** 1%. Standard errors are clustered at the metropolitan *microregião* of origin. Estimator is alternative specific conditional logit. In each column, the first set of regressors are the difference between destination and origin for each destination alternative.
In the first column are results from the estimation using the wage differences based on predictions using the sample of unmatched residents at the origin and price differences using a Hedonic price index. Column 2 and 3 present the same results, but wage differences are now based on predictions using the sample of matched residents at the origin and previous migrants at destination, price differences are computed using the Hedonic price index. In column 3, unemployment is added as regressor.

In column 3 of table 1.8, I add the difference in unemployment rate to the regression. In the migration literature, it has been suggested that not only expected wages, but also the likelihood to earn these matter for migration decisions. In the case of Brazilian workers moving out of metropolitan cities, differences in unemployment rates are not significantly related to their destination choice. This could be due to the fact that the unemployment rates in non-metropolitan *microregiões* are on average only 1 percentage point lower than in metropolises.

1.4.4 Relative effect size

Marginal effects of the alternative-specific logit model can be computed for each possible location choice, but this is computationally burdensome and ineffective in presenting the results. To illustrate and compare the effect size, we can look at one destination alternative, e.g. the one with a price difference very close to the average price difference to metropolitan origins. I take the significant regressors from the full specification with selectivity robust price and wage measures as in column 3 of table 1.8 and compute their elasticities using Greene's formula (Equation 1.11) for this example location. They are presented in table 1.9:

Table 1.9: Elasticities of independent variables for example location

	Mean	Elasticity
Other state	0.923	-2.133
Population (log)	-2.812	-1.924
Distance to origin (log)	6.926	-1.595
Rent per room (log)	-0.492	-0.381
Education quality index	-0.072	0.293
Health provision quality index	-0.048	-0.203
Homicide rate	-12.519	-0.113

The elasticities reveals that by far the largest effect on migrants' destination choice is that of a migrant leaving her or his state of birth, not quite as large but with an elasticity of greater than 1, is also the distance to origin. These variable captures the physical and social costs of moving. This confirms that migration costs in Brazil are still high and a significant factor in labour mobility (Morten and

Oliveira, 2016). The effect of population size is equally large, which reflects the large gap in terms of city sizes between the metropolis and small towns. The elasticity of price differences is much smaller than that of distance, but it is still sizeable: A 1 percent increase in the price difference between the metropolitan origin of a migrant and this specific location make it on average 0.3 percent more likely to be chosen as destination, e.g. if prices rise in metropolitan cities by 1 percent migrants are more likely to choose this location. Local amenities concerning public service provision are significant predictors for the destination choice but their effect size is smaller than that of prices, with education appearing slightly more important than health, followed by the homicide rates.

1.5 Counterfactual earnings of metropolitan out-migrants

The previous results showed that prices play a sizeable and significant role in the destination choice of metropolitan out-migrants whereas expected wages do not appear significant once I control for self-selection of migrants. It is possible that this is due to incorrect expectations of migrants about their earnings. Thus, this section focuses on the actual observed earnings of migrants at their destination in contrast to expected wages. The actual earnings are compared to a prediction of what a migrant would have earned had she not moved out of the metropolitan city, her counterfactual wage. With this comparison, this section aims to see whether and how metropolitan out-migration is associated with a wage loss or gain and what role living costs play in this question.

1.5.1 Empirical methodology

The wage return to migration is defined as the difference between income at destination, y_d, and at the origin, y_o:

$$r = y_d - y_o \tag{1.12}$$

In the empirical application, income is proxied by the log of hourly wages, W.[9] The comparison of migrant wages between origin and destination can be interpreted as an evaluation problem. Let migration be the treatment with $M_i = 1$ if the individual moved, $M_i = 0$ if not. For each individual, two outcomes in terms of wage differences, can be defined as

$$Y_i^0 = log(w_{i,0}) - log(w_{i,0}) \qquad \text{if} \qquad M_i = 0 \tag{1.13}$$

$$Y_i^1 = log(w_{i,1}) - log(w_{i,0}) \qquad \text{if} \qquad M_i = 1 \tag{1.14}$$

Thus, the wage difference due to migration can be identified for migrants as average treatment effect on the treated (ATT):

$$ATT = E(Y^1 - Y^0 \mid M = 1) = E(Y^1 \mid M = 1) - E(Y^0 \mid M = 1) \tag{1.15}$$

The first term on the right hand side is observable in the data at hand, the wages of migrants at their destination. The second term represents the counterfactual outcome, what migrants would have earned had they not migrated which cannot be observed. I only observe wages for migrants at their destination and for non-migrants at origin. If wages were estimated using OLS and then compared, the wage differences would be biased due to selection into migration arising from individual-specific unobservable characteristics.

It is necessary to account for this potential bias in the empirical estimation of migrants' counterfactual wages. This is especially important in a context where spatial wage differences have been found to reflect variation in labour force compos-

[9] I choose to look at hourly wages earned in the main job instead of total income as hourly wages in the main job should mostly reflect the return to individual characteristics based on location whereas total income also depends on household composition and other factors.

ition and industry concentration. In Brazil, the labour force is distributed unequally across space, concentrating better-educated workers in metropolitan areas and economically stronger regions. Thus, the returns to education based on observable characteristics explain around half of the spatial wage differences (Almeida dos Reis and Paes de Barros, 1991; Foguel et al., 2015; Ferreira et al., 2006). Furthermore, Brazilian workers have shown little mobility across industries so that it seems reasonable to focus on the self-selection by location and not by sector (Menezes-Filho and Muendler, 2011; Hering and Paillacar, 2015).

I therefore use the predicted wages from the matched sample of residents in metropolitan origins of migrants as described in section 1.4.1. The difference between the actual observed wages at destination and the predicted counterfactual wages at the origin is the return to migration out of metropolitan cities. Real wages are computed using the local average rent per room as denominator of actual and predicted nominal wages.

1.5.2 Results

This section presents the results of the counterfactual analysis. Table 1.10 presents the average return to migration as the difference between average actual and counterfactual wages for migrants moving out of metropolitan areas. These migrants earn significantly lower wages at their non-metropolitan destinations. Once I account for the local living costs by using real wages, the difference becomes positive. This indicates that metropolitan out-migrants lose in nominal terms, but gain in real wages due to lower living costs in non-metropolitan destinations.

In the appendix table A.2 (page 178), I present the results without applying matching. The nominal wage differences are around 0.1 log points larger than when matching is applied. This indicates an overestimation of wages at origin when not accounting for selection and it suggests that out-migrants are negatively selected from the metropolitan working population.

Table 1.10: Differences in actual and predicted wages for metropolitan out-migrants, after matching

Log (nominal hourly wages)	*N*	Mean
Observed	*15,424*	1.816
Predicted	*15,424*	2.069
Difference		-0.253***
Log (real hourly wages)	*N*	Mean
Observed	*15,424*	-2.237
Predicted	*15,424*	-2.396
Difference		0.159***

Significance levels * 10%, ** 5%, *** 1% for t-test of difference in means between observed and predicted wages. Predicted wages are based on a matched sample of metropolitan residents.

Table 1.11 documents heterogeneity in wage returns along the education level of migrants. I define high educated workers as those who completed high school or any higher level of education. Low educated workers are those who did not complete high school or any lower level of education. Results show that real wages are higher at destination than origin for both groups. For highly educated individuals leaving the big cities the real wage gains are larger, because their loss in nominal wages is relatively small. In contrast, low educated workers see a large loss in nominal terms, and a relatively smaller gain in real wages. For both groups the nominal and real wage differences are statistically significant.

Table 1.11: Differences in actual and predicted wages for metropolitan out-migrants, by education level

	High educated	
Log (nominal hourly wages)	*N*	Mean
Observed	*3,107*	2.846
Predicted	*3,107*	2.930
Difference		-0.084***
Log (real hourly wages)	*N*	Mean
Observed	*3,107*	-1.270
Predicted	*3,107*	-1.544
Difference		0.274***
	Low educated	
Log (nominal hourly wages)	*N*	Mean
Observed	*12,317*	1.556
Predicted	*12,317*	1.851
Difference		-0.295***
Log (real hourly wages)	*N*	Mean
Observed	*12,317*	-2.481
Predicted	*12,317*	-2.611
Difference		0.130***

Significance levels * 10%, ** 5%, *** 1% for t-test of difference in means between observed and predicted wages. Predicted wages are based on a matched sample of metropolitan residents.

So far, the counterfactual wage comparison has focused on the average wage return. However, the distributional graphs of actual and counterfactual wages document the return to metropolitan out-migration along the wage distribution. Figures 1.6 and 1.7 show the wage distributions of workers who have moved out of a metropolitan areas. They compare the observed wages of migrants at their destination and the predicted counterfactual wages at origin. As suggested from the results in Table 1.10, for nominal wages the distribution of observed wages lies left of the predicted earnings in metropolitan origins. Wages are generally higher in origins and out-migration implies a loss in nominal terms.

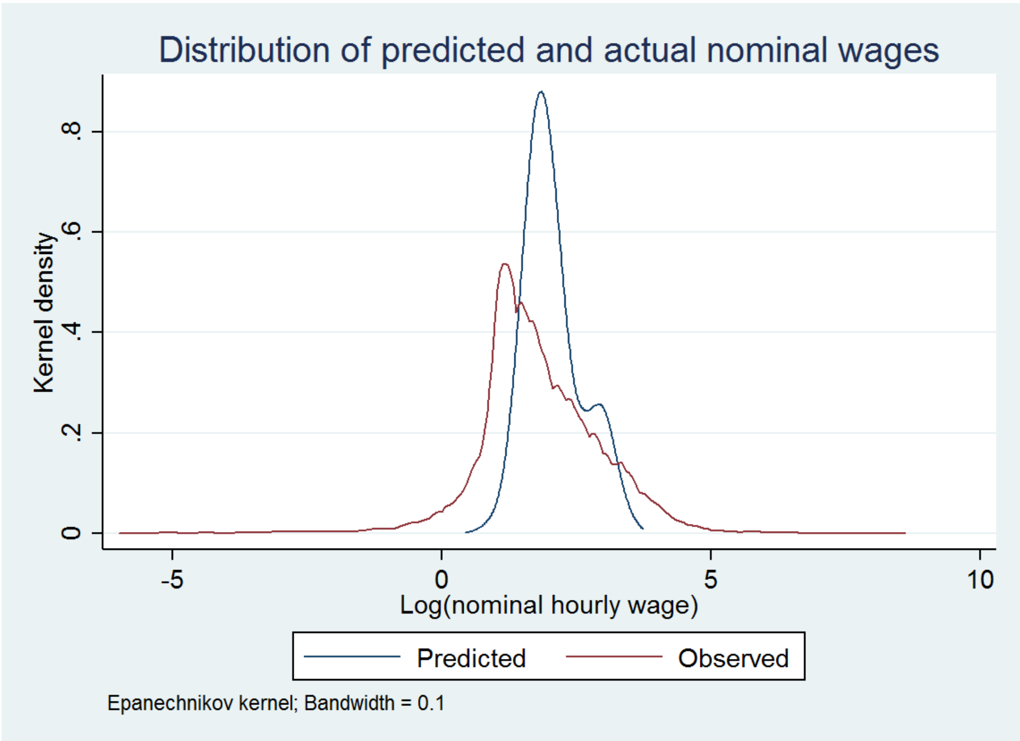

Figure 1.6: Kernel density plots of actual and predicted nominal wages of metropolitan out-migrants with matching.

For real wages (see figure 1.7) the distribution of observed wages lies now a bit to the right of counterfactual wages in metropolitan origins reflecting the positive return in real terms to leaving expensive cities.[10]

[10]The distributions are tested to be significantly different with a Kolmogorov-Smirnov test for equality of distributions. Both, the nominal and real wage distributions are significantly different.

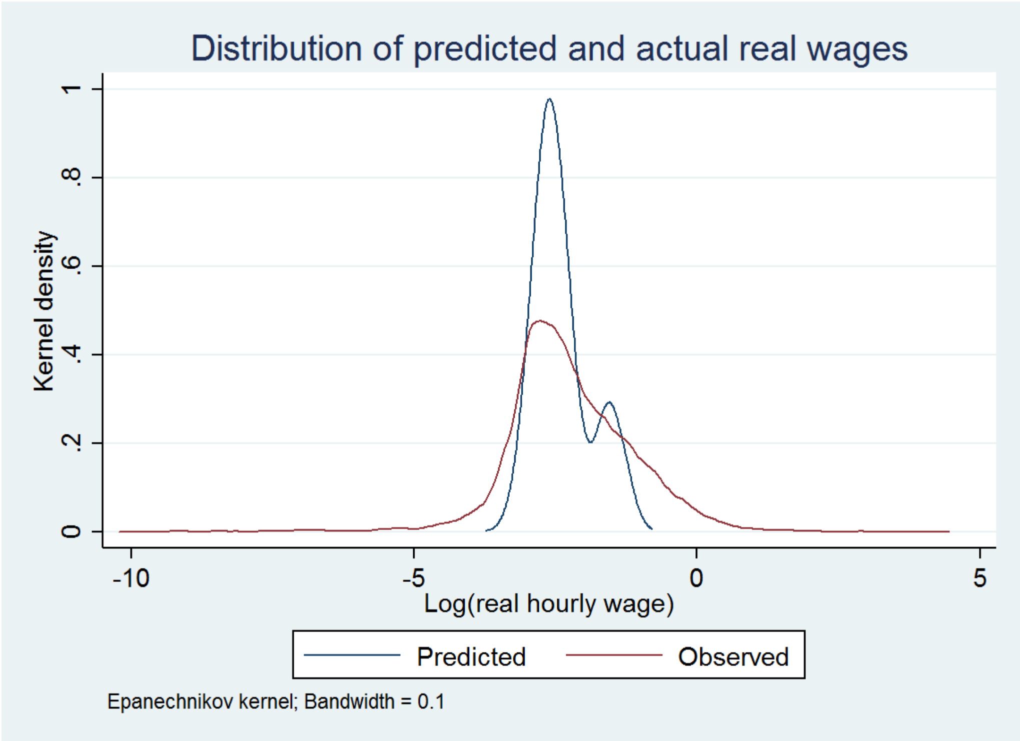

Figure 1.7: Kernel density plots of actual and predicted real wages of metropolitan out-migrants with matching.

1.5.3 Robustness: Price measures

In section 1.4.3, I computed a hedonic price measure to control for location-specific differences in housing quality. The results did not change qualitatively, but the coefficients became slightly smaller, thus indicating that this measure better captures unconditional housing prices. I therefore also compute the real wages using the hedonic price as denominator. Results are presented in table 1.12. The wage difference remains positive and statistically significant, but it is smaller by around a third than in the initial results.

Table 1.12: Differences in actual and predicted real wages for metropolitan out-migrants using hedonic prices as denominator, after matching

Log (real hourly wages)	*N*	Mean
Observed	*15,424*	-2.105
Predicted	*15,424*	-2.155
Difference		0.050***

Significance levels * 10% ** 5% *** 1% for t-test of difference in means between observed and predicted wages. Predicted wages are based on a matched sample of metropolitan residents.

Further concerns regarding the measure of living costs could arise from the fact, that low and high educated workers probably face different housing markets with different average prices. In appendix table A.3 (page 179) the results of the counterfactual analysis are presented using education-group specific living costs as denominator when computing real wages. I also use the median prices in the *microregião* instead of the mean. The signs and significance of wage differences remain the same as in the initial results, but their size changes. For high educated workers, the estimates yield larger positive wage differences when I use education-group specific living costs. The differences are 0.37 and 0.3 log points for average and median rents specific to high educated workers. For low educated workers the education-specific wage differences are smaller than unadjusted ones. Applying the median prices faced by low educated workers yields the smallest difference of 0.07 log points compared to 0.13 in the initial results. Rents are on average lower for low educated workers so that their gain from leaving metropolitan cities becomes smaller than when I did not account for the education-group specific rents. The opposite applies for high educated workers.

In the analysis of the destination choice of migrants we learned that metropolitan out-migrants face a trade-off in choosing between destinations where their loss in nominal wages is smallest, but their gain in lower living costs largest. This can explain why some individuals do not experience a positive return to migrating out of metropolitan areas across the income distribution. Some might fail to successfully evaluate their destination alternatives, some might lack the information about wages and prices at all destinations and others might just not be successful in acquiring the wage employment they had expected or they migrate for other reasons such as family. In this way, some metropolitan out-migrants lose out, while on average they gain in real wage returns.

1.6 Conclusion

The economic literature on migration in developing countries has been focusing on rural to urban movements and the growing metropolitan cities because this was the dominant observation in most countries, among them Brazil. In the decade of the 2000s, the movements of workers across Brazil have shown to lead equally out of metropolitan cities as into them.

This chapter uses the Brazilian Census data of 2010 to study this movement. I estimate the importance of real wages in the destination choice of metropolitan out-migrants and find that migrants maximize their utility by moving into smaller towns not far from their metropolitan origins. In these destinations they face lower nominal wages, but also lower prices. The counterfactual analysis reveals that on average the migrants achieve a positive return in real wages. This finding is especially strong for low educated workers who appear to lose from leaving the big cities if only nominal wages were considered. Non-metropolitan areas have on average worse quality of public service provision. The metropolitan out-migrants reveal a preference for education provision over health service emphasizing that preferences are different even between local amenities.

The findings are in line with the literature on wage returns to migration. It is confirmed that the comparison of wages conditional on individual skills is important for the destination choice, but migrants seem to consider them only with accounting for living costs (Tunalı, 2000; Dahl, 2002; Kennan and Walker, 2011; Moretti, 2011). Furthermore, selection corrected expected wages did not enter the model of destination choice significantly, which could indicate that workers have incorrect expectations about their wages and they could do even better in their destination choice. Rather, metropolitan out-migrants choose destinations that reduce their costs of moving as well as their living costs, which is why I find a positive return in real wages.

The Brazilian government invested heavily in regional development targeting those areas that were economically lagging behind. The regions that received most funding and that have been growing fastest in this period (Mata et al., 2005; Lall et al., 2009). The migration patterns I document could be a response to this regional development.

I analysed the out-migration from metropolitan cities in 2009, the year when Brazil was briefly affected by the international economic crisis following the financial crisis that had erupted in 2007. Even though not specifically identified, it is possible that the crisis played a role in the migration decisions and destination choice of metropolitan out-migrants. Metropolitan cities in Brazil concentrate many exporting industries and GDP declined in these cities in the years of the crisis. It is therefore not unlikely that these events contributed to the large migratory movements out of these primary cities. Boustan et al. (2010) and Monras (2015) showed how migration is one response of workers to economic crises in past and present times. If this was also the case in the observations of this paper, it would be another example of how the growth of smaller urban areas can offer alternative opportunities to workers responding to economic crises.

Chapter 2

Internal migration and crime: Municipal homicide rates in Brazil 2005-2010

2.1 Introduction

A lot of attention has been paid to empirical evidence for the consequences of immigration (e.g. Borjas (2003); Card (2001)). Only few studies, however, investigate the effect of internal migration on labour markets (Boustan et al. (2010); Kleemans and Magruder (2017)). They find a negative impact on residents' employment and wages. The negative effect of internal migration on local labour markets can have further consequences, such as an increase in crime. Theoretical and empirical evidence relates crime to unemployment and low income (Becker, 1968; Lin, 2008). I hence analyse in this chapter whether migration has an effect on crime rates at destinations. This question has not been investigated for internal mobility of workers, but only for international migration (e.g. Spenkuch (2011); Bianchi et al. (2012); Bell et al. (2013); Chalfin (2015)).

Literature on international migration finds that immigrants with good labour market prospects at their destination do not commit significantly more crime than native workers. There are some results suggesting that very low-skilled immigrants or those who are restricted by law from working are more likely to commit crime than residents at destinations (Spenkuch, 2011; Bianchi et al., 2012; Bell et al., 2013).

These results might also apply to internal migrants, but there are some differences. First, internal migration is less costly so that more people can move and those moving might differ in their characteristics. For example, due to lower costs poorer and less educated people might move. Secondly, internal migrants are much closer substitutes to residents in the destination labour markets than their international counterparts. They do not face language barriers or legal restrictions to their participation. Consequently, their arrival directly increases labour supply at destinations and puts downward pressure on wages and employment (Kleemans and Magruder, 2017). A decline in income and job availability can in consequence induce more participation in crime. At the same time, easier access to the labour market implies a lower likelihood to be involved in crime for the internal migrants themselves (Spenkuch, 2011; Bianchi et al., 2012; Bell et al., 2013). This could lead to a reduction in crime.

It therefore remains an empirical question whether there is an effect of internal migration on crime and in which direction it goes. If there is an effect, the subsequent question to ask is whether it differs by the structure of destination labour markets.

Assessing the impact of internal migration on crime at destinations is methodologically challenging. On the one hand, destinations with low crime are more attractive to migrants and on the other hand higher crime rates could attract specific types of migrants so that the estimated effect would be biased. To overcome this endogeneity issue, I apply an instrumental variable approach to estimate the effect of internal migration rates on homicide rates in Brazilian municipalities from 2005 to 2010. Using the nationally representative Brazilian Census survey of 2010, I reconstruct annual migration rates from people's former and current location of work and residence for the period from 2001 to 2010. Exogenous variation in the immigration rate over time comes from local labour demand shocks in the manufacturing sector in migrants' origins (Bound and Holzer, 2000; Notowidigdo, 2013; Monras, 2015; Diamond, 2016; Morten and Oliveira, 2016). This instrument is combined with the migration rates from 2001 to 2004 to establish the sorting of migrants from each origin across all possible destinations (Card, 2001; Boustan et al., 2010; Spenkuch, 2011; Bell et al., 2013; Jaitman and Machin, 2013; Chalfin, 2015; Kleemans and Magruder, 2017; Özden et al., 2017). The advantage of the Brazilian data is that it allows the analysis at a fine geographic level, the municipality, as well as of annual dynamics.

I find an elasticity of 1.2% between immigration rate and homicide rate for Brazilian municipalities in the studied period. The elasticity varies between 1% and 1.8% depending on the sub-sample applied and on how I deal with measurement issues in homicide rates. The magnitude of the effect compares to what other studies find as impact of international immigration on crime.[1] The result is only significant for municipalities with a small informal labour market and high historical crime rates. This suggests that the impact is heterogeneous across labour market structures. Criminal activity rises if migrants arrive at destinations with fewer easily accessible job opportunities and at locations with a large and established criminal sector.

Brazil provides a good case study as crime rates are among the highest in the world and explanations for its variation across locations are rare (World Bank, 2006; Reichenheim et al., 2011; Sachsida, 2013; Dix-Carneiro et al., 2017). High crime, especially homicides, comes at high economic costs. Aside from public safety expenditures, violent crime is related to other socio-economic costs, such as the negative impact on human capital (Monteiro and Rocha, 2017) or health (Manacorda and Koppensteiner, 2015). According to the Annual Brazilian Public Safety Report 2014, violent crime costs Brazil an equivalent to 5.4% of the country's GDP (Forum

[1] Özden et al. (2017) find a negative elasticity of -1.8% of international immigration on violent crime in Malaysia. Bianchi et al. (2012) find no effect on most crime, but an elasticity of 1% of immigration on robbery in Italy.

Brasileiro de Segurança Pública, 2014). Moreover, around 20% of Brazilians had moved to another municipality according to the Census of 2010[2] so that the assessment of potential impacts of internal migration is of relevance. Like in many developing countries, labour markets in Brazil are segmented with a large informal sector that offers more flexibility at low wages than the higher paying but rigid formal sector. Kleemans and Magruder (2017) find that such structures can lead to differential impacts of immigration. In their study, wages only respond in the informal sector to immigration, whereas employment effects were restricted to the formal sector. This chapter provides evidence that these labour market structures can also have differential implications for outcomes such as crime.

The chapter is structured as follows: An overview of literature on crime and migration in section 2.2 is followed by setting the context of crime in Brazil in section 2.3. Next comes a description of the data and definitions used for the analysis (section 2.4). Section 2.5 explains the empirical methodology. Thereafter, results are presented and discussed in section 2.6. Section 2.7 presents robustness checks. Section 2.8 discusses possible channels of the effects before I conclude in section 2.9.

2.2 Literature review

This chapter is related to the literature on migration and its consequences for local labour markets as well as that on crime and its determinants.

The empirical literature finds that unemployment, low wages and inequality are main determinants, especially for property crimes in the US (Gould et al., 2002; Grogger and Willis, 1998). Kelly (2000) emphasizes that only inequality matters significantly for violent crimes in the US. Evidence for developing countries is thinner despite the high prevalence of crime and inequality in some regions, such as Latin America. Fajnzylber et al. (2002) present cross-country evidence for the positive effect of inequality on crime including developing countries. They also document that crime behaves counter-cyclically and past crime is a strong predictor

[2]Author's own calculations based on the Census survey.

for current crime levels. Demombynes and Özler (2005) find that in South Africa higher inequality is related to higher property crime. Evidence for Latin America comes mainly from Mexico, where homicide rates have increased severely in the past decade. Inequality significantly predicts drug-related homicides (Enamorado et al., 2016). For Brazil, there is evidence that inequality, urbanization and unemployment are determinants of federal homicide rates (Sachsida et al., 2010; Sachsida, 2013), but also income, density of male population, drug use, firearm ownership, incarceration rate and police effectiveness appear to matter (Cerqueira, 2014b). The recent study by Dix-Carneiro et al. (2017) is the first to identify the causal effect of local economic shocks on crime in a developing country by exploiting the trade liberalization in Brazil as exogenous shock. Their findings confirm that worse employment rates increase crime in the medium run. However, this study and the previously mentioned literature do not consider the role of internal migration for variation in crime.

A fast growing strand of literature investigates the impacts of international immigration on crime rates in destination countries (Bianchi et al. (2012) for Italy, Bell et al. (2013) for UK, Spenkuch (2011); Chalfin (2014, 2015) across US counties, Özden et al. (2017) for Malaysia). They conclude that there is no or only a weak positive effect of immigration on property crime, but all emphasize that there is heterogeneity in the group of immigrants and their impact on crime. Immigration of groups with little labour market perspectives, e.g. unskilled Mexican immigrants in the US or asylum seekers in the UK, leads to higher property crime rates, whereas those immigrants with qualifications and legal access to the labour market, e.g. Polish immigrants in the UK, have no effect on property crime or they are even associated with lower crime rates relative to native residents.

Finally there is the literature that investigates the impact of internal migration on labour markets. For example, Boustan et al. (2010) find that the arrival of internal migrants had no effect on local wages in the destination cities in the US. It significantly raised out-migration of residents and reduced their time at work.

The only evidence on this topic from a developing country comes from Kleemans and Magruder (2017). They study the effect of internal migration on local labour markets in Indonesia and find that immigration increases unemployment and lowers wages for residents in the destination areas. Furthermore, they demonstrate how the impact differs for formal and informal sector workers. The formal sector is affected in terms of employment, whereas incomes are affected in the informal sector.

This chapter contributes to the literature in the following way. First, I analyse a possible determinant of crime that has not been considered so far, internal migration. In the economic model of crime (Becker, 1968), unemployment and low wages both reduce the opportunity costs of crime. Hence, I empirically test whether internal migration contributes to a reduction in the opportunity costs and consequently increases crime. Secondly, the literature on migration and crime has focused so far only on international migration even though internal migration differs in its mechanisms and the migrants' characteristics. My analysis provides evidence how these differences can imply different results. Thirdly, by focusing on Brazil this chapter enriches the literature of migration and crime as well as migration and its impact on labour markets in developing countries.

2.3 Homicides in Brazil

The literature on migration and crime focuses on property crime because it can be seen as an economic activity (Becker, 1968). Economic gains are achieved, for example, by the profit made from selling a stolen car. Furthermore, empirical studies in developed countries in most cases do not find any significant effect of international immigration on violent crimes (Bianchi et al., 2012; Bell et al., 2013). In these countries violent crime is comparatively low in contrast to Brazil, which ranks as one of the most violent countries in the world in terms of homicides. Many Latin American countries have high homicide rates, driven either by civil war-like struggles (Colombia) or wars between the state and drug gangs (Mexico) (Fernandes and de Sousa Nascimento, 2007). In Brazil, this is not the case. Violence is the result of different factors, hence the challenge to identify determinants of high homicide rates. In some areas of the country, such as large metropolitan cities or the Amazon, illegal markets have emerged over the past decades. Drug trade or trade of tropical woods are markets outside the legal order and agents use violence to enforce their rules (Chimeli and Soares, 2011). Violence between drug gangs in Rio de Janeiro regularly hit the global news and are often accompanied by crime unrelated to drugs in the surroundings of drug trade areas in the city (Fernandes and de Sousa Nascimento, 2007). Other sources of conflict are those of land, dating back to the colonisation of Brazil and today are often also encountered in the conflict with indigenous or landless population of Brazil (Hidalgo et al., 2010).

Additionally, during the dictatorship many weapons have found their way into the country and thus unreported firearm ownership is common. Small crimes like street robbery can easily result in violent and often deadly crimes. “Brazil is a society with rates of firearm victimization that surpass some countries at war” (p.228 in Fernandes and de Sousa Nascimento (2007)). Despite the introduction of a disarmament law, crime was not reduced. The state of São Paulo was the only one successful in the reduction of death due to firearm (De Castro Cerqueira and Pinho De Mello, 2012). All these points of conflict contribute to high homicide rates. Those who are mainly involved in committing homicides and those mostly victimised are young, non-white men, reflecting the underlying socio-economic issues of violence in Brazil (Reichenheim et al., 2011).

2.4 Data, variables and definitions

2.4.1 Migration

The migration data come from the Brazilian Population Census survey of 2010. Every ten years the Brazilian National Institute for Geography and Statistics (IBGE) conducts a nationally representative household survey (*Censo Demográfico* 2010, IBGE 2012). This survey comprises around 20 million individuals in all municipalities of Brazil, covering 10% of the whole population. It contains information on household composition, living conditions, labour market, education, geographic location, and on migration.

I construct a panel of annual migration between municipalities from the retrospective information about migration that the individuals provide. Each individual is asked about their former municipality of residence and the time since migration in years. For each year from 2005 to 2010 based on which year the individual stated as the year of her or his move, I aggregate the in- and out-migration rates at municipality level. The sample of migrants is restricted to working-age[3] male and female Brazilians, because the interest lies in labour market dynamics. For the aggregation from the individual to the municipality level, I apply population survey weights.

The immigration rate $\mathrm{M}_{m,t}$ is the number of immigrants arriving in municipality m in year t relative to the local population in the previous year:

$$\mathrm{M}_{m,t} = \mathrm{Immigrants}_{m,t} * 100,000/\mathrm{Population}_{m,t-1} \tag{2.1}$$

This is the immigration rate per 100,000 inhabitants. I use the rate of migration relative to population and not the absolute number of migrants, because I expect the effect of migration to be different if ten immigrants arrive in a municipality of 100 people compared to one with 100,000 people. It is also the same unit as the dependent variable, homicides per 100,000 inhabitants, which allows for an easier interpretation of the estimates.

This definition reflects the annual new arrivals of workers in a destination municipality. The measure differs from other studies which look at the change in the stock of migrants relative to the population (e.g. Bianchi et al. (2012) or Kleemans and Magruder (2017)). In contrast, I assume that internal migrants become part of the resident population after their arrival. I am therefore only interested in the impact of new arrivals on crime and not in the change of the population composition of

[3]The legal working age in Brazil is 16 years and the retirement age for men is 65 years. All individuals in the sample are currently not in school.

migrants and natives. I justify this assumption with the fact that internal migrants are the closest substitutes to natives. They share the same nationality and do not face language barriers or other such restrictions that international immigrants are often confronted with. It is therefore not of interest to look at the change in the stock of internal migrants at a destination, but instead I want to estimate how the inflow of new workers affects the destination.[4]

I restrict the definition to people who move at least 339 kilometres. This is the median distance that migrants in the data move. I do this to avoid capturing people who just move to their neighbouring town and who remain in the same local labour market. This definition also reduces the concern of spatial correlation between the instrument measured at the origins and the dependent variable observed at destinations. In the robustness checks (section 2.7.1 on page 84), I will change the distance cut-off and see whether and how this affects my results.

2.4.2 Origins and destinations

One concern in the analysis of internal migration arises from the fact that in each municipality migrants arrive, but others also leave at the same time. This would imply to measure the outcome at locations that are both, origin and destination. The studies that analysed the impact of immigration on crime or local labour markets were able to separate the origin from the destination because they looked at international immigration.

To identify the impact of internal migration on local labour markets, Kleemans and Magruder (2017) defined rural areas as the origins of migrants and urban areas as destinations. The urbanization rate in Brazil is around 80 percent and almost as high as that of the United States (Chauvin et al., 2016) so that it does not seem sensible to look at rural to urban migration. The largest share of migration would be ignored if it was restricted to rural-to-urban movements. Instead, I define sending

[4] Furthermore, data limitation does not allow to construct a stock variable. The stock of migrants would only be for those who moved between 2000 and 2010, as the survey does not include longer time periods than that. This would not measure the 'true' stock of internal immigrants.

and receiving municipalities based on past migration patterns. This reflects the locations pushing and pulling migrants within the country.

