UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

ESSAYS IN ECONOMICS OF INTERNAL MIGRATION

Ensaios em Economia da Migração Interna

Tiago Pontes Ferraz

Orientador: Prof. Dr. Raphael Bottura Corbi

São Paulo - Brasil
2023

TIAGO PONTES FERRAZ

Tese apresentada ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito parcial para a obtenção do título de Doutor em Ciências.

Área de concentração: Teoria Econômica

Orientador: Prof. Dr. Raphael Bottura Corbi

Versão Original

São Paulo - Brasil
2023

FICHA CATALOGRÁFICA
Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Ferraz, Tiago Pontes
Essays in Economics of Internal Migration / Tiago Pontes Ferraz. – São Paulo, 2023.
136 p.

Tese (Doutorado) – Universidade de São Paulo, 2023.
Orientador: Raphael Bottura Corbi.

1. Migração interna 2. Mercados de trabalho 3. Participação feminina I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II.Título.

CDD – 305.522

À minha esposa, Maristela, que sonhou junto comigo e ao meu filho, Bruno, que me faz continuar sonhando.

AGRADECIMENTOS

Agradeço a toda minha família que, de algum modo, sempre esteve presente ao longo da jornada. Em especial aos meus pais, Valter e Ana, que me ajudaram a dar os primeiros passos e me ensinaram o valor desse percurso.

Agradeço ao meu orientador, Raphael Corbi, que sempre acreditou no meu trabalho (até mais do que eu) e me fez dar sempre um passo a mais. Esse trabalho não teria acontecido sem o seu apoio.

Agradeço à professora Renata Narita, que acompanhou de perto todo este trabalho e me ensinou muito ao longo do caminho. Também à professora Fernanda Estevan, que sempre foi uma importante influência e me ajudou a ter certeza de que a academia era o caminho a seguir.

A todos os professores, pesquisadores e equipe de apoio do departamento que ao longo do tempo, cada um à sua maneira, deram alguma contribuição para que este trabalho acontecesse.

Aos meus amigos que, desde os primeiros dias no mestrado, dividiram tantas angústias e também algumas conquistas. Em especial, o (agora) professor Danilo Souza, Raphael Bruce, Ricardo Sabbadini e Michael França.

Aos meus colegas Fernando de Lollo e Mateus Villaça que acompanharam de perto o desespero do fim de prazo e me ajudaram a manter a sanidade.

Por fim, agradeço ao apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo número 2018/12249-7, pela concessão de bolsa no país; e também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), projeto 870229/1997-4, pela concessão de bolsa no país, que viabilizaram a realização deste trabalho.

“Forget safety.
Live where you fear to live.
Destroy your reputation.
Be notorious.”

Rumi

RESUMO

Esta tese de doutorado é composta por três artigos cujo tema central são os impactos econômicos da mobilidade interna de trabalhadores. O primeiro artigo busca entender como os mercados de trabalho podem se ajustar a choques de oferta, num contexto de alta informalidade e com firmas do setor formal podendo se ajustar através do pagamento de benefícios. Usando um instrumento 'shift-share' que combina choques climáticos com uma rede preexistente de migrantes em cada município de destino, o trabalho mostra que um incremento da taxa de migração interna induz um ajuste menor nos salários do setor formal do que no setor informal e também uma redução na fração de trabalhadores formais que recebem algum tipo de benefício, levando a uma queda na compensação total. Trabalhadores menos escolarizados são os mais afetados por este choque, em especial no setor informal, implicando um aumento na desigualdade salarial. O segundo artigo investiga o papel que trabalhadoras domésticas, especialmente as migrantes vindas do Semiárido, exercem sobre a participação de mulheres mais escolarizadas na força de trabalho nos municípios de destino. O trabalho adota uma estratégia empírica similar à do primeiro artigo e mostra que a chegada de migrantes em uma dada localidade alivia restrições com as quais mulheres de alta escolaridade se deparam e aumenta sua participação na força de trabalho e na probabilidade de trabalhar ao menos 40 horas/semana. Em particular, mulheres vivendo em domicílios com crianças menores de 6 anos, que são mais afetadas pela restrição. Os impactos são ainda mais fortes em locais com baixa oferta de pré-escola, mas bem menores em municípios com altas taxas de violência e entre mulheres vivendo em ambientes mais conservadores em termos de normas sociais e de gênero. Por fim, o terceiro artigo analisa os impactos da chegada de migrantes na expansão de favelas em áreas urbanas nos municípios de destino. Os resultados encontrados mostram que a entrada de migrantes internos aumentam tanto a área ocupada por favelas em municípios em regiões não-metropolitanas quanto a população vivendo nas favelas, especialmente nos municípios em áreas metropolitanas. Os efeitos são ainda maiores em localidades com pouca disponibilidade de área livre para construção.

Palavras-chaves: Migração interna, mercados de trabalho, participação feminina,

trabalho doméstico, expansão de favelas.

ABSTRACT

This doctoral thesis consists of three articles whose central theme is the economic impacts of the internal mobility of workers. The first essay investigates how labor markets can adjust to supply shocks in a context of high informality and with formal sector firms being able to adjust via non-wage benefits. Using a 'shift-share' instrument approach that combines weather-induced migration with past settlement patterns in each destination municipality, this paper shows that increasing the migration rate leads to a smaller reduction in earnings in the formal than in the informal sectors and to a reduction in the share of formal workers receiving non-wage benefits, leading to a fall in total compensation. Less-educated workers bear most of these costs, especially in the informal sector, resulting in an increase in earnings inequality. The second essay investigates the role of domestic workers, especially incoming migrants from the Semiarid region, in the labor force participation of high-educated women in the destination municipalities. This paper uses an empirical strategy similar to that adopted in the first essay and shows that the incoming migration to a given destination alleviates some constraints that high-educated women face and leads to an increase in labor force participation and the probability of working at least 40 hours per week. In particular, women living in households with children aged less than 6 years, who are more constrained. The impacts are larger in places with a lower supply of preschools, but smaller in more violent municipalities and among women living in more conservative environments regarding social and gender norms. Finally, the third essay studies the impact of incoming internal migration on the expansion of slums in the urban areas of destination municipalities. The findings show that incoming migration increases both the area occupied by slums in municipalities in non-metropolitan areas as well as the population living in these precarious conditions, especially in the metropolitan regions. The effects are larger in places where there is lower availability of usable land for construction.

Key-words: Internal migration, labor markets, female participation, domestic work, slum growth.

TABLE OF CONTENTS

1 Internal Migration and Labor Market Adjustments in the Presence of Non-wage Compensation

With Raphael Corbi and Renata Narita

1.1 Introduction

Migration within and beyond borders has important implications regarding development, demographic and economic dynamics. As a response, an extensive literature on the impacts of migration on the native population has emerged, particularly in terms of employment and wage levels. In a recent book, Borjas (2014) summarizes his vast contribution to the field and underscores the costs of immigration for competing native workers. On the other hand, a growing fraction of scholars has concluded that migration may have more nuanced effects (CARD; PERI, 2016). Card (2009) finds that immigration to the United States has only a minor impact on native wages, and Ottaviano e Peri (2012) report minor positive wage effects.

Canonical partial equilibrium models with perfect competition and substitution between natives and migrants predict full adjustment through wages when natives are immobile or lower native employment when wages are rigid (for an early example, see (ALTONJI; CARD, 1991)). Attempts to reconcile the apparently contradicting empirical evidence include expanding models to accommodate multiple outputs and technology margins (LEWIS, 2011; DUSTMANN; GLITZ, 2015) as well as recognizing that different specifications measure different parameters (DUSTMANN; SCHÖNBERG; STUHLER, 2016).

While the debate remains contentious, implicit in this discussion is a common but under-considered assumption that non-wage aspects of jobs are fixed. Allowing for adjustments along these margins may have important implications for the study of labor markets. Clemens (2021) argues they may explain existing controversies over the economics of minimum wages. Less is known about the role of non-wage adjustments in our understanding of the consequences of migration.

In this paper, we argue that adjustments in non-wage compensation are empirically relevant and thus can have important implications for studying the effects of labor supply shocks due to migration. In particular, we investigate the impacts of internal migration in Brazil on the labor market outcomes of natives in a setting where downward wage rigidity is present, non-wage benefits are a significant margin of compensation, and labor informality is pervasive. This setup allows us to study how firms and workers – when adjusting to a labor supply shock due to increased migration inflows – may circumvent the binding minimum wages by reducing non-wage benefits of formal jobs or simply lowering salaries in unregulated informal markets.

The theory is based on a simple model that generates predictions for the impact of migration on labor markets with two sectors in an economy with different levels of intersectoral linkages and with endogenous or fixed benefits. From low to medium levels of linkages, the impact of migration in terms of employment and wage drops in magnitude. Allowing firms to adjust benefits as a response to shocks also softens the impact on employment. Non-wage benefits are a relevant margin of adjustment for firms, especially in more regulated labor markets. They ease constraints and allow employers to partially absorb shocks.

Brazil provides a good environment for our investigation for three reasons. First, over 3 million people in the Brazilian Semiarid, a historical source of climate migrants, left their hometowns during our sample period of 1996-2010. Second, a within-country analysis minimizes econometric concerns about allocating migrants to particular skill groups (DUSTMANN; FRATTINI; PRESTON, 2012). Third, over 40 percent of workers are employed in the less frictional informal labor sector, where firms do not comply with labor market statutes, such as minimum wage laws and layoff regulations. The rest of the workforce participates in the formal sector where the minimum wage is binding (above 70% of the median wage) and non-wage compensation is frequently offered. Indeed, over 31 million people, or 20% of registered workers, are covered through employer-provided health insurance. After payroll expenses, this is the second highest component of total labor costs

(ANS, 2019). Also, 40% of these workers receive food subsidies, costing firms about 57% of the minimum wage per worker.[1.1.1] To the extent that workers value non-wage benefits, changes in this margin of adjustment can have important welfare implications.

To address the econometric concerns associated with the fact that migrants tend to move to areas with better labor market opportunities, we take advantage of a recent body of work that provides a clear framework for distinguishing sufficient conditions for identification and properly computing standard errors (GOLDSMITH-PINKHAM; SORKIN; SWIFT, 2020; BORUSYAK; HULL; JARAVEL, 2021; JAEGER; RUIST; STUHLER, 2018; ADAO; KOLESÁR; MORALES, 2019). In particular, we combine two extensively used identification strategies into a shift-share instrument approach. First, we exploit exogenous rainfall and temperature shocks (or "shift") at the origin to predict the number of individuals leaving each Semiarid's municipality. Then we leverage the history of the Semiarid as a large source of climate migrants and use the past settlement patterns (or "share") to allocate migration outflows to destination areas (MUNSHI, 2003; BOUSTAN; FISHBACK; KANTOR, 2010). The resulting predicted inflow of migrants is an instrument for observed migration.

Our results show that increasing the rate of migration inflows by one percentage point reduces the share of formal employment among native workers by 0.13*p.p.* while increasing the number of informal[1.1.2] jobs by 0.11*p.p.*. These results are consistent with a less than perfect adjustment of wages and benefits in the formal sector. In particular, wage rigidity in this sector limits the negative impacts for low-paid workers. Our results are also consistent with important production linkages between formal and informal sectors such that individuals who lost their formal jobs are absorbed by informal firms or self-employment, which are more competitive labor sectors. Thus the overall effect on native total employment across sectors is small or even null.

Regarding compensation, we find a decrease between 0.59% and 1.00% on

[1.1.1] Arbache e Ferreira (2001) based on various sources estimate the average cost of providing some job benefits in Brazil.

[1.1.2] Our definition of informal sector also includes self-employed workers.

average earnings in the formal sector and a negative impact on the share of formal workers receiving employer-provided health insurance in the range of $0.31 p.p.$ to $0.47 p.p.$, food vouchers from $0.33 p.p.$ to $0.69 p.p.$ and transportation subsidies from $0.37 p.p.$ to $0.57 p.p.$ Our evidence on employer-sponsored health insurance provision is complemented with firm-level administrative data on health insurance contracts matched to firm-level data on formal sector jobs. We find that firms operating in a municipality that receives more incoming migrants are less likely to provide health insurance to employees, an effect that is mostly driven by large firms. Despite declines in the provision of non-monetary benefits, which increases labor demand, employment in the formal sector still drops.[1.1.3] Wages in the formal sector reduce across the entire wage distribution but more so for higher wage percentiles which is consistent with binding minimum wages.

For individuals employed in the informal sector or self-employed we show a decrease on earnings between 0.75% and 0.99% mostly concentrated on the bottom third of the wage distribution, consistent with predictions from a two-sector labor market model where wages can freely adjust in the informal sector, and given a less educated migrant workforce that increases competition relatively more among informal workers.

We provide a simple back-of-the-envelope exercise to gauge whether changes in non-wage compensation are important relative to the changes in earnings. When accounting for food subsidies and private health insurance, a $1 p.p.$ increase in the number of migrants reduces *total compensation* by 1.06%. This highlights that, if we do not account for the non-wage compensation margin of adjustment, we will underestimate the change in overall compensation by 44% percent, assuming that workers value such benefits at their cost estimate.

To explore the mechanics behind our results, we investigate whether municipalities where the national minimum wage was more binding at baseline had stronger employment and/or nonwage compensation adjustments. We find that in places with greater minimum wage bite, the adjustment occurs mostly via informal

[1.1.3] Recent literature, as discussed in Clemens (2021), reports a negative correlation between minimum wage increases and health insurance provision, with the variation of this benefit offsetting about 15% of the cost with minimum wage increases.

earnings and formal employment. In regions where the minimum wage is less binding, we find little effect on informal employment and earnings. The adjustment in the formal sector occurs in total compensation - both on earnings and non-wage benefits.

Heterogeneity analysis shows that the estimated effects are stronger for less educated native workers, which is consistent with the fact that they directly compete with Semiarid's migrants. When compared to those with high education, less educated natives are more likely to exit the formal sector and experience a 26% greater wage reduction. Moreover, as more low-education workers earn close or equal to the minimum wage, the negative impact on the most frequently non-wage benefits provided by firms is greater for them. This suggests that welfare declines more for low-income workers, therefore, rising welfare inequality among natives.

Next, we find that unemployment and labor force participation increases by roughly $0.07 - 0.09 p.p.$, which may seem at odds with previous results since earnings fall in the informal sector and benefits drop in the formal sector. By running separate regressions for head and non-head of the household, we find that almost all the impact on the employment margins comes from the head of household while the change in unemployment and inactivity rates are led by the non-head member, consistent with the added worker effect (LUNDBERG, 1985).[1.1.4]

We then turn our attention to the long-term impacts of migration on local labor markets in Brazil. Our results indicate that the estimated effects on average earnings in the formal sector remain mostly the same, but in the informal sector decrease even further. In the case of employment, we see a larger negative impact on formal workers while there are no significant effects on informal jobs. As for non-wage benefits, the impact on health insurance is mostly the same as in the short run, while the negative effects on transport subsidies are larger and there is no significant effect on food benefits. Also, we show that a potential mechanism behind these dynamics is that short-run effects might be partially offset by further

[1.1.4] The "added worker effect"in a broader sense here refers to an increase in the labor supply of secondary earners (typically wives and children) when the primary earner (husbands) becomes unemployed or lose a formal sector job where benefits, sometimes extended to the family, are provided.

internal migration as (mainly low-education) natives respond to the adverse effects by moving to markets that were not directly targeted by migrant arrivals.

Our work is related to a broad literature that examines the impact of migration flows on labor market outcomes of natives (see (BORJAS, 2014) and (DUSTMANN; SCHÖNBERG; STUHLER, 2016) for a review). Despite the fact that migration within countries is a larger phenomenon,[1.1.5] most studies are concerned with international immigration to high-income countries, with particular attention given to Mexican immigration to the United States (BORJAS, 2003a) and, more recently, to immigration to Western Europe (DUSTMANN; FRATTINI; PRESTON, 2012). Some of these studies find that the wages of natives are harmed by immigration (BORJAS; MONRAS, 2017), while others find only a minor negative effect on native wages (CARD, 2001), or even positive (OTTAVIANO; PERI; WRIGHT, 2013; FOGED; PERI, 2016; AZOULAY et al., 2022).[1.1.6] A smaller set of studies explore environmental and other economic shocks to study the causal impact of internal migration on local labor markets in the US, China and Thailand. (BOUSTAN; FISHBACK; KANTOR, 2010; HORNBECK, 2012; IMBERT et al., 2022; BADAOUI; STROBL; WALSH, 2017).[1.1.7] More closely related to our work is Kleemans e Magruder (2018) who study the impacts of internal migration in Indonesia from a two-sector labor market perspective. They show that internal migration reduces employment in the formal sector and earnings in the informal sector.[1.1.8]

Our contribution to the economics of migration literature is fourfold. First, we show that firms systematically adjust non-wage benefits in response to labor

[1.1.5]Rough estimates indicate that global internal migration sits around 740 million (UNDP, 2009), approximately three times the estimated number of international migrants (UN DESA, 2017).

[1.1.6]Dustmann, Schönberg e Stuhler (2016) argue that such often contradictory estimates are a result of (i) different empirical specifications (sources of variation), as well as the fact that labor supply elasticity differs across different groups of natives, and immigrants and native do not compete in the labor market within the same education-experience cells.

[1.1.7]See also Molloy, Smith e Wozniak (2011) for a comprehensive literature review on the determinants of internal migration in the U.S. and Lagakos (2020) on urban-rural internal movements.

[1.1.8]This approach relates to the seminal work of Harris e Todaro (1970). A similar extension and test of this model is provided in Busso, Chauvin e Herrera L (2021) using census data from Brazil.

supply shocks. Second, accounting for such adjustments are key to understanding the effects of migration on natives. Third, we provide evidence on the effects of internal migration on local labor markets in a large developing country and show that these different adjustment patterns are relevant even in the presence of informality. Fourth, we add to a growing body of evidence that migration is a relevant coping mechanism against climate change, especially for vulnerable populations in rural areas of developing countries (SKOUFIAS; VINHA; CONROY, 2013; ASSUNçãO; CHEIN, 2016).

Non-wage benefits are also an important part of compensation in developed countries. In the US, employer-provided health insurance and other benefits account for around one-third of compensation costs (CLEMENS; KAHN; MEER, 2018). 74% of firms in Europe paid non-base wage components such as benefits and bonuses in 2013 (BABECKÝ et al., 2019). Evidence shows that firms adjust non-wage components when facing adverse economic shocks (BABECKÝ et al., 2019) or as a strategy to offset collective bargaining (CARDOSO; PORTUGAL, 2005), particularly when base wages are rigid (BABECKý et al., 2012). We add to this literature by showing that non-wage benefits are an important margin of adjustment in the case of labor supply shocks due to internal migration.

This paper is organized as follows. In the next section, we first present background information on the Brazilian Semiarid region and local labor markets. Section 1.3 outlines a simple framework for interpreting our findings. Section 3.3 describes the data and empirical framework, and reports first-stage estimates that link observed migration patterns to our predicted migration flows. Next, we present and analyze the main results on employment, wages and non-wage wage benefits in Section 3.4. We also study the sensitivity and heterogeneity of our main estimates. Finally, we interpret our main estimates in light of our simple model and conclude.

1.2 Background

In this section, we first describe the economic background and weather conditions in the Semiarid region, the functioning of local labor markets in Brazil, and a simple framework in an effort to contextualize our analysis. We then discuss the main sources of data regarding labor market outcomes, migration flows, and weather, and present some descriptive statistics.

1.2.1 Brazilian Semiarid

The Brazilian Semiarid encompasses 960 municipalities spread over 9 states, covering an area of around 976,000km^2.[1.2.1] According to the official definition by the Ministry of National Integration, a municipality qualifies as Semiarid if at least one of these three criteria holds: (i) annual average precipitation below 800 mm between 1961 and 1990; (ii) aridity index up to 0.5[1.2.2]; (iii) risk of drought above 60%[1.2.3]. The average historical precipitation in the Semiarid is about 780mm, as opposed to around 1,600 mm for the rest of the country[1.2.4], while average temperature is around 25°C. The rainy season occurs between November and April, with the highest levels of precipitation after February, when the sowing seasons typically starts.

Municipalities are relatively small with median population around 20,000 and have economies mainly based on agriculture and cattle ranching in small subsistence properties. Local economic activity is particularly susceptible to weather shocks (WANG et al., 2004), with some studies showing a loss of up to 80% of agricultural production in periods of long drought (KAHN; CAMPUS, 1992). About 80% of the children lived below the poverty line and infant mortality reached 31 per 1,000 births in 1996, compared to a national average of 25% and 15 per 1,000 births, respectively (ROCHA; SOARES, 2015). More than 80% of the adult population had less than 8 years of schooling in 1991.

Such poor socioeconomic indicators associated with periods of extreme

[1.2.1] That is roughly the same as the territory of Germany and France combined. The semiarid comprises 11 percent of the Brazilian territory and includes parts of almost all Northeastern states (except for *Maranhão*) plus the northern area of *Minas Gerais*, but it does not cover any state capital.

[1.2.2] Thornthwaite Index, which combines humidity and aridity for a given area, in the same period.

[1.2.3] Defined as the share of days under hydric deficit, using the period 1970-1990.

[1.2.4] See Figure 1.6.1.

drought have historically driven large outflows of migrants - or so-called *retirantes* - from the Semiarid to other areas of the country (BARBIERI et al., 2010). During the 1960s and 1970s, net migration out of Northeastern states (where most of the Semiarid is located) was 2.2 and 3.0 million individuals (CARVALHO; GARCIA, 2002), which correspond to net migration rates of -7.6 and -8.7%, respectively. Between 1996 and 2010, around 3.0 million people left the Semiarid alone searching for better conditions elsewhere in the country. Appendix Figure B1 shows that these migrants tend to be historically concentrated in some states. São Paulo alone harbored over 30 percent of the people arriving from the Semiarid in the last four decades. However, in relative terms, incoming migrants represented a population increase of above 2% for the top 10 receiving states.

1.2.2 Labor Markets in Brazil

A common feature of labor markets in developing countries is the existence of a two-sector economy where the informal sector accounts for one to two-thirds of the GDP (see Perry et al. (2007) and Ulyssea (2020) for a review). In Brazil, over 40% of individuals work in the informal sector (those without registration or who do not contribute to social security) including the majority of the self-employed who are not protected through social security. When firms hire workers under a formal contract they are subject to several legal obligations, such as paying minimum wages and complying with safety regulations. Registration also entitles workers to other benefits such as a wage contract, which in Brazil prevents downward adjustment, working up to 44 hours weekly, paid annual leave, paternity or maternity leave, retirement pension, unemployment insurance, and severance payments (GONZAGA, 2003; ALMEIDA; CARNEIRO, 2012; MEGHIR; NARITA; ROBIN, 2015; NARITA, 2020).

If firms do not comply with working regulations they may be caught by the labor authorities and have to pay a fine. For example, a firm is fined about one minimum wage for each worker that is found unregistered, or the firm can be fined up to a third of a minimum wage per employee if it does not comply with mandatory contributions to the severance fund (ALMEIDA; CARNEIRO, 2012).[1.2.5] On the other hand, it is a well-known fact that compliant (formal) firms are those more visible to labor inspectors and thus subject to more inspections whereas informal firms are smaller and thus difficult to get caught (CARDOSO; LAGE, 2006). There are also other expected costs for formal firms associated with labor courts in case the worker is fired and decides to file a lawsuit against the firm. Judges decide in favor of workers in nearly 80% of cases (CORBI et al., 2022). This all points to a significant cost of operating in the formal sector, particularly for smaller firms. Imperfect enforcement and costly regulation are associated with high labor informality in the country.

Finally, as there is a strong overlap between the productivity distributions of formal and informal sectors (MEGHIR; NARITA; ROBIN, 2015), even for lower percentiles of the overall distribution, both sectors should be affected by the influx of migrants. In other words, both sectors have workers who are close substitutes to the migrant workforce and thus will experience competition.

Non-wage compensation.

In our empirical analysis, we focus on three main fringe benefits we observe in the data: private health insurance, food, and transport subsidies. In Brazil, benefits became popular in the 1980s, as the provision of food subsidies and employer-provided health insurance became more frequent among private sector firms (ARBACHE, 1995). Data from PNAD surveys for 1996-2009 indicate that 39% of workers in the formal sector receive food subsidies, 36% receive transport subsidies and 21% get private health insurance through their employers. Arbache e Ferreira (2001) estimate that benefits like food subsidies for instance cost around 57% of one minimum wage (around 16% of average total compensation). Similarly, Brazilian Federal Health Agency data (ANS, 2018) show that employer-provided health insurance cost on average R$582 in 2018, which is 17% of total compensation in that same year. These numbers imply that depending on how firms opt to mix benefits in the workers' package, these expenses may add up to above 30% of the

[1.2.5] The minimum wage is above 70% of the median wage in Brazil.

total payroll cost. In the US, benefits including employer-provided health insurance account for around one-third of compensation costs (CLEMENS, 2021).

There are at least two reasons that can explain the use of fringe benefits in workers' compensation. First, these benefits in Brazil are not subject to payroll taxes and therefore reduce total labor costs. Second, labor legislation is generally more flexible regarding the provision of benefits such that it is easier to adjust benefits than wages (ARBACHE, 1995). Even though regulations for fringe benefits provision are considered less rigid than for wages, collective bargaining agreements (CBA) sometimes include clauses pertaining to these benefits. In particular, the third most common clause type among extended firm-level CBA includes wage supplements such as food subsidy (LAGOS, 2020) Also, around 10% of all formal sector firms are under CBA with a clause on health plan/insurance (MARINHO, 2020).

But how important should be adjustments in the non-wage compensation margin in Brazil compared to other settings? Non-wage benefits should play a larger role in countries where minimum wages are high. The minimum wage as a fraction of the median wage ranges is lower in the United States (32%), Japan (44%), and Mexico (46%), than in Brazil (75%) and many European countries, for example, the UK (55%) and France (61%).[1.2.6]

Although transport subsidy is a mandated benefit in Brazil since 1985, we treat this as a benefit that firms can adjust. This is likely the case since we observe that only 36% of formal sector workers report they receive this benefit. That is, firms may not fully comply with all aspects of labor regulations. Also, as transport benefit is non-wage compensation, firms do not incur payroll taxes. In addition, firms may deduct the cost with the offered subsidy from the base for income taxation as well as from their operational cost, lowering net revenues which are the base for other corporate and payroll taxes.[1.2.7] This implies that firms have incentives to offer transport benefits and a further incentive to adjust it at the intensive margin by providing better means of transportation or increasing the

[1.2.6] Brazil (Source: PNAD Continua 2015). Data for OECD countries in 2019 (Source: OECD.Stat, https://stats.oecd.org/Index.aspx?DataSetCode=MIN2AVE

[1.2.7] The income tax due cannot be reduced by more than 10%.

benefit in cash.

1.3 A Simple Theory

In this section, we describe a simple model assuming perfectly competitive labor markets to guide our analysis. We assume migrants and natives to be perfect substitutes and investigate the consequences of a migration shock that shifts the aggregate labor supply to the right. Then we introduce intersectoral linkages where formal and informal workers are substitutes.

In a standard competitive market model, where the wage is determined such as labor supply equals labor demand, migration negatively affects wages. In the extreme case in which supply is inelastic, migration has no effect on the employment of natives and all migrant workforce is absorbed. On the other hand, with an upward-sloping labor supply, the decline in wages makes jobs less attractive for some native workers such that native employment falls.

Of course, this benchmark is a vast simplification of how the labor market works in practice. Downward wage rigidities are often present in reality due to minimum wage laws and collective bargaining agreements. In this case, migration shocks can be accommodated by job losses or lower labor costs, for example, reducing non-wage benefits (MCKENZIE, 1980; CLEMENS, 2021).

Also, we have considered that the formal and informal sectors are independent, which masks important intersectoral linkages, in particular, on the production side (see (ULYSSEA, 2010; ULYSSEA, 2018) and (BOSCH; ESTEBAN-PRETEL, 2012)). When the two types of labor are highly substitutes, informal employment and wages can compensate for wage rigidities in the formal sector. This should be particularly relevant to understanding the implications of an increase in migration in the labor market where formal wages are rigid thereby increasing the importance of the informal sector as an outside option.

To address such context, we develop a simple model with informality in which the formal sector is subject to the minimum wage and offers non-wage benefits that are frequently observed in the data (e.g. health insurance and food subsidy). Our starting point is the seminal contribution of Harris e Todaro (1970). We follow their model, in which minimum wage is the main institution behind the existence of a two-sector economy. We add non-wage benefits in the formal sector as a source of adjustment of total compensation in the presence of minimum wages.[1.3.1] Appendix 1.A presents the model and here we summarize its main predictions.

In this model, the effects of migration may depend on the degree of substitution between formal and informal labor inputs in production and on the non-wage benefits margin. Considering that migration exogenously shifts the supply of workers to the informal and formal sectors at the destination, our model has clear predictions regarding the direction of effects of migration on employment by sector, unemployment, formal sector non-wage benefits, and informal wages.

Table 1.A.1 simulates the impact of a migration shock when non-wage benefits are flexible (column 2) or fixed (column 3). It does so by assuming that production linkages are low, medium, and high, in Panels A, B, and C, respectively. In the benchmark economy with medium production linkages across sectors and flexible non-wage benefits, migration increases unemployment and informal employment but decreases non-wage benefits and informal sector wages. Formal employment remains unchanged. With non-wage benefits fixed at baseline levels, our main results show that migration has now a negative impact on formal employment as expected since formal firms cannot adjust benefits after the supply shock from migration. Consequently, unemployment and the informal sector adjust more. Under lower intersectoral linkages, the qualitative results are the same with stronger effects in informal markets and formal sector benefits.

In sum, in a simple competitive model with no rigidities and absent non-wage benefits, which is likely closer to the informal (unregulated) sector case, we expect negative effects of migration on wages and possibly on the employment of natives depending on the labor supply elasticity. As we consider that formal and

[1.3.1] We abstract from other sources of labor market frictions, which are explored in much recent work on models of the labor market with monopsony to study immigration effects (e.g. (AMIOR; MANNING, 2020) and (AMIOR; STUHLER, 2022)). These are not needed to understand the mechanisms we emphasize, so we proceed with a simpler approach accounting for unemployment, two employment sectors, and intersectoral linkages.

informal labor inputs can be substitutes in production, minimum wage regulation, and the possibility of adjustment in non-wage benefits in the formal sector, we find that the model yields predictions that are largely consistent with a decline in formal employment of natives and in benefits offered in this sector. This reduction would be even larger if no adjustment in non-wage benefits were possible. On the other hand, as the informal sector absorbs a larger fraction of migrants, wages fall.

Empirically, these forces are likely to affect mainly those workers at the bottom and medium of the formal wage distribution since migrants are generally low skilled.[1.3.2] Low-skill workers receive disproportionately more generous benefits (e.g. health insurance and food vouchers) but have wages that are more prone to be affected by minimum wage policies and collective bargaining agreements. Even if many firms employ both high and low-skilled workforce which may produce spillover effects due to complementarities, the impact on high-skilled workers is likely to be of second order.

1.4 Data and Empirical Strategy

In this section, we begin by listing the main sources of data used in our analysis and showing some descriptive statistics. Then we describe the empirical framework and report first-stage estimates that link observed migration patterns to our predicted migration flows.

Migration

We draw data from three waves of the Brazilian Census (1991, 2000 and 2010), provided by the *Instituto Brasileiro de Geografia e Estatística (IBGE)*, to construct two of the main variables used in out study.[1.4.1] First, we leverage Census

[1.3.2]This follows the arguments developed by the labor market model in Card e Lemieux (2001) and Borjas (2003b). In Section 3.3, we present descriptive evidence that supports greater substitutability between migrants and less skilled natives in the labor market.

[1.4.1]As several municipalities were split into new ones during the 1990s, we aggregate our data using the original municipal boundaries as they were in 1991 (so-called "minimum comparable areas" or MCA) in order to avoid potential miscoding regarding migration status or municipality answers about the municipality of origin and year of migration to construct a measure of yearly migration outflow from each municipality in the Semiarid and a measure of inflow to each destination (all but Semiarid) during the 1996-2010 period. Second, we use the 1991 Census to build a "past settlement" measure by associating the share of migrants from each Semiarid municipality who resides in each destination. In Appendix 1.B we provide more details on how we structure our yearly migration dataset.

Weather shocks

Weather data were retrieved from the Climatic Research Unit at the University of East Anglia (HARRIS et al., 2020). The CRU Time Series provides worldwide monthly gridded data of precipitation and temperature, at the $0.5° \times 0.5°$ level ($0.5°$ is around 56km on the equator). We construct municipality-level monthly precipitation and temperature measures based on grid-level raw data as the weighted average of the municipality grid's four nodes using the inverse of the distance to the centroid as weights.[1.4.2] We define rainfall shocks as deviations from the historical average.[1.4.3]

Labor outcomes

Our primary source of data for the outcome variables is the *Pesquisa Nacional por Amostra de Domicílios - PNAD*, a household survey also collected by the *IBGE*, the bureau responsible for conducting the Census. The survey is conducted yearly, except during Census years. Thus our data spans from 1996 to 1999 and from 2001 until 2009. Interviews take place in 808 municipalities in all 27 Brazilian states and cover several dimensions such as education, labor, income, fertility, and household infrastructure. On average, around 300,000 people are interviewed in each round.

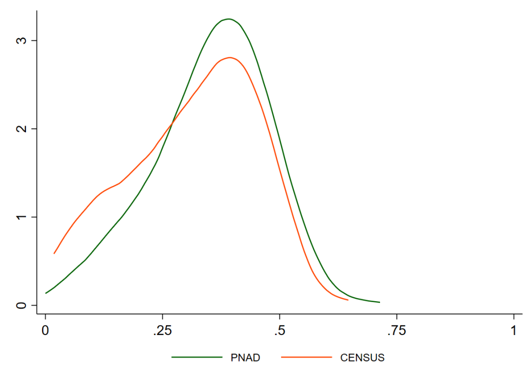
of origin. We use municipality and MCA as synonyms throughout the paper.

[1.4.2] This approach is similar to the one used by)rocha2015water.

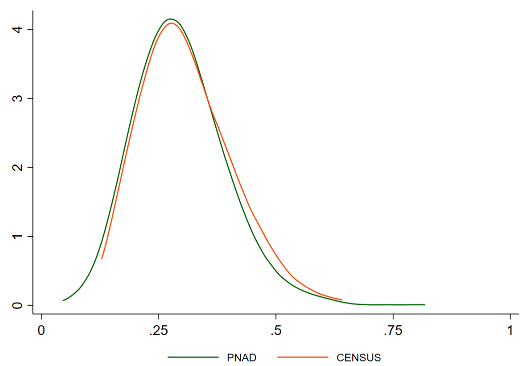
[1.4.3] See Appendix 1.C for a detailed description and discussion on this measure.

Two features of *PNAD* are worth pondering regarding the suitability of the data for our analysis. First, less than 20% of all municipalities are comprised in the survey. However, they represent nearly 80% of the destination choice of migrants from the Semiarid region and 65% of the employed population in Brazil. Second, *PNAD* is not designed to be representative at the municipality level. To reassure that our primary outcome measures from *PNAD* are valid, we compare data from *PNAD* 2009 with Census 2010. In Figure 1.4.1 we compare the distribution of municipality-level employment rate in the formal and informal sectors from the 2009 round of *PNAD* with the distribution for the same *PNAD* municipalities obtained from the 2010 Census. In Figure 1.4.2 we show the log earning distributions from *PNAD* and Census in the same years. Both measures across formal and informal labor markets overlap very significantly. Unfortunately, the Census does not provide any information on non-wage benefits.

Figura 1.4.1 – Distribution of employment rates, PNAD vs. Census



(a) Formal sector

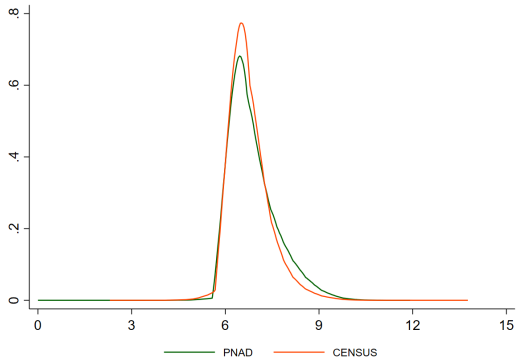


(b) Informal sector

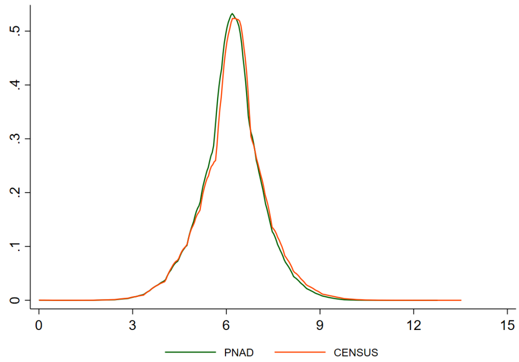
Notes: Municipality-level employment rate in the formal and informal sector from PNAD (2009) and Census (2010).

Our main outcomes comprise earnings and employment. We explore whether the worker is an employee in the registered formal sector, informal sector or self-employed; and whether she is unemployed or out of the labor force. We also use information on non-wage compensation. The survey asks specifically whether the individual received any kind of payment or help to cover expenses with food, and

Figura 1.4.2 – Distribution of log earnings, PNAD vs. Census



(a) Formal sector



(b) Informal sector

Notes: Log earnings in the formal and informal sector from PNAD (2009) and Census (2010).

transport and if the employer provides health insurance. Unfortunately data on the intensive margin of nonwage compensation is unavailable.[1.4.4]

We restricted our attention to individuals between 18 and 65 years old, living in the municipality for 10 years or more and we refer to them as *natives*. We consider as destinations all the PNAD municipalities that are not in the Semiarid in order to minimize concerns about spatial correlation in weather shocks. We pool the 13 years of individual survey data and take averages at the municipality-year level. The final destination sample has 684 unique municipalities and 8,190 municipality-year observations, averaged from 2,152,950 individuals.

Table 1.6.1 describes municipality-level data for origin (Panel A) and destination (Panel B) municipalities. Semiarid's areas show lower levels of rainfall, and slightly higher temperatures and are less populated than destination municipalities. On average, 1.0 p.p. of Semiarid's population leaves every year, resulting in an average increase of 0.30 p.p. of the labor force in the destination.

Table 1.6.2 provides descriptive statistics for destination municipalities. In our sample, 63% of individuals are employed - with 31% having a formal job, the

[1.4.4] In section 1.5.2 we provide a back-of-the-envelope calculation of the impacts of internal migration on total compensation for native workers, using the average cost of food subsidies and private health insurance.

same proportion of informal workers. The unemployment rate is 13% and 24% of individuals are not in the labor force. The average monthly earning is R$ 637.89, with the formal sector having a substantially higher average (R$ 788.22) than the informal sector (R$ 491;28).[1.4.5] Among workers employed in the formal sector, 39% receive financial help to cover expenses with food, 36% for transport and 21% for health expenditures.[1.4.6]

Finally, Table 1.6.3 compares migrants to low- and high-education natives. Migrants are slightly more educated and earn slightly less than less-educated natives. They also have a similar likelihood of working part-time and being in the formal sector when compared to low-education natives. On the other hand, high-education natives are more likely to work in the formal sector and have considerably higher pay. Appendix Table B1 shows that top occupations for migrants (e.g. typically bricklayer for men, domestic worker for women) are also top occupations for low-education natives, but not for the skilled. Also, the same five industries that concentrate over 80% of working migrants also employ a similar share of low-education workers (see Appendix Table B2). Overall, this characterization is consistent with greater substitutability between migrants and less skilled natives in the labor market.

1.4.1 Empirical Strategy

Here we first describe the empirical framework that allows us to (i) isolate the observed variation in migration induced by exogenous weather shocks, and (ii) the migration flows into destination municipalities determined by past settlements. Next, we discuss and present supportive evidence on the validity of this shift-share instrument approach based on insights from the recent econometric literature that analyzes its formal structure.

We specify a model for the changes in labor market outcomes of native individuals as a function of internal migration flows. Specifically, we assume that

[1.4.5] Earnings are measured in R$ (2012).
[1.4.6] Less than 1% of informal and self-employed workers receive any kind of non-wage compensation.

$$\Delta y_{dt} = \alpha + \beta m_{dt} + \gamma X_d + \psi_t + \epsilon_{dt} \tag{1.4.1}$$

where y_{dt} is a vector of labor outcomes at destination municipality d in year t, m_{dt} is the destination migrant inflow from the Semiarid region, X_d are destination-level controls interacted with time dummies, ψ_t absorb time fixed effects and ϵ_{dt} is the error term. The main challenge to identifying β is that the observed migration, m_{dt}, is the equilibrium between the demand and supply of migrants. Another issue is that the error term, ϵ_{dt}, may include unobserved characteristics that could be correlated with migration inflows. In particular, migrants could choose a specific destination municipality due to demand shocks leading to higher wages or job prospects. By differencing the outcome variables we can account for time-invariant unobserved characteristics that could be correlated with migrant inflows, but not the time-varying confounders which would potentially bias OLS estimates.

We account for this endogeneity problem following a two-step procedure to construct an instrumental variable for the number of migrants entering a destination. First, we predict m_{ot}, the migration outflow rate[1.4.7] from origin municipality o in year t, using weather shocks in the previous year:

$$m_{ot} = \alpha + \beta' Z_{ot-1} + \phi_o + \delta_t + \varepsilon_{ot} \tag{1.4.2}$$

where Z is a vector of rainfall and temperature shocks at the origin municipality o in the previous year, ϕ_o and δ_t are municipality and year fixed effects, respectively, and ε_{ot} is a random error term. For each year the predicted number of migrants who leave their hometowns is obtained by multiplying this predicted rate by the municipality population reported in the 1991 Census:

$$\widehat{M}_{ot} = \widehat{m}_{ot} \times P_o \tag{1.4.3}$$

In the second step, we use the past settlements of migrants from the origin o

[1.4.7] Defined as the observed number of migrants leaving the municipality divided by the population in the 1991 Census.

to municipality d in order to distribute them throughout the destination areas. More formally, we define this share of migrants from origin o settling into destination d in 1991[1.4.8]

$$s_{od} = \frac{M_{od}}{\sum_d M_{od}} \tag{1.4.4}$$

allowing us to define our shift-share instrumental variable (SSIV) as

$$\widehat{m}_{dt} = \sum_{o=1}^{O} \frac{s_{od} \times \widehat{M}_{ot}}{N_d} \tag{1.4.5}$$

where N_d is the total native population at d in 1991. Thus our instrument $\widehat{m}_{dt}$ can be thought of as a combination of exogenous shocks or 'shifts' $\widehat{M}_{ot}$ (weather-driven outflows) with 'shares' ($s_{od} \geq 0$) or 'ethnic enclaves' as in Card (2001).[1.4.9]

The validity of the shift-share instrument approach relies on assumptions about the shocks, exposure shares, or both, as discussed by recent literature which analyzes its formal structure. Goldsmith-Pinkham, Sorkin e Swift (2020) demonstrate that a sufficient condition for consistency of the estimator is the strict exogeneity of the shares. Alternatively, Borusyak, Hull e Jaravel (2021) show how one can instead use the exogenous variation of shocks for identification by estimating a transformed but equivalent regression - at the origin level in our setup - where shocks are used directly as an instrument.

Based on these insights, we leverage origin-level weather shocks[1.4.10] for identification and define the reduced-form relationship that associates labor market outcomes and the predicted migrant flow at the destination as

$$\Delta y_{dt} = \alpha + \beta \widehat{m}_{dt} + \gamma X_d + \psi_t + \epsilon_{dt} \tag{1.4.6}$$

[1.4.8] We fix our past settlement measure in 1991 across the time span of our sample so as to avoid concerns about the persistence in migrant flows as discussed by Jaeger, Ruist e Stuhler (2018). We also experimented with a specification that updates the past settlement using the data from the immediate previous Census and the results are similar.

[1.4.9] In appendix 1.C we discuss this approach in more detail.

[1.4.10] Figure 1.6.2 illustrates the variation in weather shocks that use for identification.

We follow Borusyak, Hull e Jaravel (2021) and calculate an origin-level weighted average version of equation 1.4.6, that uses the exposure shares s_{od} as weights, and results in the transformed reduced form relationship

$$\bar{y}_{ot} = \alpha + \beta \widehat{M}_{ot} + \bar{\varepsilon}_{ot} \tag{1.4.6'}$$

In Appendix 1.D we provide a detailed derivation of the transformation performed. As discussed shown in Borusyak, Hull e Jaravel (2021), the consistency of our shift-share approach is based on two conditions:

Assumption 1(*Quasi-random shock assignment*): $\mathbb{E}[Z_{\tilde{o}}|\bar{e}, s] = \mu$ for all $\tilde{o}$.

Assumption 2 (*Many uncorrelated shocks*): $\mathbb{E}[\sum_o s_o^2] \to 0$ and $Cov[Z_{\tilde{o}}, Z_{\tilde{o}'}|\bar{e}, s] = 0$ for all $\tilde{o}, \tilde{o}'$.

where $\tilde{o} = (o, t)$, $\bar{e} = \{\bar{e}_{\tilde{o}}\}_{\tilde{o}}$, $s_o = \sum_d s_{od}$ and $s = \{s_o\}_o$.[1.4.11] Assumption 1 guarantees that our shift-share IV is valid when weather shocks are as-good-as-randomly assigned, which comes from standard natural shocks arguments. Given identification, Assumption 2 gives us consistency when the number of observed shocks is large and when shocks are mutually uncorrelated given the unobservables and s_o. In the Appendix Table D1 we show that the effective sample size[1.4.12] is sufficiently large, reassuring us that exposure concentration is not a relevant issue in our setting. Also, in Appendix Tables E2-E5 we present evidence that the shocks we are using can be treated as uncorrelated, which supports the validity of Assumption 2 in our setting.

One additional advantage of using the origin-level shocks concerns hypothesis testing. Adao, Kolesár e Morales (2019) show that conventional inference in shift-share regressions is generally invalid because observations with similar exposure shares are likely to have correlated residuals, potentially leading to null hypothesis

[1.4.11]As in Borusyak, Hull e Jaravel (2021), $\bar{e}_{ot}$ correspond to the error term from equation 3.3.2 computed at the level of shocks (e.g. municipality of origin).

[1.4.12]According to Borusyak, Hull e Jaravel (2021), the effective sample size is the inverse of the Herfindahl index of concentration of migrants: $H = \frac{1}{\sum_o s_o^2}$.

overrejection. Borusyak, Hull e Jaravel (2021) show that by using the shock-level relationship instead of the destination-level one can obtain standard errors that converge to those obtained by the Adao, Kolesár e Morales (2019)'s correction procedure.

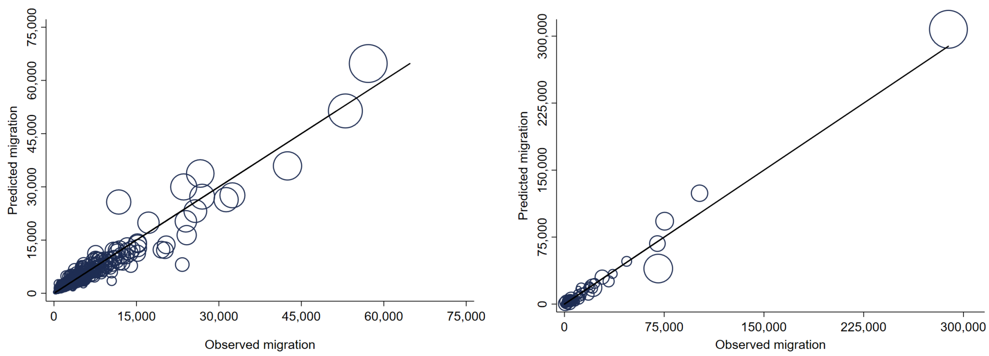
1.4.2 Weather-induced Migration

We begin the exploration of our first-stage results by estimating variations of specification 3.3.3 and report the estimates in Table 1.6.4. All regressions control for temperature shocks and the log of the total population in the previous census; and include time and municipality fixed effects. In columns (2)-(8) we include a flexible trend interacting time dummies with 1991 characteristics (age and the shares of high school and college-educated individuals). Columns (3)-(6) include up to three lags, contemporaneous and one lead of rainfall and temperature shocks. For brevity, we omit (mostly insignificant) coefficients associated with temperature shocks in Table 1.6.4. Standard errors are clustered at the grid level to account for the fact that municipalities in the same grid will have similar shocks.[1.4.13]

As expected, rainfall shocks in the previous year are negatively correlated with migration outflows indicating that Semiarid's inhabitants leave the region during drought periods. Coefficient estimates are remarkably stable across specifications and adding more lags does not change the baseline results. More important to our identification, we include as control rainfall and temperature shocks one year forward to ensure that our instrument is not contaminated by serial correlation in the weather measures. The coefficient on $rainfall_{t+1}$ reported in column (6) is small in magnitude and not statistically significant, while the coefficient for $rainfall_{t-1}$ remains almost unchanged. Our estimates indicate that a municipality where annual rainfall is 10% below the historical average will experience an increase of $1p.p.$ in migration outflow rate.

[1.4.13] Similar, but not identical, as shocks are computed by taking the average of the grid's four nodes, weighted by the inverse of the distance from each node to the municipality centroid. Therefore, two municipalities inside the same grid have different shocks because the distance to the centroid is not the same.

Figura 1.4.3 – Observed vs predicted migration



(a) Migration outflow (b) Migration inflow

Notes: This figure presents the relationship between the predicted and observed migration flows across Brazilian municipalities from 1996 to 2010. Panel (a) shows the number of migrants leaving the Semiarid region to non-Semiarid municipalities. Panel (b) shows the number of incoming Semiarid migrants for destination municipalities. The circle size represents the municipality's total population in 1991. *Data source*: Census microdata (IBGE).

Next, we distribute the predicted migration outflow shock using past settlement patterns of migrants from the origin municipality o to destination d. A *sine qua non* requirement implicit in our empirical framework is that both predicted migration outflow and inflow rates, $\widehat{m}_{ot}$ and $\widehat{m}_{dt}$ respectively, should be strongly correlated with their observed counterparts. Figure 3.5.5 illustrates that our predictions provide a strong fit of the observed migration. Panel (a) shows the relationship between the predicted and observed number of migrants leaving the Semiarid region and entering non-Semiarid municipalities, accumulated over the period 1996-2010. Panel (b) shows the predicted and observed numbers of incoming Semiarid migrants for destination municipalities.

In Appendix 1.C we describe in more detail our data source for weather shocks, discuss alternative measures of weather, and present further details about how we constructed our instrument including predicted and past settlement patterns.

Overall, this analysis shows that our strategy provides a strong first-stage as predicted migration rates, $\widehat{m}_{dt}$, are strongly correlated with observed migration. Appendix Table D1 reveals that our first-stage point estimates are close to a one-to-one relationship (0.92) - making the magnitude of reduced-form and IV estimates almost identical - and have an F-stat of 2,275.[1.4.14]

1.5 Labor Market Effects of Migration Inflows

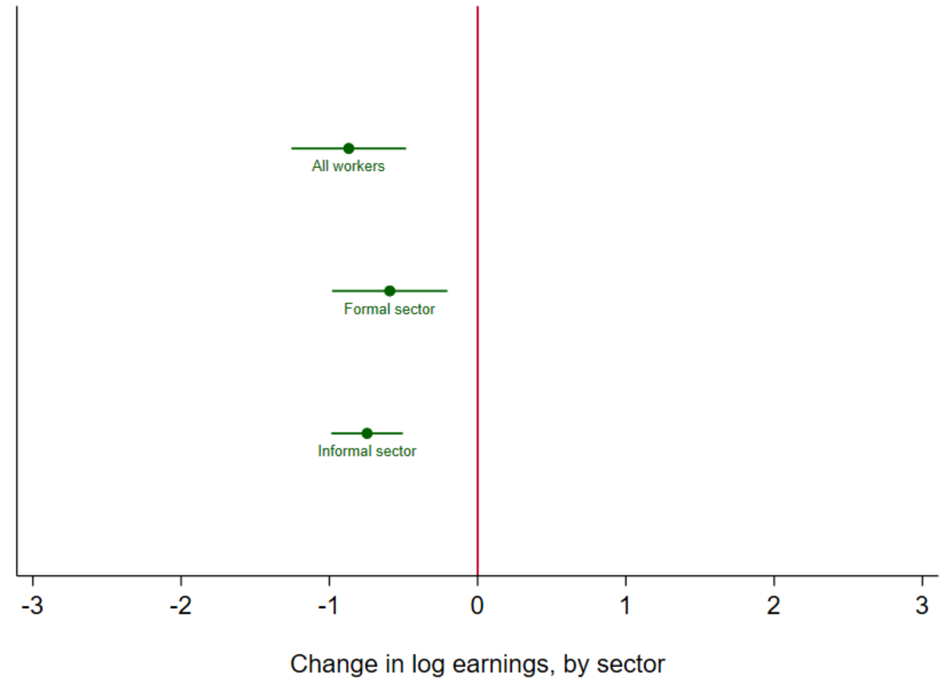
Now we turn our attention to labor markets at the destination and investigate how migration inflows affect earnings, employment, unemployment, and labor force participation of native workers. Next, we explore how labor markets adjust to migration shocks in terms of non-wage compensation.

We begin by investigating how native workers' earnings adjust to exogenous migration inflows. Table 1.6.5 reports several specifications for our SSIV estimates. Column (1) displays a flexible specification, without any control. In column (2) we include time dummies and in column (3) we also control for a vector of destination-level characteristics measured in 1991 (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors are clustered at the origin municipality level.

Panel A reveals a negative effect of the inflow of Semiarid's migrants on average log earnings for native workers. Adding covariates lowers somewhat the magnitude of our estimates but does not change substantially our main conclusions. One percentage point increase in the number of migrants reduces earnings by 0.87%. In Panel B we restricted our analysis to native workers holding a formal job, while in Panels C we focus on those in the informal sector, including workers who are self-employed. We find that a one percentage point increase in the inflow of migrants reduces the earnings of formal workers by 0.59% and by 0.75% for those employed in the informal sector.

[1.4.14] A sufficiently high F-stat avoids weak instrument concerns, especially in the light of the recent discussion in Lee et al. (2020) who show that a 5 percent test requires an F statistic of 104.7, significantly higher than the broadly accepted threshold of 10.

Figura 1.5.1 – **Effects of internal migration on earnings**



Notes: This figure plots origin-level SSIV coefficients on change in log earnings, by sector. The informal sector includes self-employed workers. Controls include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Green markers are statistically significant at the 5% level.

Figure 1.5.1 summarizes our main findings. Any downward wage restrictions, such as minimum wages or collective bargaining agreements, may alleviate the impacts of the incoming migration on earnings for natives employed in the formal sector. The negative effect on informal earnings is greater, albeit only slightly. That is consistent with the absence of downward wage rigidity in this sector such that the classic predictions from perfect competition prevail.

The magnitude of our estimates is larger than that found by Imbert et al. (2022) (0.15%). An important difference between our approaches is that their estimates come from changes over a six years period, while we use year-on-year changes. On the other hand, our estimates are lower than those from Kleemans e Magruder (2018) (0.97%). In their paper, they estimate the effects on the log income per hour - so they may also be capturing the adjustment on hours worked - while we focus on log monthly earning. Similarly, they also find a larger effect on the income of informal workers compared to those in the formal sector.

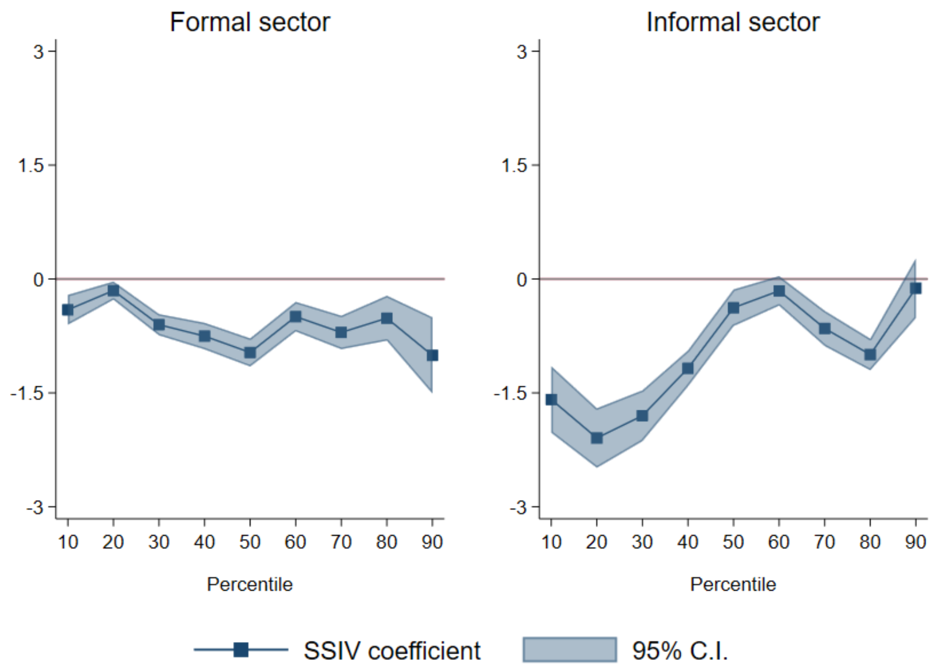
We also investigate the differential effects according to the native worker's position in the earnings distribution. We calculate the municipality-level thresholds of each decile of the earnings distribution, separately for each sector, and run a regression on the changes of each threshold. Figure 1.5.2 reports our results. For those workers employed in the formal sector, we find smaller impacts at the bottom of the distribution. This is consistent with wage rigidity in the formal sector which limits the negative impacts for low-paid workers. The entry of low-skilled migrants competing for less paid or low productivity jobs in this sector can be enough also to reverberate and have amplified effects all the way up the earnings distribution. This seems consistent with the reallocation of at least part of native workers from less productive to more productive employers in the formal sector, increasing competition at the top of the distribution.[1.5.1]

For informal workers, the impact is substantially stronger for those at the bottom third of the distribution, consistent with classic predictions from perfect competition and greater substitutability between migrants and less skilled natives in this sector. To a smaller extent, migration also affects higher earnings deciles of informal sector workers and self-employed. The negative impact of migration, in this case, may be attenuated due to some formal sector workers moving into informality or self-employment. As workers in the formal sector are more productive, on average, this increases earnings at higher percentiles in other sectors.