I distinguish origins from destinations by their past net-migration rate. I sum up the migration rates from the pre-study period, 2001 to 2004, and compute net-migration rates as the difference between in- and out-migration rate. Then I define the net-sending municipalities to be the origins, meaning those that saw more workers leaving than arriving in that period. Net-receiving municipalities are destinations.

This definition leaves me with 2,125 destination and 3,439 origin municipalities to conduct the analysis. I map these destinations and the origin municipalities in figure 2.1 to illustrate their distribution across the country.

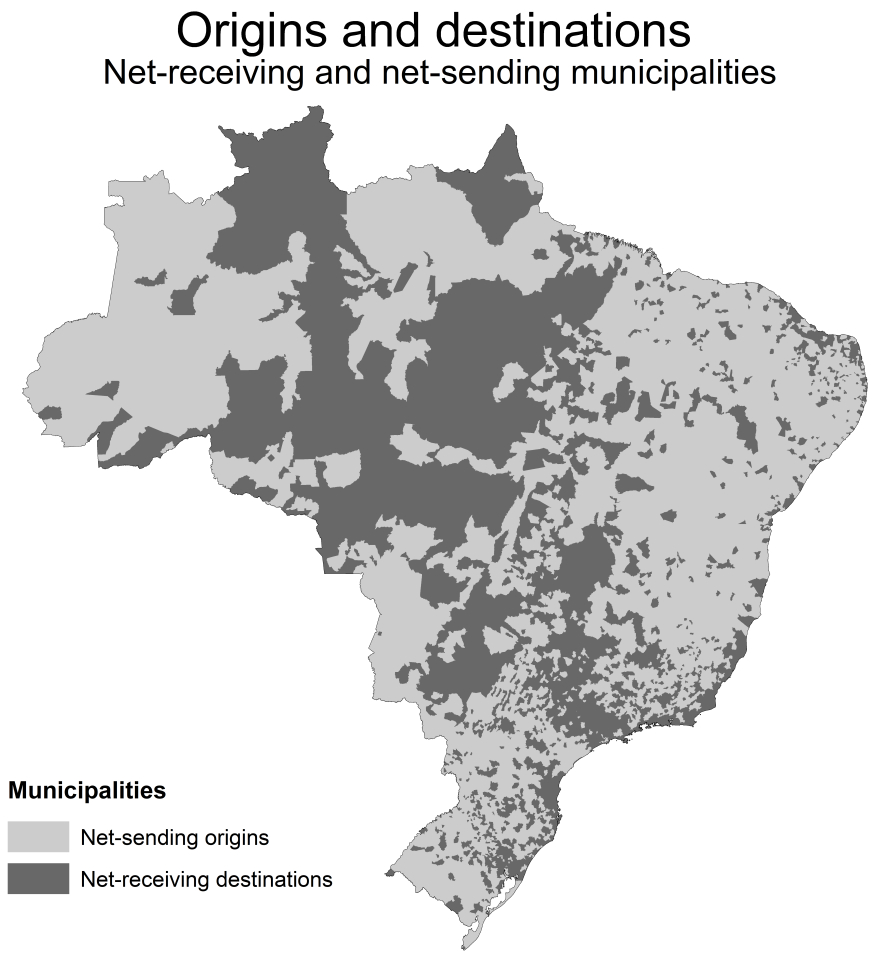

Figure 2.1: Map of origin and destination municipalities.

The darker locations are the destination municipalities. There is no clear geographical clustering. The North-eastern region overall appears to have seen more people leave over the pre-study period than the rest of the country, but even within that region there are still net-receiving municipalities.

2.4.3 Homicides

Crime rates in most studies are a combination of property crime, such as burglary and theft, and violent crime, with homicide as the most extreme case. In contrast to the US for example, there is little reliable data on property and other crimes aside from homicides available for Brazil. Dix-Carneiro et al. (2017) show highly significant correlations of homicides and various other crimes for municipalities in the two most populous states of Brazil, for which data on other crime is available. Homicide rates are therefore a strong indicator of general crime rates in Brazil and any effect found on homicide rates can be seen as a lower bound estimate for the impact on overall crime.

Homicide rates, i.e. the number of homicides per 100,000 persons living in a municipality, come from the Brazilian System of Death Registration (SIM) maintained by the Brazilian Ministry of Health (Ministério da Saúde). Data were extracted from the Department of Public Health Information (DATASUS) which is regarded as the most reliable information source on homicides in Brazil (Cerqueira, 2012). Homicides are deaths registered with the codes X85 to Y09 according to the international coding of violent deaths in the Global Burden Disease 2004 Update by the World Health Organization (Murray et al., 2013).

The main problem with this source is that it only reports homicides for municipalities where they have been reported in the system. Municipalities that did not actually experience any homicides are not listed with zeros, but they appear with missing values. They look exactly like municipalities with missing values due to misreporting. This can create a bias in the results if municipalities with no homicides reported are systematically different than those that report positive values. I test whether this is the case and how this affects the results in the robustness checks in section 2.7.5 (page 88).

The average homicide rate in Brazil from 2005 to 2010 is 25 (Table 2.1). On average each year 25 out of 100,000 people in a municipality are murdered. This rate is very high in international comparison, where the average is around 9 and the median only 3.5 in 2010 (World Bank, 2016). Homicide rates dramatically vary across the country. As I show in figure 2.2, some locations have a rate of only 10 per 100,000 while in other municipalities it is as high as 150.

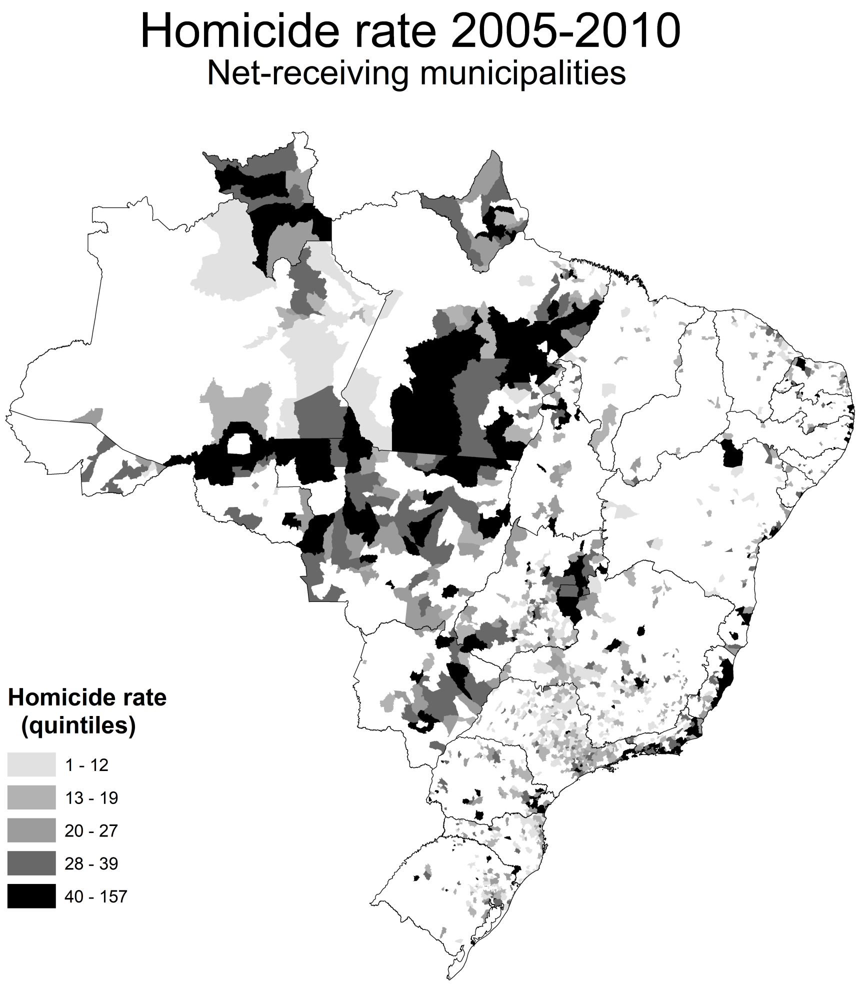

Figure 2.2: Map of homicide rate in destination municipalities.

2.4.4 Migration and crime in Brazil from 2005 to 2010

In table 2.1, I present the summary statistics of the two main variables, homicide and immigration rates, in destination municipalities. The number of observations is only 7,354 due to the missing values of homicides in some locations and years. The immigration rate into net-receiving destinations is on average 696 migrants per 100,000 inhabitants for the study period from 2005 to 2010.

Table 2.1: Descriptive statistics of main variables, destination municipality-year observations

	N	Mean	Std.dev.
Homicide rate	7,354	25.7	20.8
Immigration rate	7,354	696.5	827.8

Homicide and migration rates are the number of persons killed or of people who moved into a destination municipality per 100,000 inhabitants of that municipality.

The standard variations of the main variables presented in Table 2.1 indicate a large variation across municipalities. Figures 2.2 and 2.3 show a map of Brazil with the local variation of homicide rates (Figure 2.2 on page 71) and immigration (Figure 2.3) for the study period 2005 to 2010. The North-East experienced lower immigration in contrast to the rest of the country. The Central-West and Northern regions appear to have both higher immigration and higher crime rates than the rest of the country, but there is variation also within these regions. These maps show how much crime and immigration rates vary at municipality level within regions and states, and that high crime is not only concentrated in the big cities. This emphasizes the value of an analysis at this level.

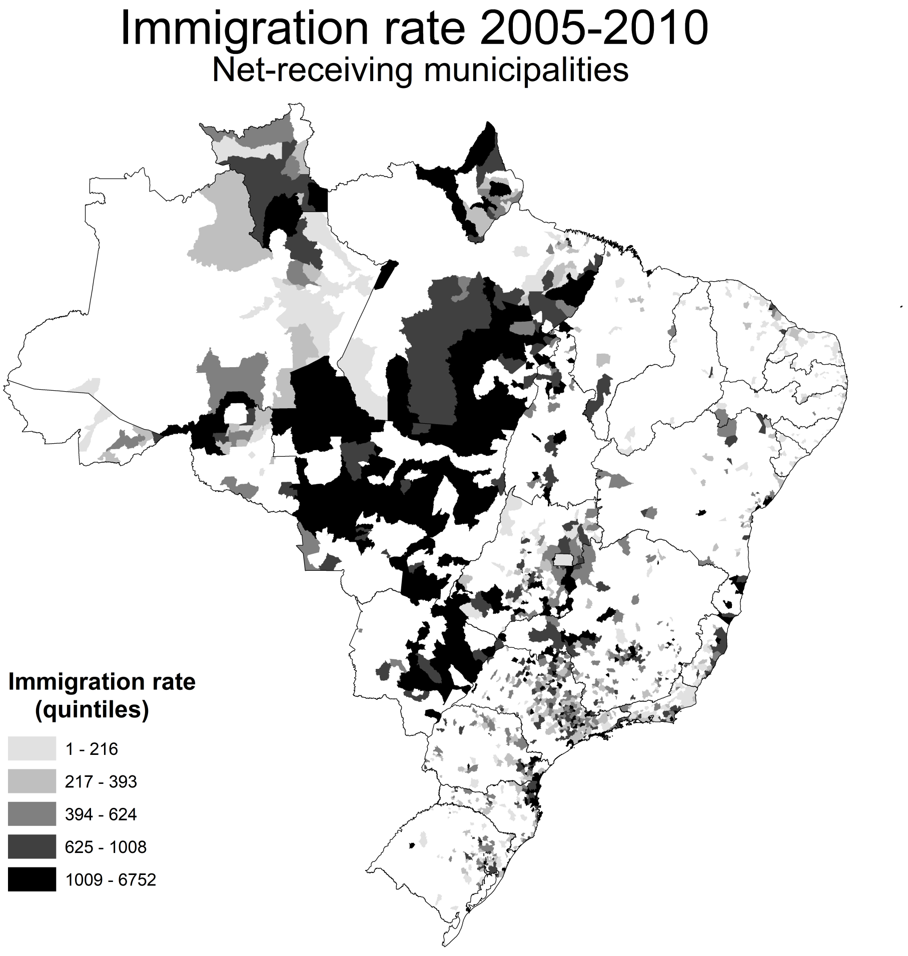

Figure 2.3: Map of immigration rate in destination municipalities.

2.4.5 Additional variables

Data on municipal population size is obtained from *Ipeadata*.[5] They are projections based on the Census from 2000 and 2010.

The data for the instrumental variable comprises employment and wages in the manufacturing sector at municipality and national level and they come from the RAIS (Annual Social Information Report)[6], a national employment registry. The federal government of Minas Gerais publishes the annual national wage growth and the total number of formal employments in each municipality by sector aggregated from the individual level RAIS data on the *Dataviva* online platform.[7]

2.5 Methodology

2.5.1 Empirical strategy

I estimate the impact of internal migration on local crime rates drawing on the literature on the effect of internal migration on local labour markets, especially Kleemans and Magruder (2017).

The sample for the analysis includes only net-receiving municipalities, the destinations of internal migrants, and the study period is from 2005 to 2010. I estimate a log-log fixed-effects model[8] of crime on immigration:

$$\ln(\mathrm{H})_{m,t} = \beta \ln(\mathrm{M})_{m,t} + \alpha_m + \delta_t + \gamma_s \mathrm{t} + \epsilon_{m,t} \tag{2.2}$$

The log of the homicide rate H in municipality m in year t is a function of the log

[5]Ipeadata is an online data pool provided by Ipea (Instituto de Pesquisa Econômica Aplicada), a Brazilian public research institute, which collects the data from several ministries and other public sources.

[6]The Annual Social Information Report (RAIS, *Relação Anual de Informações Sociais*) is collected by the Ministry of Labor and Employment and comprising around 97% of Brazilian formal enterprises.

[7]Manufacturing sector is used as broad term for the processing industry. It is category 'C' in the updated national code of economic activities (CNAE) by the Brazilian Institute for Geography and Statistics (IBGE) and it includes the two digit level from 10 ("Production of food products") to 33 ("Maintenance, reparation and installation of machinery and equipment") (IBGE 2010).

[8]I do not estimate a first-difference model due to issues of missing values so that the panel is not perfectly balanced.

of the immigration rate M into municipality m in year t. The coefficient β measures the percentage change in homicide rates associated with a one percent change in the migration rates.

I include municipality dummies, α_m, that capture any time-invariant unobservable characteristics of each locality. The year dummy δ_t controls for year-specific events that affect all municipalities.[9] ϵ is an idiosyncratic error term. I also include a state-year trend, γ_st, that captures any state-specific trends in the dependent variable. In Brazil, public safety falls under state legislation. For example, the state of São Paulo invested in policies that were successful in reducing homicide rates by almost half over the study period (see figure B.1 on page 187). The trend thus controls for policy differences between states.[10]

The errors are clustered at the level of *microregiões*. They are an agglomeration of municipalities that share a local labour market and some areas of administration. In my sample, a *microregião* comprises on average seven destination municipalities and the sample covers 486 *microregiões*. If migration is expected to affect crime rates through the labour market, it is important to allow errors to be correlated across municipalities within the same local labour market. All regressions are weighted by municipality population following the literature.[11]

Two econometric issues arise in the context of estimating β. First, there might be unobservable factors that affect homicide rates and immigration rates, for example unemployment rate or the labour market structure. Assuming that unobserved variables, such as labour market institutions, do not vary over the study period the municipality fixed effects, α_m, control for such factors.[12]

[9]Brazil's economy was hit by the global economic crisis for a short period in 2009/2010. This implied a decline in national employment and wages in the manufacturing sector (figure 2.5) and consequently a slump in the manufacturing employment share in the destination municipalities as depicted in figure B.2 (page 187).

[10]For robustness I also check for non-linearities in the trend by including a quadratic or cubic state-specific trend (see Appendix table B.2 on page 185). The linear trend appears to be the best fit as results do not change when applying a quadratic or a cubic trend.

[11]The argument comes from the health literature that shows that mortality realisation is an estimator of the underlying mortality probability. It was shown that "the variance of this estimator is inversely proportional to population size" p14, Dix-Carneiro et al. (2017). Other applications are seen for example in Bell et al. (2013).

[12]I do not include additional time-varying control variables, such as unemployment rate, because

Secondly, there is the problem of reverse causality between crime and migration. The lower crime rates are, the more attractive a location might be for migrants. This would bias the estimate downwards. Alternatively, the higher the crime rates are the larger the illegal market, which could attract specific types of migrants. In this case one would overestimate the impact. To identify the causal effect of immigration on crime I apply an instrumental variable (IV) strategy.

2.5.2 Instrument for immigration rates

For the aforementioned reasons, an instrument is required to estimate the causal effect of internal migration on crime. The IV has to be a strong predictor of migration but independent of crime rates at the destination of migrants. It has to predict that migrants leave their origin and that they choose one destination over another one (Card, 2001). In order to estimate a fixed-effect panel model, the instrument for this analysis also needs to predict variation in migration over time.

Monras (2015) and others showed that if a sector experiences a slump at national level, then wages and employment fall in locations where this sector usually employs a large share of workers. This crisis affects workers' mobility decision to stay in or leave these locations. In this analysis, I use such Bartik-style local labour demand shocks in the manufacturing sector to create exogenous variation in the migration rate (Bartik, 1991). This approach follows recent applications in the local labour market literature, such as Bound and Holzer (2000); Notowidigdo (2013); Monras (2015); Diamond (2016); Morten and Oliveira (2016). None of these studies uses the Bartik shock as instrument for migration, but they all show that it can be used to predict changes in local wages and employment.

The intuition behind the first stage is the following: A municipality m hosts immigrants from a specific origin o. If origin o is affected by a local labour demand shock S, I expect this to change the rate of migrants arriving in municipality m from origin o in the following period. The immigration rate, $\mathrm{M}_{m,t}$, into municipality

they are likely to be endogenous.

m in year t is determined as:

$$ln(\mathrm{M})_{m,t} = \mathrm{b}_1 S_{o,t-1} + \mathrm{b}_2 \mathrm{S}_{m,t-1} + \alpha_m + \delta_t + \gamma_s \mathrm{t} + u_{m,t} \tag{2.3}$$

The sector-specific local labour demand shock in year t, $S_{o,t}$, is an interaction of the employment share in a specific sector in location o in the pre-study year 2003, $\mathrm{e}_{o,2003}$, and the national wage growth of that sector, $\bar{w}_t^S$, in each year t of the analysis:

$$\mathrm{S}_{o,t} = \mathrm{e}_{o,2003} * \bar{w}_t^S \tag{2.4}$$

The local labour demand shock is sector specific. If wages at the national level in sector S fall, demand for workers in this sector is lower and people lose their jobs. The wages or other economic shocks in one municipality cannot change this national trend.[13] This is why I can assume the national wage growth to be exogenous.[14] If the employment share of the sector in a specific municipality, $\mathrm{e}_{o,2003}$, is large, this location will be affected more strongly by the changes in national trends. The mobility of workers is likely to be affected by this variation. The out-migration of workers would change the employment share. I therefore use the employment share in 2003 preceding the study period. This provides variation across origins and the national wage growth creates annual variation. Figure 2.4 maps the employment share in the manufacturing sector in 2003 in origin municipalities and figure 2.5 plots the national wage growth over the study period.

[13]In my sample of origin municipalities the local sector employment share is on average lower than in destinations (10% compared to 18% respectively). I am therefore not concerned that any of these locations could be the driver of national trends in the sector.

[14]Instead of the national wage growth Monras (2015) used a dummy indicating whether the year was before or after the financial crisis hit the US. Bound and Holzer (2000) used hours worked in a specific sector. I will use national wage growth like Diamond (2016), but in the robustness checks I will use sector employment growth like (Bartik, 1991).

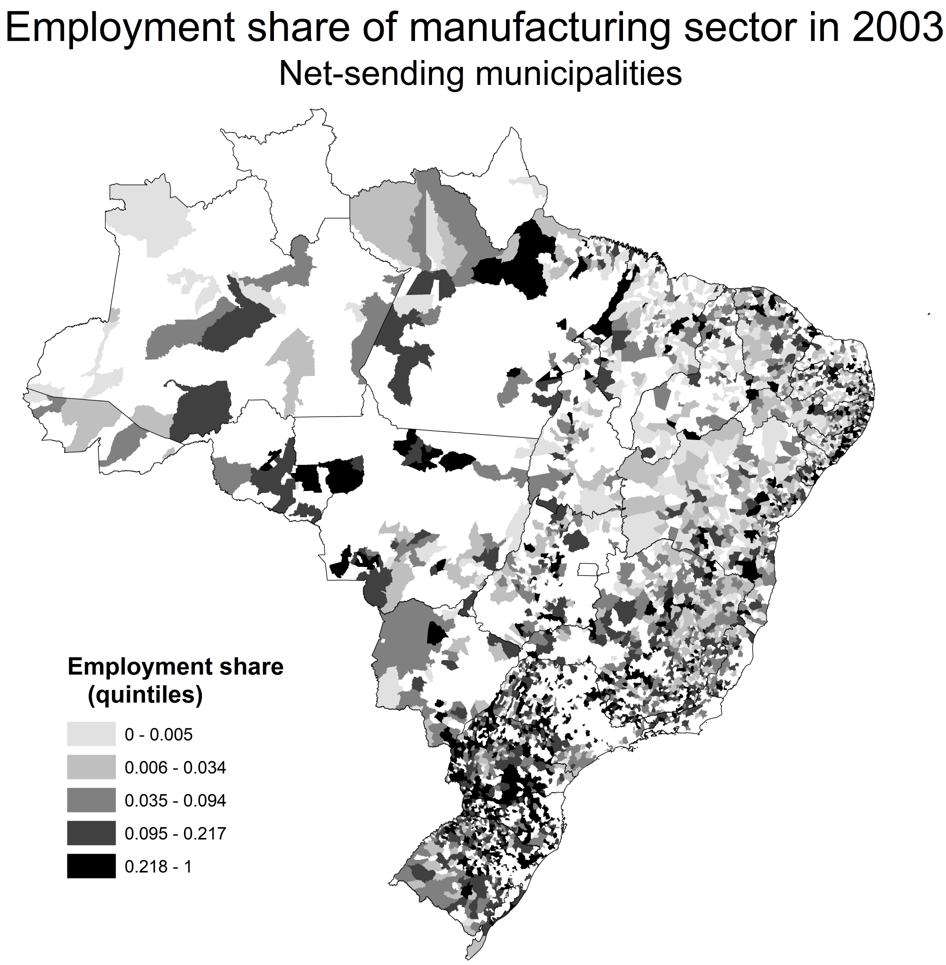

Figure 2.4: Map of instrument in sending municipalities, employment share in the manufacturing sector.

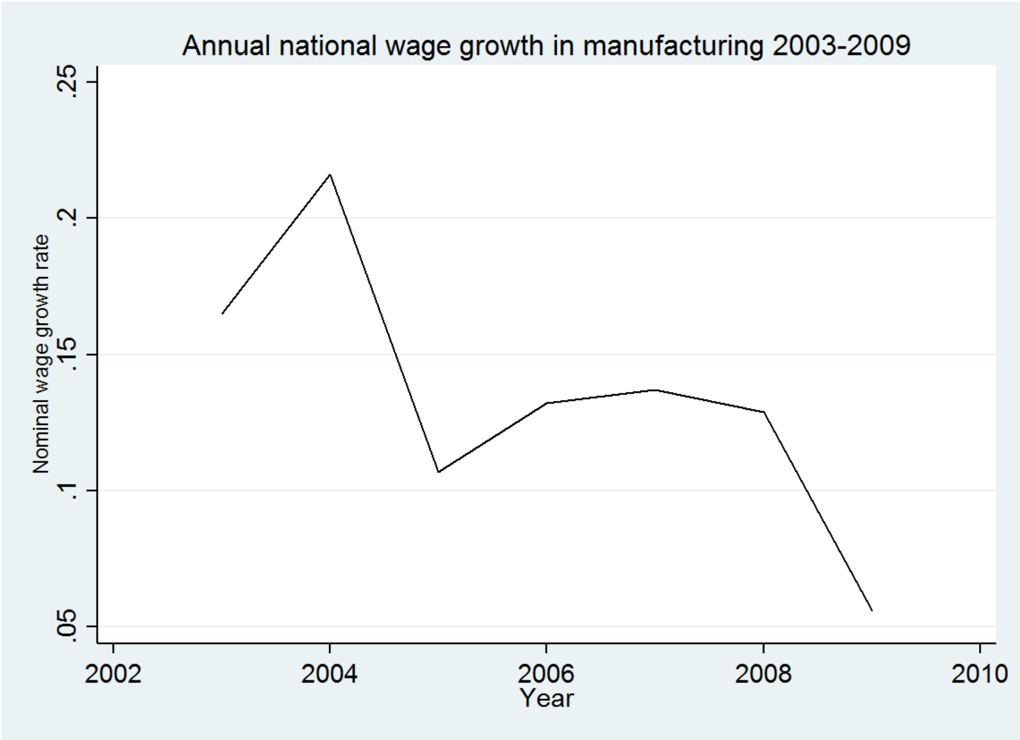

Figure 2.5: Annual national wage growth in the manufacturing sector.

In the specification 2.3, each municipality is assigned exactly one origin o as for example in Munshi (2003). Empirically, municipalities receive migrants from various origins, so that one possibility would be to use the top 3 or 4 origins in form of several instruments. In order to exploit the full information provided in the data, I follow Kleemans and Magruder (2017) and create a composite instrument of the shock across all origins.

I can therefore rewrite the first stage of the 2SLS instrumental variable estimation like this:

$$ln(\mathrm{M})_{m,t} = \mathrm{b}_1 \sum_{o=1(m \neq o)}^{O} (\mathrm{p}_{o,m} * \mathrm{S}_{o,t-1}) + \mathrm{b}_2\mathrm{S}_{m,t-1} + \alpha_m + \delta_t + \gamma_s \mathrm{t} + u_{m,t} \tag{2.5}$$

The composite measures sums up the local labour demand shock in all origins weighted by the term $\mathrm{p}_{o,m}$. This term reflects the importance of each origin o as migrant sender to each specific destination municipality m. The weight is based on migration patterns pre-dating the study period and it follows the literature exploiting the variation of historical migration flows between origin-destination pairs (Bell et al., 2013; Bianchi et al., 2012; Jaitman and Machin, 2013; Spenkuch, 2011; Chalfin, 2014, 2015; Özden et al., 2017). The idea is that the destination choice of migrants follows certain patterns that were established in the past and evolved over time due to networks (Munshi, 2003). A migrant from origin A is more likely to go to destination B than to C, if in the past more people moved from A to B than from A to C.

First, I aggregate all migrants that moved out of an origin municipality during the years 2001 to 2004. Next, I compute how many migrants moved to each specific destination municipality m. Then $\mathrm{p}_{o,m}$ is the share of the origin-destination specific migrants over all migrants that left origin o. One can think of it as the probability for migrants leaving origin o to move to destination m. The probability for migrants who move in response to a local labour demand shock S to arrive in municipality m is therefore the sum over the shocks in all origins O weighted by the destination specific migration probabilities of each origin $\mathrm{p}_{o,m}$. In section 2.7.3 (page 88), results are shown for the case in which the weights were computed using different time periods, using all data available from 2001 to 2010 or only the study period 2005 to 2010.

Finally, the first and second stage also include a control of the lagged local labour demand shock in the destination municipalities, $\mathrm{S}_{m,t-1}$, to capture potential correlation of these shocks across origin and destination municipalities. This is to ensure that the demand shock is only related to crime through the migration rate as required by the exclusion restriction.

In the second stage, homicide rates are regressed on the predicted immigration rate:

$$ln(\mathrm{H})_{m,t} = \beta_1 \widehat{\ln(\mathrm{M})}_{m,t} + \beta_2 \mathrm{S}_{m,t-1} + \alpha_m + \delta_t + \gamma_s \mathrm{t} + \epsilon_{m,t} \tag{2.6}$$

With a valid instrument, the estimation of equation 2.6 gives a consistent estimate of β, the impact of internal migration on crime.

2.6 Results

2.6.1 First stage results

Table 2.2 reports the results from the first stage of the IV two-stage least squares estimation. It is important to note that the sample is not a balanced panel at the municipality level due to missing values in the dependent variable. I return to this issue in section 2.7.5 (page 88).

The instrument is a significant predictor of immigration rates in destinations. In the first column, I included only year and municipality fixed effects whereas in the second column, I also add a state-year trend. This increases the coefficient of the instrumental variable and the Kleibergen-Paap F-test for weak identification. The latter is in both cases well above the rule-of-thumb value of 10 which confirms the relevance of the instrument for the first stage.

Table 2.2: First stage of the 2SLS regression

	Ln(Immigration rate)		
	(1)	(2)	(3)
IV: Origin labour	0.146***	0.157***	0.161***
demand shock in t-1	(0.022)	(0.023)	(0.022)
Destination labour			0.464
demand shock in t-1			(0.664)
Year fixed effects	Yes	Yes	Yes
MC fixed effects	Yes	Yes	Yes
State trends	No	Yes	Yes
N	6,589	6,589	6,589
Kleibergen-Paap F-Test	41.2	47.8	49.6

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the *microregião* level and all estimations are weighted by the municipality population. MC fixed effects refers to municipality fixed effects.

The results here show a positive coefficient of the instrument. If at national level wages in the manufacturing sector there is higher wage growth than in the previous year, more people migrate out from municipalities with a relatively larger share of the manufacturing sector and the opposite applies for slower wage growth. In the third column, a control of the lagged local labour demand shock in the destination municipalities is included to capture potential correlation of these shocks across origin and destination municipalities. If they were correlated, their inclusion in the regression should change the size of the coefficient of the immigration rate. Yet, the inclusion barely changes the estimated effect (column 3). It seems that destination labour demand variation is not a major pull factor for the sample of municipalities in this analysis.

The positive relationship between the instrument and immigration rates can be explained with the fact that, in some contexts, out-migration can be larger in response to a positive compared to a negative shock due to the presence of migration costs and credit constraints (e.g. Notowidigdo (2013)).

Guriev and Vakulenko (2015) show that the out-migration rate from less developed locations in Russia increases with an increase in income and decline in credit constraints in those areas. Hirvonen (2016) finds that Tanzanians exposed to negative income shocks are less likely to migrate at the individual and household level. Morten and Oliveira (2016) find that "37% of the total incidence of a [local] shock falls on residents [if migration is costly], compared to 1% in a model where migration is costless"[15] for Brazil. In the context of this chapter, the origins are relatively less developed municipalities (see table 2.5 on page 91) and thus I expect the majority of workers to be in the lower quantiles of the income distribution so that migration costs matter.

In the appendix table B.1 (page 184) I estimate a fixed-effects regression of local GDP, wages and formal sector employment in origin municipalities on the first lag of the instrument, the manufacturing sector labour demand shock. The estimates show a positive significant relationship for GDP and wages, but not for employment. This would support the argument that migration depends on income at origin. A positive shock increases wages in origins and GDP so that workers gain more income and can afford to move. If employment was affected, workers would be laid off and have an incentive to move in response to this negative shock. If employment increased,

[15]cited from the abstract of Morten and Oliveira (2016)

workers would have more opportunities at the origin and less incentives to move away. This would imply a negative first stage result. It is hence plausible that the first stage shows a positive correlation between local labour demand shocks at origin and immigration at destination.

2.6.2 Second stage results

I now present the main results in table 2.3. In the first column are the results of a simple fixed-effect estimation that does not account for the endogeneity of immigration and crime. In this regression, there is a weakly positive relationship between immigration and crime. Once I instrument for the immigration rate in column 2, there is a strongly significant impact of immigration rate on crime with an elasticity of around 2%. As expected, reverse causality leads to an underestimation of the relationship so that the IV-2SLS result is larger than the OLS result. In the third column, a state-year trend is included. This is important because Brazilian federal states are independently deciding about their budget for public safety and policies targeted at the reduction of crime. Once these state-specific trends are accounted for, the coefficient of immigration becomes smaller, but more precise.

Table 2.3: 2SLS estimation: Homicide rates on immigration rates 2005-2010, Second stage results

	Ln(Homicide rate)			
	OLS	IV-2SLS	IV-2SLS	IV-2SLS
	(1)	(2)	(3)	(4)
Ln(Immigration rate)	0.056*	2.341***	1.134***	1.162***
	(0.031)	(0.360)	(0.165)	(0.190)
Destination labour				0.539
demand shock t-1				(1.487)
Year fixed effects	Yes	Yes	Yes	Yes
MC fixed effects	Yes	Yes	Yes	Yes
State trend	No	No	Yes	Yes
N	6,589	6,589	6,589	6,589
Kleibergen-Paap Test		44	50	54.2

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the *microregião* level. Each regression is weighted by municipality population. MC fixed effects refers to municipality fixed effects.

Lastly, in column four, I present the results of my preferred specification that also includes the local labour demand shock in destination municipalities. Destinations have on average a larger share of employment in the manufacturing sector than origins (18% compared to 10% respectively). Thus, variations at the national level should matter even more for destinations than for origins. It is therefore important to control also for changes in the labour demand in destinations. Demand shocks at destination are insignificant and the impact of immigration on homicides is only marginally larger.