Our results for employment are summarized in Figure 1.5.3. While we find no effect on overall employment, the inflow of migrants from the Semiarid does change the composition of workers across sectors. Table 1.6.6 reports the point estimates across all specifications. Our estimates in Panels B and C imply that a one percentage point increase in the inflow reduces the share of formal employment by 0.13*p.p.*, and increases the share of informal by almost the same amount (0.11*p.p.*).

[1.5.1] Similarly, Foged e Peri (2016) show that an increase in the supply of low-skilled migrants pushed less educated native workers to pursue less manual-intensive occupations in Denmark.

Figura 1.5.2 – Effects of predicted migration along the earnings distribution

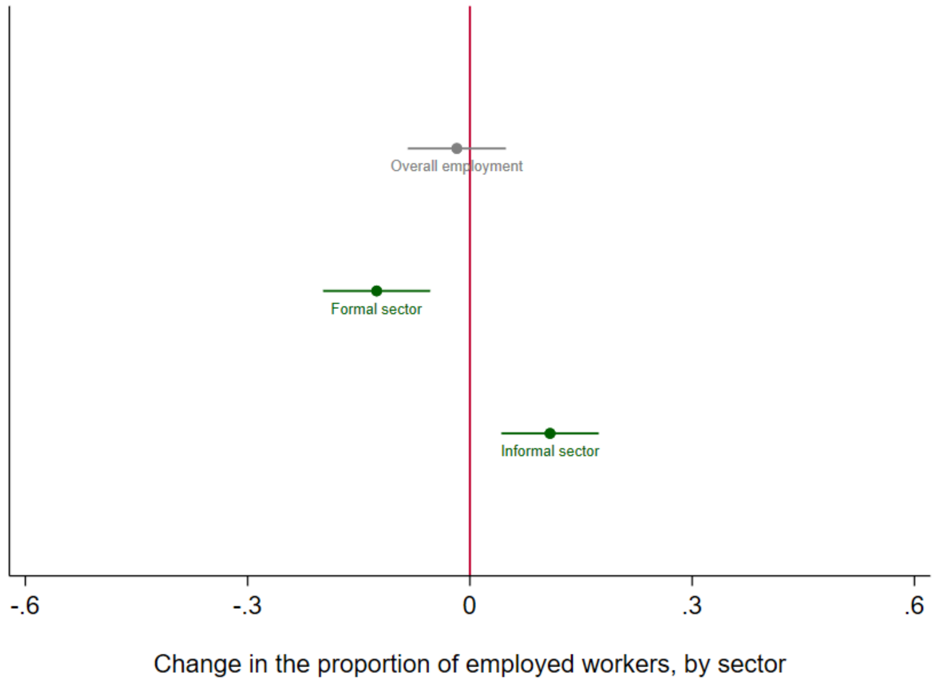


Notes: This figure plots SSIV coefficients of change in the average of log earnings, in each decile, by sector. The informal sector also includes self-employed workers. Controls include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies.

This results is largely consistent with the predicted fall (increase) in formal (informal) employment following the two-sector labor market model presented in Section 1.3. We argue that the existence of medium production linkages across formal and informal sectors, and some degree of rigidity in the adjustment of non-wage benefits can account for these findings.

Kleemans e Magruder (2018) find a larger negative effect on formal employment (-0.33p.p.) and no significant impact on the informal sector. In contrast, we find evidence consistent with the reallocation of workers between sectors. This is likely because migrants are mostly high-educated individuals in their setting, while in our case, they are more likely to be low-educated and, therefore, more likely to

Figura 1.5.3 – **Effects of internal migration on employment**



Notes: This figure plots origin-level SSIV coefficients of change in employment rate, by sector, measured as a fraction of the native working-age population in 1991. The informal sector includes self-employed workers. Controls include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Green markers are statistically significant at the 5% level.

compete with low-educated native workers.

To draw a more complete picture we also estimate the impacts on unemployment and labor force participation reported in Table 1.6.7. Migration inflows lead to an increase of 0.09*p.p.* in the unemployment rate and a decrease of 0.08*p.p.* in the proportion of out-of-labor-force individuals. What mechanism accounts for these estimates is ex-ante unclear. On one hand, increased competition in the labor market could discourage native individuals to work if wages or benefits fall, as predicted by the model developed in the Appendix 1.A. On the other hand, if the primary earner in the household loses his/her job because of the increased competition, then it is possible that other members of the household would enter the market, a phenomenon known as the added worker effect (LUNDBERG, 1985).

We test this second mechanism by running the same regressions separately for individuals identified as head or non-head of the household. According to Table 1.6.8, almost all of the employment effects come from the head of households, while the changes in unemployment and inactivity rates stem from non-head members. That suggests that the second channel prevails. Also, the symmetry between the effects on unemployment and inactivity indicates that once secondary earners enter the market, it takes time for them to find a job.

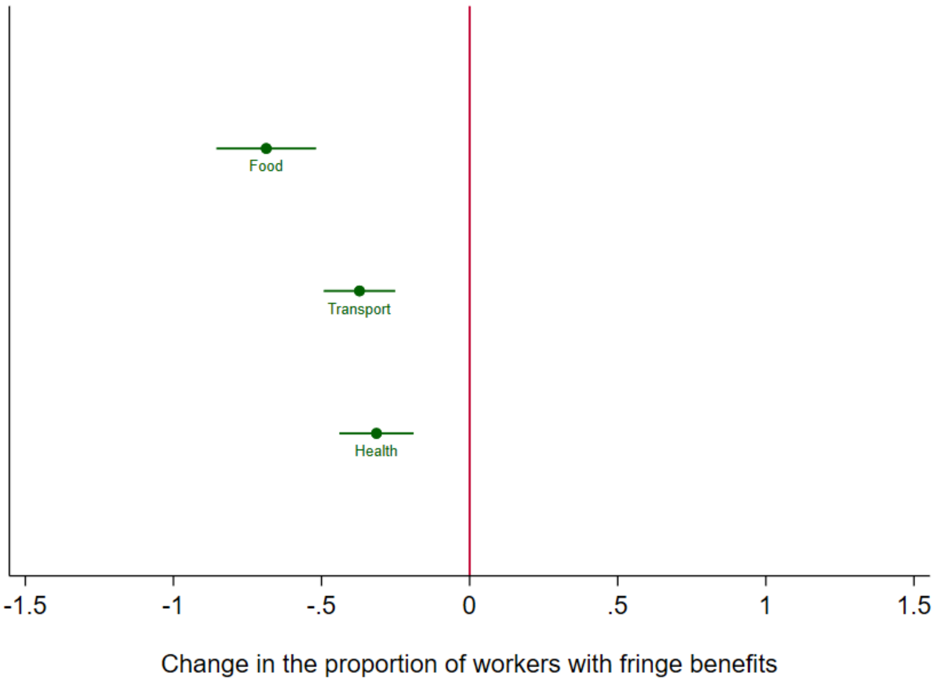
1.5.1 Non-wage Compensation.

We now explore an additional margin of adjustment due to migration shocks. As firms operating in the formal sector cannot reduce wages below the legal minimum, they may adjust to labor supply shocks by reducing fringe benefits as discussed in Section 1.2.2. We focus on individuals who are currently holding a formal job because these benefits are almost exclusively offered by formal firms. Figure 1.5.4 reports the estimates. A one percentage point increase in the predicted number of migrants reduces the share of workers receiving food subsidy between 0.33*p.p.* and 0.69*p.p.*, transport between 0.37*p.p.* to 0.52*p.p.*, and health insurance in the range of 0.31*p.p.* to 0.47*p.p.*. See Table 1.6.9 for the underlying point estimates.

Next we complement these estimates by focusing on the behavior of firms as providers of health insurance to their employees.[1.5.2] Instead of relying on survey data, here we turn to firm-level administrative data on health insurance contracts obtained from *Agência Nacional de Saúde Suplementar (ANS),* the Brazilian regulatory agency responsible for overseeing the private health industry. They provide information about every employer-sponsored contract signed going back as far as 1940. We have data on the date when the contract was signed and the firm unique identifier, which we can use to merge with *RAIS*, an employer-employee matched dataset obtained from the Ministry of Labor, that provides firm-level

[1.5.2] 20% of workers get private health insurance through their employers. In 2018, the average employer-provided health insurance benefit cost on average R$582, or 17% of total compensation in that same year (ANS, 2018). See section 1.2.2 for more details.

Figura 1.5.4 – **Effects of internal migration on non-wage compensation**



Notes: This figure plots SSIV coefficients on change in the proportions of formal sector workers who receive health insurance, food, or transport subsidies. Controls include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Green markers are statistically significant at the 5% level.

data on the near universe of formal employment contracts. We define an indicator variable $y_{idt} = \mathbb{1}\,(t \geq t^s)$ for each firm i in the destination municipality d at year t, with t^s being the year when the health insurance is hired. Then we estimate how migration inflow rates at destination municipality d affect changes in y_{idt}, that is, the variation in the likelihood that firm i provides health insurance to its employees. Formally, the estimated equation is

$$\Delta y_{idt} = \alpha + \beta \widehat{m}_{dt} + \phi_i + \delta_t + \mu_{idt} \tag{1.5.1}$$

where $\widehat{m}_{dt}$ is the predicted incoming migration rate; ϕ_i and δ_t are firm and year dummies.

In column 1 of Table 1.6.10 we find that firms operating in a municipality that receives more incoming migrants are on average less likely to provide health insurance to employees.[1.5.3] An increase of one standard deviation in migration rate of 1p.p. reported in Table 1.6.1 implies a 1.5p.p. decrease in the share of firms that provide health insurance, roughly the average of y_{idt}. In Columns 2-5 we restrict the sample to different bins of firm size. The effect is close to zero and insignificant for firms below 100 employees, but negative and of greater magnitude for larger firms. Firms above 100 employees are at least 6 times more likely to provide health insurance as part of compensation.

1.5.2 Total Compensation.

To address whether changes in non-wage compensation are important relative to the changes in earnings – and which types of compensation are more important – we would need knowledge of the willingness to pay measures for each benefit in such context. One way to calculate the change in non-wage compensation in dollar equivalents is to attach a dollar value to each type of non-wage compensation. For example, one could value food subsidy as the firm's average subsidy cost. Under this assumption, i.e., that individuals value each benefit at its cost ($value =$ average wage cost), we estimate the variation in total compensation, ΔTC_{dt}, as follows:

$$\Delta TC_{dt} = \Delta W_{dt} + \sum_i value_i * \Delta B^i_{dt} \tag{1.5.2}$$

where ΔW_{dt} is the wage variation and ΔB^i_{dt} the variation in the share of benefit type i.

We use the average cost estimates provided in Arbache e Ferreira (2001) for food subsidy (57% of one minimum wage or 45% of average formal wage in

[1.5.3] All the regressions are weighted by the number of employees in the firm in 1996, the first year in our sample.

our sample). We also use the cost of employer-provided health insurance from the Brazilian Federal Health Agency data (ANS, 2018) which is R$582 in 2018 or 52% of the average formal wage in our sample. The estimated effects on formal earnings and on the fraction of non-wage benefits are obtained from column 3 in Tables 1.6.5 and 1.6.9, respectively.

This simple exercise shows that a 1p.p. increase in the predicted number of migrants reduces total compensation by 1.06% when accounting for food subsidies and private health insurance. Compared to the effect on the formal wage, which declines by 0.59%, this means that if we do not account for the change in non-wage compensation we underestimate the change in overall compensation by 44% percent, at least when workers value such benefits at their cost estimate.

1.5.3 Exploring the Mechanism

As discussed earlier in this section, earnings fall in both sectors. However, the magnitude of this fall might depend on the extent of the labor market rigidity due to minimum wages. In places with a more binding minimum wage, the adjustment should occur mostly via formal employment and informal earnings. In regions where the minimum wage bite is lower, it works primarily through total compensation, while the impacts on employment would be limited. Finally, non-wage compensation may help introduce some wage flexibility which may, in principle, reduce job losses.

In order to provide a formal test for this mechanism, we calculate the baseline Kaitz index (minimum-to-median wage rate) in 1996 and divide the sample into two groups. Municipalities with high (low) minimum wage bite are defined as those above (below) the median of the Kaitz index. Then, we run the same specification in column 3 from Tables 1.6.5, 1.6.6, and 1.6.9 separately for each group. Tables 1.6.11 and 1.6.12 present our findings.

In regions where the minimum wage is more binding (columns 1-3), the informal sector suffers a much stronger negative effect on earnings. 1*p.p.* increase in migration rate decreases informal earnings by 1.6%. Regarding the employment margin, the magnitude of the increase in the informal sector is the same as the reduction in formal employment, consistent with native workers who lost their formal jobs migrating to the informal sector. Finally, although there is a slight increase in the provision of health insurance to formal workers, the impact on overall nonwage benefits is small. This is expected as regions with more-binding minimum wages also tend to have stronger collective bargaining and, thus, more rigid benefits (NEUMARK; WASCHER, 2006)

When the minimum wage is less binding (columns 4-6), we find little effect on informal employment and earnings. Consistent with the mechanism discussed above, the adjustment in the formal sector occurs essentially in total compensation - both on earnings and non-wage benefits. Increasing the in-migration rate by 1*p.p.* reduces formal earnings by 0.72%, and the share of formal workers receiving subsidies for food, transport and health insurance by 0.85*p.p.*, 0.48*p.p.* and 0.52*p.p.*, respectively.

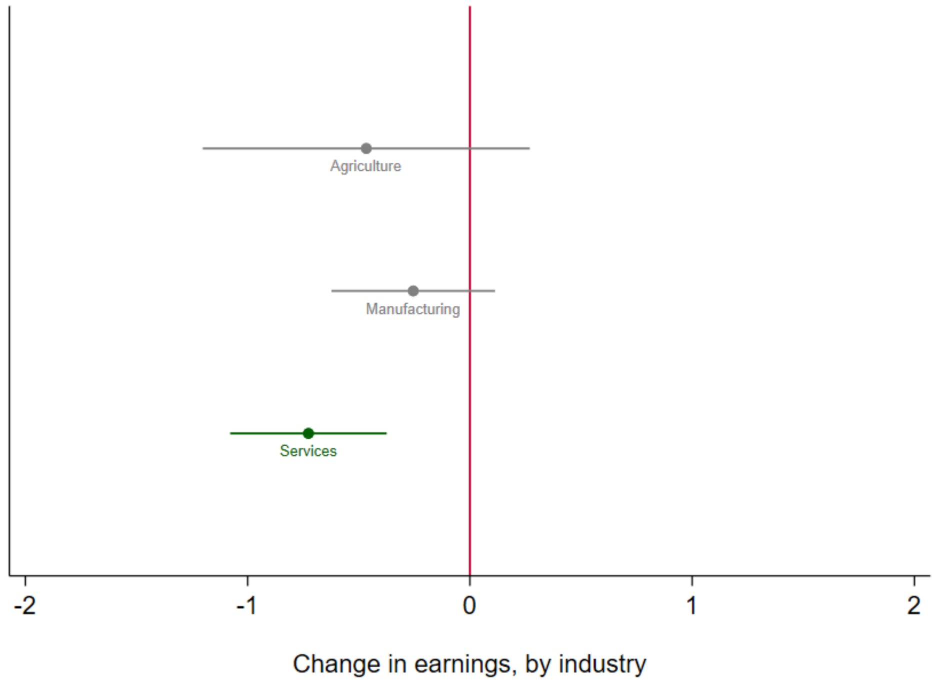
1.5.4 Sensitivity and Heterogeneity Analysis.

In this section, we summarize a series of robustness checks we have performed to assess the validity of our main findings. Then we study the heterogeneity of our main estimated effects with respect to workers' education level.

The first issue we address is whether a shift in local labor demand may be confounding our identification. If that were the case, then we should expect that migrants from other regions outside the Semiarid would be attracted to the same destinations. In other words, we should observe a positive correlation between migrant inflows from the Semiarid and that from other regions. In Table 1.6.13 we show the coefficients from a destination-level regression of the migration inflow rate of migrants from other regions on the predicted inflow rate of migrants from the Semiarid. Column (1) includes time and municipality fixed effects, while in Column (2) we add the same set of controls from our main results. Point estimates are close to zero and not statistically significant in any specification.

The second issue is that our strategy relies on the assumption that rainfall at origin municipalities affects destination labor markets only through internal

Figura 1.5.5 – **Effects of internal migration on earnings**



Notes: This figure plots origin-level SSIV coefficients on change in log earnings, by industry. Controls include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Green markers are statistically significant at the 5% level.

migration. One possible violation of this assumption would be if a negative income shock at the origin, due to low rainfall levels, had reduced trade flows with some of the destination areas, for instance. In this case, one should expect higher effects in those industries more exposed to trade shocks, like agricultural or manufactured goods. In Figure 1.5.5 we summarize the effects from regressions of changes in log earnings on the predicted migrant inflow rate, separately by industry where the individual is employed. In Table 1.6.14 we report the detailed results. There is no statistically significant effects on the earnings for workers in the agricultural or manufacturing industries. All the impact comes from those native workers employed in services, which are less likely to be affected by negative shocks at the origin municipalities.

Finally, we explore the sensitivity of our results according to the degree of aggregation of regions of origin. In Appendix 1.D we argue that the consistency of our shift-share instrument needs origin-level shocks to be mutually uncorrelated. As rainfall shocks are likely correlated across smaller geographical units, in Appendix 1.E we investigate this issue by re-constructing our instrument according to larger catchment areas of origin of a migrant - such as a microregion or a mesoregion - instead of a municipality.[1.5.4] First, we document that spatial correlation among shocks decreases dramatically as we consider larger areas. Second, Tables E2-E5 show that our results associating migration and rainfall, earnings, employment, and non-wage benefits remain virtually unchanged, indicating that spatial correlation among rainfall shocks in origin municipalities are irrelevant to our results.

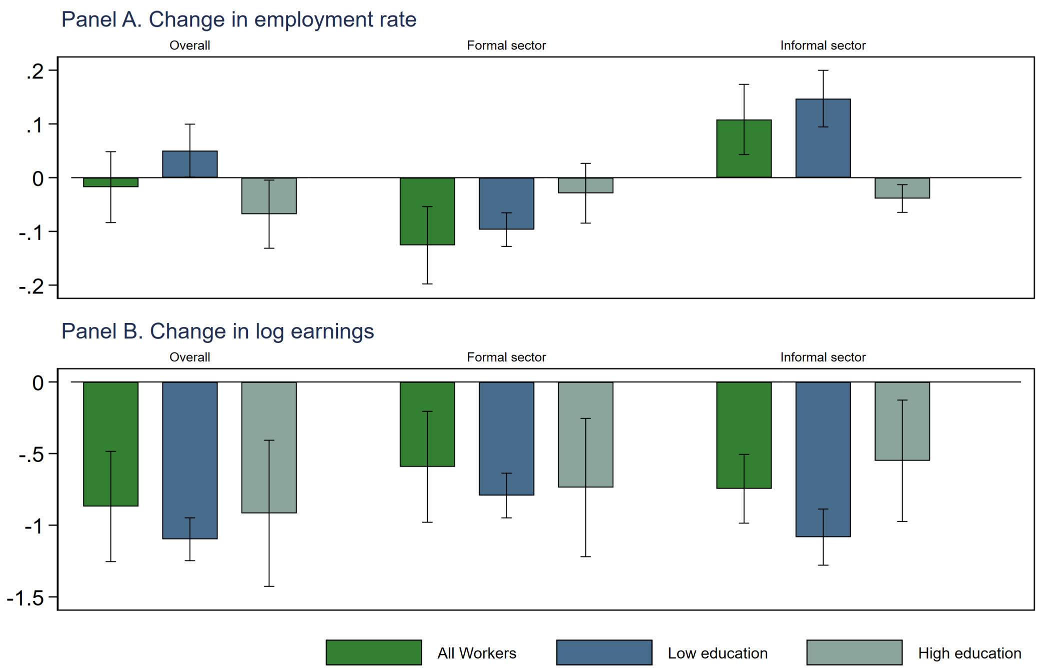
Next, we assess whether individuals with different levels of education may experience differential impacts. In particular, we expect those low-education native workers to be close substitutes for migrants. Thus we reestimate the effect of migration on local labor market outcomes of natives with low and high education, separately. We define as less educated those with up to 8 years of schooling, which is equivalent to complete elementary education. In our sample, 58% of natives are less educated.

Figure 1.5.6 illustrates the estimates by education level. Panel A shows the effect of predicted migration on the changes in employment rates, by sector and education group. Our results are consistent with less educated native individuals being more likely to exit the formal sector and become informal sector workers compared to those who have higher levels of education. In Panel B we analyze the differential effects on log earnings. In the formal sector, wages of native workers fall across education levels. However, native workers with low education have a relatively higher loss in informal and self-employment earnings, consistent with the conjecture that they compete more directly with (less educated) migrants.

In terms of adjustments on the non-wage benefits margin, it is less clear why

[1.5.4]IBGE (1990) defines microregions as "groups of economically integrated municipalities sharing borders and structure of production". Mesoregions are collections of microregions of which not all municipalities share borders. The Semiarid has 960 municipalities, 137 micro, and 35 mesoregions.

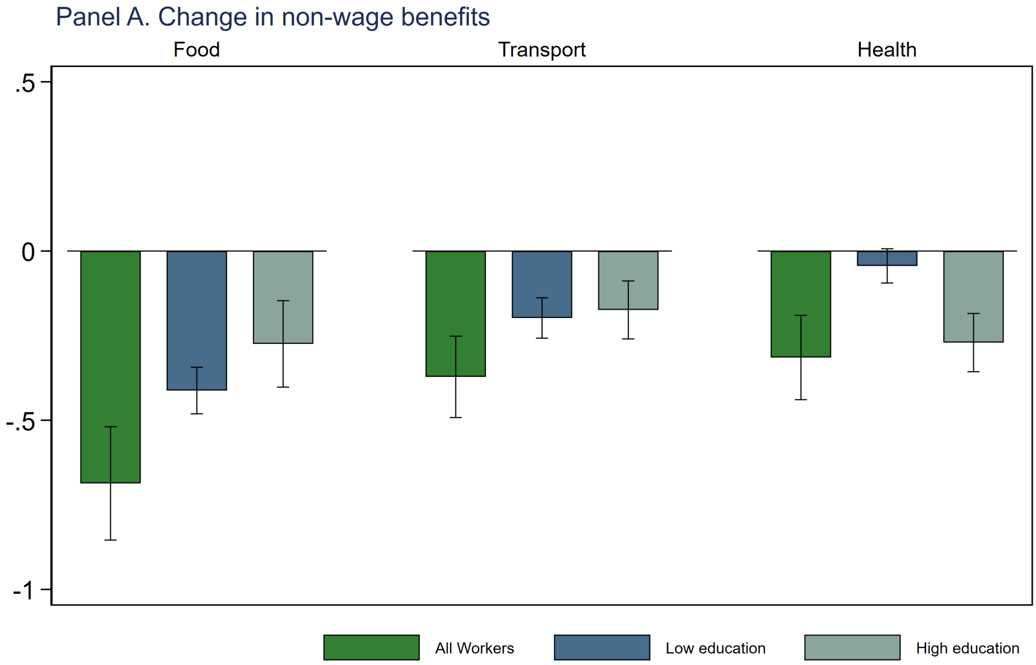
Figura 1.5.6 – **Effects of migration on employment and earnings, by education level**



Notes: This figure plots SSIV coefficients of change in labor market outcomes, by education level. In Panel A, the dependent variables are the changes in employment rates while in Panel B we present estimates for changes in log earnings, for each sector. Each bar represents the SSIV coefficient for a separate regression on the average and by education (low education = up to 8 years of schooling). All regressions are weighted by the working-age native population in 1991, include time dummies and control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. The capped lines show the 95% confidence intervals.

they should differ by worker skills. In principle, working in the same firm implies that workers of different skills are offered a common benefits package. However, if there is a positive matching in the labor market with low (high) education workers selecting into less (more) productive and small (large) firms, then we should expect the less educated workers to be the most affected as the minimum wages bind more tightly in the firms where they work. In Figure 1.5.7 we show that the negative impact on food and transport benefits are indeed stronger and relatively more precise for low-education workers. In contrast, high-education workers have a clear reduction

Figura 1.5.7 – **Effects of predicted migration on non-wage benefits, by education level**



Notes: This figure plots regression coefficients of change in non-wage benefits, by education level, against the predicted number of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991. The dependent variables are the changes in the proportions of native workers in the formal sector who received some help to cover expenses with food, transport or health insurance. Each bar represents the reduced form coefficient by education level (low education = up to 8 years of schooling). All regressions are weighted by the working-age native population in 1991, include time dummies and control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. The capped lines show the 95% confidence intervals.

in employer-provided health insurance which is consistent again with some selection of these workers in large firms which tend to offer health insurance and where there is a mix of high and some low education workforce. A possible explanation is that the inflow of migrants competing with native low-education workers in large firms pressures wages down. However, under minimum wage restrictions, the adjustment occurs through lowering health insurance.

Changes in the benefits can have important welfare implications. We found that migration lowers the provision of food and transport benefits to less educated individuals. On the other hand, we show that health insurance is not significantly changed for low-education workers on average, while it is less offered for the high-education workers. Considering that food and transport are the two most offered benefits in the data (as shown in Table 1.6.2) and to the extent that workers value these benefits, their reduction together with a stronger negative impact on earnings for the low education workers suggest that the welfare of the less educated workers declines more than for high education workers.

1.5.5 Long Run Effects.

Here we turn our attention to the dynamics of the impact of migration on local labor markets in Brazil. Short and long-run effects might differ as markets adapt to current shocks. Jaeger, Ruist e Stuhler (2018) report short-run local effects of migration inflows for the US in the 1970s that are more negative than many in the previous literature, suggesting that the initial impact on natives is potentially large. However, they also show that much of this decline is reversed in later periods.

We account for these long-run effects by calculating the long differences in the outcome variables from 1996-2001 and 2001-2009.[1.5.5] We stack the two periods and estimate the same origin-level SSIV regressions from Section 3.4.

Table 1.6.15 shows the long-term effects of the inflow of migrants from the Semiarid region on the changes in earnings and employment. In the long-run destination labor markets adjust further, resulting in more negative impacts for the native workers. The average earnings reduce by 0.66% and 1.57% among workers in the formal and informal sectors, respectively. On the employment margin, our estimates show a decrease of .26*p.p.* in the formal sector, but no significant effect in the informal sector. Such a result may be reflecting the dual nature of formal and informal markets. In the more rigid formal sector, the markets adjust more slowly than in the flexible informal sector. Table 1.6.16 shows that non-wage benefits also are an important margin of adjustment in the long run. There is no change in the proportion of workers receiving food vouchers, but the share of natives who receives

[1.5.5] PNAD data are not available for the years when the Census are collected - 2000 and 2010.

transport subsidies decreases by $0.75p.p$ and those with health insurance reduce by $0.38p.p.$.

A potential mechanism behind these dynamics is that short-run effects might be partially offset by further internal migration as natives respond to the adverse effects by moving to markets that were not directly targeted by migrant arrivals. Table 1.6.17 reports coefficients of the effect of our predicted shocks on the migration outflows of natives, according to levels of schooling. All estimates are positive but not very precisely estimated, and the magnitude is greater for natives of lower education, who are the most affected by the arrivals of Semiarid migrants.

1.6 Discussion and Concluding Remarks

In this paper, we investigate the labor market impacts of weather-induced internal migration in Brazil. We use a shift-share instrument approach combining variation in the number of people leaving their hometowns, driven by weather shocks, with past settlement patterns to exploit exogenous variation in the number of migrants entering each destination municipality.

Overall our results indicate that an exogenous supply shock of low-skill workers reduces earnings in the unregulated informal sector, especially at the bottom of the wage distribution. Earnings also drop in the formal sector, with close to zero estimates at the bottom as minimum wage restrictions and collective agreements are more binding. Adjustments in non-wage benefits such as food vouchers, transportation subsidies, and health insurance compensate for these rigidities. From a back-of-the-envelope exercise, we find that if we do not account for the non-wage compensation margin, we tend to underestimate the change in overall compensation by 44% percent, at least when we assume that workers value benefits at their cost estimate.

We also observe a decrease in the formal employment of natives due to wage rigidities and an imperfect adjustment of the benefit margin. In the informal sector, an increase in employment follows a significant fall in earnings, consistent with workers reallocating from the formal to the informal sector or self-employment.

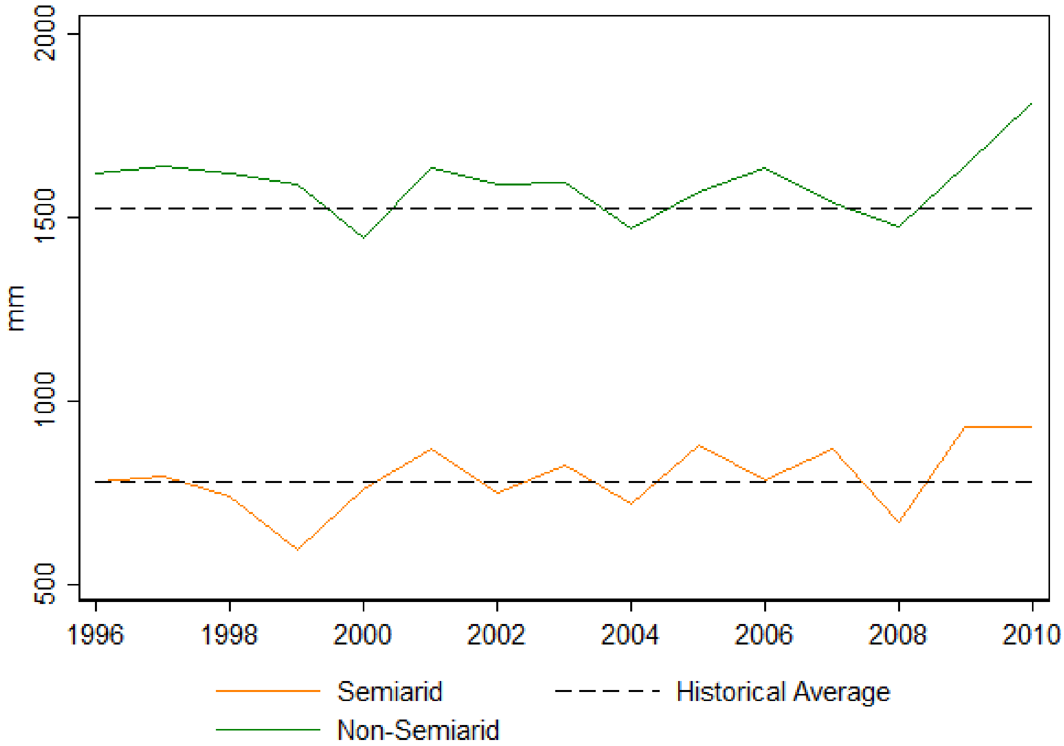
Unemployment and labor force participation also increase, in part, due to non-head members of the households joining the job market in response to migration shocks. Most of the adjustment occurs via formal employment and informal earnings in places with a more binding minimum wage. In contrast, it works primarily through total compensation with limited employment impact in regions where the minimum wage bite is lower.

As discussed in Section 1.3, our model generates predictions for the impact of migration on labor markets with two sectors in an economy with different levels of intersectoral linkages and with endogenous or fixed benefits. From low to medium levels of linkages, the impact of migration in terms of employment and wage drops in magnitude. Allowing firms to adjust benefits as a response to shocks, also softens the impacts as expected. In summary, our estimates are broadly consistent with lower/medium levels of linkages and with imperfectly flexible benefits, as formal employment drops likely due to collective agreements over non-wage benefits.

Taking stock, our findings call attention to the fact that non-wage benefits are a relevant margin of adjustment for firms, especially in labor markets where minimum wages are binding and benefits are commonly offered, such as in many European and Latin American countries. They ease constraints and allow employers to absorb part of the shocks, lowering the impact on employment.

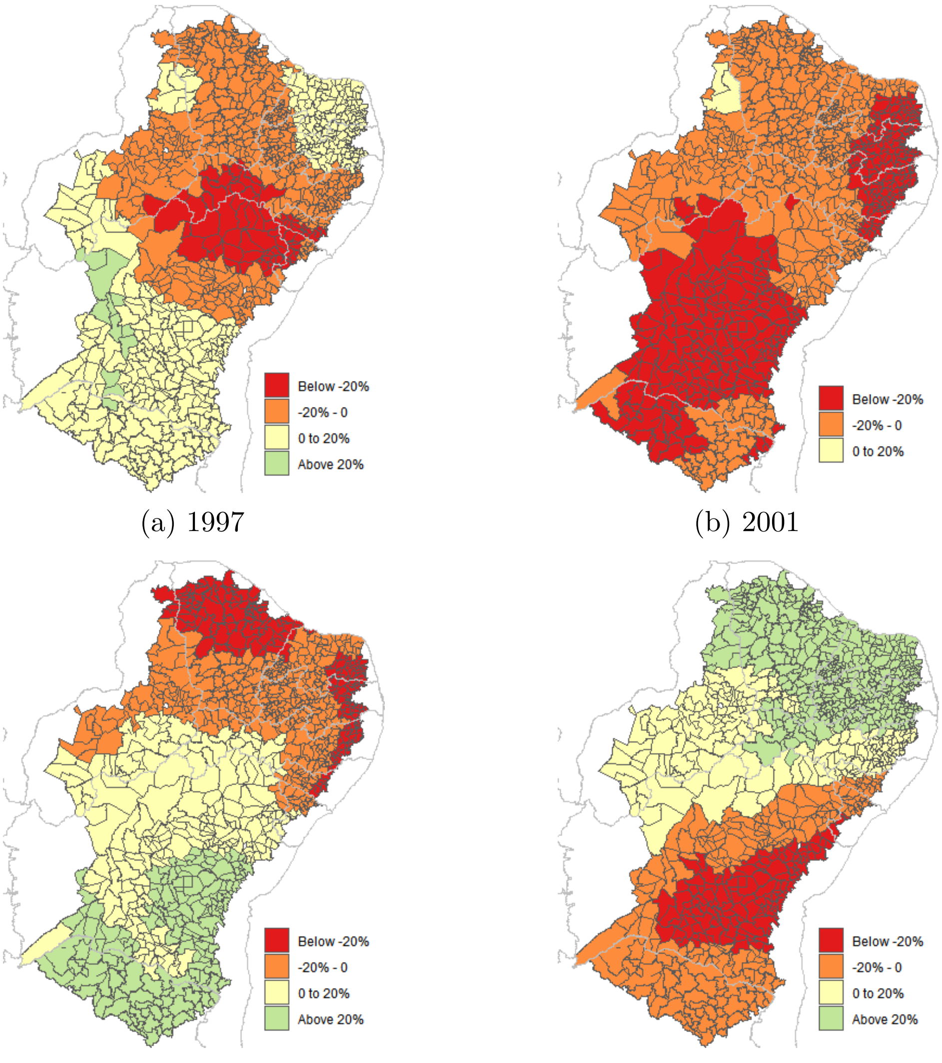
Figures and tables

Figura 1.6.1 – **Precipitation level: Semiarid vs Non-Semiarid**



Notes: This figure compares the average precipitation level for the Semiarid region and the rest of the country, from 1996 to 2010. *Data source*: CRU Time Series v4 (HARRIS et al., 2020).

Figura 1.6.2 – Precipitation levels in the Semiarid region for selected years



(a) 1997 (b) 2001

(c) 2005 (d) 2009

Notes: This figure presents the distribution of rainfall across the Semiarid region municipalities for selected years. Rainfall is measured as the log-deviations from historical averages. *Data source*: CRU Time Series v4 (HARRIS et al., 2020).

Tabela 1.6.1 – **Summary statistics: weather and migration data**

Panel A: Origin (Semiarid)	Mean	Std. Dev.	Min	Max	Obs
Annual Rainfall	782.33	248.71	165.49	1,953.17	14,400
Rainfall shock	-0.02	0.19	-0.73	0.48	14,400
Annual Temperature	25.54	1.39	21.42	28.93	14,400
Temperature shock	0.01	0.01	-0.01	0.05	14,400
Out-migration	214.16	323.66	0.00	5,773	14,400
Out-migration rate (p.p.)	1.05	0.62	0.00	7.22	14,400
Population	21,377	30,386	1,265	480,949	14,400
Panel B: Destination (Non-Semiarid)	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**
Annual Rainfall	1,610.44	401.69	660.63	3,618.55	8,190
Rainfall shock	0.04	0.16	-0.77	0.65	8,190
Annual Temperature	23.15	2.82	15.82	28.77	8,190
Temperature shock	0.03	0.02	-0.03	0.08	8,190
In-migration	146.69	896.95	0.00	25,423	8,190
In-migration rate (p.p.)	0.30	1.00	0.00	27.95	8,190
Native population	51,963	231,29	290	4,771,961	8,190

Notes: Rainfall is measured in mm. Temperature is measured in degrees Celsius. Migration outflow (inflow) rate is the share of migrants over the local (native) population.

Tabela 1.6.2 – **Summary statistics: Native individuals in destination municipalities**

	Individual Characteristics				
	Mean	Std. Dev.	Min	Max	Obs
Female	51.08	3.65	0	72.72	8,190
Black	6.23	5.98	0	53.85	8,190
Mulatto	40.32	24.48	0	100	8,190
White	52.82	25.47	0	100	8,190
Age	37.45	1.96	30.15	55	8,190
Years of schooling	6.58	1.78	0	13.52	8,190
Less than elementary	65.33	15.75	4.71	100	8,190
	Employment				
	Mean	Std. Dev.	Min	Max	Obs
Any Employment	62.72	7.95	10	100	8,190
Formal sector	31.34	11.85	0	100	8,190
Informal sector	31.38	9.05	0	81.80	8,190
Unemployed	13.05	7.73	0	80	8,190
Out of labor force	24.23	7.08	0	58.14	8,190
	Earnings				
	Mean	Std. Dev.	Min	Max	Obs
Any Employment	637.89	348.99	60.88	3,582.08	8,190
Formal sector	788.22	439.49	58.67	15,167.10	8,174
Informal sector	491.28	284.28	20	4,941.10	8,172
	Non-wage benefits				
Food	38.89	21.06	0	100	8,165
Transport	36.39	25.40	0	100	8,165
Health	20.86	16.41	0	100	8,165

Notes: Each observation is a destination municipality-year cell. Earnings are measured in R$ of 2012. The informal sector also includes self-employed workers. Non-wage benefits are calculated only for native workers employed in the formal sector.

Tabela 1.6.3 – **Comparative characteristics: Migrants vs Natives**

	Migrants		***Low-ed. natives***		***High-ed. natives***	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age	29.19	10.25	38.43	13.30	33.04	11.14
Number of children	2.13	2.98	3.31	3.07	1.39	1.58
Schooling	4.65	3.96	3.25	2.14	10.90	2.52
Earnings	765.89	1,370.52	783.83	1,516.89	1,994.34	3,300.81
Work less than 40 hours/week	0.11	0.31	0.14	0.34	0.21	0.41
Share of employment	0.67	0.47	0.56	0.50	0.71	0.46
Share of formal employment	0.64	0.48	0.60	0.49	0.81	0.39

Notes: This table compares the characteristics of migrants from the Semiarid region and native individuals in destination municipalities. We use data from the 1991 Census on individuals aged between 18-65 in municipalities covered by the *PNAD* survey. Low-education individuals are those with incomplete elementary schooling. Earnings are measured in R$ of 2010.

Tabela 1.6.4 – **Migration outflows induced by weather shocks**

	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall_{t-1}	-0.099***	-0.092***	-0.093***	-0.092***	-0.093***	-0.096***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.030)
Rainfall_{t-2}			0.008	0.022		
			(0.030)	(0.031)		
Rainfall_{t-3}				0.059**		
				(0.028)		
Rainfall_{t}					-0.047	
					(0.031)	
Rainfall_{t+1}						-0.059
						(0.036)
Observations	14,400	14,400	14,400	14,400	14,400	14,400
Municipalities	960	960	960	960	960	960
R-Squared	0.461	0.465	0.465	0.466	0.465	0.466
Time dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Temperature shocks	✓	✓	✓	✓	✓	✓
Baseline × time		✓	✓	✓	✓	✓

Notes: Each observation is an origin municipality-year cell. The dependent variable is the number of individuals who left the origin municipality divided by the total population in the 1991 Census. Rainfall is measured as the log-deviation from the historical average (for the 6 months in the crop growing season). All specifications include controls for temperature shocks, municipality and year fixed effects. Columns (2)-(6) also control for municipality-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Standard errors are clustered at the grid level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.5 – **Effects of migration on earnings**

	(1)	(2)	(3)
	A. Δ log earnings		
Migrant inflow	-1.323***	-1.252***	-0.869***
	(0.143)	(0.142)	(0.197)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
	B. Δ log earnings, formal sector		
Migrant inflow	-1.005***	-0.929***	-0.593***
	(0.171)	(0.169)	(0.198)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
	C. Δ log earnings, informal sector		
Migrant inflow	-0.986***	-0.908***	-0.746***
	(0.072)	(0.072)	(0.123)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
Time dummies		✓	✓
Baseline × time			✓

Notes: This table shows origin-level SSIV coefficients on changes in log earnings, by sector. Each observation is an origin municipality-year cell. The informal sector also includes self-employed workers. Column (2) includes time dummies while Column (3) also controls for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.6 – **Effects of migration on employment**

	(1)	(2)	(3)
	A. Δ employment rate		
Migrant inflow	-0.011	-0.019	-0.018
	(0.022)	(0.022)	(0.034)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
	B. Δ formal employment rate		
Migrant inflow	-0.312***	-0.317***	-0.126***
	(0.025)	(0.025)	(0.037)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
	C. Δ informal employment rate		
Migrant inflow	0.301***	0.298***	0.108***
	(0.029)	(0.029)	(0.034)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
Time dummies		✓	✓
Baseline × time			✓

Notes: This table shows origin-level SSIV coefficients on changes in employment rate, by sector. Each observation is an origin municipality-year cell. The informal sector also includes self-employed workers. Column (2) includes time dummies while Column (3) also controls for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.7 – **Effects of migration on unemployment and participation**

	(1)	(2)	(3)
	A. Δ unemployment rate		
Migrant inflow	0.167***	0.176***	0.094***
	(0.013)	(0.013)	(0.020)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
	B. Δ inactivity rate		
Migrant inflow	-0.155***	-0.157***	-0.077***
	(0.022)	(0.022)	(0.029)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
Time dummies		✓	✓
Baseline × time			✓

Notes: This table shows origin-level SSIV coefficients on changes in unemployment and inactivity rates. Each observation is an origin municipality-year cell. Column (2) includes time dummies while Column (3) also controls for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.8 – **Effects of migration on labor market outcomes, by status in the household**

	(1) Employment	(2) Formal	(3) Informal	(4) Unemployment	(5) Inactivity
			A. Head		
Predicted inflow	-0.028*	-0.113***	0.085***	0.018*	0.032**
	(0.015)	(0.021)	(0.019)	(0.011)	(0.013)
Observations	11,460	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955	955
			B. Non-head		
Predicted inflow	0.010	-0.013	0.024	0.076***	-0.108***
	(0.024)	(0.019)	(0.018)	(0.014)	(0.019)
Observations	11,460	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955	955
Time dummies	✓	✓	✓	✓	✓
Baseline × time	✓	✓	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients on changes in employment (by sector), unemployment, and inactivity rates. Each observation is an origin municipality-year cell. The informal sector also includes self-employed workers. In Panel A we use only individuals identified as the head of the household while in Panel B only those identified as non-head are used. All regressions include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.9 – **Effects of migration on non-wage benefits**

	(1)	(2)	(3)
		A. Food	
Migrant inflow	-0.336***	-0.369***	-0.687***
	(0.062)	(0.063)	(0.086)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
		B. Transport	
Migrant inflow	-0.523***	-0.570***	-0.372***
	(0.040)	(0.040)	(0.062)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
		C. Health	
Migrant inflow	-0.442***	-0.472***	-0.315***
	(0.043)	(0.044)	(0.064)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
Time dummies		✓	✓
Baseline × time			✓

Notes: This table shows origin-level SSIV coefficients on changes in the proportions of formal sector workers who receive health insurance, food, or transport subsidies. Each observation is an origin municipality-year cell. Column (2) includes time dummies while Column (3) also controls for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.10 – Effects of predicted in-migration on employer-provided health insurance

	(1)	(2)	(3)	(4)	(5)
Predicted inflow	-0.015**	0.003	-0.004	-0.010**	-0.048**
	(0.007)	(0.002)	(0.003)	(0.005)	(0.022)
Mean of dep. var.	0.0158	0.0131	0.0448	0.0609	0.0758
Observations	4,462,346	4,167,842	138,572	142,100	13,832
Municipalities	682	679	482	608	280
Firms	318,739	297,703	9,898	10,150	988
Time dummies	✓	✓	✓	✓	✓
Firm dummies	✓	✓	✓	✓	✓
Firm size	All firms	1 to 50	51 to 100	101 to 1,000	More than 1,000

Notes: This table shows the reduced form coefficients of changes in the probability of a firm offering health insurance to its employees on the predicted inflow of migrants from the Semiarid region. Each observation is a firm-year cell. The dependent variable is the difference in the dummy variable that is equal to one for every year greater than or equal to the year when the health insurance contract was signed. The regressor is the predicted number of migrants from the Semiarid region in each destination municipality (excluding those in the Semiarid region), measured as a fraction of the native working-age population in 1991. Our sample comprises a balanced panel of all firms included in RAIS during the period. All the regressions are weighted by the number of employees in the firm in 1996. Standard errors clustered at the municipality level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.11 – **Effects of migration on employment and earnings, by level of MW bite**

	A. Δ log earnings					
	(1)	(2)	(3)	(4)	(5)	(6)
	High MW Bite			Low MW Bite		
	Overall	Formal	Informal	Overall	Formal	Informal
Migrant inflow	-1.295***	-0.696	-1.605***	-0.655***	-0.718***	-0.181
	(0.445)	(0.497)	(0.234)	(0.176)	(0.160)	(0.135)
Observations	9,840	9,833	9,840	11,460	11,460	11,460
Municipalities	820	820	820	955	955	955
	B. Δ employment rate					
	High MW Bite			Low MW Bite		
	Overall	Formal	Informal	Overall	Formal	Informal
Migrant inflow	0.002	-0.303***	0.304***	-0.026	-0.079***	0.054
	(0.064)	(0.098)	(0.061)	(0.035)	(0.024)	(0.046)
Observations	9,840	9,840	9,840	11,460	11,460	11,460
Municipalities	820	820	820	955	955	955
Time dummies	✓	✓	✓	✓	✓	✓
Baseline × time	✓	✓	✓	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients on changes in employment rate and log earnings, for each sector and by level of minimum wage bite. Each observation is an origin municipality-year cell. The informal sector also includes self-employed workers. All regressions include time dummies, control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies, and are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.12 – **Effects of migration on non-wage benefits, by level of MW bite**

	High MW Bite			Low MW Bite		
	Food	Transport	Health	Food	Transport	Health
Migrant inflow	0.070	0.110	0.208**	-0.849***	-0.477***	-0.524***
	(0.153)	(0.126)	(0.100)	(0.086)	(0.075)	(0.064)
Observations	9,840	9,833	9,840	11,460	11,460	11,460
Municipalities	820	820	820	955	955	955
Time dummies	✓	✓	✓	✓	✓	✓
Baseline × time	✓	✓	✓	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients on changes in the proportions of formal sector workers who receive health insurance, food, or transport subsidies, by level of minimum wage bite. Each observation is an origin municipality-year cell. All regressions include time dummies, control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies, and are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.13 – **Correlation between predicted migration from the Semiarid and other regions**

	(1)	(2)
	Migrant inflow from other regions	
Predicted inflow	0.080	0.081
	(2.847)	(2.848)
Observations	8,190	8,190
Municipalities	684	684
Municipality dummies	✓	✓
Time dummies	✓	✓
Baseline × time		✓

Notes: This table shows destination-level regression coefficients of the observed inflow of migrants from other regions on the predicted number of migrants from the Semiarid, both measured as a fraction of the working-age native population in 1991. Each observation is a destination municipality-year cell. All regressions include municipality and time dummies. Column (2) controls for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.14 – **Effects of migration on earnings, by industry**

	(1)	(2)	(3)
	Agriculture	Manufacturing	Services
Predicted inflow	-0.466	-0.255	-0.831***
	(0.375)	(0.188)	(0.187)
Observations	11,447	11,460	11,460
Municipalities	955	955	955
1991 Demographics × Time	✓	✓	✓
Time dummies	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients on changes in log earnings, by industry. Each observation is an origin municipality-year cell. All regressions include time dummies and control for destination-level 1991 characteristics (log of the working-age native population; shares of population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.15 – **Long run effects**

	(1)	(2)	(3)
	Overall	Formal	Informal
	A. Δ log earnings		
Migrant inflow	-1.111***	-0.658***	-1.570***
	(0.312)	(0.253)	(0.265)
Observations	1910	1910	1910
Municipalities	955	955	955
	B. Δ employment rate		
Migrant inflow	-0.305***	-0.257***	-0.048
	(0.051)	(0.048)	(0.053)
Observations	1,910	1,910	1,910
Municipalities	955	955	955
Time dummies	✓	✓	✓
Baseline × time	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients of stacked long differences in log earnings and in the employment rate. Each observation is an origin municipality-year cell. The long difference is calculated from 1996-2001 and from 2001-2009. The instrument is the predicted migration accumulated in the same periods, measured as a fraction of the 1991 working-age native population. All regressions include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.16 – **Long run impacts: non-wage benefits**

	(1)	(2)	(3)
	Food	Transport	Health
Migrant inflow	-0.112	-0.753***	-0.384***
	(0.124)	(0.088)	(0.068)
Observations	1,910	1,910	1,910
Municipalities	955	955	955
Time dummies	✓	✓	✓
Baseline × time	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients of stacked long differences in the proportion of native formal workers receiving non-wage benefits. The long difference is calculated from 1996-2001 and from 2001-2009. The instrument is the predicted migration accumulated in the same periods, measured as a fraction of the 1991 working-age native population. All regressions include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 1.6.17 – **Effects on migration outflows of natives**

	(1)	(2)	(3)	(4)
	Panel A. Low education			
Predicted inflow	1.352	1.273		
	(0.822)	(0.775)		
Lagged pred. inflow			-0.586	-0.342
			(0.615)	(0.450)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
	Panel B. High education			
Predicted inflow	0.151	0.109		
	(0.746)	(0.750)		
Lagged pred. inflow			0.871	0.466
			(1.186)	(0.838)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
Municipality dummies	✓	✓	✓	✓
Time dummies	✓	✓	✓	✓
Baseline × time		✓		✓

Notes: This table shows regression coefficients of the number of people leaving the destination areas against the predicted number of migrants from the Semiarid region at the origin municipality level, both measured as a fraction of the native working-age population in 1991. Each observation is a destination municipality-year cell. In Columns (1)-(2) the regressor is the contemporaneous predicted migrant flow, while in Columns (3)-(4) is the same variable lagged one year. All specifications include municipality and time dummies and are weighted by the 1991 native population. Columns (2) and (4) also control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Appendix

1.A A Simple Model with Informality

In interpreting our findings, we develop a simple extension of a model with informality in which the formal sector has minimum wage but offers non-wage benefits that are frequently observed in the data (e.g. employer-provided health insurance). We follow an extension of the labor market model developed by (HARRIS; TODARO, 1970) provided in (ALMEIDA; CARNEIRO, 2012). In their model minimum wage and labor legislation are the main institutions behind the existence of a formal and an informal sector. We add non-wage benefits in the formal sector as a source of adjustment of total compensation in the presence of binding minimum wages. We abstract from other sources of frictions, which is explored in much recent work on models of the labor market with monopsony to study immigration effects (e.g. (AMIOR; MANNING, 2020) and (AMIOR; STUHLER, 2022)). They are not needed to understand the mechanism we emphasize, so we proceed with the following model.

Suppose an aggregate output function that combines both formal and informal labor inputs:

$$Y = f(L_f^d, L_i^d, \bar{K}) \tag{1.A.1}$$

where L_f^d and L_i^d are total formal and informal labor, respectively, required to production and $\bar{K}$ the fixed capital stock. $f_{L_f^d L_i^d} \neq 0$ captures production linkages. The wage or the value of a job in the formal and informal sectors are determined by marginal products in each sector, i.e.

$$W_i = \frac{\partial f}{\partial L_i^d} \tag{1.A.2}$$

$$W_f + (1-t)B = \frac{\partial f}{\partial L_f^d} \geq \underline{W}_f + (1-t)B \tag{1.A.3}$$

which yield labor demand equations assuming that the minimum wage is binding while non-wage benefits can be optimally chosen in the formal sector.

The labor market with two sectors at the destination can be represented by the following equations:

$$\begin{array}{ll}
\text{Formal labor demand} & L_f^d = a - b(\underline{W}_f + (1-t)B) + cW_i \\
\text{Informal labor demand} & L_i^d = d - eW_i + f(\underline{W}_f + (1-t)B) \\
\text{Formal labor supply} & L_f^s = g + h(\underline{W}_f + vB)(1-U) - iW_i \\
\text{Informal labor supply} & L_i^s = j + kW_i - l(\underline{W}_f + vB)(1-U) \\
\text{Equilibrium} & L_f^d = L_f^s(1-U) = L_f^*;\ L_i^d = L_i^s = L_i^* \\
\text{Labor constraint} & L_f^s + L_i^s + O = M
\end{array}$$

where W_f and W_i denote wages in the formal and informal sectors, respectively. B are non-wage benefits offered in the formal sector only. For simplicity, we also do not consider labor taxes or enforcement costs since this is not central in this paper. With the exception of the intercepts of the equations, we assume that all parameters are positive also implying that the two types of labor (formal and informal) are substitutes ($f_{L_f^d L_i^d} < 0$). Employers hiring formal workers can offer benefits (e.g. health insurance and food subsidies) at a cost that is below the wage cost ($t \leq 1$). We assume that workers value such benefits at the rate v, which can be smaller, equal or even larger than 1. The total number of individuals in the economy is M (natives plus migrants), who can either work or search for a job in the formal sector (L_f^s), work in the informal sector (L_i^s), or be out of the labor force (O). Labor markets are competitive, and equilibrium wages and quantities of labor in each sector are determined by the intersection of supply and demand.

We solve for L_f^*, L_i^*, W_i, B and U. The solution to this system is complex so we provide a numerical solution, given the above parametrization. The details of the construction of our numerical example are described in the footnote below.[1.A.1]

[1.A.1] We set our benchmark at $a = d = 1$, $g = j = 0$, $b = e = h = k = 1$ and $c = f = i = l = 0.5$. The slope restrictions are consistent with integrated formal and informal sectors but we do consider that own effects are likely larger than cross-effects determining demand and supply of labor in each sector. We also consider that offering benefits is 50% cheaper to firms consistent with fiscal exemptions on such benefits ($t = 0.5$) and that workers value non-wage benefits less

In our model, we consider that migration exogenously shifts the supply of workers to the informal and formal sectors. Suppose an equal increase of 10p.p. in the parameters g and j (intercept shifters), so that the total labor force M increases by 20p.p.

Tabela 1.A.1 – **Effects of introducing migration in a labor market with informality**

	(1) benchmark	(2) migration with B^*	(3) migration with $\overline{B}$
		Panel A: Low linkages	
L_f^*	100	100	88
L_i^*	100	115	124
W_i	100	93	89
B	100	33	100
U	100	128	156
		Panel B: Medium linkages	
L_f^*	100	100	89
L_i^*	100	109	119
W_i	100	94	90
B	100	45	100
U	100	183	283
		Panel C: High linkages	
L_f^*	100	100	100
L_i^*	100	103	100
W_i	100	96	100
B	100	96	100
U	100	98	97

Table 1.A.1 shows the effects of migration on equilibrium allocations for each economy under two different scenarios: (i) with flexible non-wage benefits,

than wages with $v = 0.5$, motivated by lack of liquidity or pension accumulation. Finally, given the above parametrization, we set the minimum wage $\underline{W}_f$ at 1, $O = 0$, and $M = 1$.

and (ii) with non-wage benefits fixed at the baseline level. Panel A considers our benchmark economy with $c = f = 0.5$ while Panels B and C allow for higher and lower linkages in production across sectors, which is done by imposing $c = f = 0.7$ and $c = f = 0.3$, respectively. All baseline economies in column (1) are normalized to 100.

In columns (2) and (3) we show the effects of introducing migration in each scenario. Panel B shows that in the benchmark economy with medium production linkages across sectors and varying benefits (column (2)), migration increases unemployment and informal employment, and drops non-wage benefits and informal sector wages. Formal employment is unchanged. With benefits fixed at the baseline level (column (3)), the overall impacts on employment and informal wages are relatively higher since benefits do not adjust. Moreover, formal employment declines.

Panel C considers an economy with higher linkages, the new equilibrium reflects a better adjustment of informal wages and non-wage benefits to the fact that there are wage rigidities in the formal sector, despite a lower value of non-wage benefits. Consequently, the effects of migration with flexible non-wage benefits are lower than in the economy in Panel B. With fixed benefits, the impacts are even lower or nonexistent for most outcomes given the fast degree of market integration.

Under lower linkages, Panel A shows that the qualitative results of Panel B are kept. However, migration induces a larger fall in informal sector wages and formal sector benefits with flexible formal sector benefits. With fixed benefits, the results in column (3) show that migration has now a negative impact on formal employment, as expected since formal firms cannot adjust benefits after the supply shock from migration. Consequently, unemployment also increases more under fixed benefits in such economy.

1.B Migrant flows from the Semiarid region

In this section, we discuss in more detail our measure of migration between cities and how we structure a yearly panel dataset from the 2000 and 2010 censuses.

1.B.1 Migration from the Semiarid region

In every round of the Census, there are two questions that allow us to track the migrants and establish their municipalities of origin and destination, as well as the year when they moved.