I find that on average an increase of immigration rate into net-receiving municipalities in Brazil in the period from 2005 to 2010 was associated with a significant increase in homicide rates of 1.2%. This effect is comparable to the magnitudes that other studies found for the impact of immigration on crime, when they looked at international immigrants. Özden et al. (2017) find a negative elasticity of -1.8% of international immigration on violent crime in Malaysia. Bianchi et al. (2012) find no effect on most crime, but an elasticity of 1% of immigration on robbery in Italy. [16]

2.7 Robustness and sensitivity of results

2.7.1 Challenges to the exogeneity of the instrument

One major challenge to identification is posed by the possibility that manufacturing sector labour demand shocks at origins predict not only out-migration from origins, but also other economic activity between origins and destinations, such as trading of goods. This could then affect local homicide rates through labour market spill-overs

[16] One major challenge in the analysis of the impact of internal migration on local outcomes is that each destination does not only receive internal migrants, but also experiences out-migration. The impact on the labour force in the destination caused by the arrival of new migrants might be much lower or not important if at the same time a significant share of workers leave this location. I therefore also tried to conduct the same analysis only using net-migration instead of immigration. The effect is still positive and significant, yet much smaller as expected. Out-migration of workers dampen the impact of the arrival of immigrants (Boustan et al., 2010). Results are presented in the appendix table B.3 (page 185), but the instrumental variable does not appear to be a good predictor for net-migration rates. One would need to find another instrument to predict the out-migration from destinations to then jointly predict the net-migration.

at the destinations.

For example, a negative shock in one origin might not only have the consequence that fewer workers leave this location but it might also cause changes in prices of goods produced and their trade. The firms affected will lower the prices of their goods in order to be able to compete in the national market. Their goods become cheaper relative to the goods produced in another municipality. Thus, internal trade from the municipality hit by the shock to an unaffected location should increase. This would put pressure on the firms in that destination and it could imply negative consequences for the local labour market there. Such negative spill-overs could in turn lead to more crime in destinations.

Ideally, I would test this relationship with data on internal trade which is not available. Instead, I conduct the same estimation with different samples for each of which migration is defined through a different minimum distance. The first estimation is done at a minimum distance of 100 kilometres and followed by estimations up to a minimum at 2,000 kilometres in steps of 100. Trade and labour market spill-overs from the local labour demand shock at origins to the destinations are much more likely under the scenario when distances are shorter. If my main results were driven by inter-municipality trade and labour demand spill-overs, I would expect the effect to be strongest for locations that are closer.

I plot the estimated coefficient of immigration on crime from each of these separate regressions in graph 2.6. The effect is smaller and insignificant for distance cut-offs below 300 kilometres. From 300 kilometres and higher, the effect is fairly robust to changes in the distance cut-off. As my main analysis was conducted using the distance cut-off of 339km, the main results should not be affected by such spill-overs.

2.7.2 National employment growth

The construction of the Bartik-style local labour demand shock requires an exogenous component that varies over time (Bartik, 1991). The most common approach in

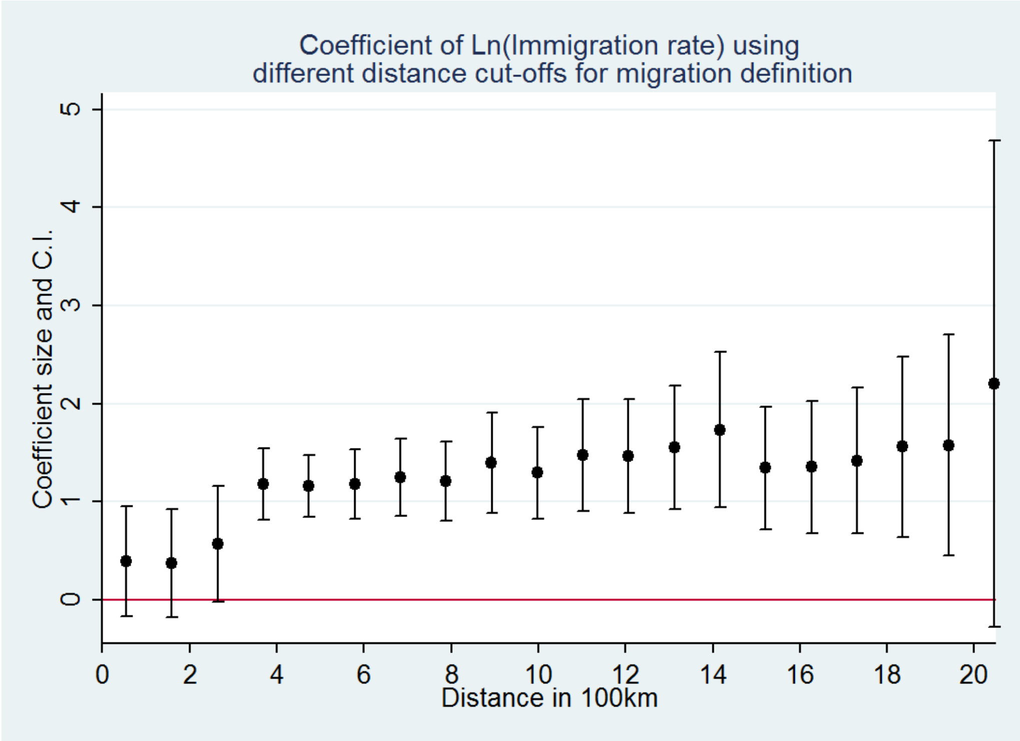

Figure 2.6: Results for different distance cut-offs in the migration definition.

the literature is to use wage or employment growth at the national level specific to a sector. So far, I constructed the instrument using national nominal wage growth. Now I use national employment growth to construct the instrument and present the estimation results in table 2.4, column 1. The elasticity reduces from 1.2% to 0.9%, but it is still positive and strongly significant.

Another option is to use the wage growth at the state level instead of the national level. This would still be exogenous to wage growth at each individual municipality, but it would allow for differential trends in greater economic areas. The results are presented in column 2 of table 2.4. The first stage F-test is very high indicating strong predictive power of this version of the instrument. The elasticity is smaller than the initial result of 1.2%, but very close to an effect of 1%, similar to the case when using employment instead of wage growth. The type of labour demand variation might induce different types of migrants to move and hence results in different effects.

Table 2.4: IV-2SLS regression: Sensitivity analysis

	Ln(Homicide rate)					
	Other instrument		Altering periods for weights		Altering periods for analysis	
	National level employment growth	State level wage growth	*2001 - 2010*	*2005 - 2010*	*2005 - 2007*	*2008 - 2010*
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Immigration rate)	0.889***	0.972***	1.161***	1.160***	0.947**	0.888***
	(0.196)	(0.147)	(0.218)	(0.221)	(0.388)	(0.252)
Destination shock	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
MC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State trend	Yes	Yes	Yes	Yes	Yes	Yes
N	6,589	6,589	6,589	6,589	3,112	2,947
Kleibergen-Paap Test	40.2	116	57.1	60.4	16	21.4

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the *microregião* level. Each regression is weighted by municipality population. MC fixed effects refers to municipality fixed effects. Destination shock indicates that the regression includes the local labour demand shock variable at destination.

2.7.3 Alternative weighting of origins

The construction of weights is based on migration rates from 2001 to 2004. Hence, one concern could be that this short period yields migration patterns specific to these years but not to the study period. As robustness check, I also compute these weights using the full period of migration data available, that is from 2001 to 2010, and only for the study period from 2005 to 2010. The results of these estimations are presented in table 2.4 in columns 3 and 4 respectively. Both estimates are not different from the original estimate of 1.2%.

2.7.4 Recall bias

The migration data is constructed from retrospective information self-reported by the migrants in 2010. As discussed by McKenzie and Sasin (2006) this can lead to a recall bias, meaning that the most recent movements are more likely to be (correctly) reported than migration a longer time ago. This implies that the estimated effect in my analysis should be different for the early years compared to the later ones. I run the regression just for the sample of most recent migration, from 2008 to 2010 (column 6, table 2.4), and for migration longer ago, from 2005 to 2007 (column 5). Despite losing power, the coefficients of immigration rates on crime are not statistically different from each other and only slightly smaller than the effect measured with the full sample due to the shorter periods analysed.[17]

2.7.5 Misreported homicide data

Homicides can be considered as being reported more accurately than other types of crimes due to their severity. However, in Brazil there are known issues with regards to reporting of homicides (Cerqueira, 2014a). For example, Brazil has a high number of police killings which are often not reported or hidden. Thus, I conduct a matching exercise to impute values for missing homicide rates. The detailed discussion of this

[17] I conduct a Chow-test whether the coefficients across these two sub-samples are different. The F-test statistic is 0.04 and the p-value for the difference in coefficients is 0.9888. I thus cannot reject the validity of pooling the sample across these years.

issue is presented in the appendix (see section B.3 on page 188). The results of this exercise are very similar to the main result with effects between 0.9 and 1.4 percent depending on the assumptions and corresponding imputations for missing values made.

2.7.6 Dynamic effects

I estimate the model including the first and second lag of the immigration rate. Results are in appendix table B.4 (page 186). Both lags are positive and significant indicating that past immigration also affects crime. This could suggest that internal immigrants need some time to find a job at destination or that the labour market adjustment takes more than one year, but it also suggests criminal inertia. Immigrant inflow increases crime and this by itself leads to higher crime rates in the future. It is important to note that the period of analysis is a short panel and it looks at annual changes. It is therefore plausible that also the lagged immigration rate are correlated with current crime rates.

2.8 Discussion of possible channels

2.8.1 Channels at the individual level

The results indicate that an increase in immigration into net-receiving municipalities leads to a small but significant rise in homicide rates in these locations. The literature on migration and crime suggests different channels to explain this result. Migrants could be criminals and thus increase the crime rates upon their arrival. To test this, one would usually look at incarceration or conviction data to compare the share of crime committed by immigrants compared to that of natives. In the context of internal migration it is difficult to distinguish migrants, because the incarceration data only reports basic demographics, but not the former and current municipality of residence.

Another explanation in the context of international migration is that hate crime against immigrants increases when more immigrants arrive. This is unlikely to be the case for internal migrants and even if hate crimes against internal migrants existed it is impossible to differentiate them from residents in Brazilian victimization data.

The literature further suggests that those immigrants with restricted access to the labour market might be more prone to commit crime due to lower opportunity costs. Internal migrants, however, do not face legal or language barriers to participate in the destination labour market.

Studies investigating the impact of internal migration on local labour markets found negative effects of migration on wages and employment of residents at destinations (Kleemans and Magruder, 2017). This could in consequence lead to higher crime committed by those residents that lose their job or receive lower wages. In order to confirm this hypothesis, I would have to estimate the impact of immigration on labour market outcomes of residents at the destinations at the individual level. Unfortunately, for this estimation I only have one cross-section of data so that the results would not be comparable to the panel estimates. The only panel of labour market outcomes at municipality level available is the RAIS data, but it only covers formal sector employment. This is not very informative for the Brazilian case, because the informal labour market is very large. I would not be able to see whether residents or migrants are unemployed as a result of the immigration or whether they work in the informal sector. Neither could I detect effects of immigration on informal sector wages. Kleemans and Magruder (2017) document that in a developing country with a large informal sector, the impact of rural to urban migration is different for formal and informal sector workers.

Given the data at hand, I investigate the possibility that due to the different characteristics of origins and destinations the increase in crime could be an export of crime from poorer origins to richer destinations via crime-prone migrants. Table 2.5 contrasts the main characteristics of net-sending origin and net-receiving destination municipalities. Destinations are overall the locations with slightly better labour market conditions. Unemployment and informality are lower, more high skilled jobs are located here, and agriculture is less important. It is plausible that internal migrants on average prefer these locations over the net-sending municipalities. It also implies that on average workers who leave the net-sending municipalities are probably less skilled and more often confronted with unemployment or low paid informal work in their origins. They could be more likely to be involved in illegal activities and their migration would simply imply an export of crime from origins to destinations.

Table 2.5: Descriptive statistics of origin and destination characteristics 2010

	Destinations *(Net-receivers)* Mean	Origins *(Net-senders)* Mean	Difference
GDP growth rate	0.05	0.046	0.004
Unemployment rate	0.057	0.062	-0.005***
Population	48,265	25,642	22,623***
Informality rate	0.57	0.68	-0.11***
Share of high skilled workers	0.31	0.27	0.04***
Share of non-white workers	0.48	0.57	-0.09***
Agricultural work share	0.31	0.38	-0.07***
Observations	*3,439*	*2,125*	

*** The difference in means between origins and destinations is statistically significant at 1%. The data for GDP growth rate come from *Ipeadata*.

I look further into the characteristics of workers who migrate in response to the sector-specific demand shocks. Those who are most involved in crime in Brazil are young, low educated men (Reichenheim et al., 2011). I run a regression at the individual level predicting the probability of a migrant being either low-educated, female, young or a young, low-skilled man on the sector-specific shocks.[18] The results, reported in table 2.6, show that the migrant workers who move in response to a local manufacturing labour demand shock in the origin municipality are significantly more likely to be low educated or to be aged above 25 years. There are no significant differences between the sexes. Consequently, the young, male and low-skilled workers are significantly less likely to move in response to the instrument.

[18]These regressions are estimated using the individual level data and for each year of migration separately.

The results from all years are reported in the Appendix table B.5 (page 186).

Table 2.6: Migrants' characteristic in response to local labour demand shocks at origin in t-1, 2010

	Probability to be			
	Low-skilled	Female	Young (16 to 25 years)	Young, male, low-skilled
Manufacturing shock	1.009**	0.333	-0.567	-0.455*
	(0.307)	(0.316)	(0.315)	(0.230)
N	44,212	44,212	44,212	44,212

Significance levels * 10% ** 5% *** 1%. These are the marginal effects from separate probit estimations of the probability of a migrant to be either low educated, female or young/male/unskilled on the local labour demand shock in a migrant's origin in t-1 for migrants who moved between 2009 and 2010. Standard errors are robust. All regressions include dummies for the state of origin.

The main results can partly be explained by the fact that relatively more low educated workers move in response to a manufacturing demand shock. These workers add to the unskilled workforce at destinations, increasing the competition for low-skilled jobs and thus the crime rate. One could argue that these unskilled migrants either are criminals themselves or at least the most likely to commit crime. Their movement from net-sending to net-receiving municipalities could thus imply an export of crime. While it is not possible to exactly test this hypothesis, those most likely to be involved in crime, the young and uneducated men, are less likely to move in response to the instrument applied here. This provides suggestive evidence against the hypothesis of crime export.

2.8.2 Labour market structure: Informal and criminal sector

Becker's economic model of crime argues that individuals are more likely to participate in criminal activities if the opportunity costs and deterrence are very low (1968). Lower deterrence implies a lower probability of getting caught and punished. Hence, the costs to participating in crime are lower. This depends on sentences in response to crime, but also on the probability of being caught based on policing measures. If the criminal sector is very large and long established, then this is often associated with less successful policing activities and low deterrence. Fajnzylber et al. (2002) documented in a cross-country panel comparison that past crime is a significant predictor for higher current crime. I thus expect that the impact of immigration on crime is larger in areas where there is a large established criminal sector.

Opportunity costs are another factor that affect crime. If immigrants indeed increase the competition in the labour market, this may increase unemployment, which may make crime a more attractive outside option. In Brazil and many other developing countries, the existence of the informal sector creates a form of buffer between formal employment and unemployment. If the informal sector is very large, immigrants as well as natives will find many opportunities in this sector whereas the formal sector is less easily accessible.

I investigate these two hypotheses in the data. In table 2.7, I present the results of my main regression for specific sub-samples. The first column shows the impact of immigration on crime in municipalities where the homicide rate has been above the median in the past (in 2000). The effect is slightly larger than the main result. In those municipalities where the past homicide rate was below the median, the effect is negative and statistically insignificant (see column 2).

In the third and fourth column, I conduct the same exercise with locations where the informal sector is above or below the median size based on the 2010 data.[19] As hypothesised, in municipalities where the informal sector is relatively small the arrival of more workers is associated with a significantly higher homicide rate and zero in places where the informal sector is large. In the last column, I look at destinations where both scenarios apply: a high past homicide rate and a small informal sector. The effect is again significant and the coefficient even larger than for the respective sub-samples. With 1.9% it is larger than the initially estimated elasticity of 1.2%.

[19] Using the informal sector size of 2010 could imply that the sector size is itself a result of immigration. It is not possible to match exactly the municipalities from the Census of 2000 to the municipalities in 2010 due to changes in administrative codes. Yet, as I do not use the informal sector size within the regression, but purely to define the sub-sample, I am less concerned that this should affect the results. Furthermore, recent analysis showed that the informal sector has been shrinking in the past years (Haanwinckel and Soares, 2016) which would work against my result.

These results suggest that the prevalence of crime in combination with low availability of outside options to the restrictive formal labour market is associated with a stronger impact of immigration on local crime rates.

Table 2.7: IV-2SLS regressions, Sub-samples

	Ln(Homicide rate)					
Sub-samples:	*Homicides in 2000*		*Informal sector size*		*Homicides high, informal sector small*	*Homicides low, informal sector large*
	High	*Low*	*Large*	*Small*		
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Immigration rate)	1.731***	-0.179	0.477	1.399***	1.962***	-0.166
	(0.479)	(0.313)	(0.931)	(0.299)	(0.603)	(0.298)
Destination shock	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
MC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State time trend	Yes	Yes	Yes	Yes	Yes	Yes
N	2,276	2,508	2,623	3,966	1,520	5,069
Kleibergen-Paap Test	28.4	20.8	1.71	40.6	22.6	26.2

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the *microregião* level and all regressions weighted by municipality population. Destination shock indicates that the regression includes the local labour demand shock variable at destination. Sub-samples are divided into those above and below the median homicide rate in 2000 (18.3), the median share of workers in the informal sector in 2010 (55%) in net-receiving municipalities (destinations).

2.9 Conclusion

This chapter aims to estimate the causal effect of internal migration on crime rates in Brazilian municipalities. In order to overcome the endogeneity of migration, I instrument it with local labour demand shocks in migrants' origins.

The results indicate a significant and positive effect of internal immigration on homicides at municipality level. A one percent increase in in-migrants relative to the local population leads on average and ceteris paribus to an increase of 1.2% in the homicide rate in a Brazilian municipality in the period from 2005 to 2010. This result is comparable in size to those obtained in studies of the impact of international immigration on local crime rates. The estimated effect is the elasticity of homicide rates to immigration rates. The average migration rate in the study period is 700 per 100,000 inhabitants, that of homicide rates is 25. If the in-flow of migrants increases to 770, homicides are expected to increase to 25.3, which is on average approximately 11 homicides per year more in a municipality than before. Due to data limitations, crime is measured with homicide rates. This is the most extreme form of crime and therefore the results are expected to reflect the lower bound for the effect of internal migration on crime.

I show that the estimated effect applies strongly to low-skilled migrants who are most likely to move in response to local labour demand shocks in the manufacturing sector at their origins. There is, however, no indication that migrants are crime exporters whose migration reduces crime in origins while increasing it at destinations. It is not possible to say more about whether it is the migrants or local residents who commit more crime due to a lack of corresponding data.

The hypothesis of this chapter was that internal migration has an impact on destination labour markets and through these affects crime. I therefore investigated heterogeneous effects with respect to labour market structures. Many developing countries (including Brazil) have segmented labour markets with a large informal sector. The effect of migration appears to depend on such local labour market structures. One may hypothesize that the more flexible, lower paying informal sector can help absorb migrants and thus reduce incentives to participate in crime. Indeed, I find that the estimated effects are the largest for municipalities with a small informal labour market.

In line with the traditional view by Fields (1975), large shares of informal work contracts in Brazil are due to the strict labour market regulations that make firing difficult and expensive, require payment of minimum wages, contribute to social protection and low working hours and that relax the role of trade unions (Barros and Corseuil, 2001; Mariano Bosch, 2007). This incentivizes employers to hire informally. Most recent evidence suggests that formally registered firms hire approximately 40-50 percent of their workforce informally, depending on firm size. The wage differential between formal and informal workers is zero within firms, conditional on individual characteristics (Ulyssea, 2010). Thus, the traditional view of informality as way to keep labour costs low seems valid in the Brazilian context. If the informal sector is very large, immigrants as well as natives will find many opportunities in this sector, whereas the formal sector will be less easily accessible. The results of Kleemans and Magruder (2017) reflect this. Informal wages are affected by migration, but not informal employment, while formal employment is directly affected by the increase in labour supply. If the informal sector is small, low-skilled workers will be confronted with a very rigid formal labour market and higher chances of staying unemployed. This reduces the opportunity costs for crime for these workers. This chapter confirms that the impact of immigration on local labour markets is different from previous model predictions and from empirical analyses that did not include an informal sector (Card and Lemieux, 2001; Borjas, 2003). In many developing economies such as Brazil, the informal sector is very large and thus should be accounted for.

In chapter 1, I provided a snapshot of the opportunities that many locations in Brazil yield for migrants to optimize their location choice. However, the results also revealed that not all migrants were able to realise positive returns to their move. This could be explained by the lack of access to jobs in highly formalised local labour markets. Improving the capacity of local markets to absorb new workers should therefore be of policy interest as migration appears to be a common strategy for Brazilians to adjust to local shocks.

I further find that the effects are largest in municipalities that have historically high crime rates, consistent with the interpretation that these are regions with low levels of deterrence. For policy makers, the focus should thus be on improving deterrence to avoid that *"crime" is an economically important activity or "industry"* (Becker, 1968, p.170).

Chapter 3

The nature and impact of repeated migration within households in rural Ghana

joint work with Julie Litchfield

3.1 Introduction

Internal migration is a common and sizeable phenomenon in many developing countries. An estimated 740 million people live outside their region of birth (Bell and Muhidin, 2009). Differences in regional economic performance induce people to leave poorer areas and move to those where more and better opportunities are located. In Ghana, around 35 percent of people in the population Census of 2010 had moved from their place of birth to another location within the country (Ghana Statistical Service, 2013). Many people move from poorer to richer regions, some move with the whole household, others send a member of the household (Litchfield and Waddington, 2003; Molini et al., 2016).

Internal migration plays an important role in poverty reduction and economic development at the individual, household and macroeconomic level. On the one hand, it contributes to structural change in the country when rural workers move into non-agricultural work in urban areas (Harris and Todaro, 1970). On the other hand, migration of a household member can insure the sending household against income shocks in the origin. Such insurance can prevent households from falling into poverty. Moreover, the income earned by the migrant member can raise consumption levels at home or even pay for investments in profitable technologies (Stark and Bloom, 1985). Additionally, geographic mobility offers young people to advance their education and gain new skills if their origins do not provide these opportunities.

Because of its size and relevance for economic development, economists study internal migration, but data limitations and methodological issues remain a challenge. One focus of research is the question whether and how internal migration affects households at origin. This chapter contributes to this strand in the literature. We investigate the impact of having a new migrant on the welfare of sending households conditional on their prior migration experience.

The engagement in migration of some village or community members was shown to significantly reduce migration costs for later migrants from that same network. This local migration experience would also increase the probability to be successful at destination in terms of finding a job. Thus, households are more likely to send a migrant if they have access to such a network of migration experience (Munshi, 2003; McKenzie and Rapoport, 2007). Households themselves can gain migration experience through their engagement in migration. Bryan et al. (2014) provide experimental evidence that the idiosyncratic migration experience of a household in contrast to that of social networks significantly predicts the repetition of migration within this household. Migration experience at the household level is hence important for future migration decisions and their impacts on the household.

Furthermore, the focus on new migrants is adequate for a setting in which households have several migrant members who move at different points in time. This is revealed by the data available in this chapter. We use primary data from a new two-wave household panel survey conducted in Ghana in 2013 and 2015.[1] The sur-

[1] As part of my PhD, I contributed to the completion of the data set by cleaning the data and ensuring that households and individuals can be matched between survey waves in close

veys were designed with the goal to collect as much information as possible about migration.

The econometric challenge of the comparison between migrant and non-migrant households is unobserved heterogeneity. There are unobservable factors that determine both, the fact that a household has a migrant and the outcome of interest, for example household income. Any result from a simple comparison of these households with and without migrants would be biased. Instrumental variables and selection models have been used to address this issue, often however restricted by the cross-sectional structure of the data employed (Litchfield and Waddington, 2003; Adams et al., 2008).

Comparing households that all have prior migration experience reduces the selection bias to some extent in the analysis of this chapter. Gibson et al. (2011) provide experimental evidence for different stages of selection, first that into migration, then into who moves. We apply entropy balancing weights (Hainmueller, 2012), similar to matching methods, and exploit the panel nature of our data to overcome remaining selection and omitted variable bias. The outcome variable of interest is an asset index.

Because there is little existing evidence on the consequences of idiosyncratic migration experience of households, we first describe migrants and their households in our new data to explore the dynamic patterns of migration. A comparison of the new migrants to those migrants who left the household before documents that new migrants are from a younger generation within households, such as children or grandchildren of the head. Their migration costs are lower, which might be related to family networks and the households' prior engagement in migration. From these observations we derive hypotheses for the impact assessment. Then we estimate how the asset welfare of households with a new migrant changes compared to those without, conditional on the fact that all households have previously had a migrant. We analyse whether there are heterogeneous effects by gender of the migrant, by

collaboration with the survey team in Ghana.

type of migration (seasonal or permanent), reason for migration (family or work), and by destination (within or across region).

We find no effect of sending a new migrant on the change in the asset index of origin households compared to those households who do not engage further in migration in the same period. This result is robust to a sensitivity analysis. Our interpretation is that the returns to migration might not show after the short period of our study. Households in our sample use their savings to finance migration. They hence do not experience a drop in their asset index. However, they also do not experience an increase in their asset index since the new migrant left. This could be, on the one hand, due to their use of savings to cover migration costs instead of investing into more assets and, on the other hand, because new migrants send only rarely and low remittances. We further suggest that due to prior engagement in migration our sample of households does not experience an initial decline in welfare. This could be caused by the migration costs or the loss in labour due to a member leaving (Taylor and López-Feldman, 2010). We however document that migration costs for new migrants are smaller than for prior migration, which indicates that migration experience at the household level reduces the costs of migration. In addition, prior to their move new migrants are either in school or doing unpaid work. It is thus less likely that their migration implies a loss in labour income for the household.

The chapter is structured as follows. In the next section, 3.2, we discuss the literature on impacts of migration on households left behind with respect to methodological challenges, knowledge gaps and evidence for our context. This is followed by the analytical framework for this study in section 3.3. Then we present the data used for the analysis (section 3.4) followed by a description of the migrants, migrant households and their prior migration experience (section 3.5). In section 3.6, we explain the methodology to estimate the impact of sending a new migrant on the welfare of origin households. In section 3.7, we provide results and robustness checks. Section 3.8 concludes.

3.2 Literature review

3.2.1 Evidence on the impact of migration on origin households

The research interest of this chapter is the short-term relationship between having a new migrant and the welfare of migrant sending households in rural Ghana. Many studies explored the more general question looking at the impact of having a migrant or not on some measure of well-being of the origin household. There exists also research that examines the effect of migration on the migrant's own welfare, e.g. Beegle et al. (2011), but this is not the focus of this chapter.

Theoretical models such as from the New Economics of Labor Migration (NELM, Stark and Bloom (1985)) cannot predict the direction of the impact of migration on origin households. The reason for this is that the impact depends on counteracting factors. For example, De Brauw and Harigaya (2007) model the impact of migration on consumption growth. It depends at the same time on the loss of farm production incurred by migration and the increase in consumption due to remittance receipt (De Brauw and Harigaya (2007), p.436) aside from the costs of moving.

Despite the use of similar outcome variables in the literature, results differ. Antman (2012) reviews the research that examines the impact of migration on the left behind family members and Mendola (2012) reviews studies looking at rural out-migration and its impacts on sending households. Both summarise mixed results from the literature. The following examples illustrate the inconclusive findings.

Empirical evidence from China by De Brauw and Giles (2012) documents an increase in consumption growth as well as "increased accumulation of housing welfare and consumer durables" (p.3). Quisumbing and McNiven (2010) consider the impact of migration and remittances on assets, consumption and credit constraints in the rural Philippines. They find that a larger number of migrant children reduces the values of non-land assets and total expenditures per adult equivalent in the origin households. However, remittances have a positive impact on housing, consumer durables, non-land assets, total (per adult equivalent) and educational expenditures. They find no effect on status of credit constraint. Mendola (2008) finds an increase in investments in agricultural production among the left behind households with international migrants in Bangladesh, but she does not find an effect for internal migration. Taylor and López-Feldman (2010) provide evidence of a positive effect of migration to the US on land productivity of migrant-sending families in Mexico. They also document an increase in per-capita income via remittances. Damon (2010) finds only weak increases in asset accumulation in El Salvador, he finds no impact of migration and remittances on investments in agricultural production.

What gives rise to these mixed results? One explanation is that the counteracting factors of costs and rewards to migration materialize at different speeds (Taylor and López-Feldman, 2010). The loss in labour is felt immediately as are the costs of paying for the migration of a household member. The returns to migration in form of remittances contribute to higher consumption levels. They delay however until the migrant arrived at the destination, found a job and earned enough income to send some of it back home. It might take even longer for remittances to accumulate enough to invest in productive assets.

Other aspects that contribute to the mixed results are the different data, definitions for migration and methodologies used. Migrants, or migrant households, are not a random sample of the population, but observable and unobservable factors determine their participation in migration. These factors can affect the outcomes of interest at the same time. In addition, the outcome itself can affect the migration decision. This is especially an issue in cross-sectional data. Aside from very few randomized control trials (Bryan et al., 2014) or natural experiments (Gibson et al., 2013, 2011; Yang, 2008), the most common approach to overcome endogeneity of migration is to use an instrumental variable (e.g. Damon (2010)). For example, De Brauw and Harigaya (2007) use historical policies and migration patterns to predict the number of migrants in households in Vietnam. Additionally, they estimate the effect of migration on consumption growth with generalized method of moments

(GMM) to control for reverse causality. Such an approach is only possible with longitudinal data or a large set of retrospective information on the relevant variables as in De Brauw and Rozelle (2008).

Only few studies consider migration experience at the household level in the form of seasonal migration. De Brauw and Harigaya (2007) and De Brauw (2010) provide evidence about the impact of seasonal migration on household welfare or agricultural production in Vietnam. While seasonal migration is most likely a repeated event, the authors do not specifically account for the repetition and potential learning process of the household. They measure the change in the number of migrants in the household without differentiating between households that have never had a migrant or those who have a migrant and send another one. Their choice to look at seasonal migration was purely motivated by pragmatic reasons due to the way migration information was reported in their data (De Brauw and Harigaya (2007), p.434).

Bryan et al. (2014) conduct a randomized control trial in a region in Bangladesh that is seasonally affected by famine to understand underused seasonal migration. Their intervention was a cash transfer to vulnerable households conditioned to finance seasonal migration of one household member. The results show significant improvements of consumption levels for the treated households. According to the authors' model, migration results in success or failure in terms of finding a job at destination and sending remittances. Households learn from this experience and it predicts their future engagement in migration. Further evidence for the role of migration experience within the family is provided by Giulietti et al. (2014). The authors develop a model that differentiates between 'weak' and 'strong' network ties and their role for migration decisions. Their findings suggest that networks at community level (weak ties) and prior migration of a family member (strong ties) act complementary, but weak ties have a have a higher impact on the migration decision. No further analysis is conducted to investigate how such different networks might impact migration and household outcomes.

In this chapter, we assess the impact of having a new migrant on origin households. We condition the analysis on prior migration experience. Thus, we contribute to the literature aiming to understand whether households learn from migration and what the implications are for future migration at household level. This chapter uses the first panel data in Ghana that contain an extensive migration module and applies a new method from the evaluation literature.

3.2.2 Migration in Ghana

Ghana is a lower-middle-income country according to the World Bank definition. It has been able to improve living standards remarkably in the past decade. The country's poverty headcount ratio decreased from 31.1 in 2005 to 24.2 in 2012 (World Bank, 2017). Despite these improvements, there remain challenges and small-scale agriculture is still the predominant income source in most regions. This gives rise to internal migration. Based on 2000 Census data, Castaldo et al. (2012) map poverty and migration rates at district level and find a clear correlation. Most people move out of the poor and into the richer regions.

Researchers document migration patterns in Ghana using various rounds of the Ghana Living Standards Survey (GLSS). Litchfield and Waddington (2003) show that in early rounds of the Ghana Living Standards Survey (GLSS) (those of 1991/92 and 1998/99) internal migration in Ghana was high and led mostly from rural to rural areas. This pattern is confirmed by Castaldo et al. (2012) for the GLSS5 in 2005. These movements were in most cases for economic reasons, to look for jobs, but around a third of migrants move also for family reasons. Molini et al. (2016) confirm with the latest GLSS6 (2012/13) that families in Ghana move to locations in hope of better prospects. Most migration in this recent survey leads again not only from rural to urban areas, but often from rural to rural areas, but into richer regions.

The evidence on impacts of migration on household welfare is mixed also for Ghana. Adams (2006) find a poverty-reducing effect of internal and international remittances at household level after controlling for selection and the application of an instrumental variable. Adams et al. (2008) show that remittances are not used differently than income from other sources. At the margin, remittance-receiving households do not spend more on consumption or investment than households that do not receive remittances. These results stand in contrast to Adams and Cuecuecha (2013) who find a marginal decrease in food consumption and an increase in investments, particularly in education, housing, and health for remittance-receiving households. They conduct the same analysis, a multinomial two-stage selection model with an instrumental variable. Their instrument draws on historical railroad networks and employment opportunities in destination countries, whereas Adams et al. (2008) relied on social networks among ethnic and religious groups. The use of different instruments could explain the contrasting results.

Ackah and Medvedev (2010) also use the GLSS5 to define determinants of internal migration at the individual and community level as well as the impact of migration on household expenditure. They apply a Heckman two-stage selection model to control for the non-randomness of migration. Migration drivers are higher education and youth, as well as worse infrastructure in home communities. Households with internal migrants are relatively better off than those without. The effect is, however, only significant for rural to urban migrants and not for those who remain in rural areas. Also applying a Heckman selection model, Mahé and Naudé (2016) find that Ghanaian internal migrants send relatively little remittances and often even receive support from their origin households using the GLSS6 (2012/13) data in combination with the Africa Sector Database (ASD). Their hypothesis is that migration is in this case often a long-term strategy based on the observations that migrants are often young members of the household moving to obtain higher education. Molini et al. (2016), exploit the GLSS6 to compare households who migrated as a whole to those who stayed. They make use of historical migration networks as instrument in a two-stage selection model. The positive impact of migration on consumption that they find is attributed to specific directions of movement, from the inland to coastal areas, and to male headed and better educated households. The authors also emphasise the absence of sectoral change in the migration strategy of households.

Due to weak instruments and bound to the use of cross-sectional data these studies lack means to control for unobservable factors that could contribute to reverse causality. It is therefore difficult to reconcile their results. This chapter contributes to the understanding of internal migration in Ghana and its consequences for origin households by using novel data. We utilize its rich questionnaire to document the diverse patterns of migration and we exploit the panel nature to reduce concerns of bias.

3.3 Analytical framework

This chapter investigates whether having a new migrant is related to a change in the welfare of the migrant's household at origin conditional on migration experience. The analysis is set in two periods, baseline and follow-up. All households have at least one member who is a migrant in the baseline period. Thus, they have previously engaged in migration, which we define as 'migration experience'. A migrant is defined in the surveys as a member of the household who is currently absent, left at least three months ago, but not more than five years.

A new migrant is defined as a household member who is present in the household in the baseline period and who then moves at least to another community and is still away in the follow-up period.[2] We look at new migrants, because it appears to be common for households to have more than one migrant and to see them move at different times. Thus, we are not interested in just the number of migrants, but in the dynamic aspect of another member migrating. Furthermore, it removes some of the selection bias of households into migration.

[2] It is possible that the new migrant had migrated in the past. In such a case, not only the household as a whole would have migration experience but also the individual migrant. The response rate to the question asking how many times a migrant moved before is unfortunately very low so that we cannot control for this in the analysis.

To give an example, imagine a household as depicted in the following table 3.3:

Household member	Migrant in baseline	Migrant in follow-up
A	1	1
B	0	0
C	0	0
D	1	0
E	0	1
Total:	2	2

This household has five members. At baseline, member A and member D are away as migrants. In the follow-up period, member A is still away as a migrant, while member D has returned to the household. Now member E is away as a migrant. If we were to compare only the total number of migrants away, we would see no difference between these two periods for this household. However, member D might have returned with money for the household, and will now contribute again to the household production (farm or business), and he or she potentially returned with new skills that could improve the returns to her or his labour. At the same time, for member E to migrate, the household had to incur some costs, maybe by selling assets or using savings. These factors have different impacts on the household welfare, so that we focus on new migrants instead of the total number of migrants. Thus, this example household would be defined as a household with migration experience and a new migrant. Member E would be this new migrant.

Different aspects determine the impact of having a new migrant. First, migration is costly and can initially lead to a decline in welfare due to the costs incurred as well as the loss in labour. Secondly, migration is beneficial when migrants send money back to their origin household and thus create another source of income. Thirdly, migration can be beneficial for the migrant him or herself directly. There might be more and better opportunities to earn an income or pursue further education at destination than at origin. Moreover, the household has one member less to care for and it might derive utility from the fact that the migrant can find a better livelihood somewhere else.

However, it is not clear in which direction the effect should work and which factor dominates. The aforementioned factors work in different directions. Additionally, in our specific case households have migration experience at baseline before they have a new migrant which can influence the effect. While sending a new migrant can incur costs, these might be lower conditional on prior migration experience of the household.

Following this discussion, we look at the impact of sending a new migrant conditional on migration experience. The sample is therefore first restricted only to households with migration experience at baseline. Then, households are assigned to a group called 'treated' and another one named 'control'. Households are in the treated group if they have at least one new migrant between the two periods. The remaining households without a new migrant between the two periods are in the control group.[3]

This definition implies that households can have more than one new migrant and they can have several baseline migrants. Our sample is restricted to those households whose new migrants were present members of the household in the baseline period.[4] Obviously, these definitions restrict the sample to a smaller set of observations than the original full survey.

3.4 Data

The data used for this analysis is a household survey collected in April/May 2013 and again at the same households in April/May 2015.[5] The data was collected by the Centre for Migration Studies (CMS), University of Ghana, Legon, through funding from the UK's Department for International Development (DFID) and made

[3]We could include households that had a return migrant at baseline, but no current migrant. They also have migration experience. However, there are no such households in our data.

[4]A special case are households that grew overall, which means that they had more members in the follow-up period than in the baseline due to new household formation. This can for example happen, when the son of the household head marries and his new wife and maybe a relative of hers join the household. If any of the newly joined household members then is a migrant in the follow-up period, we drop this household from the analysis. These households might represent a different form of household formation.

[5]In this way, the households are interviewed during the same season to avoid issues of seasonality between survey waves.

available by the Migrating out of Poverty Research Consortium, University of Sussex, UK.

In the first wave, around 1,400 households were surveyed, and in the second wave the team was able to follow up with around 1,100 of them.[6] The households are not nationally or regionally representative, but they were specifically chosen to oversample migrant sending households. While migration is a common phenomenon, it remains difficult to get a feasible sample in most nationally representative surveys.

The survey was conducted in five regions, the Northern region, the Upper East, Upper West, Brong Ahafo, and Volta region. These regions were major source areas for internal migration based on the information in the 2010 Ghana Population and Housing Census (Ghana Statistical Service, 2013). The sampling procedure followed a two-stage stratified design. Using the Census, enumeration areas (EAs) were chosen that were "proportional to the total number of out-migrants from that region" (Sugiyarto and Litchfield (2016), p.2). In the second stage, a list of households without migrants, with seasonal, returned or currently absent migrants was obtained for each Primary Sampling Unit (PSU) in each EA. Then, 4 non-migrant households and 11 migrant households were chosen at random in each of these PSUs to be interviewed.

The questionnaire was directed at the household head and asked about the demographics of each household member, their education and employment status, as well as their migration history. The questions about migration are either about current migrants or in an extra section directed towards returned migrants. These sections cover, for example, information on destination, reason for migrating, financing of the move, remittance sending, and occupation at destination.

The construction of the panel data set required a rigorous checking of data

[6]While our analysis is based on a balanced sample, we still investigate the attrition of households. Specifically, we look at how many households that were not tracked in 2015 had a migrant at baseline. These households would have been included in our sample either in control or treatment group. We then compare their baseline characteristics to those of the treated and comparison households to assess to which group they might have belonged. Comparing also their asset index indicates how their attrition might bias our results due to their attrition. See appendix C.2 on page 199 for a detailed discussion.

consistency. We were able to correct the majority of errors of misreporting caused by wrong manual entries of some enumerators in the questionnaires, e.g. skipping rows in the household roster. We were able to rely on our local partners to identify name changes based on their local knowledge and we were able to correct other errors by manually checking each individual questionnaire when we had doubts.

In the questionnaire, migrants are members who are currently not living in the household and who have been away for at least three months, but less than ten (in 2013) or five years (in 2015). This definition follows Bilsborrow et al. (1984), page 146. 60 percent of households in the sample for this analysis have only one new migrant, 25 percent have two, and the remaining 15 percent have three or more new migrants in the study period.

The outcome variables of interest are measures of economic status. The questionnaire in 2013 did not contain a consumption module and questions about income are inconsistent between the survey waves and thus not comparable. We therefore use asset information to construct an index as measure for a household's economic status.

After cleaning the data and making sure that the main variables of interest are available for all households in both survey waves, we are left with a balanced panel of 960 household-year observations. 131 migrant households are in the treated group, and 349 in the control group. The majority of households with a new migrant is located in Brong Ahafo and in the Volta region and the majority of the comparison group live in the Volta and the Northern region (Table 3.1).

Table 3.1: Sample of treatment and control households across regions in 2013

	Control		Treatment		Total	
Region	N	%	N	%	N	%
Brong Ahafo	61	17.5	40	30.5	101	21
Northern	93	26.6	19	14.5	112	23.3
Upper East	54	15.5	25	19.1	79	16.5
Upper West	43	12.3	18	13.7	61	12.7
Volta	98	28.1	29	22.1	127	26.5
Total	349	100	131	100	480	100

3.5 Descriptive statistics

The rich information about migration in this survey allows us to draw a detailed picture of migration in these areas of Ghana. We explore the characteristics of migrants and their households and we compare those migrants who had moved at baseline to the new migrants who only moved between the baseline and the follow-up survey. This comparison reveals interesting patterns. From these descriptions we can then move on to the analysis of the welfare impact of having a new migrant in section 3.6.

3.5.1 Baseline and new migrants

First, we turn to the individuals who migrate. We compare those who were migrants in the baseline (2013) and those who moved as new migrants between baseline and follow-up survey (2015). This comparison helps to document how new migrants differ from previous migrants within households with migration experience.

In our sample, we have 951 migrants in 2013, and 215 new ones in the follow-up survey. The response rates to the questions about migrants vary. We hence always report the number of responses for each question. Due to such missing values we cannot utilise all information in the impact assessment. This motivates detailed descriptive statistics which later help us explain our results. Table 3.2 provides an overview of the basic demographic characteristics of the migrants by migrant status and gender. Of the 2013 migrants, 38 percent are female, in 2015 the share of women increased to 50 percent.

Table 3.2: Demographic information of migrants, by migrant status and gender

	Baseline (2013)		New (2015)	
	Male	Female	Male	Female
N all	*592*	*359*	*107*	*108*
Age (in years)	32.4	30.7	25.6	26.8
Marital status				
N	*543*	*330*	*95*	*92*
Single (%)	44.6	42.7	68.4	47.8
Married/living with partner (%)	54	50.6	30.5	48.9
Separated/Divorced/Widowed (%)	1.5	6.7	1.1	3.3
Relation to household head				
N	*592*	*359*	*107*	*108*
Head (%)	8.3	1.9	3.7	1.9
Spouse / partner (%)	3.4	11.4	2.8	3.7
Child/adopted child (%)	52.4	49	49.5	51.9
Grandchild (%)	4.7	6.7	13.1	12
Niece/nephew (%)	5.6	7	14	13.9
Parent (%)	5.4	2.2	0.9	2.8
Sibling (%)	17.2	12.5	10.3	5.6
Son/daughter-in-law (%)	0.2	2.2	1.9	0
Sibling-in-law (%)	1.2	3.1	0.9	1.9
Parent-in-law (%)	0	2.2	0	1.9
Grandparent (%)	0.2	0.6	0	0
Other relatives (%)	1.2	1.1	1.9	2.8
Not related (%)	0.3	0	0.9	1.9
Education				
N	*520*	*296*	*97*	*89*
None (%)	14	18.6	23.7	31.5
Primary (%)	16.7	18.6	22.7	15.7
Middle/Junior (%)	31	30.4	27.8	22.5
High/Senior (%)	21.5	19.3	15.5	16.9
College/Technical (%)	16.7	13.2	10.3	13.5
Occupation prior to migration				
N	*436*	*232*	*70*	*68*
In school / education (%)	16.7	20.3	32.9	36.8
Paid employee (%)	8.9	4.7	10	5.9
self-employed (%)	35.1	27.6	24.3	17.6
Unemployed, looking for job (%)	9.9	7.8	8.6	8.8
Doing unpaid work (%)	24.1	30.2	21.4	27.9
Retired (%)	0.5	0		
Apprenticeship (%)	2.3	5.6	1.4	1.5
Others (%)	2.5	3.9	1.4	1.5
Activity prior to migration				
N	*241*	*97*	*42*	*34*
Farming (%)	43.2	34	42.9	26.5
Trading (%)	7.5	35.1	7.1	14.7
Self-employment (%)	10	17.5	2.4	8.8
Teaching (%)	9.1	5.2	7.1	14.7
Others (%)	30.1	8.2	40.5	35.3

Age and marital status

The migrants in 2013 are on average 31 years old and around half of them are married.[7] In contrast, the new migrants in 2015 are only 26 years old and a third of the men are married, but 49 percent of women are married. Separated, divorced or widowed migrants are mostly found among women who are baseline migrants. Overall, it seems that new migrants are more likely to still be single, especially men.

Position in the household

Around half of migrants are children of the household head and there is not much difference between the sexes. There is however a difference between baseline and new migrants. The former are relatively more often the head himself or his wife as well as the brother or sister of the head. It is the first or second generation in the household, who moves first. New migrants are relatively more often from the third generation, 12 percent (13 percent for men) are grandchildren of the household head, or they are relatives of second degree such as nieces or nephews. It is possible that there exist priorities in who gets to move first, starting with the head or spouse, the children or the siblings of the head, and eventually other younger relatives.

Education

The new migrants are relatively less educated and a third of women in this group have no completed education. Only 14 percent of male and 19 percent of female migrants in 2013 state to have no education at all, while around a quarter of the new migrants is uneducated. Female migrants in both groups are relatively less educated with larger shares having no or only primary education than male migrants. The most common level of education is middle/junior high school; 30 percent of baseline migrants, 28 percent of new migrant men and 23 percent of new migrant

[7] Age is measured in 2013 for baseline migrants, and in 2015 for new migrants. This is in order to avoid to make the baseline migrants artificially older by noting their age only two years after they had already been identified as migrants. However, we acknowledge that the baseline migrants might have been much younger when they moved, but the question about time since migration has a very low response rate, so that we cannot compute the age at migration.

women have completed this level. Higher levels are most likely among male baseline migrants (22 percent completed senior high school, 17 percent technical/college or tertiary education) followed by their female counterparts. Among new migrants, a slightly larger share of women achieved these higher levels of education, 30 percent, compared to 26 percent of their male counterparts.

Occupation prior to migration

Almost a third of women migrants in 2013 were doing unpaid work. While 28 percent of women did unpaid work in 2015, 37 percent of them were in education before their move. Yet, relatively more women are in education than men are before their move. Relatively fewer of the new migrant men were self-employed before migrating compared to those at baseline, 24 percent compared to 35 percent respectively. Around 10 percent of migrant men in both groups were paid employees before their move compared to only around 5 percent of female migrants.

In terms of the type of work the employed or self-employed did prior to their move, farming is the most common among men in both years with a share of around 42 percent. While women were also active in farming prior to their move, they often worked as traders or in some other type of self-employment. 35 percent of baseline migrant women were traders, but only 15 percent of new migrant women. This group was primarily active as teachers or in service work, such as hairdressing, dressmaking, domestic work, specified in the category ‘others’. For men the category ‘others’ mostly included crafts, such as masonry and carpentry, or services like driving.

Table 3.3: Migration decision and facilitation

	Baseline (2013)		New (2015)	
	Male	Female	Male	Female
Who was mainly involved in the migration decision?				
N	*461*	*251*	*85*	*80*
Self (%)	73.3	62.5	67.1	41.3
Father (%)	11.9	15.9	14.1	25
Mother (%)	3.3	4.8	2.4	7.5
Siblings (%)	1.5	3.2	1.2	2.5
Relative (%)	5.2	6.4	8.2	12.5
Community members (%)	0.2	0	1.2	0
Recruitment agent (%)	2.2	2	1.2	1.3
Others (%)	2.4	5.2	4.7	10
What was the main reason to migrate?				
N	*467*	*254*	*88*	*81*
Job transfer/opportunity (%)	17.1	13.8	15.9	6.2
Seek work/better job (%)	61	50	47.7	22.2
Study training (%)	12.6	13	12.5	25.9
To get married (%)	0.4	6.3	0	12.3
To accompany family (%)	0.2	1.2	2.3	1.2
To join family (%)	2.8	12.2	11.4	13.6
Declining yields in agriculture (%)	3.4	1.6	2.3	1.2
Civil conflict/war (%)	0.6	0.4	0	0
Family dispute (%)	0.2	0.4	1.1	0
Flood (%)	0.2	0	0	0
To join friends (%)	0.2	0	0	0
For medical treatment (%)	0	0.4	1.1	0
Others (%)	1.1	0.8	5.7	17.3
Contact at destination				
N	*481*	*259*	*87*	*83*
Yes (%)	54.3	69.1	64.4	74.7
Type of contact				
N	-	-	*56*	*61*
Father (%)			10.7	6.6
Mother (%)			7.1	9.8
Siblings (%)			17.9	14.8
Relatives (%)			55.4	55.7
Recruitment agent (%)			5.4	3.3
Other specified (%)			3.6	9.8
Job fixed up prior to moving				
N	*479*	*256*	*85*	*71*
Yes (%)	20.3	19.9	29.4	8.5

Decision makers

Now, we are turning to the migration decision, facilitation of migration and its costs. We report the descriptive statistics for these categories in table 3.3. Gender differences exist when it comes to the migration decision itself. In both years, relatively more male migrants made the decision themselves according to the household head who is answering these questions. 73 percent in 2013 and 67 percent in 2015 of male migrants contrast 63 percent and only 41 percent of female migrants respectively. The father or other relatives make the decision for female migrants relatively more often, especially for the new migrant women. In line with their younger age and lower education, it is plausible that older relatives are main decision makers when it comes to the migration of these women.

Reason for migration

Among baseline migrants, 78 percent of men and 64 percent of women moved for better job opportunities (including the categories 'job transfer/opportunity' or 'seek work/better job'). 19 percent of female migrants moved to get married, accompany or join family in contrast to only 3 percent of male migrants moving for family reasons. Around 13 percent of baseline women moved for studying or training purposes. In 2015, work is still the dominating reason to move for male migrants (63 percent), but relatively more join family than baseline migrant men (11 percent compared to 3 percent). Female new migrants move relatively more for studying (26 percent compared to 13 percent of baseline migrants). Joining or accompanying family is more common among new migrants. 14 percent of male and 15 percent of female migrants do so. 12 percent of new migrant women moved to get married whereas that was the case for only 6 percent of baseline migrant women.

Contacts at destination

Contacts at the destination can provide an important support for migrants. In our sample, women rely on networks relatively more. Almost 70 percent of women who were migrants at baseline and 75 percent of the new migrant women had a contact at destination prior to their move. In 2013, the corresponding number for men is 54 percent and 64 percent for 2015. For new migrants, we also know which contacts the migrants had at destination. Around 55 percent of times, the migrant had a relative at destination, and 18 percent of men and 17 percent of women had their parent at destination. From table 3.2 we know, that most of these new migrants are second or third generation within the household and often not direct descendants of the household head. It is therefore possible to imagine that nieces and nephews or grandchildren follow their parent who moved in the past.

Finally, we also observe whether migrants already had a job agreed before their move. This is less common, especially among female new migrants. In contrast, almost 30 percent of new migrant men state to have a job waiting for them at destination. At baseline, around 20 percent of migrants had a job fixed up prior to their move irrespective of their gender.

Financing migration (Table 3.4)

In terms of costs, female migrants pay on average less than male migrants for their move, 212 Ghanaian Cedi (GHS) at baseline and 112 for new migrants compared to 220 and 137 respectively for men. It is worth noting that new migrants pay on average less than baseline migrants do. Previously, we learned that relatively more of the new migrants have a contact at their destination and their household has prior engagement in migration. These observations suggest that costs can be reduced through migration experience.

The most common way to finance migration in 2013 were savings (70 percent) indicating that migration is an investment under credit constraints. If loans are taken then only from family. In no or very few cases formal sources for credit are used and only in very few cases migrants rely on a moneylender or recruitment agent. Around 12 percent of migration was financed by selling assets. New migrants in 2015 also rely on savings (42 percent of male and 38 percent of female migrants), but less so. Selling of assets is less likely to be used to finance the migration of a new female migrant at only 5 percent. A third of new migrant men and 42 percent of new migrant women state 'others' as source of financing. The specified sources among this category are mainly money from a parent and in some cases from the migrant her or himself. We consider this type of money as individual savings. Another source of financing are private transfers to the household from other migrants, remittances. Around 9 percent of female migrants used remittances to cover their moving costs, male migrants less so at baseline and 6 percent of new migrant men.

Table 3.4: Migration costs and means of financing

	Baseline (2013)		New (2015)	
	Male	Female	Male	Female
Migration costs				
N	*220*	*111*	*65*	*58*
in GHS of 2015	222.5	212.3	137.1	111.6
Financing of migration				
N	*371*	*173*	*79*	*79*
Savings (%)	72	67.6	41.8	38
Formal loan (%)	1.1	1.7	0	0
Loan from family (%)	7	6.9	6.3	5.1
Borrowing from money lender (%)	0.8	0.6	2.5	0
Advance from recruitment agent (%)	1.6	2.3	0	1.3
Sale of assets (%)	12.7	11	10.1	5.1
Government schemes (%)	1.6	0	0	0
Scholarship (%)	0.3	0.6	0	0
Remittances from other migrants in the HH (%)	3	9.2	6.3	8.9
Others (%)	0	0	32.9	41.8

Repeated migration, seasonality and destination (Table 3.5)

The baseline migrants have relatively more migration experience, around half moved once before their current migration. In 2015, around 70 percent of the new migrants move for the first time. Again, this is in line with the younger age of the new migrants, their unmarried status and activity prior to migration (school or unpaid work). Correspondingly, relatively fewer of the new migrants are seasonal migrants, especially of the female migrants. At baseline, 16 percent of migrants were seasonal workers, the same share of new male migrants moved seasonally, but only 9 percent of female new migrants. The new migrants moved relatively more often to another region in Ghana than to remain in their own district or region. Female migrants on average stayed closer to their origin, with only 47 percent of them leaving their region in contrast to 61 percent of male migrants. This difference could be due to those women who migrate to get married which is often tied to ethnic and family networks that might be closer to the origin community.

Table 3.5: Migration experience: repetition, seasonality, destination and occupation

	Baseline (2013)		New (2015)	
	Male	Female	Male	Female
Repeated migration				
N	*389*	*203*	*84*	*80*
First time migrants (%)	49.4	59.6	70	65
Seasonal migration				
N	*474*	*259*	*86*	*84*
Seasonal (in contrast to permanent) (%)	15.2	16.6	16.3	9.5
Destination				
N	-	-	*86*	*83*
Same district (%)			10.5	18.1
Other district, same region (%)			29.1	34.9
Other region (%)			60.5	47
Occupation at destination				
N	*353*	*182*	*54*	*51*
Farming (%)	19.8	12.1	14.8	21.6
Trading (%)	15.9	39.6	18.5	21.6
Self-employment (%)	16.1	26.4	1.9	9.8
Teaching (%)	7.9	8.2	9.3	7.8
Others (%)	40.1	13.4	55.7	39.3

Occupation at destination

At destination, the patterns of occupation change compared to what migrants did prior to their move (see section 3.5.1). Self-employment is much less common among new migrants (2 percent for men and 10 percent for women), while 16 percent of male and 26 percent of female baseline migrants work self-employed. Between 12 and 22 percent of migrants in both years work in farming at destination. This suggests that geographical mobility implies also some occupational mobility. Trading is again a common occupation for baseline migrant women and also 22 percent of new migrant women work as traders.

Remittances

Remittance sending behaviour is different between baseline and new migrants (see table 3.6). In the baseline group, relatively more men remit money to their families, 64 percent compared to 54 percent of female migrants. Among new migrants, only 41 percent of men and 39 percent of women remit. Baseline migrant men also remit larger amounts than their female counterparts (GHS 788 compared to GHS 655), but they all remit on average at least GHS 100 more than new migrants.

When asked how frequently they remit, new migrants remit relatively less frequent, half of them only on special occasions or in emergencies, whereas baseline migrants tend to remit mostly every couple of months or even monthly. New migrants are also less likely to remit goods to their origin household; only around 28 percent of them do so with no gender difference. Among baseline migrants, half of the women send goods back home and even 44 percent of men do so.

These observations that new migrants relatively more often get their migration financed from parents or paid themselves in contrast to baseline migrants, indicates that they might feel less obliged to remit money to their origin household.

Table 3.6: Remittances

	Baseline (2013)		New (2015)	
	Male	Female	Male	Female
Cash remittances				
N	*448*	*242*	*74*	*70*
Yes (%)	63.8	53.7	40.5	38.6
Amount				
N	*260*	*112*	*29*	*24*
in GHS of 2015	788.7	655.1	607.9	515.2
Frequency of remitting				
N	*267*	*120*	*29*	*26*
Weekly (%)	1.1	1.7	0	3.8
Fortnightly (%)	1.1	0	0	3.8
Monthly (%)	24.3	19.2	17.2	11.5
Every couple of month (%)	43.1	40.8	13.8	15.4
Every six months (%)	5.2	6.7	13.8	3.8
Every year (%)	6.4	9.2	3.4	11.5
Only on special occasions or emergencies (%)	18.7	22.5	51.7	50
Remittance of goods				
N	*427*	*228*	*74*	*71*
Yes (%)	44	49.6	28.4	26.8

Contact and support from origin households

We saw that family support is important for migration and its financing. Hence, migrants keep in frequent contact with their families (see table 3.7). Half of the baseline migrants contact their family at least once per week irrespective of their gender. New female migrants are even more likely to sustain frequent contact (57 percent being in contact at least weekly), as are new male migrants (53 percent). Despite the fact, that new migrants are less likely to send remittances to their origin household, they are in close contact with that household.

Finally, households sometimes also send money to the migrants to support them financially. This is relatively less common for baseline migrants, when 15 percent of male and 22 percent of female migrants received financial support from their families within the 12 months preceding the survey. Among the new migrants 26 percent of female migrants got money from home and 16 percent of male migrants.

Table 3.7: Contact and support from origin household

	Baseline (2013)		New (2015)	
	Male	Female	Male	Female
Frequency of contact				
N	*457*	*253*	*86*	*84*
More than once a week (%)	31.9	31.2	24.4	33.3
Weekly (%)	21.7	25.3	29.1	23.8
More than once a month (%)	19.9	22.9	24.4	22.6
Monthly (%)	8.3	10.3	9.3	8.3
More than once every three months (%)	5.7	3.6	3.5	3.6
More than once every six months (%)	3.9	2	1.2	1.2
More than once in a year (%)	5.9	2.8	3.5	2.4
I don't have contact with name (%)	2.6	2	4.7	4.8
Household sends money to migrant				
N	*400*	*214*	*67*	*70*
Yes (%)	15	22	16.4	25.7

These observations reveal that the new migrants in our sample are not moving for exactly the same reasons and do not share the same relationship with their origin households as the baseline migrants do.

3.5.2 Households

Before we investigate the impact of sending a new migrant on household welfare, we also look closer at the characteristics of the households that have a new migrant compared to those without. In table 3.8, we document the main characteristics of households with new migrants compared to those who do not send another migrant by 2015. All characteristics are measured at the baseline in 2013.

Household composition

Households who send a new migrant have on average between one and two more members than the comparison group. This indicates that they can afford to send members away as the remaining members are still enough to work on the family farm, in the family business or help with housework.

They have similar demographic structures measured with the dependency ratio (0.6) and female-to-male ratio (0.49). 29 percent of households that send a new migrant have a female head, 3 percentage points more than the control group. Heads of households in this group are on average 53 years old, those in treated households are on average one to two years older. Most heads are married, but 4 percentage points more among the control group. Relatively more heads in the treatment households are widowed, separated or divorced (22 compared to 18 percent). A negligible share is single (0.6 percent at baseline and 0.5 percent in 2015).

There are some differences between the groups of households with regards to their ethnicity. 29 percent of control households belong to the Mole Dagbani ethnic group and only 13 percent are of the Akan, 7 percentage points less than in the treated group. The share of Mole Dagbani is also smaller in the treatment group (24 percent) whereas the category 'others' is larger which indicates a more diverse distribution of treated households across ethnic groups.[8]

Education

Relatively more households sending a new migrant have heads with lower education, 84 percent completed no, primary school or middle school compared to 75 percent of heads in the comparison group. In terms of the highest level of education in the household, this pattern seems to reverse. It is measured as the highest level of education of any adult member in the household to capture overall education in the household as we do not have a measure of the years in education. The highest level of education within households is on average higher than that of the household heads. In a third of households lives a member with higher education such as technical college or other tertiary education. In another 31 percent, someone has completed senior high school. In the control group, these shares are very similar.

[8]The category 'others' include Ga-Dangme, Guan, Gruni, Grussi and other unspecified groups.

Table 3.8: Household characteristics at baseline, by group

	Households without new migrants (Control)	Households with new migrants (Treatment)
N	*349*	*131*
Household size (excluding currently absent migrants)	5.6	7.2
Dependency ratio	0.60	0.61
Female-to-male ratio	0.50	0.48
Female head (%)	0.26	0.29
Age of head in years	53.3	54.8
Marital status		
Single (%)	0.06	0.05
Married/ living with partner (%)	0.77	0.73
Separated/ Divorced/ Widowed (%)	0.17	0.22
Ethnicity of head		
Akan (%)	0.13	0.20
Ewe (%)	0.24	0.19
Mole Dagbani (%)	0.29	0.24
Others (%)	0.34	0.37
Education of head		
None (%)	0.41	0.41
Primary (%)	0.09	0.11
Middle/Junior (%)	0.25	0.32
High/Senior (%)	0.12	0.07
College/Technical (%)	0.12	0.08
Highest level of education in household		
None (%)	0.05	0.05
Primary (%)	0.11	0.08
Middle/Junior (%)	0.23	0.23
High/Senior (%)	0.30	0.31
College/Technical (%)	0.31	0.34
Employment status of head		
employee (%)	0.16	0.15
self-employed (%)	0.52	0.52
unpaid/unemployed (%)	0.23	0.25
inactive etc (%)	0.09	0.08
Main income source		
Public sector (%)	0.12	0.08
Private sector (%)	0.04	0.05
Own business (%)	0.28	0.26
Own farm (%)	0.42	0.51
Private transfers (%)	0.11	0.07
Others (%)	0.03	0.03
Migration experience		
Household has returnee (%)	0.17	0.24
Number of current migrants	1.9	2.1
Number of prior migration spells of current migrants	1.3	0.9
Share of seasonal migrants (%)	0.16	0.09
Share of female migrants (%)	0.35	0.41
Share of migrants with job (%)	0.60	0.66

Employment status and income source

Most household heads are self-employed, more than 50 percent in both groups. Around 24 percent of all heads are unemployed or doing unpaid work. 15 percent of them are working as paid employees. The majority of households earns the largest share of their income from their own farm. Farming is more common among the households with a new migrant than among those without, 51 percent and 42 percent respectively. Around 28 percent of households without a new migrant run their own business compared to 26 percent of households with a new migrant. 12 percent of control households rely on either public sector employment income or private transfers, which comprise remittances from migrants or other relatives. The respective share of households with new migrants is around 7 percent.

Migration experience

Around 24 percent of households with a new migrant had a member who returned to the household. In the comparison group, 17 percent of households have a returnee. Households have on average two migrants currently away in 2013 independent of the group. This is another indication for how common migration of more than one member is in our setting.

There are relatively more treated households whose baseline migrant moved for the first time. In contrast, baseline migrants in control households had migrated on average 1.3 times before. These households also have a relatively larger share of seasonal migrants in 2013, 16 percent compared to 9 percent in households that have a new migrant. It seems more common for control households to send the same member away repeatedly than to have a new migrant. This is also consistent with the difference in household size reported above.

Only a third of baseline migrants are women in households without a new migrant contrasting 41 percent in households with new migrants. The share of migrants who have a job at destination is relatively higher among households that later send a new migrant. On average, 66 percent of baseline migrants from these households have a job at destination. That is 6 percentage points more than for the comparison group.

3.5.3 Summary

In summary, there are some differences between households with a new migrant and the control group when we compare their characteristics at baseline. They differ in household size, ethnicity and livelihood. Households with new migrants are relatively larger and most live from family farm income. Additionally, their prior experience with migration appears to be successful in terms of the share of baseline migrants that have a job at destination and they are more likely to have a return migrant who potentially transmits important information for future migration.

Our sample reflects households in a setting where family farms or businesses are common, as is migration. Migration is mostly long-term and not seasonal, even though repeated migration is not unusual. The migration decision is made in a credit constraint environment. It strongly depends on the availability of savings to cover moving costs.

We observed that new migrants are different from baseline migrants. They are from a younger generation, often going to further their education or for work reasons. Fewer of them send remittances to their origin households than previous migrants. Family networks as well as frequent contact to the origin household, however, suggest strong ties between migrant-sending households and new migrants.