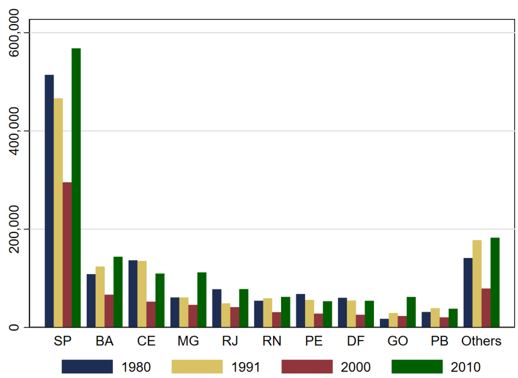
First, in the 2010 Census, respondents were asked for how many years they had lived in the current municipality (from one up to ten). With this variable, we are able to calculate the year when the individual has migrated. We consider a migrant an individual who moved to the current municipality in the previous ten years. In the 2000 Census, interviewees were asked about the municipality where they were living five years ago, instead of the last place where they lived so we can only identify migrants who came as far as 1996. This is not a major concern in our analysis as 1996 is the first year for which *PNAD* data - the source from which we draw labor market outcomes information - is available.

Second, they were asked what was the municipality where they lived before. Thus, if an individual has migrated from an origin municipality in the Semiarid region, she will be counted as a Semiarid migrant. A limitation is that we can only track one origin location for each person, probably the last municipality where she lived.

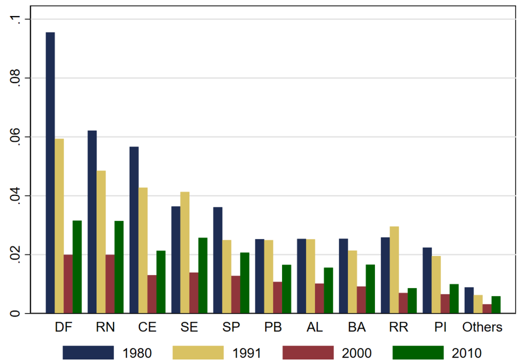
The Semiarid region has always been an important source of migrants for the rest of the country. Figure B1 shows that these migrants tend to be historically concentrated in some states. São Paulo alone harbored over 30 percent of the people arriving from the Semiarid in the last four decades. However, in relative terms, incoming migrants represented a population increase of above 2% for the top 10 receiving states.

Table 1.6.3 compares migrants to low and high-education natives. Migrants are slightly more educated and earn slightly less than less-educated natives. They also have a similar likelihood of working part-time and being in the formal sector when compared to low-education natives. On the other hand, high-education natives are more likely to work in the formal sector and have considerably higher pay. Table B1 shows that top occupations for migrants (e.g. typically bricklayer for men, domestic worker for women) are also top occupations for low-education natives, but not for the skilled. Also, the same five industries that concentrate over 80% of working migrants also employ a similar share of low-education workers (see Table B2).

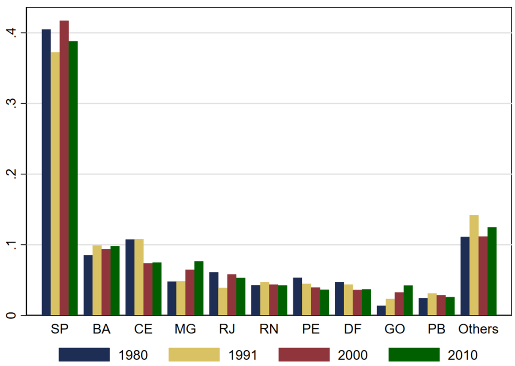
Figura B1 – Top destinations for migrants from the Semiarid region



(a) Absolute number of Semiarid's migrants



(b) Semiarid's migrants as a fraction of total population



(c) Semiarid's migrants as a share of total migration

Notes: This figure presents the main destination states chosen by migrants from the Semiarid region. Panel (a) shows the absolute number of migrants leaving the Semiarid region to non-Semiarid areas. Panel (b) presents the same inflow measured as a fraction of the total population in the state while in Panel (c) that number is measured as a share of the total number of migrants in each state. In each panel, states are ranked by the respective average across years. *Data source*: Census microdata (IBGE).

Tabela B1 – **Main occupations for employed people: Migrants vs Natives**

	Position	Occupation	Share of employment	Cumulative
Migrants	1	Domestic worker	13.8	13.8
	2	Bricklayer	9.6	23.4
	3	Non-specified occupations	9.1	32.5
	4	Salesperson	9.1	41.5
	5	Rural worker	3.6	45.2
	6	Janitor	3.0	48.2
	7	Office assistant	2.6	50.8
	8	Tailor	2.5	53.3
	9	Driver	2.3	55.5
	10	Security guard	2.0	57.5
Low-ed. natives	1	Rural worker	10.8	10.8
	2	Bricklayer	8.2	19.0
	3	Salesperson	8.1	27.0
	4	Domestic worker	7.8	34.8
	5	Non-specified occupations	6.0	40.8
	6	Driver	5.7	46.5
	7	Janitor	3.6	50.1
	8	Tailor	2.9	53.0
	9	Cook	1.7	54.7
	10	Mechanic	1.7	56.5
High-ed. natives	1	Salesperson	8.9	8.9
	2	Office assistant	7.9	16.7
	3	Non-specified occupations	4.4	21.1
	4	Tradesperson	3.1	24.2
	5	Secretary	3.1	27.2
	6	Driver	2.6	29.9
	7	Office supervisor	2.6	32.5
	8	Military	2.0	34.5
	9	Teacher	2.0	36.4
	10	Nurse	1.8	38.2

Notes: This table presents the top ten occupations for workers in the destination municipalities, using data from the 1991 Census.

Tabela B2 – **Main industries for employed people: Migrants vs Natives**

	Position	Industry	Share of employed	Cumulative
Migrants	1	Hospitality	31.0	31.0
	2	Manufacturing	19.8	50.8
	3	Retail	14.3	65.1
	4	Construction	13.0	78.2
	5	Agriculture/Mining	5.6	83.7
	6	Health/Education	5.4	89.1
	7	Transport/Communication	4.0	93.1
	8	Other Services	2.5	95.5
	9	Public Sector	2.5	98.0
	10	Professional Services	2.0	100.0
Low-ed. natives	1	Hospitality	25.5	25.5
	2	Manufacturing	18.8	44.3
	3	Agriculture/Mining	14.8	59.2
	4	Retail	12.6	71.8
	5	Construction	10.9	82.7
	6	Transport/Communication	6.0	88.7
	7	Health/Education	4.9	93.6
	8	Public Sector	3.1	96.7
	9	Professional Services	1.9	98.5
	10	Other Services	1.5	100.0
High-ed. natives	1	Health/Education	18.8	18.8
	2	Manufacturing	17.5	36.3
	3	Retail	16.8	53.1
	4	Hospitality	12.0	65.1
	5	Public Sector	9.2	74.3
	6	Professional Services	7.4	81.7
	7	Other Services	6.8	88.5
	8	Transport/Communication	4.9	93.3
	9	Agriculture/Mining	3.5	96.9
	10	Construction	3.1	100.0

Notes: This table presents the top ten industries for workers in the destination municipalities, using data from the 1991 Census.

1.C Weather shocks and predicted migration

In this section, we discuss the weather data and provide further details about how we construct our instrument. We also show that our results are robust to an alternative measure of weather shocks.

1.C.1 Weather data

Our main source for weather data comes from the CRUTS v4, a gridded dataset produced by the Climatic Research Unit at the University of East Anglia (HARRIS et al., 2020). It provides information on monthly precipitation and temperature covering the whole globe (except Antarctica) from 1901 to 2018. The grid resolution is $0.5° \times 0.5°$ (around 56km^2) and is created by interpolation from ground-based weather stations around the world.

We use the R package 'geobr' (CARABETTA; PEREIRA; GONCALVES, 2020) to download the shapefile of Brazilian municipalities and georeference the coordinates from each municipality's centroid. Then, for each municipality, we find the grid's four points that are closest to its centroid and calculate the average level of precipitation and temperature from these points, weighted by the inverse distance to the centroid.

This procedure results in a dataset of monthly averages of precipitation and temperature for each municipality, from 1901 to 2010, which we aggregate in yearly measures. Precipitation is defined as the sum of monthly levels and temperature as the average. For each municipality in the Semiarid region, we calculate the historical mean from both variables and take the natural logarithm of these variables (both levels and long-term averages).

Finally, our weather shock variables are defined as

$$Rainfall_{ot} = \ln\left(\sum_{\tau \in \{GS\}} r_{o\tau t}\right) - \ln(\bar{r}_o) \tag{C1}$$

where $r_{o\tau t}$ is the rainfall in the municipality of origin o in month τ of year t, and $\bar{r}_o$ is the municipality's historical average precipitation for the same months. The index τ covers the 6-month growing season (GS). Temperature is calculated in a similar way, but using the average instead of summation to create yearly data. In our main specifications, we use data from the Semiarid's growing season (from November to April), but results are very similar when we use the full year (see Table C1).

Tabela C1 – **Migration outflows induced by weather shocks (12 months)**

	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall_{t-1}	-0.126***	-0.107***	-0.111***	-0.117***	-0.112***	-0.109***
	(0.033)	(0.033)	(0.034)	(0.034)	(0.033)	(0.033)
Rainfall_{t}			-0.015	-0.029	-0.014	
			(0.039)	(0.041)	(0.039)	
Rainfall_{t-2}			0.037	0.059		
			(0.038)	(0.039)		
Rainfall_{t-3}				0.047		
				(0.033)		
Rainfall_{t+1}						-0.068*
						(0.037)
Observations	14,400	14,400	14,400	14,400	14,400	14,400
Municipalities	960	960	960	960	960	960
R-Squared	0.461	0.465	0.465	0.466	0.465	0.466
Time dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Temperature shocks	✓	✓	✓	✓	✓	✓
Baseline × time		✓	✓	✓	✓	✓

Notes: Each observation is a municipality-year cell. The dependent variable is the number of individuals who left the origin municipality divided by the total population in the 1991 Census. Rainfall is measured as log-deviation from the historical average. All specifications include controls for temperature shocks, municipality and year fixed effects. Columns (2)-(6) control for municipality-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Standard errors are clustered at the grid level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

1.C.2 Alternative measures of weather

One possible concern about our measure of weather is that we focus on rainfall levels, controlling for temperature variation, to predict the flow of migrants leaving the Semiarid region. This may be problematic because we cannot account for the presence of groundwater or any other factors that influence water balance. To circumvent this issue we use the Standardized Precipitation Evapotranspiration Index (SPEI) developed by (VICENTE-SERRANO; BEGUERÍA; LÓPEZ-MORENO, 2010). The SPEI is a measure that accounts for both precipitation and potential evapotranspiration, providing a measure of water balance in a given period. This index also captures deviations from the historical average (1905-2018) in the net water need for a given location. An SPEI value of -1 means that the precipitation level is one standard deviation below the historical level needed to maintain the balance given the potential evapotranspiration. According to (VICENTE-SERRANO; BEGUERÍA; LÓPEZ-MORENO, 2010), the SPEI is particularly useful to detect, monitor, and explore the consequences of global warming on drought conditions. We repeat the first step in our procedure to construct an instrument for in-migration, using the SPEI instead of rainfall and temperature shocks. Once again, we calculate the average SPEI for the 6-month growing season. Table C2 shows that we can also use this measure to predict the out-migration rate from the origin municipalities in the Semiarid region, although the estimates are noisier than those in Table 1.6.4.

We estimate the same specification of column 3 in Tables 1.6.5, 1.6.6 and 1.6.9 using this new instrument and show in Tables C3 and C4 that the results are very similar.

Tabela C2 – **Migration outflows induced by weather shocks: Standardized Precipitation Evapotranspiration Index (SPEI)**

	(1)	(2)	(3)	(4)	(5)	(6)
$SPEI_{t-1}$	-0.041***	-0.036**	-0.037**	-0.035**	-0.036**	-0.037**
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$SPEI_{t-2}$			0.040**	0.040**		
			(0.017)	(0.017)		
$SPEI_{t-3}$				0.014		
				(0.018)		
$SPEI_{t}$					-0.012	
					(0.015)	
$SPEI_{t+1}$						-0.011
						(0.016)
Observations	14,400	14,400	14,400	14,400	14,400	14,400
Municipalities	960	960	960	960	960	960
R-Squared	0.461	0.465	0.465	0.465	0.465	0.465
Time dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓	✓

Notes: Each observation is a municipality-year cell. The dependent variable is the number of individuals who left the origin municipality divided by the total population in the 1991 Census. All specifications include municipality and year fixed effects. Column (6) also controls for municipality-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Drought severity measures are the quartiles of the Aridity Index z-score. Standard errors are clustered at the grid level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela C3 – **Effects of migration on earnings and employment, using SPEI to predict out-migration**

	(1)	(2)	(3)
	Overall	Formal	Informal
	A. Change in log earnings		
Migrant inflow	-0.844***	-0.562***	-0.730***
	(0.197)	(0.198)	(0.122)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
	B. Change in employment rate		
Migrant inflow	-0.010	-0.118***	0.108***
	(0.034)	(0.037)	(0.033)
Observations	11,460	11,460	11,460
Municipalities	955	955	955
Time dummies	✓	✓	✓
Baseline × time	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients of regressions on log earnings and on the employment rate for native workers. Each observation is an origin municipality-year cell. The instrument for the migrant inflow is calculated using the SPEI to predict out-migration distributed by the 1991 share of Semiarid migrants. All regressions include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela C4 – **Effects of migration on non-wage benefits, using SPEI to predict out-migration**

	(1)	(2)	(3)
	Food	Transport	Health
Migrant inflow	-0.717***	-0.398***	-0.332***
	(0.087)	(0.063)	(0.063)
Observations	11460	11460	11460
Municipalities	955	955	955
Time dummies	✓	✓	✓
Baseline × time	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients of regressions on the proportion of native formal sector workers receiving non-wage benefits. Each observation is an origin municipality-year cell. The instrument for the migrant inflow is calculated using the SPEI to predict out-migration distributed by the 1991 share of Semiarid migrants. All regressions include time dummies and destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. All regressions are weighted by the working-age native population in 1991. Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

1.D Shift-share instrument (SSIV)

In this section, we derive the origin-level SSIV estimator, and present and discuss the identifying assumptions needed to produce a consistent estimator of the effects of the inflow of migrants from the Semiarid region on labor markets in the destination municipalities.

We start from the structural equation 3.3.2. To simplify notation we omit the time subscript t. By the Frisch-Waugh-Lovell Theorem, we can re-write it as

$$y_d^{\perp} = \beta m_d^{\perp} + \varepsilon_d^{\perp} \tag{D1}$$

where all $y_d^{\perp}$ is the vector of outcomes, $m_d^{\perp}$[1.D.1] is the observed number of Semiarid's migrants who entered the destination municipality d and $\varepsilon_d^{\perp}$ is a structural residual. All variables are residualized to remove the effects from the covariates.

In equation 3.3.5 we defined the shift-share instrumental variable (SSIV) as

$$\widehat{m}_d = \sum_{o=1}^{O} s_{od} \frac{\widehat{M}_o}{N_d} \tag{D2}$$

where $s_{od} = \dfrac{M_{od}}{\sum_d M_{od}}$ is the share of migrants from origin municipality o who lived in the destination area d in 1991 and $\widehat{M}_o$ is the predicted number of migrants leaving the Semiarid region driven by weather shocks.

The more traditional approach would be to estimate β using $\widehat{m}_d$ as an instrument for the endogenous migrant inflow $m_d^{\perp}$. In such a case, we would have

$$\hat{\beta} = \frac{\sum_d \widehat{m}_d y_d^{\perp}}{\sum_d \widehat{m}_d m_d^{\perp}} \tag{D3}$$

[1.D.1] In order to facilitate the interpretation of the coefficients we normalize this measure dividing by the working-age native population in 1991, which means $m_d = \dfrac{M_d}{N_d}$

By the definition of $\widehat{m}_d$ in equation D2 and switching the order of the summation,

$$\hat{\beta} = \frac{\sum_d \left(\sum_o s_{od}\frac{\widehat{M}_o}{N_d}\right) y_d^{\perp}}{\sum_d \left(\sum_o s_{od}\frac{\widehat{M}_o}{N_d}\right) m_d^{\perp}} = \frac{\sum_o \widehat{M}_o \left(\sum_d s_{od}\frac{y_d^{\perp}}{N_d}\right)}{\sum_o \widehat{M}_o \left(\sum_d s_{od}\frac{m_d^{\perp}}{N_d}\right)} = \frac{\sum_o s_o \widehat{M}_o \bar{y}_o}{\sum_o s_o \widehat{M}_o \bar{m}_o} \tag{D4}$$

where $\bar{y}_o = \frac{\sum_d s_{od}\frac{y_d^{\perp}}{N_d}}{\sum_d s_{od}}$ is a weighted average of the residualized outcome, normalized by the native population, which uses as weights the destination's average exposure to the shocks $s_o = \sum_d s_{od}$. The same result is valid for the endogenous variable $m_d^{\perp}$, meaning that we can estimate the following IV regression at the origin municipality level:

$$\bar{y}_o = \beta \bar{m}_o + \bar{\varepsilon}_o \tag{D5}$$

using the predicted number of migrants from the Semiarid region, $\widehat{M}_o$, as instrumental variable and weighting by the average exposure s_o.

This derivation is almost identical to that presented by (BORUSYAK; HULL; JARAVEL, 2021), except for the fact that we need to divide both variables by the predetermined native population. Their equivalence result shows that the parameter β can be estimated at the level of the identifying variation, which in our case is the origin municipality hit by weather shocks.

Tabela D1 – **SSIV First Stage**

	(1)	(2)	(3)
First stage coefficient	0.912***	0.910***	0.925***
	(0.015)	(0.015)	(0.019)
F-statistic	3,462	3,464	2,275
Observations	11,460	11,460	11,460
Municipalities	955	955	955
Effective sample size	7,301	7,301	7,301
Time dummies		✓	✓
Baseline × time			✓

Notes: This table shows the SSIV first stage coefficients of the origin-level weighted average of the endogenous inflow of migrants at the destinations against the predicted number of migrants from the Semiarid region. Each observation is an origin municipality-year cell. The F-statistic is calculated as the square of the coefficient t-statistic (see (BORUSYAK; HULL; JARAVEL, 2021)). The effective sample size is the inverse of the HHI of the origin-level exposure. Column (2) includes time dummies while Column (3) also controls for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Regressions are weighted by the working-age native population in 1991. Standard errors cluster by the municipality of origin in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

1.E Spatial correlation in weather shocks

Weather events are likely correlated across space. Figure 1.6.2 shows that precipitation levels in the Semiarid are similar among nearby municipalities. Potentially, this could invalidate the consistency of our estimator given by *Assumption 2 (Many uncorrelated shocks)* discussed in Appendix 1.D. Here we investigate this issue by re-constructing our instrument according to different degrees of aggregation of regions of origin of a migrant - such as a microregion or a mesoregion - instead of a municipality. (IBGE, 1990) defines microregions as "groups of economically integrated municipalities sharing borders and structure of production". Mesoregions are collections of microregions of which not all municipalities share borders.[1.E.1] Brazil has 5,565 municipalities, 361 micro, and 87 mesoregions overall. The Semiarid has 960 municipalities, 137 micro, and 35 mesoregions.

The intuition behind this exercise is that even if weather shocks are spatially correlated among contiguous municipalities, such a correlation should decrease as we consider larger areas. Table E2 displays Moran's index of spatial correlation of rainfall shocks for each of the three geographic aggregates in columns 1-3.[1.E.2] As expected, neighboring municipalities display a correlation above 0,94, but it decreases rapidly as we aggregate up to micro and meso regions, to 0,16 and 0,07, respectively.

Table E2 also shows the association between rainfall shocks and migration outflows. Column 1 is identical to Table 1.6.4 for reference. Columns 2 and 3 report

[1.E.1]Table E1 reports summary statistics of our main variables for both levels of aggregation.
[1.E.2]Moran's I is calculated according to the following formula:

$$I = \frac{1}{\sum_i \sum_j w_{ij}} \times \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\frac{1}{N}(y_i - \bar{y})^2} \tag{E1}$$

Essentially, it is a correlation coefficient weighted by an appropriate matrix that models how different units are related across space. We use a row-standardized contiguity matrix with the queen criterion, meaning that two localities i and j sharing either borders or vertices are considered 'neighbors' and the entry w_{ij} has a positive value. Row-standardization ensures that weights are positive and no greater than 1. Non-adjacent pairs receive a zero weight. As discussed by (BEENSTOCK; FELSENSTEIN et al., 2019), Moran's I can be calculated for each period and averaged out with panel data.

almost identical point estimates and precision, indicating that we do not lose any significant information by aggregating origin areas. Next we estimate our main specification from Column (3) in Tables 1.6.5,1.6.6 and 1.6.9 using instruments corresponding to micro and mesoregion-level aggregation. Tables E3-E5 show that our results associating migration and earnings, employment and non-wage benefits are very similar to the municipality-level estimates, although standard errors increase substantially, as one would expect considering that there are fewer units from which we can leverage variation. All those results indicate that spatial correlation among rainfall shocks in origin municipalities is not a source of relevant bias in our setting.

Tabela E1 – **Summary statistics: Micro- and meso-regions in the Semiarid**

Panel A - Micro-regions	Mean	Std. Dev.	Min	Max	Obs
Rainfall shock	-0.01	0.20	-0.70	0.47	2,055
Temperature shock	0.01	0.01	-0.01	0.05	2,055
Out-migration	1,500.70	1,371.95	6.00	9,685.00	2,055
Out-migration rate (p.p.)	1.08	0.41	0.12	3.12	2,055
Population	148,981.55	128,183.19	4,968	752,719	2,055
Area	7,150.16	7,857.60	84.94	55,358.33	2,055
Number of municipalities	8.20	4.56	2.00	26.00	2,055
Panel B: Meso-regions	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**
Rainfall shock	-0.02	0.20	-0.69	0.44	525
Temperature shock	0.01	0.01	-0.01	0.05	525
Out-migration	5,874.18	5,766.16	51.00	34,800.00	525
Out-migration rate (p.p.)	1.08	0.37	0.24	2.32	525
Population	583,156.36	524,776.40	15,499	2,349,152	525
Area	27,986.83	30,649.61	84.94	124,505.71	525
Number of municipalities	37.20	21.51	10.00	118.00	525

Notes: Rainfall is measured in mm. Temperature is measured in degrees Celsius. Migration outflow (inflow) rate is the share of migrants over the local (native) population. The area is measured in km^2.

Tabela E2 – **Migration outflows induced by weather shocks according to different aggregation levels**

	(1)	(2)	(3)
	Municipality	Micro-region	Meso-region
Rainfall_{t-1}	-0.099***	-0.094***	-0.099***
	(0.028)	(0.032)	(0.025)
Observations	14,400	2,055	525
Origins	960	137	35
R-Squared	0.461	0.764	0.866
Moran's I	0.947	0.158	0.075
Time dummies	✓	✓	✓
Origin dummies	✓	✓	✓
Temperature shocks	✓	✓	✓

Notes: Each observation is a region-year cell. The dependent variable is the number of individuals who left the origin region divided by the total population in the 1991 Census. Rainfall is measured as log-deviation from the historical average. All specifications include controls for temperature shocks, municipality and year fixed effects. Moran's I show the spatial correlation in rainfall shocks among origin regions. Standard errors are clustered at the respective region level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela E3 – **Effects of migration on earnings according to different aggregation levels**

	(1)	(2)	(3)
	A. Δ log earnings		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.869***	-0.846***	-0.871
	(0.197)	(0.302)	(0.550)
Observations	11,460	1,644	420
Regions	955	137	35
	B. Δ log earnings, formal sector		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.593***	-0.558*	-0.556
	(0.198)	(0.290)	(0.527)
Observations	11,460	1,644	420
Regions	955	137	35
	C. Δ log earnings, informal sector		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.746***	-0.745***	-0.769**
	(0.123)	(0.201)	(0.343)
Observations	11,460	1,644	420
Regions	955	137	35
Demographics	✓	✓	✓
Time dummies	✓	✓	✓

Notes: This table shows origin-level SSIV coefficients on changes in log earnings, by sector. Each observation is an origin region-year cell. The informal sector also includes self-employed workers. All specifications include time and control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Column (1) replicates the same results from Column (3) of Table 1.6.5. In columns (2) and (3) we aggregate the origin-level shocks at the micro- and meso-region levels, respectively. All regressions are weighted by native working-age population in 1991. Standard errors clustered at the respective aggregation level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela E4 – **Effects of migration on employment according to different aggregation levels**

	(1)	(2)	(3)
	A. Δ employment rate		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.018	-0.003	0.007
	(0.034)	(0.058)	(0.091)
Observations	11,460	1,644	420
Regions	955	137	35
	B. Δ formal employment rate		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.126***	-0.117**	-0.125
	(0.037)	(0.055)	(0.098)
Observations	11,460	1,644	420
Regions	955	137	35
	C. Δ informal employment rate		
	Municipality	Micro-region	Meso-region
Predicted inflow	0.108***	0.114**	0.133
	(0.034)	(0.057)	(0.095)
Observations	11,460	1,644	420
Regions	955	137	35
Demographics	✓	✓	✓
Time dummies	✓	✓	✓

Notes: This table shows origin level SSIV coefficients of Δ the proportions of employed natives, by sector. Each observation is an origin region-year cell. The informal sector also includes self-employed workers. All specifications include time and control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Column (1) replicates the same results from Column (3) of Table 1.6.6. In columns (2) and (3) we aggregate the origin-level shocks at the micro- and meso-region levels, respectively. All specifications use the same set of controls defined in Table 1.6.6. All regressions are weighted by native working-age population in 1991. Standard errors clustered at the respective aggregation level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela E5 – **Effects of migration on non-wage benefits according to different aggregation levels**

	(1)	(2)	(3)
	A. Food		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.687***	-0.658***	-0.688***
	(0.086)	(0.134)	(0.216)
Observations	11,460	1,644	420
Regions	955	137	35
	B. Transport		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.372***	-0.305***	-0.290*
	(0.062)	(0.104)	(0.157)
Observations	11,460	1,644	420
Regions	955	137	35
	C. Health		
	Municipality	Micro-region	Meso-region
Predicted inflow	-0.315***	-0.289***	-0.312*
	(0.064)	(0.097)	(0.173)
Observations	11,460	1,644	420
Regions	955	137	35
Demographics	✓	✓	✓
Time dummies	✓	✓	✓

Notes: This table shows origin level SSIV coefficients of change in the proportions of formal sector workers who receive health insurance, food, or transport subsidies. Each observation is an origin region-year cell. All specifications include time and control for destination-level 1991 characteristics (log of the working-age native population; shares of the population aged 15-25, 26-50, 51-65 and older than 65; share of the non-white population; share of the population with a college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing; logs of the average household income and size; and the shares of households with access to electricity and piped water) interacted with time dummies. Column (1) replicates the same results from Column (3) of Table 1.6.9. In columns (2) and (3) we aggregate the origin-level shocks at the micro- and meso-region levels, respectively. All regressions are weighted by native working-age population in 1991. Standard errors clustered at the respective aggregation level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

2 Women's Career and Care Responsibilities: The Role of Migrant Domestic Workers in Brazil

With Raphael Corbi and Renata Narita

2.1 Introduction

While several studies have shown that low-skilled immigrant inflows help highly educated women in developed countries to combine their work and childcare responsibilities (EAST; VELÁSQUEZ, 2022; CORTÉS; PAN, 2019)[2.1.1], much less is known about how domestic workers enable women in developing countries to pursue careers while caring for children. Much of the existing literature has focused on immigrant inflows as a mechanism to alleviate the constraints that prevent high-educated women to enter the labor market. Cortes e Tessada (2011) show that high-educated women in the U.S. living in cities with larger concentrations of low-educated immigrants work more hours than similar women living in areas with fewer immigrants. Similar results were found in Spain (FARRÉ; GONZÁLEZ; ORTEGA, 2011), Italy (BARONE; MOCETTI, 2011) and Hong Kong (CORTES; PAN, 2013). In more recent work, Cortés e Pan (2019) show that when there are more low-educated immigrants working in the service sector, high-educated women become more likely to work in occupations that reward long hours of work, leading to decreases in the gender wage gap. If all these effects were driven by complementarities in the labor market as opposed to home production, then we should expect fertility to decrease, but instead, there is some evidence of low-educated immigration leading to increases in the fertility of high-educated women (FURTADO, 2016). As a whole, this literature implies that the more low-educated workers are available to work in the care sector, the easier it is for high-educated women to combine their work with caregiver roles.

[2.1.1] See also Furtado (2015) for an overview.

We examine these issues in the context of Brazil, a large developing country where poverty and income inequality have fallen quite substantially in recent decades, but where gender norms and family values have adjusted more slowly[2.1.2]. Over the last decades, Brazilian working women are increasingly entering professions dominated by men. Still, women in the formal sector work fewer hours and earn significantly lower wages than their male counterparts, despite achieving higher levels of education. They also spend significantly more time on household chores than men (AGÉNOR; CANUTO, 2013) even though many middle-class families in Brazil have some help with domestic chores.

To explore how care responsibilities affect women's career paths, we consider the impact of weather-driven migration flows of low-educated women from Brazil's Semiarid region, who tend to work in urban households as housekeepers and home care workers, on labor market outcomes of high-skilled women. Between 1982 and 2010, about 4.3 million women left the Semiarid region, a historical source of migrants, to settle in other regions in Brazil. Among the migrants who were able to find a job (1.6 million), 29% were occupied as domestic workers in the destination areas.