From these findings we cannot clearly predict the relationship of migration and household welfare, nor can we hypothesise its direction. In some cases, new migrants might be sent to diversify income sources and it is seen as an investment expecting returns to the household in form of remittances. In this case, we would expect to see a negative impact of the initial investment costs due to our short panel period as remittances usually delay to arrive and materialise in origin households (Taylor and López-Feldman, 2010). In other cases, it could be possible that migrants are already successful at their destination and are sending remittances that improve the household welfare.

Other migrants moved financially supported from their families to pursue more education or find new opportunities in other locations. This could be in line with human capital models of migration (Sjaastad, 1962). In these cases, it would be possible to find a negative effect on welfare of origin households due to the incurred migration costs and the loss in labour, but it is also possible that due to prior migration experience there is no impact on the origin households. This could even imply a positive impact as fewer members in the household leave more financial resources available for those who stay.

3.6 Methodology

Theoretically, there are no clear answers to the question whether migration has a positive or negative effect on the welfare of left-behind households. The New Economics of Labour literature (Stark and Bloom, 1985; Taylor, 1999) suggests that the migration decision is part of the overall household strategy in a context of market imperfections, but it cannot provide clear predictions for the impact of this decision (Mendola, 2012). As documented in the descriptive part migrants move for different reasons, which might imply different costs and different remittance sending behaviour. Additionally, prior experience with migration at the household level is also expected to affect the costs and migrants' remittance behaviour.

It remains an empirical question to study how having a new migrant relates to the welfare of origin households conditional on prior migration experience.

3.6.1 Empirical Strategy

We estimate the impact of having a new migrant on household welfare in the following specification:

$$Y_{i,t} = \beta_1 2015_t + \beta_2 \text{NewMig}_i * 2015_t + \beta_3 X_{i,t} + \beta_4 LM_{c,t} + H_i + \epsilon_{i,t} \tag{3.1}$$

Our interest is to see how the welfare of households is affected when they have a new migrant. With two time periods, we regress the outcome variable Y for household i on the treatment status of household i, NewMig_i and other variables. NewMig_i is a dummy indicating whether the household has a new migrant or not. This dummy is zero for all control households and it is 1 for the treated households, thus the subscript t. This term is interacted with a dummy indicating the second survey year. We also control for the general change of welfare over time by including the variable $Year_t$. Further controls are household characteristics, $X_{i,t}$, and labour market properties that vary over time, $LM_{c,t}$. The specification further includes household fixed effects, H_i to capture any unobservable characteristics of the households that do not vary between the survey waves.

The parameter of interest is β_2, the coefficient of the indicator whether a household has a new migrant or not interacted with the indicator for the second survey year. It measures the effect of having a new migrant between the two survey waves on the welfare of the origin household compared to those households that did not see another member migrate. It should be interpreted as an average treatment effect on the treated (ATT).

The time-varying household characteristics, $X_{i,t}$, are the dependency ratio, whether the household has a returned migrant and the employment status of the household head (unemployed/unpaid work, self-employed, employed or inactive). These can all affect household welfare and they can change within the time period under investigation. If a household has another child or if one of the older members becomes too old to work, the welfare might decline, as per capita income declines. Similarly, if a household head becomes unemployed this affects household welfare negatively. Finally, a migrant who returns to the origin household can, on the one hand, bring home money and invest it in assets to increase welfare or, on the other hand, the returnee might have failed at destination and now presents an additional burden to the household.

The local labour market variable, $LM_{c,t}$, is the employment rate in a community *c*. It is measured as the share of individuals who work as wage employees relative to the local labour force. This is included because a household seeking to diversify its income sources will consider local opportunities, where household members could earn a wage.[9]

3.6.2 Dependent variable: Asset index

As outcome variable we construct an asset index. Starting from Sahn and Stifel (2000) researchers used the rich information on assets available in many developing country household data sets to construct an index as welfare measure. The main argument for the use of the asset information instead of conventional measures such as consumption or income is that the latter are much more volatile and more difficult to measure. For a long-term assessment of the economic status of households, assets have been proven to be more stable and more reliable measures. Filmer and Pritchett (2001), McKenzie (2005), and Booysen et al. (2008) all used asset indices to compare poverty reductions in various countries and the use of such welfare indices has been increasing since the concept of multi-dimensional poverty was introduced (for a discussion see Ravallion (2011)).

It is important to note that a welfare index is a relative, not an absolute measure. It is very useful for comparisons of welfare between groups and/or over time. A detailed explanation of the method applied to construct the index (Multiple Correspondence Analysis) can be found in the appendix C.1 on page 193.

An asset index is a composite measure using information about asset ownership

[9]This measure is obtained using all individuals in our data in each community. Based on their main activity we define those who are employed and we sum all who are either employed, unemployed, doing unpaid work or self-employed. This captures how common paid employment is in a community and thus reflects the local opportunities for wage work outside the family farm or business. It is important to note that this measure is not correctly measuring the true employment rate, because our data is not representative of the local population. We looked into the possibility to obtain local labour market information from other locally representative datasets. However, we cannot use Census data because it is only available for one year before our survey was conducted so that we cannot control for variation over time. Alternatively, we could use the Ghana Living Standard Survey or the Ghana Socioeconomic Panel Survey (Institute of Statistical, Social, and Economic Research (ISSER),University of Ghana and Economic Growth Center (EGC), Yale University, 2015), but only half of the districts in our survey are covered in these surveys and neither of these datasets is available for the years of our survey.

and/or other welfare indicators in survey data. The researcher is interested in one continuous measure that captures the welfare of a household. In its simplest format, we can think of an asset index as the sum of its weighted components:

$$A_i = p_1 a_{1,i} + p_2 a_{2,i} + \ldots + p_k a_{k,i} \tag{3.2}$$

The asset index of household i is the sum of each of the individual asset indicator dummies, a_k weighted by an asset specific weight, p_k. Each indicator is equal to 1 if the household owns this specific asset, 0 otherwise. There are different possibilities to assign weights. The simplest, but most arbitrary, is to assign equal weights for each indicator. Ideally, one would use the price of each asset as weight. That is most times impossible due to lack of data. Alternatively, there are three statistical methods used in the literature to retrieve the indicator weights, Principal Component Analysis (henceforth PCA), Factor Analysis (FA), and Multiple Correspondence Analysis (MCA). These methods follow the same idea, but differ in their assumptions and restrictions imposed on the data. We apply the non-parametric and least restrictive method of MCA.

We use assets which are comparable to those found in the most commonly used household surveys in developing countries, the Demographic and Health Surveys (DHS). These are indicators of housing quality. They comprise the number of rooms, dwelling ownership, the presence of a bathroom and a toilet, main source of drinking water, and the floor and wall material.[10]

In table 3.9, we tabulate the ownership of each of these indicators by year and treatment status and describe the major changes observed.

[10]We were not able to include landownership, neither as a simple dummy variable whether a household owns land or not, nor in terms of land size. The reason for this is that this question is only available in the 2015 survey.

Table 3.9: Asset ownership by group and year

	Control		Treatment	
	2013	2015	2013	2015
N	*349*		*131*	
Number of rooms				
1	0.10	0.08	0.04	0.07
2	0.17	0.15	0.15	0.15
3	0.22	0.20	0.21	0.24
4	0.15	0.19	0.18	0.15
5 or more	0.35	0.37	0.42	0.37
Dwelling ownership				
Owned	0.83	0.85	0.89	0.90
Rented	0.17	0.09	0.11	0.05
Other	0.00	0.05	0.00	0.05
Bathroom	0.96	0.93	0.95	0.98
Toilet	0.37	0.36	0.41	0.42
Main source of drinking water				
Pipe borne water inside	0.12	0.18	0.11	0.16
Pipe borne water outside	0.29	0.30	0.21	0.18
Borehole	0.32	0.30	0.34	0.35
Dug well	0.13	0.09	0.13	0.14
Tanker service	0.00	0.01	0.00	0.00
Stream/river/lake	0.09	0.09	0.15	0.12
Rain water	0.01	0.00	0.01	0.01
Bottled or sachet water	0.05	0.01	0.05	0.04
Other	0.00	0.01	0.00	0.00
Floor material				
Mud	0.20	0.19	0.28	0.16
Raw wood, boards	0.00	0.01	0.00	0.00
Cement/concrete	0.77	0.77	0.69	0.80
Burnt brick	0.01	0.00	0.02	0.01
Terrazo	0.00	0.01	0.00	0.02
Floor tile	0.00	0.01	0.01	0.02
Polished wood	0.01	0.00	0.01	0.00
Wall material				
Bamboo or other organic materials	0.04	0.04	0.05	0.05
Cloth, cardboard, cans	0.01	0.00	0.00	0.00
Zinc	0.05	0.11	0.02	0.16
Raw wood	0.00	0.00	0.00	0.00
Mud, adobe, cane wall	0.36	0.35	0.40	0.32
Block, bricks, stone, prefabricated material, polished wood	0.50	0.49	0.50	0.46
Other	0.03	0.01	0.03	0.01

The ownership status and presence of a bathroom or toilet are relatively stable. There are some larger changes between years for floor and wall material and smaller changes for the number of rooms and the source of drinking water. These changes also differ between treatment and control group which is important for our identification strategy. If all changes would go in the same direction we would not be able to identify an effect of having a new migrant on the change in the index.

Figure 3.1 presents the asset index in 2013 of households with a new migrant and of those without, figure 3.2 depicts the same for 2015.

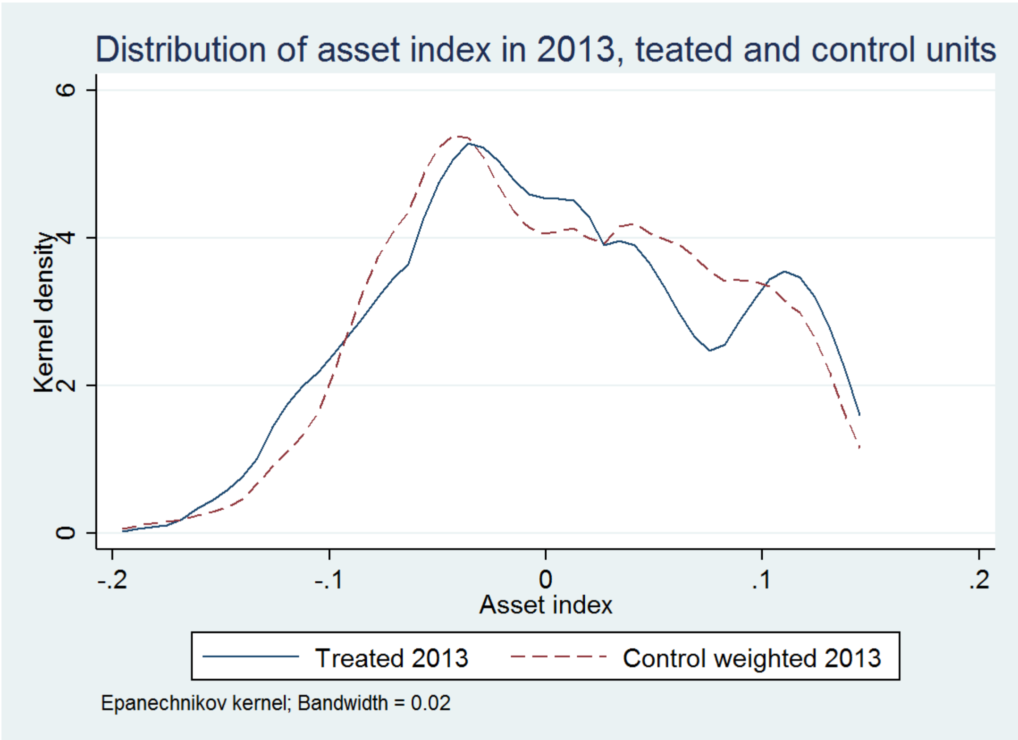

Figure 3.1: Asset index of treated and control households in 2013

These figures illustrate that the distribution of the asset index overlap in 2013, but they shift apart in 2015. It seems that households without a new migrant have a higher distribution of the index. Note that the distribution for control households are weighted to make households comparable applying a method which is described in section 3.6.3. This explains the overlap in the baseline year (figure 3.1).

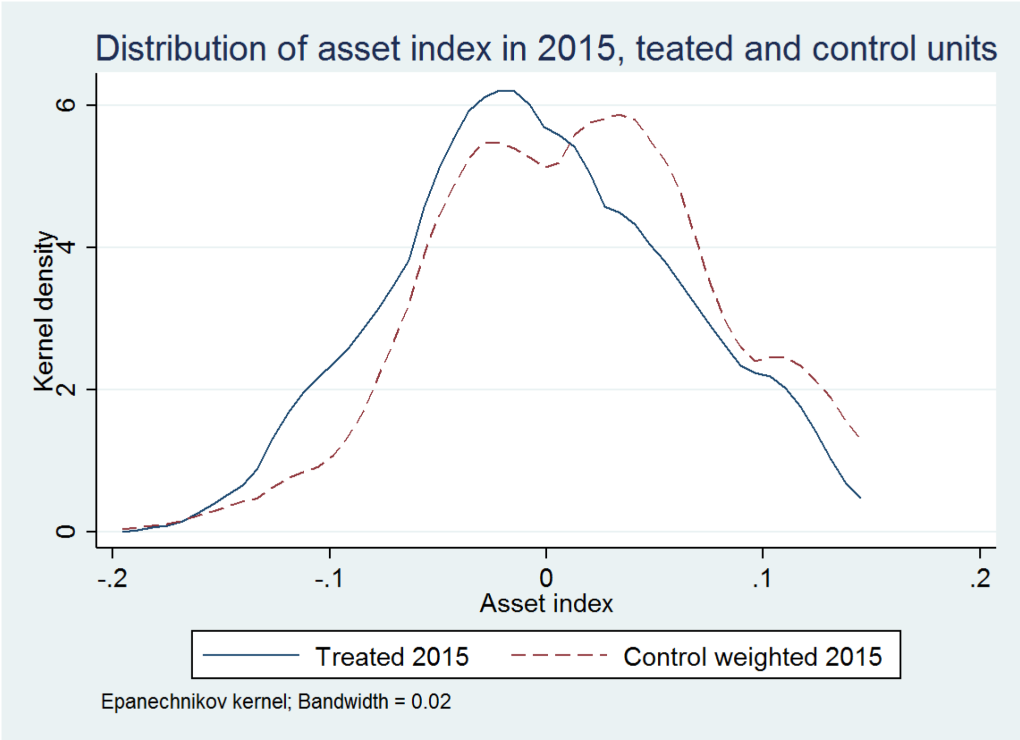

Figure 3.2: Asset index of treated and control households in 2015

3.6.3 Identification strategy

Several issues challenge the empirical identification of the impact of migration on households left behind.

First, we can think of factors that simultaneously affect both the migration decision and the outcome. For example, risk aversion of a household might prevent it from engaging in migration or in more profitable but riskier technologies in their farm or business. Hence, such households would be less likely to have a new migrant and would remain at a lower welfare level. Such omitted variables would bias the coefficient of interest. In the given example, we would overestimate a negative effect of having a new migrant. We cannot foresee the direction of the effect, but it would be biased upwards.

By modelling a fixed-effects specification, we capture any unobservable and time-invariant factors at the household level and the vector of household characteristics accounts for observable characteristics that vary over time. We assume that unobservable factors such as risk aversion are not varying over two years.

Secondly, the migration decision could be influenced by the outcome variable. This is especially a problem with cross-sectional data (Antman, 2012). The change in asset ownership in the period preceding our baseline could affect the treatment status of households. We cannot exploit previous data to control for this, but by balancing households on baseline characteristics we only compare those that look similar and thus capture any effect the prior welfare change had on households (see detailed discussion of the weighting method in section 3.6.3).

The specification in equation 3.1 assumes that new migrants are randomly allocated across treated and comparison households. Migration is however not a random process, but instead a strategic decision determined by observable (e.g. education, income) and unobservable (e.g. risk attitude, motivation) characteristics of the household. We will have to address the selection bias arising from this non-random treatment assignment.

Depending on the main drivers of selection of a new migrant, the bias could lead both ways, upwards or downwards. For example, the literature on self-selection finds that highly educated migrants tend to be positively selected and thus lead to overestimations of the outcomes of these migrants; the opposite applies to unskilled migrants (Borjas, 1987). In our context, not only migrant characteristics, but also household structure and prior migration experience are important determinants of the selection into having new migrants.

For unbiased identification, natural experiments (among others Gibson et al. (2011)) or randomized control trials (Bryan et al., 2014) are the ideal approach. Without such settings at hand, many researchers rely on either instrumental variables or matching approaches to reduce the issue of selection bias. Common instrumental variables for migration are historical road or rail networks that led to location-specific migrant networks (for example Woodruff and Zenteno (2007)). For Ghana, Adams and Cuecuecha (2013) rely on ethnic groups and their social networks that make it more or less likely for households to send migrants.

The common instruments could be used to predict whether households engage in migration or not, but less so to predict whether conditional on prior migration experience households have new migrants. We expect that this decision depends primarily on household-level characteristics. We therefore rely on a weighting method based on the assumptions of matching approaches.

The weighting method makes the comparison group look like the treated group in terms of observable characteristics at baseline. This approach assumes selection on observables. It means that conditional on observable characteristics, having a new migrant is as good as random (Wooldridge, 2010). This balance is achieved for observable characteristics that are expected to influence the likelihood to be a treated household and the outcome variable (Imbens, 2015). Once these observables are balanced, the selection bias is reduced (Heckman et al., 1998).

Entropy balancing weights

Conventional matching methods, such as Propensity Score Matching (PSM), require the researcher to define a specification for the propensity score, which then leads to balanced treatment and comparison groups after matching (Imbens, 2015). This process involves several stages and adjustments of the specification and sometimes the improvement of balance in one covariate goes hand in hand with worsening the balance of another covariate (Iacus et al., 2012). Due to this laborious process the matching method is prone to model dependence and researcher discretion (King and Nielsen, 2016). For example, in many cases only the means of matching variables between treated and control group are compared, not accounting for potential differences in the distribution of variables (Lee, 2013).

To simplify this process, some researchers have developed matching methods that achieve balance before the matching itself. One is the entropy balancing developed by Hainmueller (2012). This approach defines weights for each observation that ensure a predefined balance of covariates. The balance can be defined in terms of the first, second and even higher order moments of the covariates. The main advantages of this method are that balance checks become redundant, the majority of observations are retained, the computation of the weights is fast, and the method can be combined with many other matching and regression methods, similarly to inverse probability weighting methods and regression adjustment procedures (Imbens, 2015).

Entropy weights, w, minimise the entropy distance metric which is defined as:

$$\min_{w_i} H(w) = \sum_{i|D=0} w_i log(\frac{w_i}{q_i}) \tag{3.3}$$

and which is subject to balance (Equation 3.4), and normalizing constraints (Equations 3.5 and 3.6 respectively):

$$\sum_{i|D=0} w_i c_r i(X_i) = m_r \qquad \text{with} \qquad r \in 1, ..., R \qquad \text{and} \tag{3.4}$$

$$\sum_{i|D=0} w_i = 1 \qquad \text{and} \tag{3.5}$$

$$w_i \geq 0 \qquad \text{for all} \qquad i \qquad \text{such that} \qquad D = 0 \tag{3.6}$$

q_i is a base weight defined as 1 over the number of control units. $c_{ri}(X_i)$ "are a set of R balance constraints that are imposed on the covariate moments of the reweighted control group" (Hainmueller and Xu, 2013). Finally, it computes a set of weights that minimize the first equation (3.3) subject to the balance constraint, the normalisation constraint, and the non-negativity constraint.

The procedure is easily implemented in Stata using the command *ebalance*. The command first defines the first moment of the covariates using only the treated units. Then the control units are re-weighted so that their mean is equal to that of the treated units for the chosen covariates complying with the normalizing constraints (3.5 and 3.6). The same procedure applies to higher moments. It is important to note that one has to consider the sample at hand when using this method. Entropy balancing is a useful method only if the treated and control units do not look radically different and there can only be as many balance conditions as control observations. Like in other matching methods this implies the assumption of common support. Observations that make it impossible to achieve the balance defined by the researcher are dropped and weights are only computed for the remaining observations.[11]

Once the weights are computed, they are applied to estimate equation 3.1 with weighted least squares (WLS). This approach works like any Regression Adjustment method (Wooldridge, 2010).

Variables to balance

The decision which variables to include in the entropy balancing weight computation follows the suggestions about PSM by Imbens (2015). We include all variables that we consider substantive for having a new migrant or for the outcome. We also include squared terms of continuous variables. Smith and Todd (2005) stress the importance to include a rich set of such covariates, preferably past measures of the outcome variable and to ensure that one compares units within the same labour market or, more generally speaking, from the same geographical context.

Region dummies should capture any such factors that relate to migrant networks, regional development and economic opportunities. Most importantly, we control for the household size and dependency ratio of elderly and children to adult members to capture the household structure. These variables are important for the household decision about migration as well as the household's welfare. Another important characteristic is the main household income source, that is whether the household earns its living from agriculture, employment, its own business, public or private transfers. We also control for the employment status of the household head (employed, self-employed, unemployed or inactive) to capture economic activity. As a measure for human capital in the household, we include the highest level of education of adult members in the household. Many studies show that education is an important

[11] In our case, we drop 91 observations, 22 treated and 69 control households. Around a third of these are dropped due to missing values for some of the covariates that we required to be balanced. Others had extreme values for some covariates, e.g. a dependency ratio of 5.

predictor for households' welfare. It is also related to migration decisions as higher educated people have higher expected incomes at home as well as at possible destinations (Sjaastad (1962)). We include a dummy for female household heads, shown to be a strong predictor for household welfare in the rural context as well as reflecting a households' options for migration decisions (Adams and Cuecuecha, 2013). In addition, age and marital status of the household head are added to control for the life-cycle of a household (Lipton, 1980). Ethnicity was found to be an important factor in creating and maintaining migrant networks in Ghana (Awumbila et al., 2016). Such networks are important determinants for migration decisions as they reduce the risk and costs associated with migration (Carrington et al., 1996), which is why we include the ethnicity of the household head. We also include our measure of community employment rate. We choose this measure, because if a household seeks to diversify its income sources, it will also consider other opportunities in the community where household members could earn a wage (Bazzi, 2017).

Economic welfare is an important predictor for migration decisions and it is our outcome variable. In a credit constraint context, only households at a certain level of wealth are able to afford migration (McKenzie and Rapoport, 2007). Thus, only households with a similar level and distribution of welfare should be compared. While we do not have information on economic welfare pre-dating our baseline as suggested by Smith and Todd (2005), we include a rich set of asset indicators and information on asset purchases. Asset indicators are those that are used to construct the asset index. Asset purchase is a dummy that is equal to 1 if a household has purchased a specific asset within the past five years before the baseline survey, 0 otherwise.[12] In this way, we can capture a certain level of wealth and investment behaviour of the household that pre-dates the baseline survey.

[12] These assets are electric household goods, white household goods, livestock, generator, car, computer, electronic appliances, other investments, agricultural land, agricultural machinery, non-agricultural land, new house.

Balance statistics for treatment and control group

Here we present an overview of the balanced characteristics of treated and control households. The summary statistics provide evidence that the balance is achieved using the entropy weights. Figure 3.3 plots the kernel density of household size in 2013 for treatment and control group. The latter is represented once without applying the entropy balancing weights, and then with weighting.

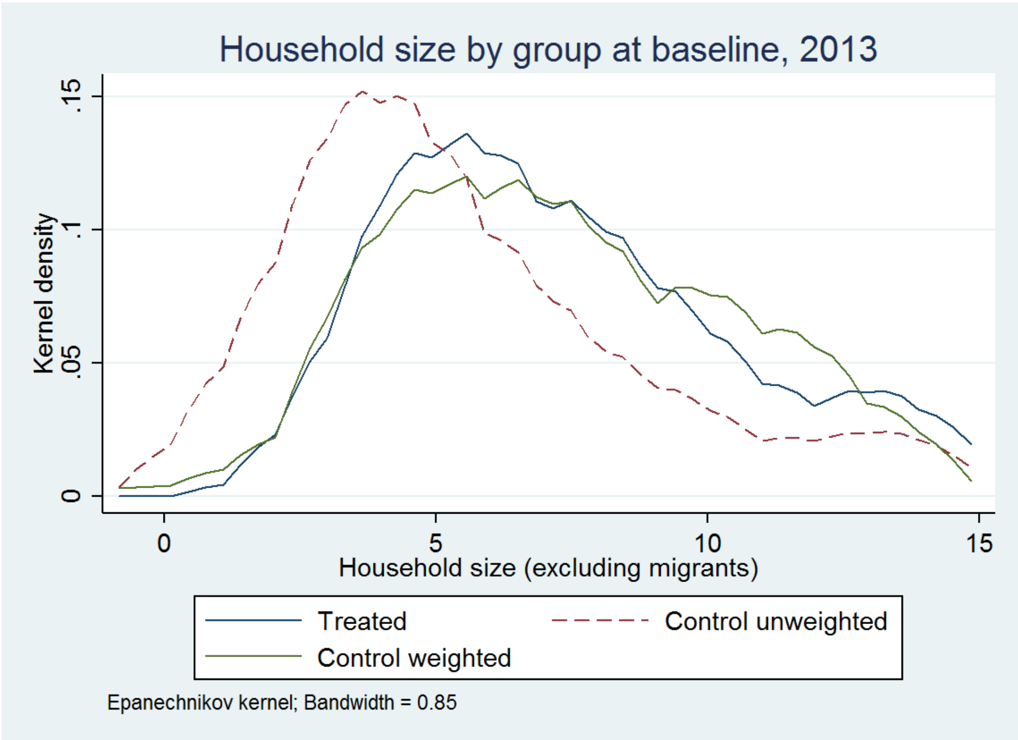

Figure 3.3: Kernel density of household size in 2013, by treatment groups

Without the weights, the dashed line shows a very different distribution. Control households are on average smaller than treatment households so that sending a new migrant is much more likely if there are more members that could make this choice. Thus, it is important to make households more comparable concerning this variable. The overlap between the treated distribution and the weighted control distribution confirm that the balance is achieved using the entropy weights.

In table 3.10, we show the mean and variance of the variables that were included in the construction of the entropy balancing weights with the weights applied to the control group. Using the weights leads to identical means of all variables and the variance is in some cases only slightly different. The last column lists the standardised differences between treated and control observations. They are all smaller than (+/-) 0.01. The entropy balancing weights construct a comparable sample of households to reduce the selection bias.

Even though we are not able to include the change in the outcome variable for the years before our analysis, we included information on the asset purchases within the two years prior to the baseline survey. Households purchased larger assets within a two year period preceding our survey. It is therefore plausible to expect also further changes in assets.

Table 3.10: First and second moments of covariates after applying entropy balancing weights, by group in 2013

	Mean		Variance		Standardised
	Treated	Control	Treated	Control	difference
Dependency ratio	0.660	0.658	0.846	0.844	0.002
Female household head	0.299	0.298	0.211	0.210	0.001
Highest level of education in household					
Primary	0.075	0.075	0.070	0.069	0.000
Middle/Junior	0.224	0.224	0.175	0.174	0.001
High/Senior	0.313	0.313	0.217	0.216	0.001
College/Technical	0.343	0.343	0.227	0.226	0.001
Ethnicity of head					
Akan	0.194	0.194	0.158	0.157	0.001
Ewe	0.194	0.194	0.158	0.157	0.000
Mole Dagbani	0.231	0.231	0.179	0.178	0.001
Main income source					
Private sector	0.052	0.052	0.050	0.050	0.000
Own business	0.269	0.268	0.198	0.197	0.001
Own farm	0.500	0.499	0.252	0.251	0.003
Private transfers	0.075	0.075	0.070	0.069	0.000
Others	0.030	0.030	0.029	0.029	0.000
Asset purchases in preceding 2 years					
Electronic goods	0.403	0.402	0.242	0.241	0.002
White goods	0.187	0.186	0.153	0.152	0.000
Livestock	0.284	0.283	0.205	0.204	0.001
Generator	0.022	0.022	0.022	0.022	0.000
Car	0.067	0.067	0.063	0.063	0.000
Computer	0.052	0.052	0.050	0.050	0.000

Continued on next page

Table 3.10 – continued

	Mean		Variance		Standardised
	Treated	Control	Treated	Control	difference
Electric Appliances	0.082	0.082	0.076	0.076	0.000
Other Investments	0.104	0.105	0.094	0.094	-0.001
Agricultural land	0.224	0.224	0.175	0.174	0.001
Agricultural machinery	0.022	0.022	0.022	0.022	0.000
Non-agricultural land	0.127	0.127	0.112	0.111	0.000
New house	0.313	0.313	0.217	0.216	0.001
Household size (excl. migrants)	7.299	7.280	9.640	9.615	0.006
Age of household head	55.276	55.136	218.021	217.450	0.009
Marital status					
Married/ living with partner	0.739	0.737	0.194	0.194	0.004
Separated/ Divorced/ Widowed	0.216	0.216	0.171	0.170	0.001
Employment status of head					
self employed	0.522	0.521	0.251	0.250	0.003
unpaid/unemployed	0.246	0.246	0.187	0.186	0.001
inactive etc.	0.090	0.090	0.082	0.082	0.000
Community employment rate	0.090	0.090	0.005	0.005	0.003
Household has returnee	0.246	0.246	0.187	0.186	0.001
Household receives remittances	0.545	0.543	0.250	0.249	0.003
Number of current migrants	2.090	2.084	1.842	1.837	0.004
Number of rooms (Base = 1)					
2	0.149	0.149	0.128	0.127	0.000
3	0.201	0.201	0.162	0.161	0.001
4	0.179	0.179	0.148	0.147	0.000
5 or more	0.425	0.424	0.246	0.245	0.002
Dwelling ownership(Base = Owned)					
Rented	0.119	0.119	0.106	0.105	0.000
Bathroom	0.403	0.402	0.242	0.241	0.002
Main source of drinking water (Base = pipe borne water inside)					
Pipe borne water outside	0.209	0.209	0.167	0.166	0.001
Borehole	0.343	0.343	0.227	0.226	0.001
Dug well	0.127	0.127	0.112	0.111	0.000
Tanker service	0.000	0.000	0.000	0.000	
Stream/river/lake	0.149	0.149	0.128	0.127	0.000
Rain water	0.007	0.007	0.007	0.007	0.000
Bottled or sachet water	0.052	0.052	0.050	0.050	0.000
Floor material(base = Polished wood)					
Mud	0.291	0.291	0.208	0.207	0.001
Raw wood, boards	0.000	0.000	0.000	0.000	
Cement/concrete	0.679	0.677	0.220	0.219	0.004

Continued on next page

Table 3.10 – continued

	Mean		Variance		Standardised
	Treated	Control	Treated	Control	difference
Burnt brick	0.015	0.015	0.015	0.015	0.000
Floor tile	0.007	0.007	0.007	0.007	0.000
Wall material (base = others)					
Bamboo or other organic materials	0.060	0.060	0.057	0.056	0.000
Cloth, cardboard, cans	0.022	0.022	0.022	0.022	0.000
Zinc	0.396	0.395	0.241	0.240	0.002
Mud, adobe, cane wall	0.493	0.491	0.252	0.251	0.002
Block, bricks, stone, prefabricated material, polished wood	0.030	0.030	0.029	0.029	0.000
Access to public services					
Electricity	0.634	0.633	0.234	0.233	0.003
Natural gas	0.142	0.142	0.123	0.122	0.000
Safe drinking water	0.694	0.692	0.214	0.214	0.004
Sewerage system	0.067	0.067	0.063	0.063	0.000
Garbage collection	0.112	0.112	0.100	0.100	0.000
Telephone	0.291	0.291	0.208	0.207	0.001
Region(Base = Brong Ahafo)					
Northern	0.142	0.142	0.123	0.122	0.000
Upper East	0.201	0.201	0.162	0.161	0.001
Upper West	0.134	0.134	0.117	0.117	0.000
Volta	0.224	0.224	0.175	0.174	0.001

3.7 Results

3.7.1 Main results

How does having a new migrant affect the asset welfare of households left behind conditional on prior migration experience? To answer this question we estimate weighted least squares regressions applying the entropy balancing weights. Table 3.11 presents the results. The coefficient of interest is the dummy variable of having a new migrant. This estimates the average effect on the change in the asset index for households with a new migrant between baseline and the follow-up survey compared to households without a new migrant.

In column 1, we show results without applying entropy balancing weights sug-

Table 3.11: Effect of having a new migrant on asset index, weighted least squares

	Asset index		
	(1)	(2)	(3)
New Migrant*2015	-0.011	-0.017	-0.016
	(0.007)	(0.011)	(0.011)
Household has return migrant (=1)			-0.015*
			(0.008)
Dependency ratio			0.002
			(0.004)
Employment status of household head (base = inactive/others)			
Employee			0.014
			(0.015)
Self-employed			-0.001
			(0.016)
Unpaid work / unemployed			-0.003
			(0.018)
Local employment rate			0.138
			(0.104)
Entropy balancing weights	No	Yes	Yes
Observations	*960*	*960*	*960*
Adjusted R-squared	0.584	0.522	0.528
Number of clusters	93	93	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level.

gesting that they might be biased due to selection. The effect of migration on household welfare could be driven by the fact that only households who are less likely to improve their welfare due to household characteristics sent a new migrant because of these same characteristics. We then apply balancing weights to the regression in column 2. The coefficient becomes larger but remains insignificant.

In column 3, time-varying household and local labour market characteristics are included that we consider relevant for the welfare of households. Of all control variables, only that indicating whether a household had a return migrant or not is significant.[13] Households are on average slightly worse off if they had a migrant return to their home.

The inclusion of time-varying covariates improves the precision of the estimates minimally, as indicated by a higher adjusted R-squared statistic. The coefficient of interest becomes minimally smaller. On average and everything else constant, sending a new migrant does not change the asset index of households significantly compared to those who do not send another migrant.

Next, we interact the number of new migrants and its squared term with the treatment dummy (see table 3.12). With this interaction we want to estimate the effect of the intensity of the treatment on the outcome. Around 40 percent of treated households have two or more new migrants. Thus, the effect might differ depending on the number of new migrants. Yet, there is again no significant effect when we allow for variation in the number of new migrants.

We now look further into the role of migrant characteristics. Table 3.13 lists the coefficients of the main estimation, each time interacting the treatment dummy with a migrant feature. These characteristics are whether the new migrant is female or whether they are seasonal migrants. Finally, we also differentiate between the effect of new migrants who move within the same region and those moving to another region.

[13]There might arise the concern that the measure of local employment is not well defined. When we drop this variable from the estimation, results remain unchanged (see appendix table C.4 on page 204).

Table 3.12: Number of new migrants, weighted least squares

	Asset index
Number of new migrants*2015	-0.008
	(0.009)
(Number of new migrants)^2*2015	0.001
	(0.002)
Entropy balancing weights	Yes
Other controls	Yes
Observations	*960*
Adjusted R-squared	0.525
Number of clusters	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level.
Other controls include whether the household has a returned migrant, employment status of the household head, dependency ratio and community employment rate.

Table 3.13: Interaction of treatment with the characteristics of new migrants

	Dependent variable: Asset index		
Migrant characteristics (X):	Female migrant	Seasonal migrant	Remained in region
New Migrant * X * 2015	-0.009	0.010	-0.013
	(0.011)	(0.014)	(0.021)
New Migrant * 2015	-0.010	-0.017	-0.005
	(0.014)	(0.012)	(0.022)
Entropy balancing weights	Yes	Yes	Yes
Other controls	Yes	Yes	Yes
Observations	*960*	*960*	*960*
Adjusted R-squared	0.528	0.528	0.528
Number of clusters	93	93	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level. Other controls include whether the household has a returned migrant, employment status of the household head, dependency ratio and community employment rate.

None of these interactions shows a significant effect on the asset index. There are three possible explanations for the fact that we do not find an impact of having a new migrant on households' asset index. One refers to the outcome variable used, one to the role of migration experience and the other to the sample investigated.

First, considering that asset indices are less volatile than for example consumption measures, it might be due to their stable nature that we do not find a significant effect in the short period of two years. We emphasise that the estimated effect is that of households sending a new migrant compared to those who do not. Hence, even a zero effect does not imply that there was no change in the asset index, but it means that the index of treated households changed in the same direction and magnitude as that of the control group. The distributional graphs of the welfare index (figure 3.1 and 3.2 in section 3.6.2) indicated some changes in the welfare of households. It appears, however, not to be significantly different between the groups once we control for observable and unobservable household characteristics. Booysen et al. (2008) also point out that because assets are more durable than other consumption goods, they tend to show an increase in asset wealth more than a reduction of the same. As our coefficients are negative, it is possible that we cannot find a significant effect due to this issue.

Secondly, we suggest that migration of a new migrant might be less costly than first-time migration. If we consider migration as an investment, then we would expect an initial decline in welfare and in the longer run an increase as suggested by Taylor and López-Feldman (2010). We do not observe that households with a new migrant experience a decline in welfare that could have been caused by the cost of migration and the loss of a working household member. In the descriptive statistics we saw that the average costs of migration for baseline migrants in 2013 was above 200 Ghanaian Cedis (in 2015 prices) compared to on average 120 Ghanaian Cedis for new migrants by 2015 (see table 3.4). This documents that costs for new migrants are relatively lower than for previous migrants.[14] Similar to the reduction of migration

[14]Using the information on previous migration we find that migrants who move the first time - independent of whether they are new or baseline migrants - pay on average more than those who

costs with the growth of social migrant networks, the migration experience at the household level itself can reduce costs of migration (McKenzie and Rapoport, 2007). This could be happening through similar channels, such as information transfer and family connections at the destination to find a job.

Another reason for not finding an effect might be that we are looking at the wrong sample. Some of the new migrants move for family reasons, such as marriage or joining other family members, while the majority moves for work. These reasons can have quite different implications for household welfare. We therefore estimate the effect of a new migrant including the interaction of the treatment with an indicator for those households whose new migrant moves for family reasons. Table 3.14 shows the results. They do not change neither for the main estimate, nor when we look at specific characteristics of the migrant, for example gender. All we observe is that the coefficient of the interaction that indicates households with a new migrant moving for family reasons is positive, while the overall treatment effect is negative. Both are however always insignificant.

Table 3.14: Having a new migrant by reason for migration, weighted least squares

	Dependent variable: Asset index			
Migrant characteristics:	All	Female migrant	Seasonal migrant	Remained in region
New Migrant * X * 2015		-0.011	0.011	-0.014
		(0.012)	(0.013)	(0.021)
New Migrant * 2015	-0.019	-0.012	-0.020	-0.006
	(0.012)	(0.014)	(0.012)	(0.023)
New Migrant moves	0.011	0.015	0.013	0.014
for family reason * 2015	(0.017)	(0.019)	(0.018)	(0.018)
Entropy balancing weights	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Observations	*960*	*960*	*960*	*960*
Adjusted R-squared	0.521	0.528	0.528	0.528
Number of clusters	93	93	93	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level. Other controls include whether the household has a returned migrant, employment status of the household head, dependency ratio and community employment rate.

moved the second time or more often (see appendix table C.2 on page 204).

We also conduct a Chow test of stable coefficients across the sub-samples of family reason and work reason. We cannot reject the hypothesis that the sample should remain pooled and we should not separately estimate the effect (see test statistics in appendix table C.3 on page 204).[15] The results presented here can be challenged concerning methodological concerns, which we address in the next section.

3.7.2 Sensitivity analysis

One concern is measurement error in the asset index. The measurement error could be even larger as it is a linear variable constructed from individual factor variables. In consequence, the estimates are still unbiased and consistent, but less precise which could explain the insignificant results (Wooldridge (2010), pp.287). We would be concerned if there was a reason to think that measurement error in the index was systematically related to the independent variables in our model.

We therefore estimate the main regression and exclude each time one component of the index to see how sensitive the results are to this.[16] We find stable results across index compositions presented in table 3.15.

[15] This tests whether all coefficients of the sub-sample with family migrants are equal to zero and should thus not be treated separately from the pooled sample.

[16] We also change the variables we determine to be balanced with the entropy balancing weights. Instead of the individual asset components, we include the asset index and its squared term at baseline. When we run our main regression using these weights, the results become weakly significant, but coefficient size only changes by 0.001 and is only significant at 10-percent level (see in appendix table C.5 on page 205).

Table 3.15: Sensitivity of results of asset index using different ways to construct the asset index, weighted least squares

	Dependent variable: Asset index						
	Exclude specific item from asset index construction						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Number of rooms	Dwelling ownership	Bathroom	Toilet	Drinking water	Floor material	Wall material
New Migrant * 2015	0.019	-0.017	-0.017	-0.015	-0.020	-0.013	-0.009
	(0.014)	(0.012)	(0.012)	(0.011)	(0.015)	(0.009)	(0.008)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entropy balancing weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	*960*	*960*	*960*	*960*	*960*	*960*	*960*
Adjusted R-squared	0.515	0.473	0.524	0.47	0.462	0.544	0.485
Number of clusters	93	93	93	93	93	93	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level. Other controls include whether the household has a returned migrant, employment status of the household head, dependency ratio and community employment rate.

3.7.3 Community shocks

One major concern challenging our identification strategy is that of unobserved shocks experienced by the households between the two survey waves. A shock could reduce household welfare and at the same time motivate people to leave their home or deter migration, as savings would be used to cover the damages of the shock instead of financing migration. This could affect whether we observe an impact of having a new migrant on welfare of households left behind.

In 2015, the enumerators interviewed village elders to collect information about the communities. These surveys included questions about shocks experienced by the village, and how many people were affected by it. The questions were asked open ended, so that the respondent could name any type of shock that s/he considered relevant. The most commonly named shocks are droughts, flooding or crop infestation by insects. We identified the communities where at least 50 percent of inhabitants were affected by such a shock.

In table 3.16, we present the results of the main specification, only that we include a dummy variable indicating a major shock at the community level and interact this with the treatment indicator. This interaction captures the impact of households that experienced a shock and have a new migrant in 2015.

The impact of having a new migrant on the asset index remains insignificant. Neither the coefficient of the shock variable nor its interaction with the treatment are significant. We note that there are fewer observations in these regressions due to missing values for the shock variables in six communities. We ran the main regression including a dummy for these communities. The dummy is positive and significant. On average, households in those communities for which we do not have any information about shocks, experience an increase in their asset index (see appendix table C.7 on page 206). We suggest that their missing information concerning shocks actually means that they did not experience any shock which could explain their higher asset index. If we include them in the estimation replacing their missing value of the shock with a zero, the main results are still insignificant (see appendix table C.7 on page 206).

Table 3.16: Effect of new migrant on household welfare controlling for major shocks in community, weighted least squares

	Asset index
New Migrant * 2015	-0.021
	(0.018)
New Migrant * Shock * 2015	0.015
	(0.023)
Shock	-0.018
	(0.017)
Entropy balancing weights	Yes
Other controls	Yes
Observations	*902*
Adjusted R-squared	0.521
Number of clusters	87

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level.
Other controls include whether the household has a returned migrant, employment status of the household head, dependency ratio and community employment rate.

After this test, one could still argue that an unobserved idiosyncratic shock at the household level interferes with our results. For example, a household would normally have experienced an increase in its asset index, but due to a negative shock interfered with this trajectory, for example a household member falling sick and not being able to earn income. Instead of investing in better walls or expanding the rooms of the house, the money is used to send another member as new migrant to find an income somewhere else or to pay for the medical bills. We document, however, in table 3.3 (page 117) that only in very few cases a new migrant had moved due to negative events, such as declining yields in agriculture, a family dispute, a flood or for medical treatment. Besides from lack of evidence that the reason of migration is an idiosyncratic shock, new migrants barely send remittances. If they had been sent to support the household through a crisis, one would expect regular remittances and maybe also higher amounts.

3.8 Conclusion

This paper documents the dynamic nature within households of internal migration in rural Ghana. Using a new dataset from 2013 and 2015, we show that many households with migrants at the baseline send a new migrant by 2015. Looking more closely at these migrants and their households, we provide an insight into the nature of such repeated migration. Within the same household, migrants move for different reasons, at different times and their connection with the origin household differs as well.

This motivates the question how households with prior migration experience are affected if they have a new migrant. There are hypotheses for positive, negative or no effect due to the variety of factors involved and their counteracting impacts.

We find that having a new migrant does not have an impact on the welfare measured with the asset index of origin households compared to those without a new migrant. We suggest that this is partially due to the stable nature of such an index over the short period of our analysis. In order to identify an impact, the households in our sample would have needed to invest in their housing to different amounts between treated and control group. However, their investment priorities might lie somewhere else, for example in their farm or business. Unfortunately, the questions about other forms of investment were not consistent between the two survey waves and those that were, had very low response rates so that we cannot provide an answer to this hypothesis.

Another insight we gain is that new migrants pay relatively less for their migration than baseline migrants. This indicates that migration becomes cheaper with the migration experience of the household so that a negative effect of migration incurred by moving costs might not materialize in this case. Furthermore, we observed that new migrants are in many aspects different from baseline migrants. Among the differences are for example the fact that new migrants are from a younger generation, coming straight from school and often not sending any remittances or only for special occasions. This also supports the zero effect we find for the asset index. Households with prior migration experience might not send a new migrant in expectation of future remittances and income diversification. Instead, the new migrants might move primarily to improve their own situation.

These unanswered hypotheses point at the limitations of this study. The effect we estimate is that of only two years or less since a new migrant left the household. The comparison of studies using longitudinal data from longer periods with those of short periods indicates that the positive returns to migration might only present itself after a certain period (Davis et al., 2010; Taylor and López-Feldman, 2010). More data collection is required to confirm our results over the longer run.

Conclusions

This thesis documents and analyses patterns of internal migration in two developing countries, Brazil and Ghana. Despite its important role for economic development data limitations restrict our understanding of the dynamics of internal migration at the individual, household and local level. I hence use the most recent nationally representative population survey in Brazil and a new household panel study primarily focused on migration in Ghana to examine causes and consequences of internal migration. Here, I summarise the main findings of each empirical chapter, discussing their shortcomings and offering an outlook for future research.

In the first empirical chapter, I asked why a substantial share of Brazilian workers moved out of metropolitan cities. This question was addressed with two empirical approaches. First, estimating the role of wage and price differences as well as local amenities for the destination choice found that smaller towns offer an attractive alternative to big cities due to their lower living costs. These smaller towns have been growing rapidly in the 2000's which also implies growing job opportunities. Secondly, the return to moving out of the metropolises is found to be positive in real wages based on a counterfactual wage comparison. This result was especially meaningful for low educated workers who experience a large loss in nominal wage terms, but gain significantly in real terms. Low educated workers form the largest share of migrants in Brazil. They seem to be pushed out of big cities due to high living costs.

An important policy implication of these findings is that affordable housing should be incorporated in plans that aim to manage the agglomeration in smaller towns. The congestion in large metropolitan cities and resulting high prices indicate that urbanization was not accompanied by any plans to address the challenges faced by the largest share of the work force, the low educated workers.

The chapter exploited a large and rich data set to provide evidence for a direction of movement that has not been studied before. Yet, the possibilities for further analysis with this data are limited. Due to its cross-sectional nature it is not possible to look at dynamic aspects of metropolitan out-migration. An aspect worth further research is the question whether the metropolitan migrants have incorrect or insufficient information to form their expectation about wages so that they do not choose the optimal destination with regards to nominal wages. It could be valuable to investigate how migration patterns have changed in the past decade also in other developing economies. Cases comparable to Brazil based on its size, economic development and regional diversity in economic performance could be India, Nigeria, South Africa or Indonesia.

Exploiting the Brazilian Census survey further, chapter 2 asked whether and how internal migration affects local crime rates. The finding is a positive effect of immigration on local homicide rates. Exploring possible channels of this effect, I provide indicative evidence that an increase in labour supply only leads to higher crime in specific labour market structures. It suggests that a large informal sector can usually absorb surplus labour supply due to its flexibility and that criminal inertia attracts more crime. Policy makers should therefore reflect on the rigidity of the formal sector as well as how to break the prevalence of crime.

To better understand underlying mechanisms and who actually commits the additional crime, longitudinal data at the individual level are required. Additionally, theoretical models aiming to explain the impact of immigration on destination labour markets or crime should incorporate an informal sector. This would be especially important for future research in developing countries where the informal sector is often very large.

Migration is often repeated within households in developing countries. Using a new household panel survey, the third empirical chapter documented that many households in Ghana have more than one migrant member. These migrants move at different times and for various reasons. Given that some households already have experience with migration, we ask, how new migrants differ from the previous migrants and which households repeat migration with another member. We describe these patterns of migration. Furthermore, we assess whether having a new migrant relates to changes in asset welfare. While we do not find an effect, the descriptive analysis offers some insights that could explain this result.

We draw a multi-faceted picture of migration based on the rich questionnaire of the survey. In many cases, though, response rates were low and it is not possible to fully exploit as much information as would be desirable. For example, an interesting question to ask would be, which factors of prior migration experience of the household and previous migrants determine the decision for a new migrant to move. Such factors could be whether past migrants found a job or send remittances. Another limitation is that the data only covers a period of two years, which might be too short to assess the full impact of migration on origin households. More data will be collected to extend the survey and we aim to exploit the additional information and longitudinal nature.

It is important to understand the reasons for internal migration at the individual and household level as well as its interaction with local labour markets for any policies that intend to reduce poverty or that aim at regional development. This thesis contributes to this understanding. Moreover, it offers the insight that more data collection including well-defined questions on migration and covering longer periods are essential to further our knowledge about geographic mobility.

Bibliography

Ackah, C. and D. Medvedev (2010). Internal Migration in Ghana: Determinants and Welfare Impacts. Policy Research Working Paper Series 5273, The World Bank, Washington, D.C. Cited on 107

Adams, R. H. (2006). Remittances and Poverty in Ghana. Policy Research Working Paper Series 3838, The World Bank, Washington, D.C. Cited on 21, 106

Adams, R. H. and A. Cuecuecha (2013). The Impact of Remittances on Investment and Poverty in Ghana. *World Development 50*, 24–40. Cited on 21, 107, 136, 140

Adams, R. H., A. Cuecuecha, and J. Page (2008). Remittances, consumption and investment in Ghana. Policy Research Working Paper Series 4515, The World Bank, Washington, D.C. Cited on 21, 101, 107

Aguayo-Téllez, E., M.-A. Muendler, and J. P. Poole (2010). Globalization and Formal Sector Migration in Brazil. *World Development 38*(6), 840–856. Cited on 20, 22

Almeida dos Reis, J. G. and R. Paes de Barros (1991). Wage inequality and the distribution of education. *Journal of Development Economics 36*(1), 117–143. Cited on 52

Antman, F. (2012). The impact of migration on family left behind. *International Handbook on the Economics of Migration* (6374), 1–34. Cited on 103, 136

Aroca Gonzalez, P. and W. F. Maloney (2005). Migration, Trade, and Foreign Direct Investment in Mexico. Policy Research Working Paper Series 3601, The World Bank, Washington, D.C. Cited on 20

Asselin, L.-M. (2009). Composite Indicator of Multidimensional Poverty. Technical Report January 2009, Institut de Mathématique C.F. Gauss, Quebec. Cited on 194

Awumbila, M., L. Boakye-Yiadom, E.-M. Egger, J. Litchfield, J. K. Teye, and C. Yeboah (2016). Gains and Losses from Internal Migration: Evidence from Migrant-Sending Households in Ghana. Working Paper 44, Migrating out of Poverty Research Program Consortium, Falmer. Cited on 140

Barham, B. and S. Boucher (1998). Migration, remittances, and inequality: estimating the net effects of migration on income distribution. *Journal of Development Economics 55*(2), 307–331. Cited on 21

Barnow, B. S., G. G. Cain, and A. S. Goldberger (1980). Issues in the analysis of selectivity bias. In E. Stromsdorfer and G. Farkas (Eds.), *Evaluation Studies*, Volume 5. San Francisco: Sage. Cited on 172

Barros, R. and C. Corseuil (2001). The impact of regulations on Brazilian labor market performance. Research Network Working Paper R-427, Inter-American Development Bank, Washington, DC. Cited on 97

Bartik, T. J. (1991). Who Benets From State and Local Economic Development Policies. Technical report, W.E. Upjohn Institute for Employment Research, Kalamazoo, Mich. Cited on 76, 77, 85

Bazzi, S. (2017). Wealth Heterogeneity and the Income Elasticity of Migration. *American Economic Journal: Applied Economics 9*(2), 219–255. Cited on 140

Becker, G. S. (1968). Crime and Punishment : An Economic Approach. *Journal of Political Economy 76*(2), 169–217. Cited on 16, 60, 65, 92, 98

Beegle, K., J. De Weerdt, and S. Dercon (2011). Migration and Economic Mobility in Tanzania Evidence from a Tracking Survey. *Review of Economics and Statistics 93*(3), 1010–1033. Cited on 103

Bell, B., S. Machin, and F. Fasani (2013). Crime and immigration: evidence from large immigrant waves. *The Review of Economics and Statistics 95*(4), 1278–1290. Cited on 17, 60, 61, 62, 64, 65, 75, 80

Bell, M. and S. Muhidin (2009). Cross-National Comparisons of Internal Migration. Technical Report 2009/30, United Nations Development Programme, New York City. Cited on 15, 99

Bianchi, M., P. Buonanno, and P. Pinotti (2012). Do immigrants cause crime? *Journal of the European Economic Association 10*(6), 1318–1347. Cited on 60, 61, 62, 64, 65, 67, 80, 84

Bilsborrow, R., A. Oberai, and G. Standing (1984). Migration surveys in low-income countries: guidelines for survey and questionnaire design. an ilo-wep study. In *Migration surveys in low-income countries: guidelines for survey and questionnaire design. An ILO-WEP study.* London: Croom Helm. Cited on 112

Blackwell, M., S. Iacus, G. King, and G. Porro (2009). cem: Coarsened exact matching in Stata. *The Stata Journal 9*(4), 524–546. Cited on 173

Booysen, F., S. van der Berg, R. Burger, M. von Maltitz, and G. du Rand (2008). Using an Asset Index to Assess Trends in Poverty in Seven Sub-Saharan African Countries. *World Development 36*(6), 1113–1130. Cited on 131, 148, 194, 198

Borges, D., D. Miranda, T. Duarte, F. Novaes, K. Ettel, T. Guimarães, and T. Ferreira (2012). Mortes Violentas No Brasil: Uma Análise do Fluxo de Informações. In *Homicidios no Brasil: Registro e Fluxo de Informacoes*, pp. 333–412. Brasília. Cited on 188

Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *The American Economic Review 77*(4), 531–553. Cited on 20, 136

Borjas, G. J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics 118*(4), 1335–1374. Cited on 60, 97

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*(2), 159–185. Cited on 20

Bound, J. and H. J. Holzer (2000, jan). Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s. *Journal of Labor Economics 18*(1), 20–54. Cited on 62, 76, 77

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 20*(4), 719–746. Cited on 59, 60, 62, 64, 84

Brown, R. P. and E. Jimenez (2008). Estimating the net effects of migration and remittances on poverty and inequality: Comparison of Fiji and Tonga. *Journal of International Development 20*, 547–571. Cited on 21

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Escaping Famine Through Seasonal Migration. *Econometrica 82*(5), 1671–1748. Cited on 100, 104, 105, 136

Caliendo, M. and S. Kopeinig (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys 22*(1), 31–72. Cited on 173

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Market Impacts of Higher Immigration. *Journal of Labor Economics 19*(1), 22–64. Cited on 60, 62, 76

Card, D. and T. Lemieux (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics 116*(2), 705–746. Cited on 97

Carrington, W. J., E. Detragiache, and T. Vishwanath (1996). Migration with Endogenous Moving Costs. *The American Economic Review 86*(4), 909–930. Cited on 140

Castaldo, A., P. Deshingkar, and A. Mckay (2012). Internal Migration, Remmitances and Poverty: Evidence from Ghana and India. Working Paper 7, Migrating out of poverty Research Programme Consortium, Falmer. Cited on 106

Cerqueira, D. (2012). Mortes violentas não esclarecidas e impunidade no Rio de Janeiro. Texto para discussão, Ipea, Brasília. Cited on 70, 188

Cerqueira, D. (2014a). Mapa dos Homicídios Ocultos no Brasil. Texto para Discussão 1884, Ipea, Brasília. Cited on 88, 188, 189

Cerqueira, D. R. d. C. (2014b). *Causas e consequências do crime no Brasil* (33. Prêmio BNDES de Economia ed.). Rio de Janeiro: BNDES. Cited on 64

Chalfin, A. (2014). What is the contribution of Mexican immigration to U.S. crime rates? Evidence from rainfall shocks in Mexico. *American Law and Economics Review 16*(1), 220–268. Cited on 64, 80

Chalfin, A. (2015). The Long-Run Effect of Mexican Immigration on Crime in US Cities: Evidence from Variation in Mexican Fertility Rates. *American Economic Review Papers and Proceedings 105*(5), 220–225. Cited on 60, 62, 64, 80

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio (2016). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India, and the United States. Technical Report 22002, National Bureau of Economic Research, Cambridge, MA. Cited on 68

Cheng, W., J. Morrow, and K. Tacharoen (2012). Productivity As If Space Mattered: An Application to Factor Markets Across China. CEP Discussion Papers 1181, Centre for Economic Performance, London. Cited on 20

Chimeli, A. B. and R. R. Soares (2011). The use of violence in illegal markets: Evidence from mahogany trade in the Brazilian Amazon. Technical report, IZA, Bonn. Cited on 66

Christiaensen, L., J. De Weerdt, and R. Kanbur (2017). Where to Create Jobs to Reduce Poverty. Cities or Towns? Cited on 20

Christiaensen, L., J. De Weerdt, and Y. Todo (2013). Urbanization and poverty reduction: the role of rural diversification and secondary towns. *Agricultural Economics 44*(4-5), 435–447. Cited on 24

Christiaensen, L. and Y. Todo (2014). Poverty Reduction during the Rural-Urban Transformation. The Role of the Missing Middle. *World Development 63*, 4358. Cited on 16

Cochrane, S. G. and D. R. Vining (1988). Recent Trends in Migration between Core and Peripheral Regions in Developed and Advanced Developing Countries. *International Regional Science Review 11*(3), 215–243. Cited on 19

Dahl, G. B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica 70*(6), 2367–2420. Cited on 20, 21, 58

Damon, A. L. (2010). Agricultural Land Use and Asset Accumulation in Migrant Households: the Case of El Salvador. *The Journal of Development Studies 46*(1), 162–189. Cited on 104

DataViva (2016). DataViva. http://dataviva.info. Cited on 24

Davis, B., G. Carletto, and P. C. Winters (2010). Migration, Transfers and Economic Decision Making among Agricultural Households: an Introduction. *The Journal of Development Studies 46*(1), 1–13. Cited on 155

De Brauw, A. (2010). Seasonal Migration and Agricultural Production in Vietnam. *The Journal of Development Studies 46*(1), 114–139. Cited on 105

De Brauw, A. and J. Giles (2012). Migrant Labor Markets and the Welfare of Rural Households in the Developing World: Evidence from China. IZA Discussion Paper Series 6765, Institute for the Study of Labor, Bonn. Cited on 103

De Brauw, A. and T. Harigaya (2007). Seasonal Migration and Improving Living Standards in Vietnam. *American Journal of Agricultural Economics 89*(2), 430–447. Cited on 103, 104, 105

De Brauw, A. and S. Rozelle (2008). Migration and household investment in rural China. *China Economic Review 19*(2), 320–335. Cited on 105

de Brito Ramalho, H. M. and V. dos Santos Queiroz (2011). Migração interestadual de retorno e autoseleção: Evidências para o Brasil. *Pesquisa e Planejamento Econômico 41*(3), 369–396. Cited on 174

De Castro Cerqueira, D. R. and J. M. Pinho De Mello (2012). Menos Armas, Menos Crimes. Technical Report 1721, IPEA, Brasília. Cited on 66

de Oliveira, L. A. P. and A. T. R. de Oliveira (2011). Reflexões sobre os deslocamentos populacionais no Brasil. Technical report, Instituto Brasileiro de Geografia e Estatística - IBGE, Rio de Janeiro. Cited on 20

De Vreyer, P. and G. Spielvogel (2009). Spatial externalities between Brazilian municipios and their neighbours. Document de Travail DIAL 2005-11, DIAL - Developpement Institutions & Analyses de long terme, Paris. Cited on 22

Deaton, A. and O. Dupriez (2011). Spatial price differences within large countries. Working Papers 1321, Princeton University, Woodrow Wilson School of Public and International Affairs, Princeton, NJ. Cited on 38

Demombynes, G. and B. Özler (2005). Crime and local inequality in South Africa. *Journal of Development Economics 76*(2), 265–292. Cited on 64

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *The American Economic Review 106*(3), 479–524. Cited on 62, 76, 77

Dix-Carneiro, R., R. R. Soares, and G. Ulyssea (2017). Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization. Technical Report 23400, National Bureau of Economic Research, Cambridge, MA. Cited on 62, 64, 70, 75

dos Santos Júnior, E. d. R., N. Menezes-Filho, and P. C. Ferreira (2005). Migração, Seleção e Diferenças Regionais de Renda no Brasil. *Pesquisa e Planejamento Econômico 35*(3), 299–331. Cited on 22

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial Sorting. *Journal of Political Economy 122*(3), 554–620. Cited on 20

Enamorado, T., L. F. López-Calva, C. Rodríguez-Castelán, and H. Winkler (2016). Income inequality and violent crime: Evidence from Mexico's drug war. *Journal of Development Economics 120*, 128–143. Cited on 64

Fafchamps, M. and F. Shilpi (2013). Determinants of the Choice of Migration Destination. *Oxford Bulletin of Economics and Statistics 75*(3), 388–409. Cited on 20, 36, 37

Fajnzylber, P., D. Lederman, and N. Loayza (2002). What Causes Violent Crime? *European Economic Review 46*, 1323–1357. Cited on 63, 93

Fally, T., R. Paillacar, and C. Terra (2010). Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics 91*(1), 155–168. Cited on 22

Fay, M. and C. Opal (2000). Urbanization without Growth: A Not-So-Uncommon Phenomenon. Policy Research Working Paper Series 2412, Washington, D.C. Cited on 19

Fernandes, R. C. and M. de Sousa Nascimento (2007). Mapping the Divide. Firearm Violence and Urbanisation in Brazil. In Keith Krause (Ed.), *Small Arms Survey 2007: Guns and the City*, Chapter 7, pp. 226–255. Geneva: Small Arms Survey. Cited on 65, 66

Ferreira, F. H. G., P. G. Leite, and J. A. Litchfield (2006). The Rise and Fall of Brazilian Inequality: 1981-2004. Policy Research Working Paper Series 3867, Washington, D.C. Cited on 52

Fields, G. S. (1975). Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs. *Journal of Development Economics 2*(2), 165–187. Cited on 97

Filmer, D. and L. H. Pritchett (2001). Estimating Wealth Effects Without Expenditure Data or Tears: An Application to Educational Enrollments in States of India. *Demography 38*(1), 115–132. Cited on 131, 193

Foguel, M. N., I. Gill, R. Mendonça, and R. Paes De Barros (2015). The Public-Private Wage Gap in Brazil. Ipea Discussion Paper 95, Ipea, Brasília. Cited on 52

Forum Brasileiro de Segurança Pública (2014). Anuário Brasileiro de Segurança Pública 2014. Technical report, Forum Brasileiro de Segurança Pública, São Paulo. Cited on 62

Fu, Y. and S. A. Gabriel (2012). Labor migration, human capital agglomeration and regional development in China. *Regional Science and Urban Economics 42*(3), 473–484. Cited on 20

Ghana Statistical Service (2013). 2010 Population and Housing Census. National Analytical Report. Technical report, Ghana Statistical Service, Accra. Cited on 99, 111

Ghana Statistical Service (2015a). Consumer Price Index (CPI) April 2015. Technical report, Ghana Statistical Service, Accra. Cited on 199

Ghana Statistical Service (2015b). Consumer Price Index (CPI) May 2015. Technical report, Ghana Statistical Service, Accra. Cited on 199

Giannetti, M. (2003). On the mechanics of migration decisions: skill complementarities and endogenous price differentials. *Journal of Development Economics 71*, 329–349. Cited on 19, 21

Gibson, J., D. McKenzie, and S. Stillman (2011). The Impacts of International Migration on Remaining Household Members: Omnibus Results from a Migration Lottery Program. *The Review of Economics and Statistics 93*(4), 1297–1318. Cited on 101, 104, 136

Gibson, J., D. McKenzie, and S. Stillman (2013). Accounting for selectivity and duration-dependent heterogeneity when estimating the impact of emigration on incomes and poverty in sending areas. *Economic Development and Cultural Change 61*(2), 247–280. Cited on 104

Giulietti, C., J. Wahba, and Y. Zenou (2014). Strong versus Weak Ties in Migration. Technical Report 8089, IZA Institute for the Study of Labor, Bonn. Cited on 105

Gould, E. D., B. a. Weinberg, and D. B. Mustard (2002). Crime Rates and Local Labor Market Opportunities in the United States: 19791997. *Review of Economics and Statistics 84*(1), 45–61. Cited on 63

Greene, W. (2000). *Econometric analysis* (4th ed.). London: Prentice-Hall International (UK). Cited on 37, 42

Grogger, J. and G. H. Hanson (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics 95*(1), 42–57. Cited on 20

Grogger, J. and M. Willis (1998). The Introduction of Crack Cocaine and the Rise in Urban Crime Rates. Nber working paper series, National Bureau of Economic Research, Cambridge, MA. Cited on 63

Guriev, S. and E. Vakulenko (2015). Breaking out of poverty traps: Internal migration and interregional convergence in Russia. *Journal of Comparative Economics 43*, 633–649. Cited on 82

Haanwinckel, D. and R. R. Soares (2016). Workforce Composition, Productivity, and Labor Regulations in a Compensating Differentials Theory of Informality. Technical Report 9951, IZA Institute for the Study of Labor, Bonn. Cited on 93

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis 20*(25-46). Cited on 101, 137