We exploit variation in weather conditions over time and across origin municipalities in the Semiarid region to identify the causal effects of increasing the supply of migrant workers on the female labor force participation in the destination areas. As in Corbi, Ferraz e Narita (2021), we use rainfall and temperature shocks at the origin areas to predict the number of migrants leaving each Semiarid's municipality. Then, we allocate these predicted flows in the destination areas using the preexisting share of migrants from those same origins.

Our results show that increasing the number of migrants from the Semiarid region by one percentage point increases the labor force participation rate of high-educated women by 0.21*p.p.*, but it has no effect on the probability of working at least 40 hours per week. We also find that the impact on labor force participation is slightly larger (0.36*p.p.*) for women living in households with young children (less than 6 years old), but there is no significant impact for women living in

[2.1.2] This is not an exclusive Brazilian feature. See Jayachandran (2015) for a discussion about the literature on the persistence of gender norms when economic conditions change.

households without children[2.1.3]. Results are much stronger if we consider only the periods 1980-1991 and 1991-2000, which is not totally surprising considering that the share of incoming migrant women employed as domestic workers dropped almost 50% between 2000 and 2010. Also, these impacts are substantially higher in municipalities with a lower presence of preschools. All this evidence suggests that it is the woman's role as a caregiver, more than unpaid domestic work, the main constraint that hinders her path toward professional achievements. finally, we also found that living in more conservative environments or in places with high violence prevents them from fully benefiting from incoming migration.

In addition to this introduction, in Section 2.2 we provide some background information on the migration from the Semiarid region and the relationship between domestic workers and female employment in Brazil. Section 2.3 describes the dataset and the empirical strategy we use to identify the causal effects of incoming migration on the labor market outcomes of women in the destination municipalities. In Section 2.4 we present and discuss our results and conclude in Section 2.5.

2.2 Background

In this section, we first describe the economic background and weather conditions in the Semiarid region and contextualize the relationship between domestic workers and female employment in the destination municipalities.

2.2.1 Brazilian Semiarid

The Brazilian Semiarid encompasses 960 municipalities spread over 9 states, covering an area of around 976,000km^2.[2.2.1] According to the official definition by

[2.1.3] We choose the cutoff of 6 years old because this is the age when schooling begins to be mandatory in Brazil. Our hypothesis is that when children are in school, women are less constrained to enter the labor force, at least during school hours.

[2.2.1] That is roughly the same as the territory of Germany and France combined. The semiarid comprises 11 percent of the Brazilian territory and includes parts of almost all Northeastern states (except for *Maranhão*) plus the northern area of *Minas Gerais*, but it does not cover any state capital.

the Ministry of National Integration, a municipality qualifies as Semiarid if at least one of these three criteria holds: (i) annual average precipitation below 800 mm between 1961 and 1990; (ii) aridity index up to 0.5[2.2.2]; (iii) risk of drought above 60%[2.2.3]. The average historical precipitation in the Semiarid is about 780mm, as opposed to around 1,600 mm for the rest of the country, while the average temperature is around 25°C. The rainy season occurs between November and April, with the highest levels of precipitation after February, when the sowing season typically starts.

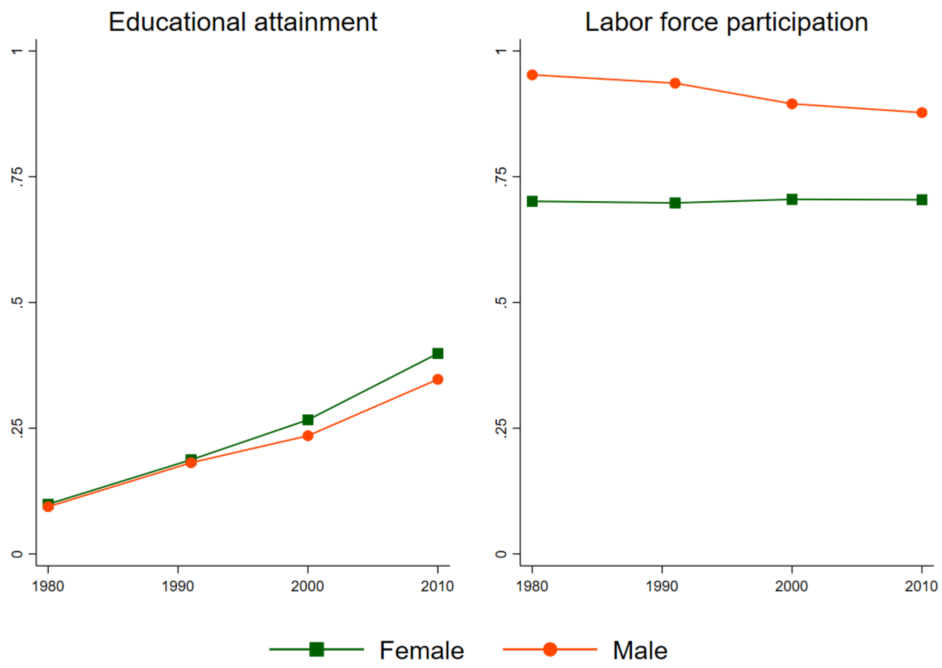
Municipalities are relatively small with a median population of around 20,000 and have economies mainly based on agriculture and cattle ranching in small subsistence properties. Local economic activity is particularly susceptible to weather shocks (WANG et al., 2004), with some studies showing a loss of up to 80% of agricultural production in periods of long drought (KAHN; CAMPUS, 1992). About 80% of the children lived below the poverty line and infant mortality reached 31 per 1,000 births in 1996, compared to a national average of 25% and 15 per 1,000 births, respectively (ROCHA; SOARES, 2015). More than 80% of the adult population had less than 8 years of schooling in 1991.

Such poor socioeconomic indicators associated with periods of extreme drought have historically driven large outflows of migrants - or so-called *retirantes* - from the Semiarid to other areas of the country (BARBIERI et al., 2010). During the 1960s and 1970s, net migration out of Northeastern states (where most of the Semiarid is located) was 2.2 and 3.0 million individuals (CARVALHO; GARCIA, 2002), which correspond to net migration rates of -7.6 and -8.7%, respectively. Between 1980 and 2010, over 4 million people left the Semiarid alone searching for better conditions elsewhere in the country. These migrants tend to be historically concentrated in some states. São Paulo alone harbored over 30 percent of the people arriving from the Semiarid in the last four decades. However, in relative terms, incoming migrants represented a population increase of above 2% for the top 10 receiving states.

[2.2.2] Thornthwaite Index, which combines humidity and aridity for a given area, in the same period.

[2.2.3] Defined as the share of days under hydric deficit, using the period 1970-1990.

Figura 2.2.1 – Gender gaps in educational attainment and labor force participation



Notes: This figure shows the gender gap in educational attainment and labor force participation. The left panel presents the share of people aged 25 to 65 with a complete high school education or higher, by gender. The right panel shows the fraction of each these groups in the labor force.

2.2.2 Domestic workers and female employment

In the last decades, women in Brazil had experienced a widening gap in educational attainment relative to men. Between 1980 and 2010, the share of high-educated (at least complete high school) women in the age group 25-65 rose from 9.8% to 39.8%, an increase about 9% higher than that observed by men in the same age group. Meanwhile, the labor force participation of those high-educated women remained essentially flat during this period (see Figure 2.2.1).

There are several aspects that can contribute to reconciling those facts. Traditional gender norms can pose some difficulties for women to advance in their careers. While part of this may operate through bias against women among bosses, coworkers, and even customers in the workplace, there is a large and growing literature, especially focusing on developed countries, suggesting that the difficulty in combining work and family responsibilities can be a major driver of gender gaps in professional achievement[2.2.4].

The more entrenched the norms, especially those related to women's role as caregivers, the more constrained their decision to enter the labor force would be. There is some evidence that in Brazil, as well as other developing countries, such norms seem to be an important mechanism constraining women's advance in the workplace. In 2014, women dedicate 25 hours per week, on average, to domestic chores, while men devote only 11 hours per week (PINHEIRO et al., 2016). Also, using data from three waves of the World Value Survey (WVS), between 1989 and 2009, we see that 32% of the Brazilian interviewees agreed with the statement "When jobs are scarce, men should have more rights to them than a woman" (INGLEHART et al., 2014).

There is a lot of room for family policies aiming to help increase female labor force participation, like expanding the supply of child care (LEFEBVRE; MERRIGAN, 2008; LEFEBVRE; MERRIGAN; VERSTRAETE, 2009; BAKER; GRUBER; MILLIGAN, 2008; BAUERNSCHUSTER; SCHLOTTER, 2015); paid family leave (JONES; WILCHER, 2019; REY; KYRIACOU; SILVA, 2021); and tax credits (BLUNDELL; HOYNES, 2004; BLUNDELL et al., 2000) [2.2.5].

We argue that an influx of low-educated internal migrants entering a given destination area can be a factor that helps to lift some of those constraints and allow high-educated women to enter the labor force. While low-educated international migration to Brazil has been rather low in recent years[2.2.6], internal migration is quite substantial, and many of the incoming migrant women work as housekeepers and childcare providers in large cities. Figure 2.2.2 shows that between 1980 and 2000, over 40% of the low-educated incoming migrant women, coming from municipalities in the Semiarid region, were employed as domestic workers. And,

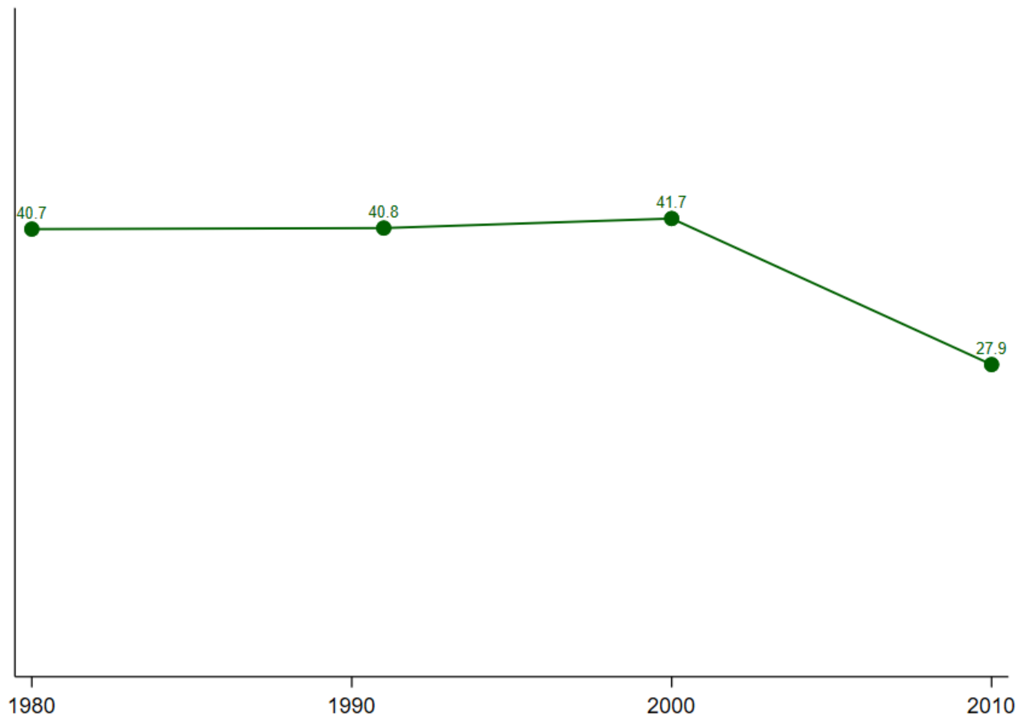
[2.2.4] For example, Kleven, Landais e Søgaard (2019) show that women's earnings drop significantly after childbirth and never catch up, while there is no impact on men's earnings.

[2.2.5] See Olivetti e Petrongolo (2017) for a comprehensive literature review.

[2.2.6] The number of international migrants as a share of the total population has dropped from 1.9% in 1960 to 0.4% in 2010 (Calculations by the authors using data available on the Migration Policy Institute's Migration Data Hub: https://www.migrationpolicy.org/programs/migration-data-hub).

even though that proportion has declined, in 2010 more than one in four of those women were still employed as domestic workers.

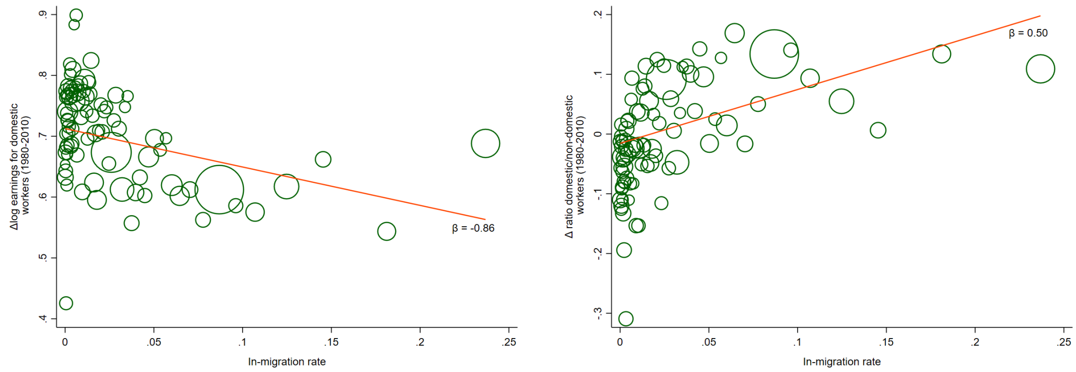
Figura 2.2.2 – Share of domestic workers among Semiarid migrants' female low-educated workers



Notes: This figure shows the evolution of the share of domestic workers in the total employment of low-educated women who migrated from the Semiarid region into other destination municipalities.

All that influx of migrants from the Semiarid into the destination areas is likely to affect the market for domestic workers, increasing the supply and, therefore, diminishing the cost of hiring low-educated workers. In Figure 2.2.3 we show that this mechanism seems to operate in our context. We estimate the relationship between the long-term (1980-2010) changes in the log earnings (2.2.3a) and the ratio of domestic to non-domestic employment (2.2.3b) among low-educated workers against the accumulated observed migrant inflow, measured as a share of the native population in 1980. To provide a better visualization we collapsed those variables into 100 bins, based on the distribution of the migration rate. As expected, there is a negative (positive) relationship between the earnings (employment) of low-educated native workers and the accumulated migration in destination areas.

Figura 2.2.3 – Incoming migration and the labor market for domestic workers



(a) Earnings (b) Employment

Notes: This figure shows the relationship between the inflow of migrants and changes in the log earnings and employment of low-educated domestic workers at the destination municipalities. All variables are collapsed into 100 bins, based on the distribution of the observed migration rate. The vertical axes are the average long differences (1980-2010) in (a) the log earnings of low-educated female domestic workers and (b) the ratio of domestic to non-domestic employment. The horizontal axis is the average accumulated migration rate. The β coefficients are estimated before collapsing the variables. Circle size is proportional to the average population size in each bin.

2.3 Data and Empirical Strategy

In this section, we begin by listing the main sources of data used in our analysis and showing some descriptive statistics. Then we describe the empirical framework and report first-stage estimates that link observed migration patterns to our predicted migration flows.

Migration

We extend the data collected by Corbi, Ferraz e Narita (2021) to the period 1982-2010. Migration data were drawn from four waves of the Brazilian Census

(1980, 1991, 2000, and 2010), provided by the *Instituto Brasileiro de Geografia e Estatística (IBGE)*.[2.3.1] Using the answers about the municipality of origin and year of migration, we construct a measure of the yearly outflow from each municipality in the Semiarid and a measure of the inflow to each destination (all but Semiarid) from 1982 to 1991 and from 1996 to 2010[2.3.2]. Then, we use the 1980 Census to build a "past settlement" measure by associating the share of migrants from each Semiarid municipality who resides in each destination.

Weather shocks

The measures of weather shocks were constructed using historical information on rainfall and temperature, retrieved from the Climatic Research Unit at the University of East Anglia (HARRIS et al., 2020). The CRU Time Series (CRU-TS)[2.3.3] provides worldwide monthly gridded data of precipitation and temperature, at the 0.5° × 0.5° level (0.5° is around 56km on the equator). We construct municipality-level monthly precipitation and temperature measures based on grid-level raw data by overlaying the Brazilian municipalities' shapefile to the CRU-TS raster and taking the average of each cell covering the municipality polygon. Then, we aggregate both variables in yearly measures. Precipitation is defined as the sum of monthly levels and temperature as the average. For each municipality in the Semiarid region, we calculate the historical mean from both variables and take the natural logarithm of these variables (both levels and long-term averages).

Finally, our weather shock variables are defined as

$$Rainfall_{ot} = \ln\left(\sum_{\tau=1}^{12} r_{o\tau t}\right) - \ln(\bar{r}_o) \tag{2.3.1}$$

[2.3.1]As several municipalities were split into new ones during the 1990s, we aggregate our data using the original municipal boundaries as they were in 1980 (so-called "minimum comparable areas" or MCA) in order to avoid potential miscoding regarding migration status or municipality of origin. We use municipality and MCA as synonyms throughout the paper.

[2.3.2]The Census round realized in 2000 only asked in which city the respondent lived 5 years past. Therefore, we can only track the migrants until 1996

[2.3.3]Data can be obtained directly from CRU website: https://crudata.uea.ac.uk/cru/data/hrg/. We used version v.4.05, which span 1901-2018.

where $r_{o\tau t}$ is the rainfall in the municipality of origin o in month τ of year t, and $\bar{r}_o$ is the municipality's historical average precipitation. Temperature is calculated in a similar way, but using the annual average instead of summation to create yearly data.

We define rainfall and temperature shocks in each municipality as deviations from its historical average[2.3.4].

Labor outcomes

We also use the Census to gather information on labor market outcomes for native individuals. To alleviate concerns that migration could affect the decision of getting more education, we restrict our sample to individuals between 25 and 65 years old, living in the municipality for 10 years or more (natives), who were not enrolled in school. We also drop destination municipalities in the Semiarid region to avoid concerns about spatial correlation in the shocks.

We create dummy variables indicating whether the individual is in the labor force and whether she works at least 40 hours/week[2.3.5]. We also create indicators to define as high-educated those individuals with a complete high school education or higher. Then, we take averages at the municipality-year level and calculate the first difference for each outcome. The final sample has 3,060 unique destination municipalities and 9,180 municipality-year observations.

Table 2.5.1 describes municipality-level data for origin (Panel A) and destination (Panel B) municipalities, collapsed into the Census-year cells. Semiarid's areas show much lower levels of rainfall, and slightly higher temperatures and are less populated than destination municipalities. On average, 12.6% of the 1980 Semiarid's population leaves every decade, resulting in an average increase of 1.1*p.p.* of the labor force in the destination.

Table 2.5.2 provides some descriptive statistics for destination municipalities. In our sample, about 14% of the women in destination municipalities are high-

[2.3.4] The historical average is calculated using the period 1901-2010.

[2.3.5] The standard working time in Brazil was 48 hours per week until the new Constitution in 1988 reduced it to 44 hours per week (GONZAGA; FILHO; CAMARGO, 2003).

educated; 37% live with children aged less than 6 years old; 34% are in the labor force and almost half of them (47%) work at least 40 hours per week. Using high-educated women as the reference group, labor force participation is very high (74%), and the probability of working more at least 40 hours per week is slightly higher (53%).

2.3.1 Empirical Strategy

In this section, we describe the empirical framework that combines exogenous shocks in weather conditions with the predetermined pattern of settlement of migrants from the Semiarid region, which allows us to establish a causal relationship between migration inflows and labor market conditions for women in destination areas.

We specify a model for the changes in labor market outcomes of native individuals in destination municipalities as a function of internal migration flows. Specifically, we assume that

$$\Delta y_{dt} = \alpha + \beta m_{dt} + \gamma X_d + \psi_t + \epsilon_{dt} \tag{2.3.2}$$

where y_{dt} is a vector of labor outcomes at destination municipality d in census-year t, m_{dt} is the accumulated migrant inflow from the Semiarid region entering the destination area during the period covered by the Census round t, X_d are destination-level baseline controls, ψ_t absorb time fixed effects and ϵ_{dt} is the error term. There are two main concerns to establish β as a causal parameter. First, the observed inflow of migrants, m_{dt}, is an equilibrium result between demand and supply. Second, time-varying unobserved characteristics can be correlated with incoming migration, rendering OLS estimates biased.

To overcome these issues we create a shift-share instrumental variable (SSIV) in two steps. First, we use the variation in weather conditions at the origin to predict the outflow rate[2.3.6] of migrants from the Semiarid region. More specifically,

[2.3.6]We define this rate as the observed number of migrants leaving each municipality divided by we estimate the following regression:

$$m_{oy} = \alpha + \beta' Z_{oy-1} + \phi_o + \delta_y + \varepsilon_{ot} \tag{2.3.3}$$

where Z is a vector of rainfall and temperature shocks, defined as the log deviations from its historical averages, at the origin municipality o in the previous year; ϕ_o and δ_y are municipality and year fixed effects, respectively; and ε_{oy} is a random error term. For each year the predicted number of migrants who leave their hometowns is obtained by multiplying this predicted rate by the municipality population reported in the 1980 Census:

$$\widehat{M}_{oy} = \widehat{m}_{oy} \times P_o \tag{2.3.4}$$

In the second step, we use the past settlements of migrants from the origin o to municipality d in order to distribute them throughout the destination areas, defining our SSIV as

$$\widehat{M}_{dt} = \sum_{t \in C} \left(\sum_{o=1}^{O} s_{od} \times \widehat{M}_{oy} \right) \tag{2.3.5}$$

where $C = \{1991, 2000, 2010\}$ are the three rounds of the Census we can use to track the migrants; and s_{od} is the share of migrants from origin municipality o who lived in the destination area d in 1980.

To make interpretation easier we divide the predicted number of incoming migrants by the native population in each destination municipality in 1980, P_d[2.3.7], resulting in

$$\widehat{m}_{dt} = \frac{\widehat{M}_{dt}}{P_d} \tag{2.3.6}$$

its population in the 1980 Census

[2.3.7] We highlight that the denominator P_d is only a normalization that helps interpret the coefficients of interest. It does not play any role in identification.

as the SSIV for the observed incoming migration rate in destination municipality d.

In other words, our instrument $\widehat{m}_{dt}$ is a combination of exogenous shocks or 'shifts' $\widehat{M}_{ot}$ (weather-driven outflows) and exposure 'shares' ($s_{od} \geq 0$) or past settlement patterns.

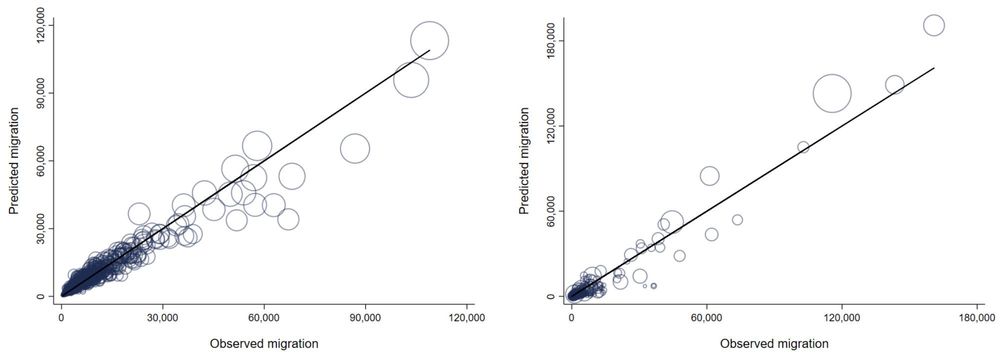
2.3.2 Prediction of the Weather-induced out-migration

We begin the exploration of our first-stage results by estimating variations of specification 2.3.3 and report the estimates in Table 2.5.3. All regressions control for temperature shocks and include time and municipality fixed effects. In columns (2)-(7) we include a linear trend interacting time dummies with 1980 characteristics (log of the local population; age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing). All regressions are weighted by the harvested agricultural area[2.3.8], measured as a fraction of the municipality's total area. Columns (3)-(6) include up to three lags, contemporaneous and one lead of rainfall and temperature shocks. For brevity, we omit (mostly insignificant) coefficients associated with temperature shocks in Table 2.5.3. Standard errors are clustered at the municipality level. In column (7) we use a more flexible form, allowing differential effects for shocks below and above the historical average.

As expected, rainfall shocks in the previous year are negatively correlated with migration outflows indicating that Semiarid's inhabitants leave the region during drought periods. Coefficient estimates are remarkably stable across specifications and adding more lags does not change the baseline results. More important to our identification, we control for rainfall and temperature shocks one year forward to ensure that our instrument is not contaminated by serial correlation in the weather measures. The coefficient on $rainfall_{t+1}$ reported in column (6) is small in magnitude and not statistically significant, while the coefficient for $rainfall_{t-1}$

[2.3.8] We calculate the average harvested area for each municipality, between 1974-1980. We dropped crops that, for a given year, represent less than 25% of the cumulative harvested area.

Figura 2.3.1 – Observed vs predicted migration



(a) Migration outflow (b) Migration inflow

Notes: This figure presents the relationship between the predicted and observed migration flows across Brazilian municipalities from 1982 to 2010. Panel (a) shows the number of migrants leaving the Semiarid region to non-Semiarid municipalities. Panel (b) shows the number of incoming Semiarid migrants for destination municipalities. In Panel (b) we dropped two destination municipalities with predicted inflow over 200,000 people, to make the visualization clearer. The circle size represents the municipality's total population in 1980. *Data source*: Census microdata (IBGE).

remains almost unchanged. Our estimates indicate that a municipality where annual rainfall is 10% below the historical average will experience an increase of 1.2*p.p.* in migration outflow rate. Finally, our estimates in column (7) show that droughts are the main driver of out-migration in the Semiarid region.

As mentioned in section 2.3.1, after leveraging variation in weather shocks to predict the number of migrants leaving the Semiarid region, we use the past settlement of pattern of these migrants to allocate them to the destination areas. One important criterion to ensure the validity of our empirical strategy is that both predicted migration outflow and inflow rates, $\widehat{m}_{ot}$ and $\widehat{m}_{dt}$ respectively, should be strongly correlated with their observed counterparts.

Figure 2.3.1 reveals that our predictions provide a good fit for the observed migration. Panel (a) shows the relationship between the predicted and observed number of migrants leaving the Semiarid region and entering non-Semiarid municipalities, accumulated over the period 1982-2010. Panel (b) shows the predicted and observed numbers of incoming Semiarid migrants for destination municipalities.

Overall, this analysis shows that our strategy provides a strong first-stage as predicted migration rates, $\widehat{m}_{dt}$, are strongly correlated with observed migration.

In almost all specifications we show the first stage F-stat larger than 150.[2.3.9]

2.4 Migration Inflows and Female Labor Force Participation

In this section, we establish the relationship between labor market outcomes for female natives and the incoming migration at destination areas.

Table 2.5.4 reports the main results for our SSIV estimates. Columns (1)-(3) present the effects on the labor force participation rate, while columns (4)-(6) show the impacts on the share of individuals working at least 40 hours per week. All regressions include time dummies and control for destination-level 1980 characteristics (log of the native population; average age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing) and are weighted by the working-age native population in 1980. Standard errors are clustered at the destination municipality level.

Panel A shows that increasing the inflow of migrants from the Semiarid region by one percentage point in the destination areas raises the labor force participation of high-educated women by 0.21*p.p.* and a small increase (0.11*p.p.*) in the share of low-educated women working at least 40 hours per week. As an important robustness check, we show in Panel B that there is no effect on the outcomes for native high-educated men, reassuring us that is unlikely that our results are driven by some local demand shock instead of the supply of migrant domestic workers.

From now on, we focus on high-educated women as they are the group more likely to benefit from the arrival of incoming migrants. In Table 2.5.5 we show that high-educated women with children aged less than 6 years old, the age of mandatory school enrollment, are more likely to be impacted by the increasing availability of

[2.3.9] A sufficiently high F-stat avoids weak instrument concerns, especially in the light of the recent discussion in Lee et al. (2020) who show that a 5 percent test requires an F statistic of 104.7, significantly higher than the broadly accepted threshold of 10.

migrant domestic workers. In Panel A, we use the full sample, stacking the first differences for all the Census periods (1980-1991, 1991-2000, 2000-2010) and show that for women living in households with children, raising the number of incoming migrants by one percentage point increases the labor force participation rate by 0.36*p.p.*, while there is no significant effect for women without care responsibilities. We also find an increase of 0.16*p.p.* in the probability of increasing the number of hours worked, although this estimate is less precise. In Panels B, C, and D we show the same analysis separately for each period. Unsurprisingly, between 1980 and 2000, when the proportion of migrants from the Semiarid employed as domestic workers is higher, the impacts are substantial, but they drop in the last period. In the period 1980-1991, both labor force participation and the share of women working at least 40 hours per week increased substantially (around 1*p.p.*), while in the next period, only the attachment to the labor force increased (1*p.p.*)

One way to alleviate the constraints women face is to provide child care outside the household. In places where such facilities are available, we should expect the importance of incoming domestic workers to be lower[2.4.1]. To test this hypothesis we run the same analysis separately for municipalities with low and high presence of preschool in 1980[2.4.2]. We use the share of children aged less than 6 years old who are enrolled in preschool as a proxy for the presence of daycare and divide the destination municipalities into two groups: low-presence are those below the median, while high-presence are those above the median share of enrollment. Comparing column (2) from Panels A and B from Table 2.5.6, we show that the magnitude of the impact from an inflow of migrants is much larger for high-educated women with children in destinations with low presence of preschool (0.59*p.p.*) than that in municipalities with high availability of preschool (0.33*p.p.*).