Hainmueller, J. and Y. Xu (2013). ebalance: A Stata Package for Entropy Balancing. *Journal of Statistical Software 54*(7). Cited on 138

Ham, J. C., X. Li, and P. B. Reagan (2011). Matching and semi-parametric IV estimation, a distance-based measure of migration, and the wages of young men. *Journal of Econometrics 161*(2), 208–227. Cited on 20

Harris, J. R. and M. P. Todaro (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review 60*, 126–142. Cited on 15, 100

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an Econometric Evaluation Estimator. *Review of Economic Studies 65*, 261–294. Cited on 137

Henderson, J. V. (1986). Urbanization in a developing country. *Journal of Development Economics 22*(2), 269–293. Cited on 20

Hering, L. and R. Paillacar (2015). Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil. Policy research working paper series, The World Bank, Washington, D.C. Cited on 52

Hidalgo, F. D., S. Naidu, S. Nichter, and N. Richardson (2010). Economic Determinants of Land Invasions. *Review of Economics and Statistics 92*(3), 505–523. Cited on 66

Hirvonen, K. (2016). Temperature Changes, Household Consumption, and Internal Migration: Evidence from Tanzania. *American Journal of Agricultural Economics 98*(4), 1230–1249. Cited on 82

Iacus, S. M., G. King, and G. Porro (2011). Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association 106*(493), 345–361. Cited on 173, 175

Iacus, S. M., G. King, and G. Porro (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis 20*(1), 1–24. Cited on 137

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics 86*(1), 4–29. Cited on 172

Imbens, G. W. (2015). Matching Methods in Practice: Three Examples. *Journal of Human Resources 50*(2), 373–419. Cited on 137, 138, 139

Institute of Statistical, Social, and Economic Research (ISSER),University of Ghana and Economic Growth Center (EGC), Yale University (2015). Ghana Socioeconomic Panel Survey 2009 - 2010. Technical report, Accra. Cited on 131

Instituto Brasileiro de Geografia e Estatística (IBGE) (2010). Censo Demográfico 2010: Características urbanísticas do entorno dos domicílios. Technical report, Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro. Cited on 30

Instituto Brasileiro de Geografia e Estatística (IBGE) (2012). Censo Demográfico 2010: Microdados. www.ibge.gov.br. Cited on 23

Jaitman, L. and S. Machin (2013). Crime and immigration : new evidence from England and Wales. *IZA Journal of Migration 2*(19), 1–23. Cited on 62, 80

Kelly, M. (2000). Inequality and Crime. *Review of Economics and Statistics 82*(4), 530–539. Cited on 63

Kennan, J. and J. R. Walker (2011). The Effect of Expected Income on Individual Migration Decisions. *Econometrica 79*(1), 211–251. Cited on 21, 58

King, G. and R. Nielsen (2016). Why Propensity Scores Should Not Be Used for Matching. Technical report, Cambridge, MA. Cited on 137

Kleemans, M. and J. Magruder (2017). Labor Market Changes In Response To Immigration: Evidence From Internal Migration Driven By Weather Shocks. *The Economic Journal (forthcoming).* Cited on 17, 60, 61, 62, 63, 65, 67, 68, 74, 79, 90, 97

Lall, S. V., C. Timmins, and S. Yu (2009). Connecting Lagging and Leading Regions The Role of Labor Mobility. Policy Research Working Paper Series 4843, The World Bank. Cited on 20, 22, 59

Lechner, M. (2000). An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany. *The Journal of Human Resources 35*(2), 347. Cited on 172

Lee, W.-S. (2013). Propensity score matching and variations on the balancing test. *Empirical Economics 44*, 47–80. Cited on 137

Li, C. and J. Gibson (2014). Spatial Price Differences and Inequality in the Peoples Republic of China: Housing Market Evidence. *Asian Development Review 31*(1), 92120. Cited on 47

Lin, M.-J. (2008). Does Unemployment Increase Crime?: Evidence from U.S. Data 19742000. *Journal of Human Resources 43*(2), 413–436. Cited on 60

Lipton, M. (1980). Migration from Rural Areas of Poor Countries: The Impact on Rural Productivity and Income Distribution. *World Development 8*, 1–24. Cited on 140

Litchfield, J. and H. Waddington (2003). Migration and Poverty in Ghana : Evidence from the Ghana Living Standards Survey. Sussex Migration Working Paper 10, Poverty Research Unit at Sussex, Falmer. Cited on 99, 101, 106

Lokshin, M., M. Bontch-Osmolovski, and E. Glinskaya (2007). Work-related Migration and Poverty Reduction in Nepal. Policy Research Working Paper Series 4231, The World Bank, Washington, D.C. Cited on 20, 21

Mahé, C. and W. Naudé (2016). Migration, occupation and education: Evidence from Ghana. UNU-MERIT Working Papers 2016-018, UNU-MERIT, Maastricht. Cited on 107

Manacorda, M. and M. F. Koppensteiner (2015). Violence and Birth Outcomes : Evidence from Homicides in Brazil. Iza discussion papers, Institute for the Study of Labor, Bonn. Cited on 62

Mariano Bosch, Edwin Goni, W. M. (2007). The Determinants of Rising Informality in Brazil: Evidence from Gross Worker Flows. IZA Discussion Paper Series 2970, IZA Institute for the Study of Labor, Bonn. Cited on 97

Mata, D. D., U. Deichmann, J. V. Henderson, S. V. Lall, and H. G. Wang (2005). Examining the growth patterns of Brazilian cities. Technical Report 3724, The World Bank, Washington, D.C. Cited on 19, 59

Matano, A. and P. Naticchioni (2012). Wage distribution and the spatial sorting of workers. *Journal of Economic Geography 12*(2), 379–408. Cited on 20

McCormick, B. and J. Wahba (2005). Why do the young and educated in LDCs concentrate in large cities? Evidence from migration data. *Economica 72*(285), 39–67. Cited on 20

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics 3*(4), 303–328. Cited on 36

McKenzie, D. and H. Rapoport (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics 84*, 1–24. Cited on 100, 140, 149

McKenzie, D. and M. J. Sasin (2006). Migration, Remittances, Poverty, and Human Capital: Conceptual and Empirical Challenges. Policy research working paper, World Bank, Washington, D.C. Cited on 88

McKenzie, D. J. (2005). Measuring inequality with asset indicators. *Journal of Population Economics 18*(2), 229–260. Cited on 131

Mendola, M. (2008). Migration and technological change in rural households: Complements or substitutes? *Journal of Development Economics 85*, 150–275. Cited on 104

Mendola, M. (2012). Rual Out-Migration and Economic Development at Origin: A Review of the Evidence. *Journal of International Development 24*, 102–122. Cited on 103, 129

Menezes-Filho, N. A. and M.-A. Muendler (2011). Labor Reallocation in Response to Trade Reform. NBER Working Paper Series 17372, NBER, Cambridge, MA. Cited on 52

Ministério do Planejamento (2010). Balanço 4 Anos 2007 - 2010. PAC Programa de Aceleração do Crescimento. Report, Ministério do Planejamento, Brasília. Cited on 20

Mion, G. and P. Naticchioni (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics 42*(1), 28–55. Cited on 20

Molini, V., D. Pavelesku, and M. Ranzani (2016). Should I Stay or Should I Go? Internal Migration and Household Welfare in Ghana. Policy Research Working Paper Series 7752, World Bank, Washington, D.C. Cited on 99, 106, 107

Monras, J. (2015). Economic Shocks and Internal Migration. Discussion paper series, IZA Institute for the Studies of Labor, Bonn. Cited on 59, 62, 76, 77

Monteiro, J. and R. Rocha (2017). Drug Battles and School Achievement: Evidence from Rio de Janeiro's Favelas. *The Review of Economics and Statistics 99.* Cited on 62

Moretti, E. (2011). Local Labor Markets. In *Handbook of Labor Economics, Volume 4b*, Chapter 14, pp. 1237–1313. Amsterdam: Elsevier. Cited on 20, 21, 22, 58

Morten, M. and J. Oliveira (2016). Paving the Way to Development: Costly Migration and Labor Market Integration. NBER Working Paper Series 22158, National Bureau of Economic Research, Cambridge, MA. Cited on 38, 49, 62, 76, 82

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics May*, 549–599. Cited on 79, 80, 100

Murray, J., D. R. D. C. Cerqueira, and T. Kahn (2013). Crime and violence in Brazil: Systematic review of time trends, prevalence rates and risk factors. *Aggression and Violent Behavior 18*(5), 471–483. Cited on 70, 188

Notowidigdo, M. J. (2013). The Incidence of Local Labor Demand Shocks. Technical report, University of Chicago Booth School of Business and NBER. Cited on 62, 76, 82

Özden, c., M. Testaverde, and M. Wagner (2017). How and Why Does Immigration Affect Crime? Evidence from Malaysia. *The World Bank Economic Review* (lhx010), 1–20. Cited on 62, 64, 80, 84

Quisumbing, A. and S. McNiven (2010). Moving Forward, Looking Back: the Impact of Migration and Remittances on Assets, Consumption, and Credit Constraints in the Rural Philippines. *Journal of Development Studies 46*(1), 91–113. Cited on 103

Ravallion, M. (2011). On Multidimensional Indices of Poverty. Policy research working paper series, The World Bank, Washington, D.C. Cited on 131

Reichenheim, M. E., E. R. De Souza, C. L. Moraes, M. H. P. De Mello Jorge, C. M. F. P. Da Silva, and M. C. De Souza Minayo (2011). Violence and injuries in Brazil: The effect, progress made, and challenges ahead. *The Lancet 377*(9781), 1962–1975. Cited on 62, 66, 91

Ribeiro Justo, W., R. de Alencar Ferreira, C. F. de Lima, and G. Nunes Martins (2010). Os determinantes da migração e da migração de retorno intermunicipal no Brasil. In *Proceedings of the 38th Brazilian Economics Meeting, ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia, 163.* Cited on 174

Rodriguez, E. R. (1998). International Migration and Income Distribution in the Philippines. *Economic Development and Cultural Change 46*(2), 329–350. Cited on 21

Rosenbaum, P. R. and D. B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*(1), 41–55. Cited on 172

Sachsida, A. (2013). Evolucao e determinanted da taxa de homicidios no Brasil. Texto para discussão, Ipea, Brasília. Cited on 62, 64

Sachsida, A., M. J. C. de Mendonça, P. R. A. Loureiro, and M. B. S. Gutierrez (2010). Inequality and criminality revisited: Further evidence from Brazil. *Empirical Economics 39*(1), 93–109. Cited on 64

Sahn, D. and D. Stifel (2000). Poverty comparison over time and across countries. *World Development 28*(12), 2123–2155. Cited on 131, 194

Santos, C. and P. C. Ferreira (2007). Migração e distribuição de renda no Brasil. *Pesquisa e Planejamento Econômico 37*(3), 405–426. Cited on 22

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *The Journal of Political Economy 70*(2), 80–93. Cited on 15, 21, 38, 129, 140

Smith, J. A. and P. E. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics 125*(1), 305–353. Cited on 139, 140

Soares, R. R. (2004). Crime reporting as a measure of institutional development. *Economic Development and Cultural Change 52*(4), 851–871. Cited on 190

Spenkuch, J. L. (2011). Understanding the impact of immigration on crime. MPRA Paper 31171, Universität München, Munich. Cited on 60, 61, 62, 64, 80

Stark, O. and D. E. Bloom (1985). The New Economics of Labor Migration. *The American Economic Review 75*(2), 173–178. Cited on 15, 100, 103, 129

Sugiyarto, E. and J. Litchfield (2016). Migrating Out of Poverty Ghana Household Survey 2013 User Guide. Technical report, Migrating out of Poverty Research Programme Consortium, Falmer. Cited on 111

Taylor, J. E. (1999). The new economics of labour migration and the role of remittances in the migration process. *International migration (Geneva, Switzerland) 37*(1), 63–88. Cited on 129

Taylor, J. E. and A. López-Feldman (2010). Does Migration Make Rural Households More Productive? Evidence from Mexico. *Journal of Development Studies 64*(1), 68–90. Cited on 102, 104, 128, 148, 155

Tunalı, I. (2000). Rationality of Migration. *International Economic Review 41*(4), 893–920. Cited on 21, 58

Ulyssea, G. (2010). Firms, Informality and Development: Theory and evidence from Brazil. *The American Economic Review Resubmitted.* Cited on 97

Woodruff, C. and R. Zenteno (2007). Migration networks and microenterprises in Mexico. *Journal of Development Economics 82*, 509–528. Cited on 136

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2 ed.). Cambridge, MA; London: The MIT Press. Cited on 137, 139, 150

World Bank (2006). Crime, Violence and Economic Development in Brazil: Elements for Effective Public Policy. Technical Report 36525, World Bank, Washington, D.C. Cited on 62

World Bank (2016). World Development Indicators. http://data.worldbank.org/. Cited on 71

World Bank (2017). Country Profile Ghana. http://www.worldbank.org/en/country/ghana. Cited on 106

Yang, D. (2008). International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks. *The Economic Journal 118*(April), 591–630. Cited on 104

Yap, L. (1976). Internal Migration and Economic Development in Brazil. *The Quarterly Journal of Economics 90*(1), 119–137. Cited on 22, 25

Appendix A

Appendix to Chapter 1

A.1 Matching

A.1.1 Coarsened exact matching (CEM)

Matching cannot eliminate the selection bias, but it can substantially reduce it. For a valid matching estimator, two main assumptions have to be made: First, the Conditional Independence Assumption (CIA, Lechner (2000))[1] which implies that conditional on observable characteristics, the migrants' wages had they not moved would be the same as those of the non-migrants at origin. Secondly, the assumption of common support which means that for each treated unit there must be at least one untreated unit to match.

For the ATT, the CIA can be relaxed to just the control group. The matching has to be based on variables that affect both, the migration decision M and the outcome Y, but which are in turn not affected by the treatment M. The outcome of non-migrants has to be independent of the treatment.

$$(Y^0) \amalg M \mid X \tag{A.1}$$

Thus, the conditioning variables included in X should be either time-invariant,

[1]The CIA was first established by Rosenbaum and Rubin (1983) named "ignorable treatment assignment", it is often also referred to as "unconfoundedness" (Imbens, 2004) or "selection on observables" (Barnow et al., 1980).

such as gender and race, or measured before the treatment, e.g. education or age (Caliendo and Kopeinig, 2008). With respect to the outcome, potential equilibrium effects in the labour market due to labour mobility should be considered. In the case of migrants moving out of the big metropolitan areas of Brazil it seems valid to assume that the relatively small share of out-migration should not affect the wages of non-migrants.

Coarsened exact matching (CEM) was developed as reaction to the balancing problem of commonly used propensity score matching (PSM) methods (Blackwell et al., 2009). The exact matching on coarsened variables, e.g. by education group instead of years of education, allows the researcher to choose the balance resulting from matching ex ante. The choice of imbalance for one variable does not affect imbalance of others. Thus, CEM reduces the error in estimating the average treatment effect due to selection bias and it bounds the degree of model dependence in the main analysis and the data is automatically restricted to common support. The large dataset of the Census at hand allows for this exact and full matching, not having to fight with the trade-off between bias and variance as with PSM.

More general, CEM belongs to the Monotonic Imbalance Bounding Class (MIB) (Iacus et al., 2011). Matching methods belonging to this class are defined to make no assumptions about the data and therefore suitable to use for observational data. The focus of MIB lies on in-sample imbalance in order to reduce model dependence. The level of imbalance is chosen before the matching and thus the matched sample results from the applied algorithm. Imbalance is bounded and it is generalized from mean imbalance to apply to any function. This implies that not only the mean, but other moments or interactions of covariates are balanced if the researcher considers them important. Most importantly, as a consequence of these properties, changing the bound for one variable will not affect the balance of the other covariates.

CEM fulfils these properties and is applied in three steps. First, covariates for matching are coarsened according to the researcher's discretion. Coarsening simply means to define broader categories of information than the exact data provides. For example, instead of matching on age in months, I match in age groups of 10 years because in the analysis of workers' wages, the difference of one or two months or even years should not matter too much. Secondly, the coarsened variables are matched exactly by each category and observations are allocated in strata within which the covariate takes the exact same value as defined by the coarsening. Finally, strata that only comprise of control units are dropped and the remaining strata containing treated and control units are used to compute weights for each observation. Let T^s be a treated unit in stratum s. Then m_T^s is the number of treated units in stratum s and matched units are $m_T = \sum_{s \in S} m_T^s$; C^s, m_C^s, and $m_C = \sum_{s \in S} m_C^s$ for control units respectively. Unmatched observations receive a weight equal to zero. Weights for treated units are 1, and for control units they are

$$w_i = m_C/m_T * m_T^s/m_C^s \tag{A.2}$$

The weights are essentially the ratio of treated to control units within each stratum normalized by the overall ratio of treated to control units in the matched sample. These weights are then applied to predict the counterfactual wages of metropolitan out-migrants as described next.

A.1.2 Application of CEM

The sample used for the wage predictions in metropolitan origins consists of 683,517 non-migrant residents in the metropolitan cities of origin and 16,172 migrants who have moved within the past year before the survey from one of the 22 metropolitan cities to any of the 529 non-metropolitan *microregiões*. Migrants are men and women who live and work now in a different *microregião* than one year before and different than their place of birth. In this way, it is possible to exclude return migrants[2] or commuters to the city of former residence or to a close-by metropolitan city.

One of the issues without panel data at hand is that I cannot know exactly what

[2]Return migrants are expected to have a different motivation for their movement due to different expectations and information about their destination (for Brazil de Brito Ramalho and dos Santos Queiroz (2011); Ribeiro Justo et al. (2010).

the migrants' characteristics were before they moved. However, by looking only at most recent migrants, many characteristics are very unlikely to have changed through migration and can thus be used for matching. The variables used for the matching are sex, age (25 to 65 years in groups of 10 years), race (white non-white), education level, industry and city of origin. Age of migrants and non-migrants at time of migration for the matching can be calculated by simply subtracting one year from their current age. Education is measured in levels of none, fundamental, middle and higher education. As the sample is restricted to workers who are at least 25 years and currently not in education, I can assume that their level of education has not changed since they moved.

The city of origin, in which workers lived, captures local labour market characteristics as well as overall living conditions, which are not observed in specific detail in retrospective, but which are unlikely to have changed dramatically within one year.

I test the initial balance of the unmatched sample using the multivariate balance statistic suggested by Iacus et al. (2011). This measure compares the distributions of several covariates between treated and control group. It not only compares the mean, but the whole distribution and it does so in a multivariate setting because even if the mean of each covariate is not statistically different between treated and control units, the multivariate distribution might still be imbalanced.

In short, the multivariate imbalance measure compares multivariate histograms by comparing each bin. A multivariate distance, L_1, of 1 implies that the distributions between treated and control are completely separate, while $L_1 = 0$ implies complete overlap. The measure is relative, so that we first present the imbalance statistic of the unmatched sample as point of comparison (see figure A.1).

The first column of the table lists the univariate L_1 distance for each variable, the second column the average difference, while the next columns look at the differences along the distribution of the variable. From this table, we can see that the variable age and industry ("activ_group") show some imbalance along their distribution.

```
Multivariate L1 distance: .83541258

Univariate imbalance:

                  L1      mean       min       25%       50%       75%       max
     agemig   .10634   -2.2895        -1        -1        -3        -3        -1
        sex    .0979    -.0979         0         0         0         0         0
 educ_level   .07352   -.12922         0        -1         0         0         0
      white   .00576    .00576         0         0         0         0         0
       city   .14503    .72656         0         0         4         0         0
    marital   .01909    -.0098         0         0         0         0         0
activ_group   .15388   -.96689         0        -3        -1         0        -1
```

Figure A.1: Balancing statistics for unmatched sample

Some differences remain also for the metropolitan city and the education level.

Matching results are presented below in figure A.2.

```
Matching Summary:
-----------------
Number of strata: 719
Number of matched strata: 474

                 0        1
      All   683517    16172
  Matched   680706    15424
Unmatched     2811      748


Multivariate L1 distance: .4590945

Univariate imbalance:

                 L1      mean       min       25%       50%       75%       max
    agemig   .03923   -.14602         0         0         0        -1         0
       sex    .0996    -.0996         0         0         0         0         0
educ_level   .03914   -.03914         0         0         0         0         0
     white   .00839   -.00839         0         0         0         0         0
      city   .00472    .00472         0         0         0         0         0
   marital   .02335   -.02335         0         0         0         0         0
```

Figure A.2: Matching summary of Coarsend Exact Matching

First, we can see that most observations were matched. The multivariate L_1 distance measure indicates that the matching specification improves the balance between samples resembled also by very small univariate differences.

A.2 Additional tables

Table A.1: Coefficients and t-statistics of prediction of wages for migrants based on residents at destination, OLS

	Log(hourly wage)	
	Coefficient	*t-statistic*
Age	0.036	67.013
Age squared	-0.029	-44.827
Female	-0.266	-206.027
White	0.108	76.766
Education (Base = none)		
Primary, secondary incomplete	0.300	171.025
Secondary, higher incomplete	0.530	342.234
Higher complete	1.212	540.371
Mean (Log(hourly wage)) =	1.521	

OLS estimates weighted with population weights. Sample are non-migrant residents in the destinations.

Table A.2: Observed and predicted wage differences based on unmatched sample

Log (nominal hourly wages)	*N*	Mean
Observed	*15,424*	1.816
Predicted	*15,424*	2.077
Difference		-0.262***
Log (real hourly wages)	*N*	Mean
Observed	*15,424*	-2.237
Predicted	*15,424*	-2.387
Difference		0.151***
	High skilled	
Log (nominal hourly wages)	*N*	Mean
Observed	*3,107*	2.846
Predicted	*3,107*	2.960
Difference		-0.114***
Log (real hourly wages)	*N*	Mean
Observed	*3,107*	-1.270
Predicted	*3,107*	-1.514
Difference		0.244***
	Low skilled	
Log (nominal hourly wages)	*N*	Mean
Observed	*12,317*	1.556
Predicted	*12,317*	1.855
Difference		-0.230***
Log (real hourly wages)	*N*	Mean
Observed	*12,317*	-2.481
Predicted	*12,317*	-2.608
Difference		0.127***

Significance levels * 10%, ** 5%, *** 1% for t-test of difference in means between observed and predicted wages. Predicted wages are based on unmatched sample.

Table A.3: Observed and predicted real wage differences using different measures of living costs

Log (real hourly wages)

	High skilled	
Skill-specific mean rents	*N*	Mean
Observed	*3,107*	-1.556
Predicted	*3,107*	-1.925
Difference		0.369***
Skill-specific median rents	*N*	Mean
Observed	*3,107*	-1.476
Predicted	*3,107*	-1.776
Difference		0.300***
	Low skilled	
Skill-specific mean rents	*N*	Mean
Observed	*12,317*	-2.433
Predicted	*12,317*	-2.514
Difference		0.081***
Skill-specific median rents	*N*	Mean
Observed	*12,317*	-2.341
Predicted	*12,317*	-2.411
Difference		0.070***

Significance levels * 10%, ** 5%, *** 1% for t-test of difference in means between observed and predicted wages. Predicted wages are based on matched sample. Rent for room is aggregated at the *microregião* level. Skill-specific rents are the rent per room aggregated only for the high or low skilled observations respectively in a *microregião* applying population survey weights. Once the mean is aggregated, in another case the median. Lastly, I also use the median to aggregate the hedonic housing price measure. These different price measures are used as denominator to compute real hourly wages.

Table A.4: Regression of housing prices on housing characteristics to predict hedonic price index, OLS estimates

	log(rent per room)
Urban area	0.256***
	(0.005)
Type of dwelling (Base = House)	
Townhouse/condominion	0.146***
	(0.003)

Continued on next page

Table A.4 – continued

	log(rent per room)
Flat	0.396***
	(0.002)
Hut	0.196***
	(0.006)
Wall material (Base = Bricks coated)	
Bricks not coated	-0.160***
	(0.002)
Wood	-0.265***
	(0.003)
Plaster coated	-0.461***
	(0.015)
Plaster not coated	-0.521***
	(0.020)
Wood unprepared	-0.344***
	(0.010)
Straw	-0.073
	(0.155)
Others	-0.146***
	(0.015)
Bathroom (Base = none)	
1	-0.213***
	(0.006)
2	-0.095***
	(0.006)
3	0.047***
	(0.007)
4	0.220***
	(0.012)
5	0.355***
	(0.027)
6	0.517***
	(0.054)
7	0.430***
	(0.119)
8	1.046***
	(0.237)
9 or more	0.356***
	(0.083)

Continued on next page

Table A.4 – continued

	log(rent per room)
Sanitation (Base = General sanitation network)	
Septic sump	-0.089***
	(0.002)
Rudimentary sump	-0.200***
	(0.002)
Ditch	-0.225***
	(0.005)
River, lake or sea	-0.152***
	(0.004)
Other	-0.212***
	(0.009)
Waste water (Base = General distribution network)	
Well on property	0.007**
	(0.003)
Well outside property	-0.088***
	(0.005)
Carro-pipa	-0.072***
	(0.014)
Rain water cisterne	-0.074***
	(0.028)
Rain water other	-0.097
	(0.068)
Rivers, lakes etc.	-0.081***
	(0.023)
Other	-0.155***
	(0.010)
Well in village	0.165**
	(0.066)
Canalization access (Base = Yes, in min. 1 room)	
Yes, only on the property	-0.052***
	(0.004)
No	-0.148***
	(0.006)
Garbage collection (Base = Collected directly)	
Collected in collective	-0.054***
	(0.002)
Burnt	-0.229***
	(0.008)

Continued on next page

Table A.4 – continued

	log(rent per room)
Buried	-0.017
	(0.043)
Tossed in public area	-0.229***
	(0.008)
Tossed in river, lake or sea	-0.195***
	(0.036)
Other	0.005
	(0.027)
Electricity provision (base = Yes by company)	
Yes, other	-0.094***
	(0.010)
No electricity	-0.238***
	(0.021)
Constant	4.235***
	(0.016)
Dummies for *microregião*	Yes
Observations	927,192
R-squared	0.539

Table A.5: Variables and data sources used in chapter 1

Variable	Description	Source
Variables for descriptive statistics and destination choice model, at level of microregião		
Housing prices	Average rent	Census, IBGE
Education provision quality index	Index from 0 to 1, computed based on: subscription rate of pre-school children, drop-out rate, Rate of teachers with higher education, average daily teaching hours, results of the IDEB (Indicator of development of education in Brazil)	FIRJAN*
Health provision quality index	Index from 0 to 1, computed based on: Number of pre-natal consultations, deaths due to badly defined causes, child-deaths due to evitable causes	FIRJAN*
Homicide rate	per 100,000 inhabitants in 2008	Ipeadata
GDP	Log of GDP in 2009	Ipeadata
Distance between origin and destination	geodesic distance as indicator for fixed moving costs, author's calculation from coordinates	Census, IBGE
Additional variables for wage regression, at individual level		
Partner participation	Dummy whether partner is working	Census, IBGE
Proportion of children in household	Number of children relative to number of adults in household	Census, IBGE
Marital status	Separated/divorced/widowed, single, married	Census, IBGE
Sector	public, private, informal, self-employed	Census, IBGE
Industry	21 industries according to International Standard Industrial Classification of all Economic Activities (ISIC)	Census, IBGE
Federal state	27 states	Census, IBGE
Variables for matching, on individual level		
Age	At time of migration, i.e. one year ago	Census, IBGE
Race	White and non-white	Census, IBGE
Education level	Primary, middle, high-school, college	Census, IBGE
Microregião of origin/residency	City of origin for migrants and city of residency for comparison group of non-migrants	Census, IBGE
Other variables for descriptive statistics of microregiões		
Formal sector wages	Average annual formal sector wages in *microregiões*	RAIS**

** FIRJAN (Industrial federation in Rio de Janeiro state)*
*** RAIS (National formal sector firm and employment registry)*

Appendix B

Appendix to Chapter 2

B.1 Additional tables

Table B.1: OLS regression of the local economy on the instrument in net-sending municipalities, 2005-2010

	Ln(GDP)	Ln(Wages)	Ln(Employment rate)
Manufacturing labour demand shock in t-1	0.638**	0.821***	0.212
	(0.313)	(0.119)	(0.272)
Year fixed effects	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	Yes
N	16,033	16,033	16,033
R^2	0.0718	0.875	0.285

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the *microregião* level. Wages and employment rate are only for the formal sector.

Table B.2: 2SLS estimation: Homicide rates on immigration rates 2005-2010, Second stage results

	Ln(Homicide rate)	
	Quadratic trend	Cubic trend
Ln(Immigration rate)	1.162***	1.163***
	(0.190)	(0.191)
Destination shock	Yes	Yes
Year fixed effects	Yes	Yes
MC fixed effects	Yes	Yes
Quadratic state tend	Yes	No
Cubic state trend	No	Yes
N	6,589	6,589
Kleibergen-Paap Test	54	53.7

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the microregion level. Each regression is weighted by municipality population. MC fixed effects refers to municipality fixed effects. Destination shock indicates that the regression includes the local labour demand shock variable at destination.

Table B.3: 2SLS estimation: Homicide rates on net-migration rates 2005-2010, Second stage results

	Ln(Homicide rate)
Ln(Net-migration rate)	0.170***
	(0.027)
Destination shock	Yes
Municipality fixed effects	Yes
Year fixed effects	Yes
State-year trend	Yes
Observations	4838
Kleibergen-Paap F-test	161

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the microregion level. The regression is weighted by municipality population. Destination shock indicates that the regression includes the local labour demand shock variable at destination.

Table B.4: 2SLS estimation: Homicide rates on immigration rates with lagged immigration rates 2005-2010, Second stage results

	Ln(Homicide rate)	
Ln(Immigration rate)	0.219	0.720***
	(0.215)	(0.224)
L1.Ln(Immigration rate)	0.049*	0.163***
	(0.030)	(0.053)
L2.Ln(Immigration rate)		0.122**
		(0.050)
Municipality fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
State-year trend	Yes	Yes
Observations	5254	4039
Kleibergen-Paap F-test	61.4	43.8

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the microregion level. The regression is weighted by municipality population. Destination shock indicates that the regression includes the local labour demand shock variable at destination.

Table B.5: Migrants' characteristic in response to local labour demand shocks at origin, by year

Year	*Probability to be*				*N*
	Low-skilled	Female	Young (16 to 25 years)	Young, male, low-skilled	
2009	0.317*	0.327*	-0.670***	-0.550***	33,191
	(0.161)	(0.164)	(0.163)	(0.120)	
2008	0.313*	0.034	-0.772***	-0.253*	34,237
	(0.152)	(0.154)	(0.152)	(0.105)	
2007	0.465**	-0.034	-0.738***	-0.217*	31,533
	(0.163)	(0.166)	(0.159)	(0.106)	
2006	0.599**	-0.102	-0.813***	-0.369**	27,023
	(0.213)	(0.218)	(0.204)	(0.137)	
2005	0.368**	0.017	-0.486***	-0.280***	18,759
	(0.130)	(0.132)	(0.122)	(0.081)	

Significance levels * 10% ** 5% *** 1%. These are the marginal effects from separate probit estimations of the probability of a migrant to be either low educated, female or young/male/unskilled on the local labour demand shock in a migrant's origin in t-1 by year of migration. Standard errors are robust. All regressions include dummies for the state of origin.

B.2 Additional figures

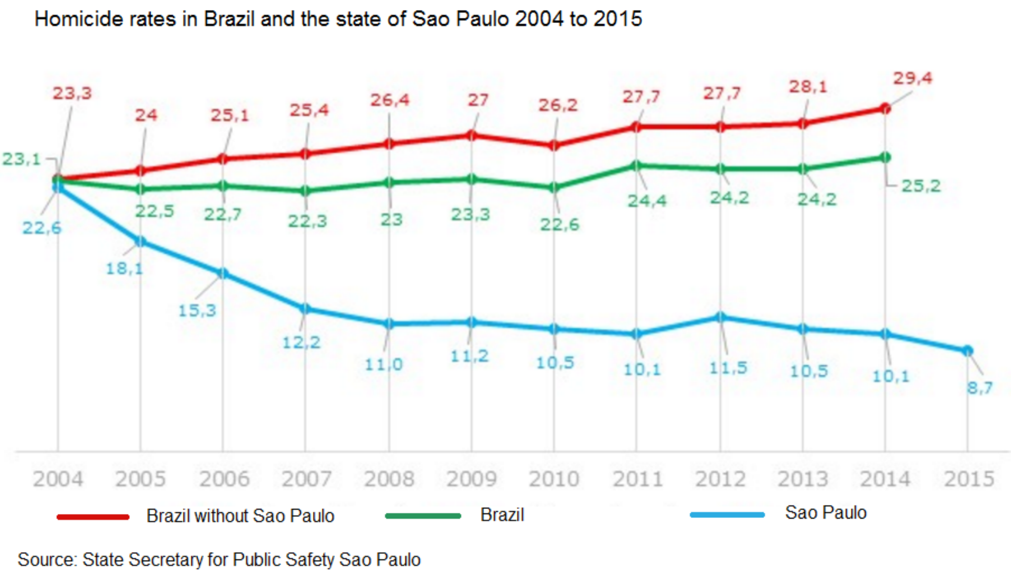

Figure B.1: Homicide rates in São Paulo State 2004 to 2015.

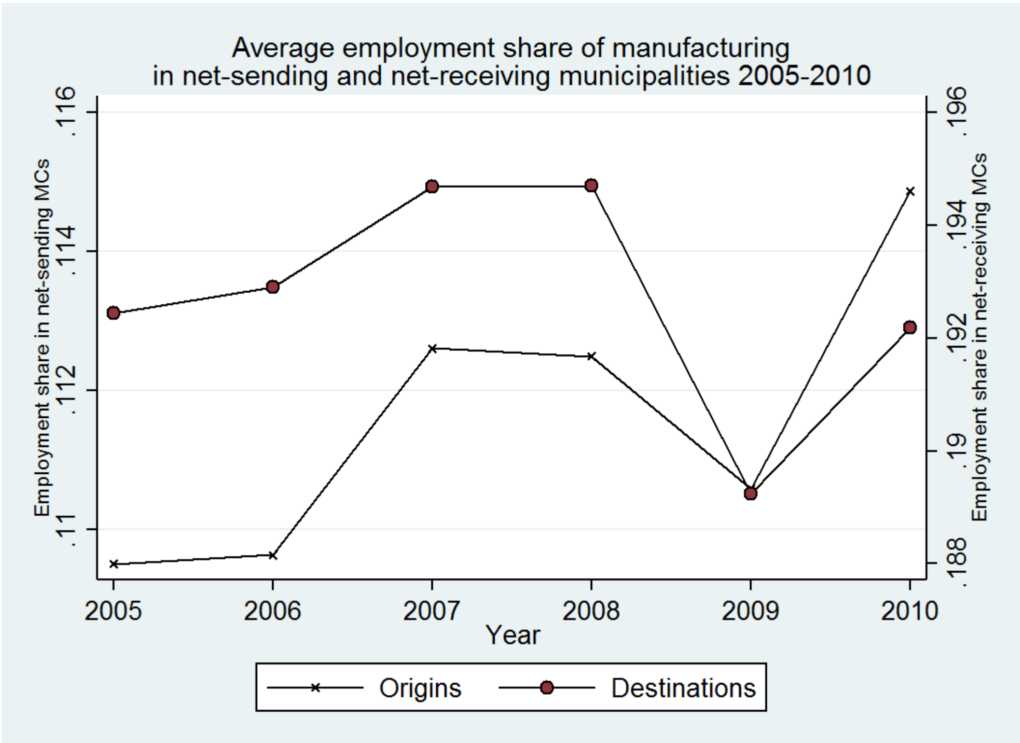

Figure B.2: Manufacturing employment share in destinations and origins 2005-2010.

B.3 Misreported homicide data

Working with crime data yields the issue of wrongly reported data. A common concern is that crime rates are often under-reported, because criminals try to hide their action. Homicide rates are less prone to this issue due to the severity of the event, but drug-related homicides or police killings are likely to lead to under-reporting (Cerqueira, 2012). The process of reporting and categorizing a death in Brazil is as follows: By law, a doctor has to fill out the declaration of death for any dead individual and in case of an external and non-natural cause of death, this doctor has to be from the Legal Medical Institute. She or he has to define whether the death was due to homicide, suicide or undetermined causes. One copy of the declaration is then handed to the next-in-kin of the dead person, another to the municipal Secretariat of Health, where administrators fill in the ICD-10 code (International Code of Death, World Health Organization). They have to follow up with doctors from the Legal Medical Institute and/or the police in case the declaration is not clear enough to define the ICD-10 code (Borges et al., 2012). Thus, it is crucial that the police passes on information to health institutions properly. This is often done poorly, so that in 2010 around 37% of deaths reported did not report the location of death and around 7% of homicides were filed in the category of 'death due to unknown cause' but are significantly more likely to have been homicides (Cerqueira, 2014a). These incorrectly reported homicides are thus not included in the data used for this analysis. Additionally, Brazil has a high number of police killings which are often not reported or hidden, as only in few states there is actual data available for these events. In 2010 alone, around 1.6 people in 100,000 were killed by police in the states where data was available (Murray et al., 2013). Some of these deaths might show up under the category of 'death due to unknown cause' but they are less likely to be reported as homicides and are thus excluded from this analysis.

The analysis so far ignored the fact, that almost one third of the observations are dropped due to missing values of homicide rates. I can either assume that these missing values are actually zero homicides which do not appear in the original data from the Ministry of Health Death Registry. In the database where homicide data is stored, municipalities that do not actually experience any homicides, i.e. they should appear with a zero, are not listed and thus show up as missing values. They cannot be differentiated from municipalities that report no homicides to the registry even though they have occurred. It is a major concern that some municipalities systematically non-report or under-report homicides. There are two reasons that could give rise to under-reporting of homicides. Firstly, homicides are a crime and as such mostly occur in a hidden and obscure environment so that not all events are discovered and reported. Secondly, homicide figures can be seen as political instrument and there is an incentive for police stations or other public institutions to under-report in order to appear in a better public light. These issues give rise to the assumption that the homicide rates are on average under-reported. For the case of Brazil there is evidence that some states have higher tendency to under-report homicides (Cerqueira, 2014a).

Systematic misreporting of homicides raises a valid concern of selection bias. The assumption that the missing values for homicides are actually zeros is motivated by the fact, that municipalities with missing homicide rates are on average much smaller than municipalities that report homicide rates. They have a population of around 7,800 people compared to almost 52,000 on average in the municipalities without missing values (Table B.6). Hence, the number of homicides in the municipalities with missing values is only between one and two, very close to zero. It is also important to note that municipalities with missing data report homicide values in some years, but not in others. This is why I can report a homicide rate for these municipalities in the first column of table B.6.

However, it is necessary to further check whether the zero assumption is plausible and how robust the results are if the assumption is challenged. The concern of non-reporting becomes an issue for the identification strategy of this paper if municipalities that have missing homicide values are systematically different from those that do report and if that difference is related to immigration rates. Table B.6 shows demographic, economic and social characteristics of all municipalities grouped by their reporting of homicides in the period from 2005 until 2010. In the last column are the p-values of the t-test of statistical difference in the means of each variable between the two groups. The municipalities are statistically different in all characteristics but public security spending at state level. The municipalities with missing values are very small, less urbanized and more remotely located than those that do not have missing values. Despite remoteness municipalities with missing homicide values have higher scores in health indices. They show better institutional quality with regards to factors that can affect the reporting of homicides, because the health institutions bear responsibility in the reporting process (Soares, 2004). This supports the zero homicide assumption.

Table B.6: Characteristics of destination municipalities with and without missing homicide values

	Missing	Non-missing	Difference	P-value of t-test
Homicide rate	27.7	25.7	2	0.000
Homicides N	**1.5**	22.2	-20.7	0.000
Immigration rate	642.8	752.1	-109.3	0.000
Population 2010	**7,505**	76,852	-69,347	0.000
Public security spending	17.81	17.78	0.03	0.452
Health quality index (0 to 1)	0.81	0.79	0.01	0.000
Distance to state capital	257.4	205.9	51.5	0.000
Urbanization rate	0.62	0.75	-0.13	0.000
N	*5,396*	*7,354*		

One possibility would be to model the selection of municipalities into reporting of missing and non-missing values. But this is not possible because it is not clear for which municipalities there is a selection problem in contrast to those that actually have zero homicides so that an exclusion restriction - even if available - would not work. Therefore, I conduct a simple matching exercises to estimate bounds for the main results and present these in the robustness checks in table B.7. I impute the homicide rates of other municipalities to those which have missing values. In order not to just randomly impute values, I match municipalities based on characteristics that are likely to affect the quality of reporting: population, distance to the state capital, the federal state, local GDP, health institutional quality index[1], and homicide of the previous year. The estimated propensity score is then used to match municipalities to their nearest five neighbours. Then I will separately use the highest and lowest homicide values of the five nearest neighbours for each municipality with missing homicide values to replace these and run the 2SLS regression. The coefficients should indicate, how much the main result changes if the assumption of zero homicides is wrong and present a worst case scenario, in which homicides are actually higher than reported or a less concerning setting in which homicide rates are not zero, but rather low.

In table B.7, I present the result of the 2SLS regression using the IV as in the main estimation, but with a different sample. In column 1 missing homicide values are treated as very small values between zero and one. If I used zeros they would become missing once I take the logarithm of homicide rates. As shown above, it is plausible to think that the homicide rates are very small values in years when only one or two homicides occurred. The effect is naturally smaller as there are many such small values. The instrument still passes the first stage F-test rule of thumb. This reduces the concern that there is a selection bias which is related to the instrument and would thus interfere with the identification strategy. In the other two columns are the coefficients when missing homicide values are replaced by the values of homicide rates from matched neighbours. In the second column, the effect is very close to the main result. In the worst case scenario, that missing homicide values are actually high values, the effect becomes only slightly larger. Even though the matching exercise did not impute zeros by construction, it did not systematically change the result of the analysis. The assumption that missing homicide values are actually mis-reported zeros or values close to zero seems plausible. The imputed

[1]This index is compiled by the Federal Industrial Association of the State of Rio de Janeiro (FIRJAN) from data on the number of pre-natal consultations, number of deaths due to badly defined causes, number of infant deaths due to evitable causes and whether hospitalization is sensible to basic care. The index is a value between 0 and 1, with 1 indicating the highest level of development. The fact, that the index includes the number of deaths due to badly defined causes makes it an important variable for the matching, because this indicates the quality of the institutions involved in the reporting of homicides.

values of the lowest nearest neighbours are on average 9.9, for the highest, 44.9. The latter is much higher than the average homicide rate in the missing sample for the years, in which these municipalities do not report missing values (23.2). Even if I assume that this scenario is unrealistic, it is only 0.1 percentage points higher than the initial estimate of 1.2% so that I am not concerned with this issue.

Table B.7: Robustness of results to changes in homicide values

Effect of migration on homicide rates, 2SLS estimation			
	Ln(Homicide rate)		
Imputation for missing values:	Between 0 and 1	Lowest nearest neighbour	Highest nearest neighbour
Ln(Immigration rate)	0.902***	1.090***	1.338***
	(0.229)	(0.206)	(0.212)
Destination shock	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
MC fixed effects	Yes	Yes	Yes
State trend	Yes	Yes	Yes
N	11,492	7,927	7,927
Kleibergen-Paap Test	44.9	46.1	46.1
Mean of imputed homicide rate		9.2	44.9

Significance levels * 10% ** 5% *** 1%. Standard errors are clustered at the microregion level. Each regression is weighted by municipality population. MC fixed effects refers to municipality fixed effects. Destination shock indicates that the regression includes the local labour demand shock variable at destination.

Appendix C

Appendix to Chapter 3

C.1 Multiple Correspondence Analysis to construct asset index

An asset index is a composite measure using information about asset ownership and/or other welfare indicators in survey data. The researcher is interested in one continuous measure that captures the welfare of a household. In its simplest format, we can think of an asset index as the sum of its weighted components as specified in equation C.1:

$$A_i = p_1 a_{1,i} + p_2 a_{2,i} + \ldots + p_k a_{k,i} \tag{C.1}$$

The asset index of household i is the sum of each of the individual asset indicator dummies, a_k weighted by an asset specific weight, p_k. Each indicator is equal to 1 if the household owns this specific asset, 0 otherwise. There are different possibilities to assign weights. Filmer and Pritchett (2001) apply Principle Component Analysis (PCA) to retrieve weights for each asset to construct an asset index that reflects household wealth. PCA uses the cross-sectional covariance of all individual assets to define the linear combination of these assets with the largest variance. This would be the first principal component. The linear combination orthogonal to this first one would be the second principal component and so on. This method thus assumes that the variation in assets explained by all components is complete and accurate. The PCA also imposes the linear constraint of equal distance between categories and ordered categories (Booysen et al., 2008). It is however arbitrary to judge whether a cement wall is of better or lower quality than a burnt brick wall, and even if that was established it is not clear that the difference in quality would be the same between a burnt brick and a cement wall as between a cement and a mud wall.

In response to the strong assumptions of PCA, Sahn and Stifel (2000) applied the less restrictive method of Factor Analysis (FA). This method also assumes that the cross-sectional variance-covariance matrix of individual asset indicators can reveal a common factor. It can therefore be used to extract the linear combinations of those individual assets that reflect this factor the best. In contrast to PCA, it allows individual assets to explain variances and not all indicators have to explain the full covariance matrix. This approach however requires imposing a linear relationship and assumes normality for model estimation (Asselin, 2009). This assumption and the arbitrary choice of rotation methods in the second step of FA seem restrictive and arbitrary (Booysen et al., 2008).

The third approach, used in chapter 3, is the non-parametric approach to let the data reveal the individual asset weights without the imposition of any functional form, that is Multiple Correspondence Analysis (MCA). MCA is also more suitable to a setting with categorical variables than PCA. Our asset measures are always expressed in categorical variables. Asselin (2009) explains in detail the advantage of MCA over FA or PCA to construct multidimensional poverty measures. Booysen et al. (2008) for example use MCA to construct an asset index to compare poverty dynamics across time and countries.

The construction of an asset index using MCA can be described in three steps. First, we construct an indicator matrix I as form of normalization of the data. This means that we generate a dummy for each of the variable categories that is equal to one if the household, for example, has wooden walls and zero for all other types of walls and so on. It is important to note that for each asset variable the categories are mutually exclusive and exhaustive. A household cannot have a one for both, wooden walls and brick walls; neither can the household have only zeros for the categories of wall quality. The result of this data manipulation is the indicator matrix I in which the rows represent a household observation for the year 2013, and the columns represent an asset category.

In the second step, we apply MCA to this indicator matrix. The columns are vectors for each sub-category of each asset item, e.g. brick wall, cement wall, mud wall all comprise one column vector each for the asset item 'wall material'. The MCA computes the principal coordinates of each of these categories. These coordinates capture the contribution of each category in the first one-dimensional axis of best fit, which explains most of the variation in the indicator matrix I and can therefore be understood as the unobserved underlying factor. In other words, these coordinates are the average of the normalized scores of each population unit in this category. These coordinates will serve as the weights for each indicator category in the construction of the asset index.

Thirdly, we apply these weights to each asset category to compute the asset index to each household-year observation as in equation C.1. We present the results from the MCA displaying the principal coordinates and their contribution to the overall category variance in the appendix figures C.1 and C.2 (page 196 and 197).

```
Multiple/Joint correspondence analysis          Number of obs     =        593
                                                Total inertia     =  .01661842
    Method: Burt/adjusted inertias              Number of axes    =          2
```

Dimension	principal inertia	percent	cumul percent
dim 1	.0051666	31.09	31.09
dim 2	.0021568	12.98	44.07
dim 3	.0009358	5.63	49.70
dim 4	.0005249	3.16	52.86
dim 5	.000335	2.02	54.87
dim 6	.0002993	1.80	56.67
dim 7	.000156	0.94	57.61
dim 8	.0001396	0.84	58.45
dim 9	.0000911	0.55	59.00
dim 10	.0000541	0.33	59.33
dim 11	.0000215	0.13	59.46
dim 12	4.02e-06	0.02	59.48
dim 13	1.95e-06	0.01	59.49
dim 14	5.68e-07	0.00	59.50
Total	.0166184	100.00	

Figure C.1: Explanatory power of each dimension of MCA

Categories	mass	overall quality	%inert	dimension_1 coord	sqcorr	contrib
roomn1						
0	0.032	0.562	0.002	0.016	0.280	0.002
1	0.003	0.562	0.020	-0.183	0.280	0.018
roomn2						
0	0.029	0.018	0.003	0.000	0.000	0.000
1	0.005	0.018	0.016	-0.002	0.000	0.000
roomn3						
0	0.027	0.022	0.006	-0.008	0.016	0.000
1	0.007	0.022	0.021	0.028	0.016	0.001
roomn4						
0	0.029	0.023	0.003	0.004	0.007	0.000
1	0.006	0.023	0.017	-0.019	0.007	0.000
roomn5						
0	0.022	0.128	0.015	-0.021	0.036	0.002
1	0.013	0.128	0.025	0.034	0.036	0.003
q95_1						
0	0.005	0.611	0.079	-0.235	0.222	0.056
1	0.029	0.611	0.014	0.043	0.222	0.010
q95_2						
0	0.029	0.611	0.014	0.043	0.222	0.010
1	0.005	0.611	0.079	-0.235	0.222	0.056
q97a_1						
0	0.001	0.011	0.004	0.004	0.000	0.000
1	0.033	0.011	0.000	-0.000	0.000	0.000
q97b_1						
0	0.022	0.742	0.014	0.088	0.736	0.033
1	0.013	0.742	0.024	-0.153	0.736	0.057
q98_1						
0	0.031	0.623	0.004	0.038	0.608	0.008
1	0.004	0.623	0.033	-0.292	0.608	0.065
q98_2						
0	0.025	0.037	0.009	0.006	0.007	0.000
1	0.009	0.037	0.025	-0.018	0.007	0.001
q98_3						
0	0.023	0.143	0.013	-0.035	0.128	0.005
1	0.012	0.143	0.024	0.066	0.128	0.010
q98_4						
0	0.031	0.121	0.002	-0.011	0.115	0.001
1	0.004	0.121	0.017	0.091	0.115	0.006
q98_5						
0	0.034	0.308	0.000	0.001	0.300	0.000
1	0.000	0.308	0.002	-0.289	0.300	0.002
q98_6						
0	0.031	0.108	0.002	-0.011	0.108	0.001
1	0.004	0.108	0.017	0.090	0.108	0.006
q98_7						
0	0.034	0.143	0.000	0.001	0.122	0.000
1	0.000	0.143	0.003	-0.135	0.122	0.001
q98_8						
0	0.033	0.088	0.001	0.002	0.018	0.000
1	0.002	0.088	0.011	-0.048	0.018	0.001

Figure C.2: Summary of first dimension of MCA

q99_1							
	0	0.027	0.630	0.021	-0.082	0.515	0.035
	1	0.008	0.630	0.071	0.277	0.515	0.117
q99_2							
	0	0.034	0.192	0.000	-0.000	0.035	0.000
	1	0.000	0.192	0.004	0.138	0.035	0.000
q99_3							
	0	0.009	0.632	0.072	0.252	0.492	0.113
	1	0.025	0.632	0.026	-0.092	0.492	0.041
q99_4							
	0	0.034	0.319	0.000	0.001	0.005	0.000
	1	0.000	0.319	0.012	-0.052	0.005	0.000
q99_6							
	0	0.034	0.064	0.000	0.000	0.045	0.000
	1	0.000	0.064	0.003	-0.137	0.045	0.000
q99_7							
	0	0.034	0.069	0.001	-0.004	0.068	0.000
	1	0.001	0.069	0.032	0.238	0.068	0.007
q101_1							
	0	0.033	0.232	0.001	-0.008	0.166	0.000
	1	0.002	0.232	0.016	0.162	0.166	0.009
q101_2							
	0	0.034	0.286	0.000	-0.001	0.038	0.000
	1	0.000	0.286	0.004	0.090	0.038	0.000
q101_3							
	0	0.033	0.423	0.001	-0.001	0.003	0.000
	1	0.002	0.423	0.019	0.025	0.003	0.000
q101_5							
	0	0.021	0.551	0.029	-0.112	0.551	0.052
	1	0.013	0.551	0.048	0.184	0.551	0.085
q101_6							
	0	0.018	0.661	0.043	0.157	0.618	0.086
	1	0.016	0.661	0.048	-0.173	0.618	0.095
q101_7							
	0	0.033	0.013	0.001	-0.003	0.013	0.000
	1	0.002	0.013	0.030	0.064	0.013	0.001

Figure C.3: Summary of first dimension of MCA, continued

The comparison of asset indices for the same household over time yields two potential problems. The first is that the coordinates used as weights should be consistent over time to make the index comparable between periods. The coordinates are retrieved from the data and result from the cross-sectional variation of assets across households. If we now pooled the two survey waves to compute the coordinates, this variation would be different than that of one cross-section and some of the variation would only reflect variation over time. Therefore, we will rely only on the data of the base year 2013 to retrieve the coordinates as in Booysen et al. (2008). Then we use these to compute the index in both years. In the appendix table C.6

(page 206), we also present the main results using the pooled sample to compute the asset index. The results barely change.

The second issue is that prices for assets might change over time and in response to this the demand for assets and the distribution of assets across households might change.[1] There is, though, no reason to think that households with a new migrant would react differently than control households to price changes in their asset purchase behaviour.

C.2 Attrition

Some households of the baseline survey were not successfully tracked in the follow-up two years later. These were around 300 households. For the analysis in chapter 3 it is important to understand, whether some of these households would have been part of the sample of interest and what implications their attrition has for the analysis. Of the 300 households that attrited, 167 had a migrant in 2013. They would have been part of the analysis either as treated or control units. In order to understand whether these households would have been more or less likely to be in the treatment group and whether they are substantially different from our sample we compare the baseline characteristics of households. The comparison is between the 167 attrited households with migration experience, the control and the treatment households.

Table C.1 shows results of a Logit and of a Multinomial Logit (MNL) estimation of the status of a household on baseline characteristics. The most important difference between sample and attrited households is their size as we see in the first row of the regressions. Attrited families are significantly smaller than those that were successfully re-interviewed. Furthermore, treated households are significantly larger than both control and attrited observations. There are a few other weakly significant coefficients. The regional differences are notable. Attrition is highest in

[1]While there has been high inflation in Ghana between 2013 and 2015 there is no data on the price changes for each individual asset (Ghana Statistical Service, 2015b,a). Moreover, it is difficult to measure the market price of a mud wall or a brick wall, as we would need to decide whether to measure only the material or also the service to build the wall.

the Volta region, 46 percent of attrited units were located here, 25 percent in Brong Ahafo, and only 5 percent in the Upper West. In the Logit estimation the base category of regions is Brong Ahafo, so that only the coefficient for Upper West appears significantly. In contrast, treated households are much more likely to be located in Brong Ahafo than in Volta or Northern region as we observe in the MNL results.

Table C.1: Likelihood for household to attrite, Logit and Multinomial Logit results

	Logit	MNL (Base = Control)	
		Attrited (N = 167)	Treated (N = 170)
Household size	-0.129***	-0.084**	0.202***
	(0.040)	(0.042)	(0.038)
Dependency ratio	0.136	0.112	-0.119
	(0.165)	(0.164)	(0.194)
Age of head in years	-0.009	-0.011	-0.007
	(0.008)	(0.009)	(0.007)
Number of current migrants in 2013	-0.115	-0.098	0.048
	(0.094)	(0.095)	(0.075)
Female head	-0.204	-0.121	0.420
	(0.289)	(0.294)	(0.279)
Highest level of education in household (Base = None)			
Primary	0.920	0.833	-0.382
	(0.866)	(0.874)	(0.671)
Middle/Junior	1.254	1.194	-0.280
	(0.799)	(0.803)	(0.603)
Senior Secondary	1.152	1.107	-0.232
	(0.807)	(0.812)	(0.605)
Higher	1.220	1.259	0.078
	(0.827)	(0.832)	(0.605)
Main occupation of head (Base = Inactive/Other)			
Employee	-0.685	-0.603	0.540
	(0.442)	(0.451)	(0.559)
Self-employed	-0.748*	-0.685*	0.435
	(0.383)	(0.390)	(0.465)
Unpaid work / unemployed	-1.314**	-1.247**	0.478
	(0.512)	(0.523)	(0.470)

Continued on next page

Table C.1 – continued

	Logit	MNL (Base = Control)	
		Attrited	Treated
		(N = 167)	(N = 170)
Marital status of head (Base = Single)			
Married/living with partner	0.974*	0.998*	0.137
	(0.567)	(0.577)	(0.511)
Separated/Divorced/Widowed	0.864	0.904	0.194
	(0.613)	(0.625)	(0.557)
Region (Base = Brong Ahafo)			
Northern	-0.310	-0.710*	-1.984***
	(0.369)	(0.389)	(0.399)
Upper East	-0.698	-0.704	-0.014
	(0.459)	(0.477)	(0.370)
Upper West	-1.216**	-1.347**	-0.499
	(0.572)	(0.579)	(0.391)
Volta	0.119	-0.075	-0.847***
	(0.275)	(0.285)	(0.326)
Household has seasonal migrant	-0.464*	-0.424	0.222
	(0.275)	(0.278)	(0.240)
Household has returnee	-0.327	-0.236	0.445
	(0.357)	(0.365)	(0.297)
Main income source(Base = Public sector)			
Private sector	0.713	0.866*	0.688
	(0.509)	(0.525)	(0.588)
Own business	0.378	0.421	0.192
	(0.436)	(0.441)	(0.485)
Own farm	-0.022	0.161	0.786
	(0.480)	(0.488)	(0.503)
Private transfers	0.478	0.523	0.173
	(0.480)	(0.487)	(0.534)
Others	-0.015	0.048	0.317
	(0.755)	(0.779)	(0.738)
Asset purchases in preceding 2 years			
Electronic goods	0.239	0.226	0.050
	(0.252)	(0.257)	(0.243)
White goods	-0.129	-0.094	0.043
	(0.305)	(0.314)	(0.319)
Livestock	-0.180	-0.170	0.048
	(0.323)	(0.327)	(0.274)
Generator	-0.750	-0.620	0.514

Continued on next page

Table C.1 – continued

	Logit	MNL (Base = Control)	
		Attrited	Treated
		(N = 167)	(N = 170)
	(1.137)	(1.160)	(0.694)
Car	1.213**	1.633***	1.384**
	(0.507)	(0.567)	(0.557)
Computer	-0.700	-0.784	-0.190
	(0.500)	(0.522)	(0.608)
Electric Appliances	0.094	-0.007	-0.551
	(0.293)	(0.301)	(0.365)
Other Investments	0.263	0.293	0.199
	(0.367)	(0.376)	(0.383)
Agricultural land	-0.412	-0.342	0.203
	(0.356)	(0.364)	(0.275)
Agricultural machinery	0.915	0.911	0.107
	(0.856)	(0.877)	(0.733)
Non-agricultural land	-0.080	-0.171	-0.442
	(0.343)	(0.355)	(0.351)
New house	-0.456	-0.306	0.607**
	(0.286)	(0.292)	(0.255)
Constant	-1.227	-1.177	-2.665**
	(1.170)	(1.177)	(1.133)
Observations	*699*	*699*	*699*
log likelihood	-296	-573	-573
Likelihood Ratio Chi2	81.64	178.8	178.8

These observations suggest that attrited households were less likely to be among the treated group. They appear however still different from the control group in the sample, so that we cannot assume that they are missing at random. To have a better sense of how their exclusion from the analysis might affect the results of the impact assessment, we look at the distribution of the asset index. It is important to note that the index is constructed based on the cross-sectional distribution of assets in the baseline sample. Thus, the distribution changes once we include the attrited household in the construction of the index. Figure C.4 plots the kernel density of the asset index at baseline. We differentiate between the attrited households, the control and the treated. The latter two groups show two different graphs. One is the asset index when we construct it including the attrited households, the other one (indicated with ‘sample’) when we construct it only for the households included in the analysis. This is the index used in the analysis of chapter 3.

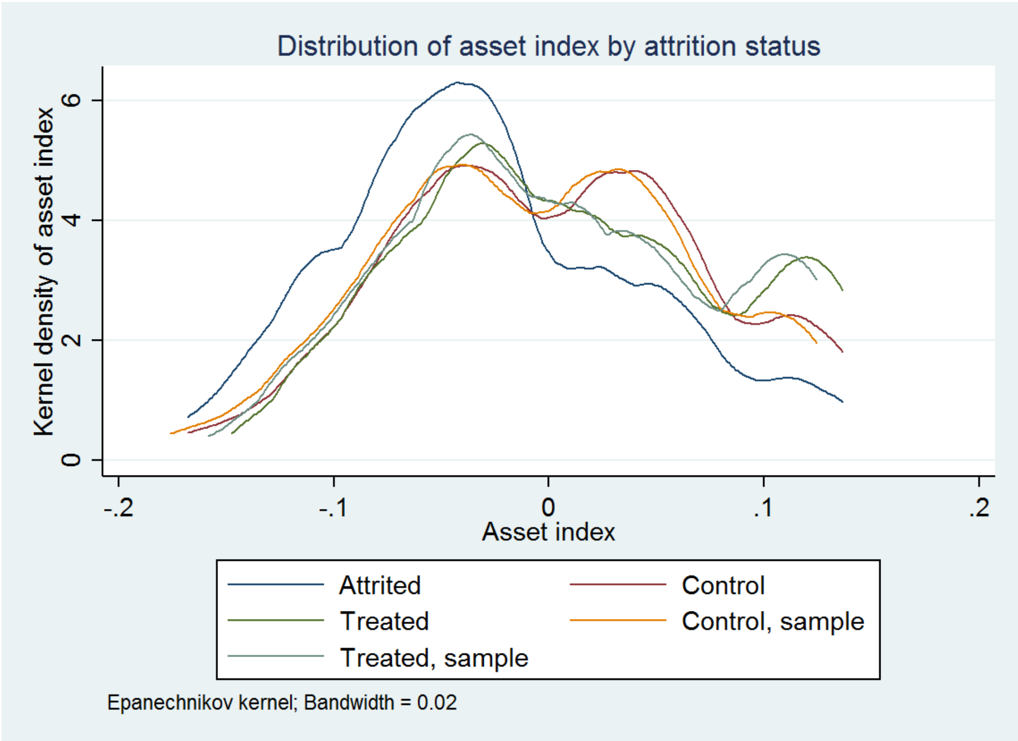

Figure C.4: Distribution of asset index for attrited, treated and control households in 2013

The exclusion of the attrited observations from the index construction implies a change in the distribution of the index. However, the distribution of treatment and control units are very similar. The attrited households have a lower asset index distribution. This could result from their smaller size which is associated with assets of housing quality. For the results of the analysis, we can only conclude, that the initial asset index distribution would look different for the control group, it would be lower. However, the application of balancing weights would ensure that the analysis would be based on a comparable sample. The analysis looks at the impact of having a new migrant on the change in the asset index between survey years. The data does not allow to see whether attrited households are on a positive or negative growth trajectory regarding their asset index. If we assume that they are in the control group (based on their observables) a positive change in their asset index would bias our results upwards, we would possible find a significant negative effect, and vice versa.

C.3 Additional tables

Table C.2: Migration costs by number of times migrant moved before

	New migrant in GHS of 2015	*N*	Baseline migrant in GHS of 2015	*N*
First time	160.3	*74*	331.0	*137*
Moved at least once before	78.2	*41*	142.3	*132*

Table C.3: Results of Chow test; H_0 = Coefficients are stable across sub-samples

F(8, 464) =	1.13
Prob > F =	0.3389

Table C.4: Effect of having a new migrant on asset index dropping local employment rate from control variables, weighted least squares

	Asset index
New Migrant * 2015	-0.016
	(0.011)
Household has return migrant (=1)	-0.015*
	(0.009)
Dependency ratio	0.001
	(0.004)
Occupation of household head (base = inactive/others)	
Employee	0.015
	(0.015)
Self-employed	0.001
	(0.015)
Unpaid work / unemployed	-0.002
	(0.018)
Entropy balancing weights	Yes
Observations	*960*
Adjusted R-squared	0.524
Number of clusters	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level.

Table C.5: Effect of having a new migrant on asset index, weighted least squares. Entropy balancing weights constructed including the asset index and its squared term at baseline instead of individual asset indicators.

	Asset index		
	(1)	(2)	(3)
New Migrant * 2015	-0.011	-0.018*	-0.017*
	(0.007)	(0.010)	(0.010)
Household has return migrant (=1)			0.002
			(0.005)
Dependency ratio			0.008
			(0.014)
Occupation of household head (base = inactive/others)			
Employee			-0.006
			(0.016)
Self-employed			-0.004
			(0.017)
Unpaid work / unemployed			-0.018**
			(0.008)
Local employment rate			0.125
			(0.094)
Entropy balancing weights	No	Yes	Yes
Observations	*960*	*960*	*960*
Adjusted R-squared	0.584	0.546	0.552
Number of clusters	93	93	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level.

Table C.6: Effect of having a new migrant on asset index, weighted least squares. Asset index constructed from data pooling both survey waves.

	Asset index
New Migrant * 2015	-0.016
	(0.010)
Entropy balancing weights	Yes
Other controls	Yes
Observations	*960*
Adjusted R-squared	0.539
Number of clusters	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level.
Other controls include whether the household has a returned migrant, occupation of the household head, dependency ratio and community employment rate.

Table C.7: Effect of new migrant on household welfare controlling for sample that did not respond to shock question, weighted least squares

	Asset index	
New Migrant * 2015	-0.016	-0.024
	(0.011)	(0.017)
Community not in sample	0.023**	
	(0.011)	
Entropy balancing weights	Yes	Yes
Other controls	Yes	Yes
Observations	*960*	*960*
Adjusted R-squared	0.529	0.532
Number of clusters	93	93

Significance levels * 10% ** 5% *** 1%. Fixed effects estimator. S.E. clustered at community level.
Other controls include whether the household has a returned migrant, occupation of the household head, dependency ratio and community employment rate.