We provided some evidence that there is an important economic mechanism explaining the impacts of internal migration on high-educated female workers.

[2.4.1] At least for women living in households with children. In theory, it would also be possible that women without children benefited from domestic workers alleviating the burden of other domestic chores.

[2.4.2] Ideally we should use the availability of daycare as a measure of how constrained women are in a given destination. Unfortunately, there is no data available for this period, forcing us to use the presence of preschool as a proxy for daycare.

However, other constraints could be preventing women to benefit from such a mechanism. For example, in places where gender norms are more entrenched women could be less responsive to these economic incentives. In order to assess this mechanism, we repeat the same analysis using two distinct groups. In the first, we use only households where the head, the spouse, or the parents/in-laws are evangelical, while in the second group, we use only households where they identified as having another religion or none at all. We use this definition because the evangelicals in Brazil tend to be more conservative than other religious denominations (CORBI; SANCHES et al., 2022). Table 2.5.7 reveals that in places where the population is more religious, high-educated women with children increase their labor force participation way more than in destination municipalities with a less religious population. One caveat is that, although the magnitude of the effect is larger, the estimate is less precise.

Finally, although our estimates show an important role of the inflow of migrants from the Semiarid region on the labor force participation of high-educated women in destination municipalities, it is not obvious that such effects should always be positive. One possible reason is that, even though gender norms are more likely to constrain the prospect of women, other situations may reduce their likelihood of participating in the labor force—for instance, the perception of violence, especially against women. Velásquez (2020) shows that, in Mexico, local violence forces women to reduce the number of hours worked or exit the labor force entirely. We provide a test for this mechanism in Table 2.5.8, where we divide destination municipalities by their ranking in the distribution of the homicide rate in 1980. Low-violence municipalities are those below the median of the homicide rate, while high-violence destinations are those above the median. Our estimates suggest that this mechanism seems to be operating in our context. The labor force participation of high-educated women increases more ($0.57 p.p.$) in less violent municipalities than in destinations with more cases of violence against women ($0.34 p.p.$).

2.5 Discussion and Concluding Remarks

In this paper, we investigate the effects of increasing the inflow of migrants from the Semiarid region on the labor force participation of high-educated women in destination municipalities. We use a shift-share instrument approach combining variation in the number of people leaving their hometowns, driven by weather shocks, with past settlement patterns to exploit exogenous variation in the number of migrants entering each destination municipality.

Our estimates indicate that an exogenous supply shock of low-educated workers, more likely to be employed as domestic workers, increases the labor force participation rate of high-educated women, especially those who are more constrained by their care responsibilities.

Such effects are amplified in destination municipalities with a lower presence of preschools but reduced in places with more violence and for women in more conservative environments.

Figures and tables

Tabela 2.5.1 – **Summary statistics: weather and migration data**

Panel A: Origin (Semiarid)	Mean	Std. Dev.	Min	Max	Obs
Annual Rainfall	1,332.57	309.16	550.58	2,739.67	2,607
Rainfall shock	-0.02	0.08	-0.24	0.15	2,607
Annual Temperature	25.53	1.42	13.80	28.35	2,607
Temperature shock	0.01	0.01	-0.01	0.04	2,607
Out-migration	2,498.88	3,494.79	55	44,698	2,607
Out-migration rate (p.p.)	12.63	5.27	0.42	47.03	2,607
Population	24,443.47	36,301.41	1,265.00	556,642	2,607
Panel B: Destination (Non-Semiarid)	Mean	Std. Dev.	Min	Max	Obs
Annual Rainfall	2,600.08	594.64	0.00	5,392.33	9,024
Rainfall shock	0.02	0.06	-0.28	0.19	8,976
Annual Temperature	22.64	3.19	0.00	28.38	9,024
Temperature shock	0.03	0.02	-0.01	0.07	8,976
In-migration	400.12	4,529.95	0.00	230,124	9,024
In-migration rate (p.p.)	1.08	3.48	0.00	120.86	9,024
Population	48,072.19	253,241.25	716	11,253,503	9,024

Notes: Rainfall is measured in mm, while the temperature is measured in degrees Celsius, both during the rainy season (November-April) in the Semiarid region. Rainfall and temperature shocks are measured as the log difference from their respective historical averages. The migration outflow (inflow) rate is the share of migrants over the local (native) population in 1980. All statistics are collapsed into the Census-year level.

Tabela 2.5.2 – **Summary statistics: Women in destination municipalities**

	All women				
	Mean	Std. Dev.	Min	Max	Obs
Age	41.57	2.5	25	48.04	15,951
Black	6.04	5.45	0	100	15,951
White	51.93	27.83	0	100	15,951
High-education	14.09	10.98	0	100	15,951
Children < 6 years	36.83	13.34	0	100	15,951
Labor Force	34.09	14.67	0	100	15,951
Work at least 40 hours/week	47.32	24.40	0.00	100.00	15,951
	Only high-educated women				
	Mean	Std. Dev.	Min	Max	Obs
Children < 6 years	35.97	18.04	0.00	100.00	15,421
Labor force	73.66	15.50	0.00	100.00	15,421
Work at least 40 hours/week	52.94	22.77	0.00	100.00	15,405

Notes: Each observation is a destination municipality-year cell. In Panel A we consider all women in destination municipalities, while in Panel B the reference group is high-educated (complete high school or higher) women.

Tabela 2.5.3 – **Migration outflows induced by weather shocks**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rainfall$_{t-1}$	-0.119***	-0.124***	-0.129***	-0.128***	-0.127***	-0.122***	
	(0.046)	(0.047)	(0.047)	(0.048)	(0.047)	(0.047)	
Rainfall$_{t-2}$			-0.074	-0.061			
			(0.054)	(0.055)			
Rainfall$_{t-3}$				0.126**			
				(0.049)			
Rainfall$_{t}$					-0.080		
					(0.059)		
Rainfall$_{t+1}$						-0.087	
						(0.053)	
Negative shocks$_{t-1}$							0.129**
							(0.064)
Positive shocks$_{t-1}$							-0.118
							(0.091)
Observations	21,725	21,725	21,725	21,725	21,725	21,725	21,725
Municipalities	869	869	869	869	869	869	869
Time dummies	✓	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓	✓
Baseline × time		✓	✓	✓	✓	✓	✓

Notes: Each observation is an origin municipality-year cell. The dependent variable is the number of individuals who left the origin municipality divided by the total population in the 1980 Census. Rainfall shocks are measured as the annual log-deviation from the historical average. All specifications include controls for temperature shocks, municipality and year fixed effects. Columns (2)-(6) also control for municipality-level 1980 characteristics (log of the local population; age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing) interacted with time dummies. All regressions are weighted by the harvested agricultural area, measured as a fraction of the municipality's total area. Standard errors are clustered at the municipality level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 2.5.4 – **Effects of migration on labor force participation and hours worked**

	Panel A: Females					
	Labor force participation			Work 40 hours/week or more		
	All Women (1)	High-education (2)	Low-education (3)	All Women (4)	High-education (5)	Low-education (6)
Migrant inflow	0.034	0.210**	0.004	0.105**	0.041	0.108*
	(0.088)	(0.098)	(0.102)	(0.044)	(0.083)	(0.057)
Observations	9186	8789	9181	9186	8776	9181
F-stat (IV)	155	154	155	155	154	155
	Panel B: Males					
	Labor force participation			Work 40 hours/week or more		
	All Men (1)	High-education (2)	Low-education (3)	All Men (4)	High-education (5)	Low-education (6)
Migrant inflow	0.151	0.070	0.142	0.084	-0.005	0.061
	(0.100)	(0.062)	(0.093)	(0.052)	(0.049)	(0.058)
Observations	9185	8564	9184	9185	8547	9184
F-stat (IV)	155	153	155	155	153	155

Notes: This table shows SSIV coefficients on changes in the municipality-level labor force participation rate and the share of individuals working at least 40 hours per week. The dependent variables are the stacked first differences (1980-1991, 1991-2000, 2000-2010) for each outcome. In Panel A, the sample uses only females; in Panel B, it has only males. High-education individuals are those who completed a high school education or higher. All regressions include time dummies and control for destination-level 1980 characteristics (log of the native population; average age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing) and are weighted by the working-age native population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 2.5.5 – **Effects of migration on labor force participation and hours worked for women with children**

	Panel A: Stacked differences (1980-2010)					
	Labor force participation			Work 40 hours/week or more		
	All Women (1)	With children (2)	No children (3)	All Women (4)	With children (5)	No children (6)
Migrant inflow	0.210**	0.363***	0.060	0.041	0.161*	-0.056
	(0.098)	(0.112)	(0.101)	(0.083)	(0.084)	(0.113)
Observations	8789	8390	7801	8776	8297	7645
F-stat (IV)	154	150	151	154	150	151
	Panel B: Δ 1980-1991					
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant inflow	0.589**	0.947***	0.408	0.890***	1.071***	0.681*
	(0.239)	(0.284)	(0.277)	(0.242)	(0.256)	(0.381)
Observations	2665	2329	1985	2657	2296	1942
	Panel C: Δ 1991-2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant inflow	0.836***	1.007***	0.559**	-0.406	-0.314	-0.445
	(0.209)	(0.299)	(0.249)	(0.271)	(0.291)	(0.355)
Observations	3058	3001	2797	3053	2948	2709
	Panel D: Δ 2000-2010					
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant inflow	0.040	0.078	-0.105	-0.141**	0.021	-0.178*
	(0.140)	(0.131)	(0.143)	(0.070)	(0.082)	(0.100)
Observations	3066	3060	3019	3066	3053	2994

Notes: This table shows SSIV coefficients on changes in the municipality-level labor force participation rate and the share of individuals working more than 48 hours per week, for high-educated women with children aged less than 6 years. In Panel A the dependent variables are the stacked first difference for each outcome for all periods (1980-1991, 1991-2000, 2000-2010). In Panels B-D we present each Census period separately. All regressions time dummies and control for destination-level 1980 characteristics (log of the native population; average age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing) and are weighted by the working-age native population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 2.5.6 – **Effects on labor force participation and hours worked, by the presence of preschool**

	Labor force participation			Work 40 hours/week or more		
Panel A: Low presence of preschool						
	All Women (1)	With children (2)	No children (3)	All Women (4)	With children (5)	No children (6)
Migrant inflow	0.122	0.593***	0.086	-0.088	-0.061	0.149
	(0.171)	(0.201)	(0.214)	(0.120)	(0.229)	(0.246)
Observations	4368	4154	3844	4360	4105	3775
F-stat (IV)	70	58	64	70	58	63
Panel B: High presence of preschool						
	Labor force participation			Work 40 hours/week or more		
	All Women (1)	With children (2)	No children (3)	All Women (4)	With children (5)	No children (6)
Migrant inflow	0.218**	0.332***	0.037	0.107	0.241***	-0.041
	(0.092)	(0.101)	(0.108)	(0.090)	(0.089)	(0.123)
Observations	4421	4236	3957	4416	4192	3870
F-stat (IV)	129	129	127	129	129	128

Notes: This table shows SSIV coefficients on changes in the municipality-level labor force participation rate and the share of women working at least 40 hours per week, for high-educated women with children aged less than 6 years. In Panel A our sample uses only municipalities below the median of the share of children below 6 years old enrolled in preschool, while in Panel B the sample is restricted to municipalities above the median. All regressions include time dummies and control for destination-level 1980 characteristics (log of the native population; average age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing) and are weighted by the working-age native population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 2.5.7 – **Effects on labor force participation and hours worked, by the level of religiosity**

	Panel A: Evangelical household					
	Labor force participation			Work 40 hours/week or more		
	All Women (1)	With children (2)	No children (3)	All Women (4)	With children (5)	No children (6)
Migrant inflow	0.157	-0.029	-0.227	-0.106	-0.309	-0.098
	(0.139)	(0.221)	(0.262)	(0.140)	(0.291)	(0.310)
Observations	5359	3916	3027	4949	3352	2624
F-stat (IV)	147	164	157	147	163	153
	Panel B: Non-evangelical household					
	Labor force participation			Work 40 hours/week or more		
	All Women (1)	With children (2)	No children (3)	All Women (4)	With children (5)	No children (6)
Migrant inflow	0.222**	0.380***	0.066	0.040	0.185**	-0.060
	(0.099)	(0.116)	(0.100)	(0.084)	(0.089)	(0.113)
Observations	8779	8337	7740	8762	8221	7564
F-stat (IV)	154	150	150	153	150	151

Notes: This table shows SSIV coefficients on changes in the municipality-level labor force participation rate and the share of women working at least 40 hours per week, for high-educated women with children aged less than 6 years. In Panel A our sample uses only women in households where either the head, the spouse, or the parents/in-laws are evangelical, while in Panel B we use only households where they declared another or none religion. All regressions include time dummies and control for destination-level 1980 characteristics (log of the native population; average age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing) and are weighted by the working-age native population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 2.5.8 – **Effects on labor force participation and hours worked, by the incidence of violence**

	Panel A: Low violence					
	Labor force participation			Work 40 hours/week or more		
	All Women (1)	With children (2)	No children (3)	All Women (4)	With children (5)	No children (6)
Migrant inflow	0.232	0.575**	-0.267	-0.070	0.048	-0.089
	(0.226)	(0.277)	(0.289)	(0.245)	(0.344)	(0.484)
Observations	4524	4227	3823	4515	4155	3722
F-stat (IV)	42	40	37	42	40	37
	Panel B: High violence					
	Labor force participation			Work 40 hours/week or more		
	All Women (1)	With children (2)	No children (3)	All Women (4)	With children (5)	No children (6)
Migrant inflow	0.224**	0.344***	0.094	0.041	0.140*	-0.083
	(0.093)	(0.100)	(0.099)	(0.083)	(0.079)	(0.124)
Observations	4265	4163	3978	4261	4142	3923
F-stat (IV)	138	135	138	138	135	138

Notes: This table shows SSIV coefficients on changes in the municipality-level labor force participation rate and the share of women working at least 40 hours per week, for high-educated women with children aged less than 6 years. In Panel A we use only municipalities below the median of the homicide rate in 1980. In Panel B there are only municipalities above the median. All regressions include time dummies and control for destination-level 1980 characteristics (log of the native population; average age; share of the non-white population; share of the population with an elementary, high school, and college education; share of women in the total and employed populations; shares of employment in agriculture and manufacturing) and are weighted by the working-age native population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

3 Internal Migration and Slum Expansion in Brazil

With Raphael Corbi

3.1 Introduction

Rural-urban migration is a well-known phenomenon that helps explain at least part of the rapid pace of urbanization over the last decades. Between 1980 and 2010 the number of people living in urban areas around the world increased from 1.74 to 3.57 billion, with most of this growth happening in developing countries. However, such a rapid expansion can pose some challenges to these countries. The more recent estimates suggest that about 1.2 billion people were living in slums around the world in 2018 (ROSER; ORTIZ-OSPINA, 2013). Although the precise definition can vary according to the country, usually slums are characterized by a combination of overcrowded occupation, low quality of housing material, and poor access to urban services (CAVALCANTI; MATA; SANTOS, 2019).

The role that these informal arrangements play in developing economies is far from settled in the literature. On the one hand, slums can be seen as a transitory step in a process of economic progress and accommodate incoming rural migrants seeking better prospects in the labor market (GLAESER, 2011). On the other hand, there is nothing to ensure that such arrangements are only temporary. Marx, Stoker e Suri (2013) argue that slums can result in poverty traps, because of under-investment either in human or physical capital; or because of political rents that can be extracted via patronage for example.

There are different explanations for the mechanisms driving the emergence of slums. One argument is that households have a preference to live in the formal city, but some of them may need to resort to slums because they are unable to afford the higher costs (LALL et al., 2007; FERREIRA; MONGE-NARANJO; PEREIRA, 2016; CAVALCANTI; MATA; SANTOS, 2019). Those costs are higher because of low availability of usable land (either because of physical or institutional constraints, like local regulations) or they could reflect better amenities present in the cities. Whatever the reason, slum dwellers would move out of the slums as soon as they could. There is some empirical evidence in Franklin (2018) and Belchior, Gonzaga e Ulyssea (2023)[3.1.1] that public housing programs actually incentivize dwellers to reallocate to better[3.1.2] neighborhoods. Alternatively, slum dwellers could optimally choose to live in worse conditions to increase the consumption of other goods. If their property rights over the land are weaker, they would have no incentive to make the investments needed to improve their living conditions. This argument is central in the work of Soto e Diaz (2002) who views the enforcement of property rights on land as a tool to allow access to credit markets[3.1.3].

The first mechanism above seems to be at odds with some of the data we have. One should expect that under some improvement in economic conditions, the expansion of slums would at least be stagnant. However, between 1985 and 2010, a period when GDP per capita increased substantially[3.1.4], the area of slums across Brazilian municipalities increased at an annual rate of 2.3%, jumping from 753 to 1,345km^2[3.1.5]. At the same time, the number of people living in slums also increased substantially, from 3.8 million in 1991 to 11.3 million in 2010[3.1.6], an annualized growth rate of 3.7% compared to only 0.9% annual growth in the total population. During this same period, about 2.2 million people left their hometowns every year and moved into urban areas, most of them settling in destination municipalities in metropolitan regions.

[3.1.1] Although not restricted to slum dwellers, the program they evaluate, *Minha Casa Minha Vida*, in *Rio de Janeiro* had a focus on people living in slums with high environmental risks.

[3.1.2] Better at least in some dimensions, e.g., house quality and reduced rent costs. But, as Belchior, Gonzaga e Ulyssea (2023) show the new locations also have a lower average income, lower labor market access, worse employment outcomes, and higher crime rates.

[3.1.3] There are also some papers that took this theory to empirical tests, like Field (2005), Galiani e Schargrodsky (2010) who found that land titling increased investments on the household quality but had weaker impacts on access to credit.

[3.1.4] In particular, 2000 and 2010 when Brazil showed a strong improvement in the economic conditions of its poorest population.

[3.1.5] Just to get an idea of the magnitude, the city of London has an area of about 1,572km^2. Just the increase in the area occupied by slums in Brazil corresponds to about 75% of the area of New York City.

[3.1.6] In relative terms, such an increase represents a growth from 3.7% to 6.0% of the total population during this period.

We investigate the relationship between incoming migration and the expansion of slums in the destination municipalities where those migrants arrive, from 1985 to 2010, in Brazil. We combine satellite and Census data to measure the expansion of slums and a shift-share instrument to address the concerns about the endogeneity of the migrant flows. To construct the instrument, we exploit the variation in the past settlement of migrants combined with variation in local income at the origin municipalities leveraged from shocks in agricultural prices[3.1.7].

We use two different measures of slum expansion, both important in the eyes of the policymaker. The first is a measure of urban sprawl, which we construct using data from *Mapbiomas*, a network of NGOs, universities, and tech companies, that produces an annual land use land cover (LULC) time series, using satellite images (Landsat) and a random forest algorithm to classify them[3.1.8]. We combine these images with the shapefile of slum areas provided by *IBGE*, the government agency responsible for the Census, to obtain a high-resolution measurement of the area occupied by slums, which we aggregate at the municipality level. The second is a measure of population density within the slums, which we construct using the Census microdata provided by *IBGE*. In order to construct this measure, we harmonize the Census tracts over the years[3.1.9], identify those that belong to slum areas, and aggregate the total population within these tracts at the municipality level.

Regarding the measure of urban sprawl, our estimates show that increasing the number of incoming migrants by one percent leads to an increase in the share of the area covered by slums equivalent to 22% of a standard deviation (SD) in destination municipalities outside the metropolitan areas, but no effect on municipalities that belong to the metropolitan regions. Turning to the measure of population density, we also find an increase of 0.42% in the number of people

[3.1.7] We create the push-shocks using data on land use, crop suitability, and agricultural productivity coupled with international price shocks similar to Imbert e Ulyssea (2022) and Imbert et al. (2022).

[3.1.8] Data can be obtained directly from the Mapbiomas website: https://brasil.mapbiomas.org/. See Jr et al. (2020) for more details on the LULC model. In this paper, we used Collection 6, which covers the period 1985-2020.

[3.1.9] The layout of these tracts have changed substantially over the years, especially because of the creation of new municipalities.

living in slums. In this case, the effect is larger for municipalities in metropolitan areas (0.52%) than for destinations outside those regions (0.19%). In both margins, the effects we find are larger for municipalities where the share of free usable land is lower, suggesting that the availability of land may be an important driver.

Our contribution to the literature is threefold. First, we provide two more accurate measures of the expansion of slums in Brazil, between 1985 and 2010[3.1.10], dealing with the problem of underestimation of the population in slums in 2000, first identified by Mation, Nadalin e Krause (2014). Although there are several papers related to the determinants of slum formation and its expansion in Brazil, none of them dealt with this problem before (FERREIRA; MONGE-NARANJO; PEREIRA, 2016; BRUECKNER; MATION; NADALIN, 2019; CAVALCANTI; MATA; SANTOS, 2019; LALL et al., 2007).

Second, we relate to a growing literature that uses satellite images and machine learning algorithms to detect the presence of slums, a procedure that had been used in other developing countries (GECHTER; TSIVANIDIS, 2020; HENDERSON; REGAN; VENABLES, 2021; MARX; STOKER; SURI, 2013; LALL; LEBRAND; SOPPELSA, 2021) but is novel in the Brazilian context and is also relevant for policymakers to inform their land-use decisions.

Finally, as far as we know this is the first study to provide a causal estimate of the relationship between internal migration and slum expansion in Brazil. There are some other works that focused on the estimation of housing supply elasticity, mostly in developed countries (BAUM-SNOW; HAN, 2019; GREEN; MALPEZZI; MAYO, 2005). More related to our study is Dutta, Gandhi e Green (2021) which investigate the effects of internal migration on the housing supply in India. But, they use Census data about the quality of the construction material to identify informal households. Also, Niu, Sun e Zheng (2021) study the impacts of internal migration on informal housing in China, using the shares of village areas in the urban built-up area as a proxy for the informal housing supply elasticity. As we understand, our measures of slum growth are more accurate, because they are based on the official definition of the agency responsible for the Census in Brazil,

[3.1.10] The measure of the area occupied by slums ranges from 1985 to 2010, but the measure of population covers only 1991 to 2010.

which considers not only the material characteristics of the household but also regulatory issues, like land tenure, and characteristics of the surroundings, like the suitability of the topography for constructions and the provision of basic public services.

3.2 Background

In this section, we discuss some facts about the flows of internal migration, the formation of slums in Brazil, and how they are connected.

3.2.1 Internal migration and slum growth in Brazil

Our study focuses on migration to urban areas (either rural-urban or urban-urban flows) because the emergence of slums is typically an urban phenomenon. Over the last decades, a large inflow of migrants has been moving around the country. Between 1991 and 2010, almost 54 million people moved into urban areas of the destination municipalities. As we illustrate in Figure 3.5.1, the majority of these migrants reallocate within the state borders, although there is also a substantial number of them moving across states.

Most of those migrants settled either in the large cities around the metropolitan areas or, more recently, in the smaller cities in the agricultural frontier in the center of the country, as we can see in Figure 3.5.2.

Although the slums in Brazil have an ancient history, there is not much systematic documentation of them until the 1980s. The first known Brazilian slum is called *Morro da Providência*, which appears in *Rio de Janeiro*[3.2.1] in the Census realized em 1920. At the time, *Morro da Providência* had 839 households plus 6 business constructions (IBGE, 1953a). There were also other settlements that could be characterized as slums in the city. The government also made other inquiries about the living conditions in the slums of *Rio de Janeiro* in 1933 and 1947-1948.

[3.2.1] Back then Rio de Janeiro was the Federal Capital of Brazil, which explains why the first record of slums appears there.

But, only in 1950, the official Census included a more systematic investigation of those settlements in the Federal Capital, that lasted until 1970. In Table 3.2.1 we show that at the time internal migrants had an important presence in the slums of *Rio de Janeiro*[3.2.2]. Today, most of the slums in Brazil are located in metropolitan areas, although they had been growing also in non-metropolitan regions (see Figure 3.5.4).

Tabela 3.2.1 – Population in the slums of *Rio de Janeiro*

	1950	1960	1970
Population in slums	169,305	335,063	584,983
Internal migrants	56,402	115,740	283,067
Share of migrants	33.3%	34.5%	48.4%

Notes: Population and internal migration measured as the absolute number of people. iInternal migrants are those who were born outside the state of *Guanabara*. *Source*: (IBGE, 1953a; IBGE, 1953b; BEZERRA; CRUZ, 1982)

The measurement of slums in Brazil represents in itself a challenge, considering that it was only in 1980 that the Census started to investigate the characteristics of the slums in the whole country. In 1950, IBGE defined the criteria to classify a settlement as a slum[3.2.3], which remained the same until now:

(i) at least 51 households;

(ii) illegal occupation of land;

(iii) irregular urbanization, narrow roads with irregular alignment, uneven lots, construction outside urban standards;

(iv) precarious provision of essential services;

(v) topography unsuitable for habitation (due to high slope, or propensity to flooding)

[3.2.2]Actually, at the time it was the state of *Guanabara*, which merged with *Rio de Janeiro* em 1975 and kept the denomination of the latter.

[3.2.3]In 1970, *IBGE* started to refer to these areas as *Aglomerados Urbanos Especiais*. Since 1991, they altered the name to *Aglomerados Subnormais*.

Despite the fact that objective criteria were defined, putting them into practice is not a trivial task. In a large country like Brazil, the necessary decentralization of this process leaves a lot of room for errors in the classification, made by the enumerators of the Census. Table 3.2.2, adapted from Mation, Nadalin e Krause (2014), helps to illustrate some of these problems. According to the official data from the Census, in 2000 the Brazilian slums accommodated only about 6.5 million people but in 2010 that number increased to 11.4 million, an increment of 75%. But, as pointed out by Mation, Nadalin e Krause (2014), such a large increase is in contrast to the improvement in other social indicators displayed by the country during this period.

Tabela 3.2.2 – Population in the slums of Brazil

	1980	1991	2000	2010
Total population	62,391	80,885	169,799	190,756
Population in slums	2,224	4,084	6,536	11,432
Share of population in slums	3.6%	5.0%	3.8%	6.0%

Notes: Population measured in thousands of people. *Source*: Adapted from (MATION; NADALIN; KRAUSE, 2014).

The main issue, also highlighted by Costa e Nascimento (2005) and Taschner (2001), is the fact that slum areas need to have at least 51 households[3.2.4] and need to be classified before the data collection. That means that local knowledge is necessary to delineate correctly the slum areas beforehand. In practice, it is possible that some of these slums started with fewer households at the time of the interview and grew over time or that the delineation process has divided one single slum into two or more tracts with less than 51 households.

It was only in 2010 that *IBGE* started to improve their ability to detect slums, using satellite images and digitized street maps to delimit the census tracts in slum areas. This process allowed them to obtain a more accurate view of the regularity of constructions and streets. They also realized several meetings with

[3.2.4]There are two reasons to justify this threshold. The first is a logistical matter: each census tract is only by one enumerator; the second is to guarantee the privacy of the respondents, since at this level only a reduced questionnaire is available.

local authorities to validate the areas previously identified as slums (MATION; NADALIN; KRAUSE, 2014).

To conclude this section, we present a stylized fact about the relationship between the incoming migration at urban destination areas and the expansion of slums. In Figure 3.5.3 we show that there is a positive association between the (endogenous) number of incoming migrants and both the area occupied by and the population living in slums in Brazil. Although it is not causal, this figure suggests both the area (extensive) and the population (intensive) are important margins to describe the mechanism.

3.2.2 A Tale of Two Cities

As we mentioned in Section 3.1, we create two different but related measures of slum formation, one for the geographic dispersion of slums and the other for the population density within these informal settlements. Our main reason for doing so is that the initial conditions could lead two different places to evolve in very different paths. We illustrate this with two emblematic examples, the municipalities of *Ilhéus* and *Cubatão*.

Ilhéus is a city on the coast of the state of *Bahia*, in the Northeast region of Brazil. It has an area of about 1,600km^2 and a population density of 104.67 people per km^2, according to the 2010 wave of the Census. Based on the measure from *Mapbiomas*, its built-up area in 2010 is about 1.4% of the total area. Even if we take into account the area not suitable for construction[3.2.5] more than 98% of its area is considered free[3.2.6]. Unsurprisingly, *Ilhéus* is one of the places where slums have expanded spreading the urban sprawl. Between 1985 and 2010 the municipality's area occupied by slums more than tripled, reaching almost 5km^2.

On the other hand, *Cubatão* is a city in the metropolitan region of the *Baixada Santista* in the mountain range near the coast of the state of São Paulo. It

[3.2.5] For instance, the area identified as water bodies or beaches;

[3.2.6] At this point we cannot identify protected areas, like nature reserves. That means that for some municipalities the free area could be overestimated. In the near future, we intend to improve this measure.

has an area of around 142km^2 and a population density of 830.91 people per km^2. According to our measures, about 18% of its territory is built-up area and only 36% of it could be considered free in 2010. In *Cubatão* the expansion occurred mostly by increasing the population density within the slums. Between 1991 and 2010 the area occupied by slums increased 52%, while the population living in these areas more than tripled. In 2010, there were over 49 thousand people living in slums in *Cubatão*, around 41% of the total population.

3.3 Data and Empirical Strategy

In this section, we begin by listing the main sources and describing the construction of the data used in our analysis. Then we show some descriptive statistics and describe the empirical framework that link observed migration patterns to our predicted migration flows.

Migration

Migration data were drawn from four waves of the Brazilian Census (1980, 1991, 2000, and 2010), provided by the *Instituto Brasileiro de Geografia e Estatística (IBGE)*. The Census asks how long the individual has lived in the current municipality and which municipality she previously lived in. We use that information to construct a yearly panel of migration flows from each origin into each destination, from 1982 to 1991 and from 1996 to 2010[3.3.1]. We also use the 1980 wave of the Census to build a "past settlement" measure which is the share of migrants from each origin residing in each destination municipality.

As several municipalities were split into new ones since the 1980s, we aggregate our data using the original municipal boundaries as they were in 1980 (so-called "minimum comparable areas" or MCA) in order to avoid potential miscoding regarding migration status or municipality of origin. We refer to these

[3.3.1] The Census round realized in 2000 only asked in which city the respondent lived 5 years past. Therefore, we can only track the migrants until 1996.

areas simply as municipalities. Our sample contains 3,990 municipalities, 535 of them in metropolitan areas.

We calculate the incoming migration in each destination as the accumulated flow of migrants entering the municipality along the period covered by each Census wave. In our sample, the average number of incoming migrants is 4,504, about 15.1% of the average population in 1980 in the overall sample but 17,063 in metropolitan areas, equivalent to 18.4% of the average population in those destination municipalities (see Panels A and B of Table 3.5.1).

Slums

We use two distinct but related measures of slum formation. The first is a measure of expansion at the extensive margin, i.e. the area occupied by slums, which we create using data from *Mapbiomas*. They provide a series of raster images from Landsat satellites covering the entire country, one for each year from 1985 to 2020, at a 30m resolution. They applied a random forest algorithm to classify the pixels from those images according to the category of use or cover identified by the model. There are 34 possible categories, including forests, wetlands, grasslands, pasture, and urban areas (built-up) among others. We also use the shapefile of slum areas (*Aglomerados Subnormais*) provided by *IBGE* in the 2010 wave of the Census to identify the slum areas across the Brazilian municipalities. A census tract belongs to a slum area if the characteristics presented in Section 3.2.1 are predominant. There are 316,574 census tracts in the 2010 wave of the Census covering all the municipalities of Brazil, of which 15,868 (5.0%) are slum areas. For each year, we overlay the slum shapefile on the raster images and count the number of pixels classified as a built-up area within each slum census tract and aggregate the total number of slum pixels in each municipality. We also calculate the total area of each municipality aggregating the total number of pixels within its boundaries, regardless of pixel classification. Therefore, our measure of slums is the share of the municipality's area classified as built-up within the areas identified as slums in 2010.

Our approach has some advantages, but also some limitations that we need to discuss. The main advantage is that we are using the shapefile provided in the 2010 Census, when *IBGE* improved significantly its capacity to detect slums, allowing us to get a more accurate measure of these areas. On the side of limitations, we need to assume that the boundaries of slums in 2010 are stable over time, which may be a strong assumption. Suppose a certain area was a slum in the past, but in 2010 it was classified as part of the formal city. Then, our measure of slums would be underestimated. Another possibility is that the area was part of the formal city and after some time became a slum remaining so until 2010, rendering an overestimated measure. However, the latter seems less likely to happen because of land tenure. If an area belonged to the formal city, then its owners had the tenure of the land. It is hard to imagine that they would lose their rights and the land would be occupied by squatters.

We also create another measure, one of expansion at the intensive margin, i.e. the number of people living in slums. We use data from the universe of respondents of the Census[3.3.2] to calculate the total number of people living within each census tract in slum areas. But, because the boundaries of the tracts can vary over time, we use a crosswalk spreadsheet provided by *IBGE* to harmonize these census tracts between 1991 and 2010[3.3.3]. Once again, we use the 2010 map of slum areas to classify each of these compatible census tracts. The 316,574 census tracts of 2010 are mapped into 187,549 harmonized tracts, of which 10,188 are within the slum areas of 2010. Then, we calculate the total population aggregating the number of people living within each harmonized division to the municipality they belong to. In our sample, the average destination municipality has 2.9% of the population living in slum areas, 6.7% in metropolitan areas (see Panels A and B of Table 3.5.1).

Because we are using only the shapefile of slum areas in 2010, the same

[3.3.2]Since 1980, IBGE releases two types of Census data: a representative sample of the universe of respondents, covering a wide range of topics, disaggregated at the level of the individual but only available for weighting areas (a collection of census tracts); and a more restrict set of topics, aggregated at the finer census tract level to protect the privacy of respondents.

[3.3.3]Unfortunately it was not possible to reconcile the 1980 Census, due to inconsistencies in the data provided by *IBGE*.

concerns highlighted about the measure of the area occupied by slums apply to this measure of the population living in precarious conditions. Unfortunately, there are no maps available for the previous waves of the Census, making it impossible to delineate the correct boundaries of the harmonized census tracts we created. Between 2000 and 2010, the vast majority of the changes (98.6%) in the boundaries were defined by the split of the existing tracts. In this case, it would be simply a matter of aggregating the divided areas back to their previous boundaries. However, between 1991 and 2000, the situation is a bit more complicated. About one-third of the census tracts in 2000 were created by the consolidation of two or more areas from 1991. Sometimes, the entire tracts were aggregated but other times it involved the aggregation of only part of the existing areas. Therefore, without knowing the correct shape of the areas in the previous Census it is impossible to delineate the boundaries of these harmonized tracts.

After reclassifying the census tracts in slum areas, we end up with a more accurate measure of the population living in precarious conditions. As shown in Table 3.3.1, we estimate the 2000 slum population in 9.5 million people, rendering a much more modest increment of 18% in the period 2000-2010. That also means that between 1991 and 2000 the population living in slums more than doubled in Brazil. However, for this particular period, those numbers seem more plausible because until 1994 the country had to deal with a process of hyperinflation that reached over 2,000% in 1993 and a currency crisis in 1999, after the stabilization of the inflation in 1994.

Tabela 3.3.1 – Population in the slums of Brazil

	1991	2000	2010
Total population	146,825	169,799	190,756
Population in slums	3,805	9,507	11,303
Share of population in slums	2.6%	5.6%	5.9%

Notes: Population measured in thousands of people. *Source*: Adapted from (MATION; NADALIN; KRAUSE, 2014).

Agricultural price shocks

We create a measure of shocks on agricultural income at the origin municipalities due to variations in the international prices of agricultural commodities, similar to Imbert e Ulyssea (2022). We collect monthly data on the prices of 12 commodities[3.3.4] from the World Bank (The Pink Sheet), from 1970 to 2011. A price shock for crop c and month m in year y is defined as the residual from an AR(1) process. We aggregate these shocks at the origin municipality level using the share of each crop π_{oc} in the value of agricultural production and the share of the harvested area h_{oc} of crop c to account for differentials in the productivity of agricultural inputs. More formally, price shocks are defined as

$$s_{oy} = \sum_{m} \sum_{c} (\pi_{oc} \times h_{oc} \times \varepsilon_{cmy}) \tag{3.3.1}$$

where ε_{cmy} is an AR(1) residual, π_{oc} comes from the 1980 Agricultural Census, and h_{oc} comes from the Municipal Agricultural Survey both published by *IBGE*[3.3.5].

3.3.1 Empirical Strategy

Our empirical strategy relies on the assumption that incoming migration is exogenous to slum expansion at the destination municipalities. In this section, we describe the empirical framework that combines exogenous shocks on agricultural income at the origin municipalities and the past settlement patterns, which allows us to establish a causal relationship between migration inflows and slum expansion in destination areas.

We specify a model for the changes in the share of slum areas or population in a destination as a function of the incoming migration flow. Specifically, we

[3.3.4] We use the prices for banana, cocoa, coffee, cotton, maize, orange, rice, soybeans, sugar, tobacco, wheat, and wood.

[3.3.5] The Municipal Agricultural Survey provides detailed annual information on the agricultural production of each municipality, by crop. We take the harvested area for all crops, averaged between 1975 and 1980, and divide it by the municipality's total area.

assume

$$\Delta y_{dt} = \alpha + \beta m_{dt} + \gamma X_d + \psi_t + \epsilon_{dt} \tag{3.3.2}$$

where y_{dt} is either the share of the area occupied by slums or the share of the population living in these areas at destination municipality d in census-year t, m_{dt} is the accumulated migrant inflow from the Semiarid region entering the destination area during the period covered by the Census round t, X_d are destination-level baseline controls, ψ_t absorb time fixed effects and ϵ_{dt} is the error term.

By estimating a model in first differences with covariates, we account for the time-invariant characteristics and for differential trends based on initial conditions that vary across municipalities. However, two main concerns remain in order to establish β as a causal parameter. First, people can be attracted to places with more affordable housing costs or better amenities, but at the same time, the incoming migration will affect these same equilibrium objects. Second, time-varying unobserved characteristics can be correlated with incoming migration, rendering OLS estimates biased.

To overcome these issues we create a shift-share instrumental variable (SSIV) in two steps. First, we use the price shocks at the origin to predict the outflow of migrants leaving their hometowns. More specifically, we estimate the following regression:

$$M_{oy} = \alpha + \beta' s^p_{oy} + \phi_o + \delta_y + \varepsilon_{ot} \tag{3.3.3}$$

where s^p_{oy} is the price shock at origin municipality o in year y, defined in equation 3.3.1; ϕ_o and δ_y are municipality and year fixed effects, respectively; and ε_{oy} is a random error term. For each year the predicted number of migrants who leave their hometowns is obtained by multiplying this predicted rate by the municipality population reported in the 1980 Census:

$$\widehat{M}_{oy} = \widehat{m}_{oy} \times P_o \tag{3.3.4}$$

In the second step, we use the past settlements of migrants from the origin o to municipality d in order to distribute them throughout the destination areas, defining our SSIV as

$$\widehat{M}_{dt} = \sum_{t \in C} \left(\sum_{o=1}^{O} s_{od} \times \widehat{M}_{oy} \right) \tag{3.3.5}$$

where $C = \{1991, 2000, 2010\}$ are the three rounds of the Census we can use to track the migrants; and s_{od} is the share of migrants from origin municipality o who lived in the destination area d in 1980.

To make interpretation easier we divide the predicted number of incoming migrants by the native population in each destination municipality in 1980, P_d[3.3.6], resulting in

$$\widehat{m}_{dt} = \frac{\widehat{M}_{dt}}{P_d} \tag{3.3.6}$$

as the SSIV for the observed incoming migration rate in destination municipality d.

In other words, our instrument $\widehat{m}_{dt}$ is a combination of exogenous shocks or 'shifts' $\widehat{M}_{ot}$ (weather-driven outflows) and exposure 'shares' ($s_{od} \geq 0$) or past settlement patterns.

3.3.2 Prediction of the out-migration

We begin the exploration of our first-stage results by estimating variations of specification 3.3.3 and report the estimates in Table 3.5.2.

In Columns (1)-(3) we estimate it using a two-way fixed effects (TWFE) model. In Columns (4)-(6) we try an alternative estimation using a poisson pseudo-maximum likelihood (PPML) (GOURIEROUX; MONFORT; TROGNON, 1984),

[3.3.6]We highlight that the denominator P_d is only a normalization that helps interpret the coefficients of interest. It does not play any role in identification.

using the absolute number of migrants as the dependent variable. All regressions include year and origin-municipality fixed effects. In columns (2)-(3) and (4)-(6) we include a linear trend interacting time dummies with 1980 characteristics (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection). In Columns (3) and (6) we use a more flexible specification, allowing for differential effects from positive and negative shocks. Standard errors are clustered at the municipality level.

As expected, the agricultural price shocks are negatively correlated with migration outflows indicating that migration to urban areas in destination municipalities increases when a negative shock hits the origin area. One standard deviation decrease in the price shock increases the out-migration rate by $0.63p.p.$ Our results are robust to the inclusion of baseline characteristics interacted with year dummies - columns (2)-(3) - and to an alternative estimation method - columns (5)-(6). Also, columns (3) and (6) show that negative income shocks are the main driver of out-migration. This is an interesting result because it is not clear *a priori* which effect should prevail. If individuals face liquidity constraints, then a transitory positive income shock could allow them to pay for the costs of migrating. On the other hand, a more permanent negative shock could provide an incentive for wealthier individuals to move, by reducing their opportunity cost (BAZZI, 2017).

As mentioned in section 3.3.1, after leveraging variation in agricultural price shocks to predict the number of migrants leaving their origin municipalities, we use the past settlement of pattern of these migrants to allocate them to the destination areas. One important criterion to ensure the validity of our empirical strategy is that both predicted migration outflow and inflow rates, $\widehat{m}_{ot}$ and $\widehat{m}_{dt}$ respectively, should be strongly correlated with their observed counterparts.

Figure 3.5.5 reveals that our predictions provide a good fit for the observed migration. Panel (a) shows the relationship between the predicted and observed number of migrants leaving their origin municipalities and moving into urban areas at the destination, accumulated over the period 1982-2010. Panel (b) shows the predicted and observed numbers of incoming migrants at the destination municipalities.

3.4 Migration Inflows and Slum Expansion

In this section, we discuss the effects of incoming migration on the expansion of the area occupied by slums and the number of people living in these areas.

First, in Table 3.5.3 we show the impacts of incoming migration on the changes in the share of the area occupied by slums. The dependent variable is the stacked first differences (1991-2000 and 2000-2010)[3.4.1] of the total number of pixels classified as built-up regions that are inside the slum areas divided by the available[3.4.2] area within the municipality. The right-hand side is the inverse hyperbolic sine (IHS) of the number of incoming migrants instrumented by the IHS of the predicted incoming migration. In Columns (2)-(5) we include time dummies to control for aggregate shocks that affect all destination municipalities. In Columns (3)-(6) we allow a differential trend for municipalities with different baseline levels of slums. In Columns (4)-(5) we also include baseline destination-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection). Finally, in Column (5) we include a state-specific linear trend. All regressions are weighted by the total population in 1980.

We adopt the more stringent specification (4) as our preferred and focus on it to interpret the results. Panel A includes all destination municipalities, while in

[3.4.1] Even though we have data on the measure of the area occupied by slums since 1985, we use only the period 1991-2010 to maintain the comparability with the measure of population, which is only available for this latter period. We also have some additional results stacking the periods 1985-1991, 1991-2000, and 2000-2010. See Tables 3.5.7 and 3.5.8.

[3.4.2] Defined as the total number of pixels within the municipality polygon minus the number of pixels classified as built-up area - either in slum areas or not - and minus the total number of pixels classified as non-usable area (water, sand/beaches, and mining)

Panels B and C we divide the sample between metropolitan and non-metropolitan areas, respectively. We find no effect at the extensive margin neither in the overall or in the metropolitan area samples. The coefficient of the baseline share of slum areas suggests that there are other factors explaining the growth in the area occupied by slums in these contexts. However, in non-metropolitan regions, we find a positive and significant impact of the incoming migration on the expansion of the area occupied by slums. Although the coefficient is of a small magnitude, it represents an increase of 22% of a standard deviation (SD).

Analyzing the effects on the intensive margin, we find more robust results. In Table 3.5.4 we estimate the effects of incoming migration on the changes in the number of people living in slum areas. The dependent variable is the stacked first differences (1991-2000 and 2000-2010) of the IHS of the total population in slums. The right-hand side is the IHS of the number of incoming migrants instrumented by the IHS of the predicted incoming migration. All the controls are the same as in the previous table. Since the IHS approximates a logarithm, we interpret our coefficients as an elasticity. Increasing the number of incoming migrants by 1% leads to an increase in the number of people living in slums of 0.43% in the overall sample. This effect is much higher in metropolitan areas (0.52%) than in non-metropolitan areas (0.19%).

We also hypothesize a possible heterogeneous effect, regarding the scarcity of usable land. In places where land is more scarce, we would expect a larger effect at the intensive margin, i.e. city densification. We show in Table 3.5.6 that this is exactly what we find. We split the sample based on the availability of usable land, by taking the share of the municipality area that is available (total number of pixels within the municipality minus the total non-usable pixels) in 1985 and dividing it into two groups. We define as lower (higher) free area those municipalities below (above) the median. All the effects are coming from places where the usable land is more scarce.

3.5 Discussion and Concluding Remarks

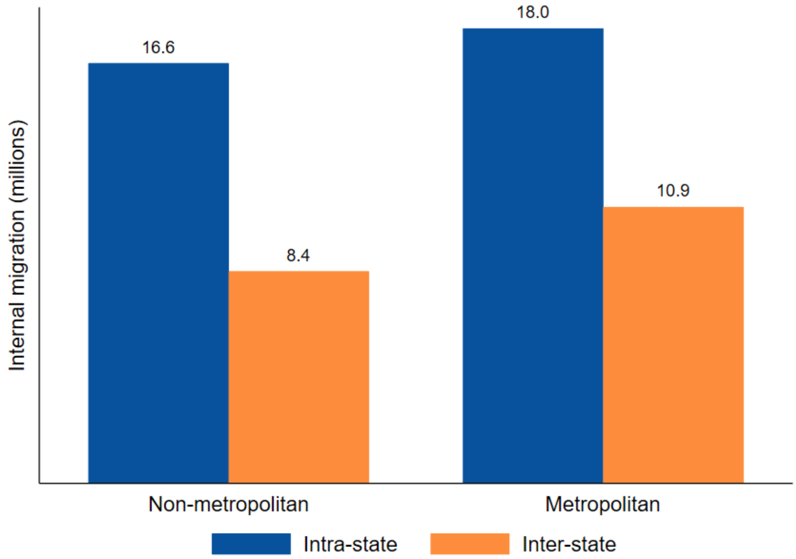
In this paper, we investigate the impacts of incoming migration on the expansion of slums in Brazil, using a shift-share instrument exploiting exogenous agricultural price shocks to account for the endogeneity of the migrant flows.

We provide more accurate estimates of the number of people living in the slum areas, by reclassifying the data available in the Census.

Our estimates suggest that increasing the number of incoming migrants causes a small rise in the share of the area occupied by slums in non-metropolitan regions, but a substantial increase in the number of people living in those areas, especially in the metropolitan areas. Such effects come exclusively from places with smaller fractions of usable land.

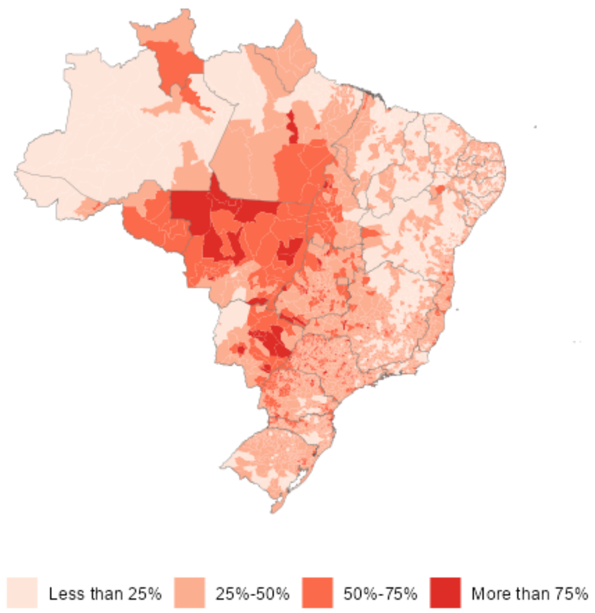
Figures and tables

Figura 3.5.1 – Internal migration in Brazil 1991-2010



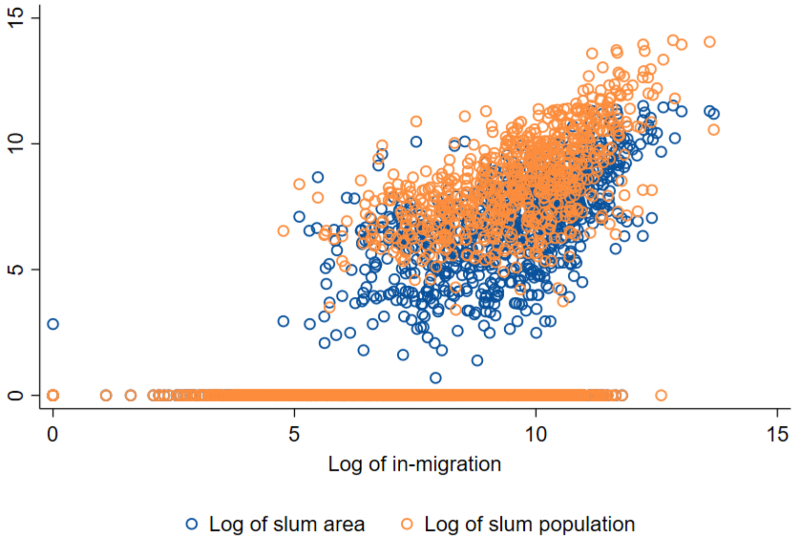
Notes: This figure presents the total number of people moving into urban areas of the destination municipalities within and across states, and divided by type of region. *Data source*: Census microdata (IBGE).

Figura 3.5.2 – Accumulated incoming migration (1991-2010)



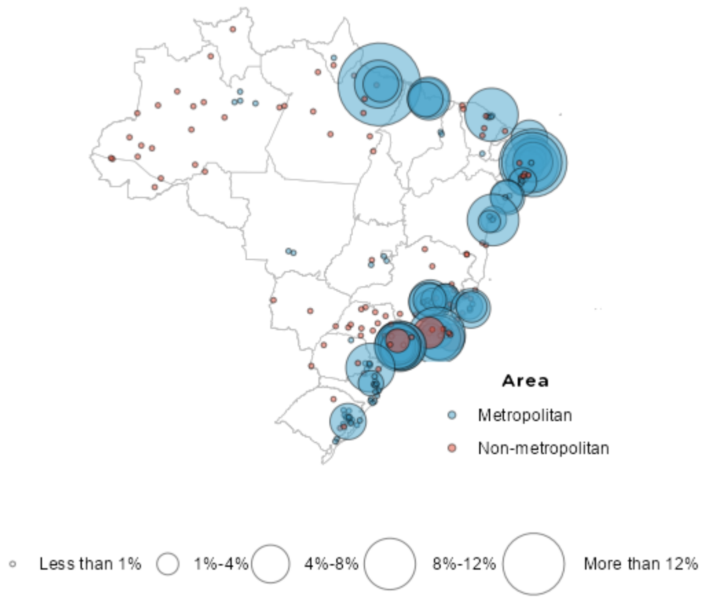
Notes: This figure presents the accumulated inflow of internal migrants entering each destination municipality, measured as a fraction of the total population in 2010. *Data source*: Census microdata (IBGE).

Figura 3.5.3 – Internal migration and slum growth in Brazil

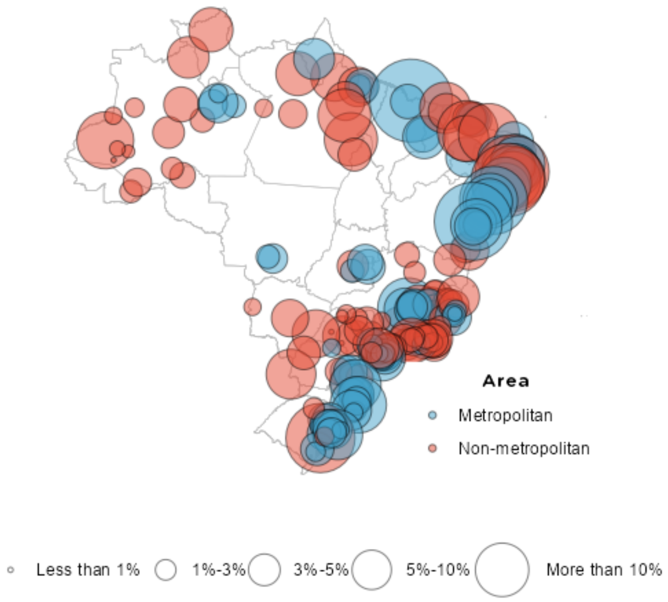


Notes: This figure presents the relationship between log of the area occupied by and the log of the population in slums, and the log of the incoming migration at destination municipalities. *Data source*: Census microdata (IBGE).

Figura 3.5.4 – Occupied area and growth rate of slums in Brazil



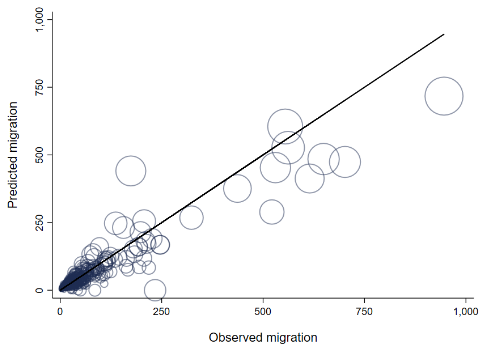
(a) Share of the area occupied by slums in 2010



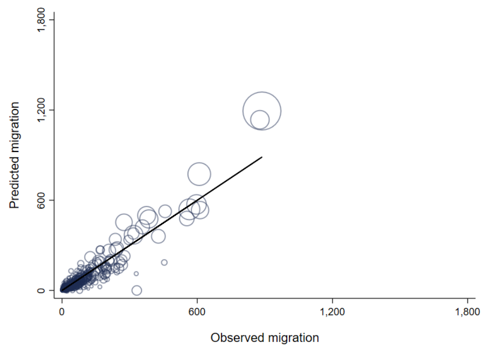
(b) Annual growth rate of slums

Notes: This figure presents the relationship between the predicted and observed migration flows across Brazilian municipalities from 1982 to 2010. Panel (a) shows the number of migrants leaving their origin municipalities to settle in urban areas at the destination. Panel (b) shows the number of incoming migrants entering the destination municipalities. We dropped destination municipalities with a predicted inflow of over 2,000,000 people, to make the visualization clearer. The circle size represents the municipality's total population in 1980. *Data source*: Census microdata (IBGE).

Figura 3.5.5 – Observed vs predicted migration (1,000 people)



(a) Migration outflow



(b) Migration inflow

Notes: This figure presents the relationship between the predicted and observed migration flows across Brazilian municipalities from 1982 to 2010. Panel (a) shows the number of migrants leaving their origin municipalities to settle in urban areas at the destination. Panel (b) shows the number of incoming migrants entering the destination municipalities. We dropped destination municipalities with a predicted inflow of over 2,000,000 people, to make the visualization clearer. The circle size represents the municipality's total population in 1980. *Data source*: Census microdata (IBGE).

Tabela 3.5.1 – **Summary statistics**

	Mean	Std. Dev.	Min	Max	Obs	Municipalities
	Panel A. All municipalities					
Observed in-migration	4,504	18,916	0	875,368	11,970	3,990
Predicted in-migration	4,123	23,794	0	1,476,883	11,970	3,990
Area (Km^2)	212,688	842,717	361.2	16,550,012	11,970	3,990
Share of slums	0.0007	0.0065	0	0.2282	11,970	3,990
Share of free area	0.8932	0.1931	0	1	11,970	3,990
Population (1980)	29,827	172,714	732	8,493,217	11,970	3,990
Share of population in slums	0.0047	0.0294	0	0.6115	11,970	3,990
	Panel B. Only metropolitan areas					
Observed in-migration	18,019	47,270	14	875,368	1,605	535
Predicted in-migration	17,063	62,388	57.8683	1,476,883	1,605	535
Area (Km^2)	113,915	357,968	1533.1	3,823,816	1,605	535
Share of slums	0.0049	0.0171	0	0.2282	1,605	535
Share of free area	0.7925	0.2530	0	0.9999	1,605	535
Population (1980)	97,577	461,308	1,481	8,493,217	1,605	535
Share of population in slums	0.0242	0.0670	0	0.6115	1,605	535
	Panel C. Only non-metropolitan areas					
Observed in-migration	2,411	5,895	0	144,328	10,365	3,455
Predicted in-migration	2,119	4,638	0	88,076	10,365	3,455
Area (Km^2)	227,982	893,630	361.2	16,550,012	10,365	3,455
Share of slums	0.0000	0.0006	0	0.0257	10,365	3,455
Share of free area	0.9088	0.1770	0	1	10,365	3,455
Population (1980)	19,336	26,336	732	348,542	10,365	3,455
Share of population in slums	0.0017	0.0153	0	0.4927	10,365	3,455
	Panel D. Only municipalities with slums					
Observed in-migration	35,565	63,102	0	875,368	792	264
Predicted in-migration	34,288	85,775	0	1,476,883	792	264
Area (Km^2)	440,537	1,408,133	1744.9	12,246,109	792	264
Share of slums	0.0104	0.0231	0	0.2282	792	264
Share of free area	0.7198	0.2720	0	1	792	264
Population (1980)	200,333	643,314	3,672	8,493,217	792	264
Share of population in slums	0.0668	0.0892	0	0.6115	792	264

Notes: The migration inflow rate is the share of migrants over the total population in 1980. The share of slum areas is the total number of pixels classified as slum areas divided by the total number of pixels within the municipality. The share of free area is the number of pixels that are neither classified as built-up or non-usable areas (water, sand/beaches, and mining) divided by the total number of pixels within the municipality. The share of the population living in slums is relative to the total population in each Census wave. All statistics are collapsed into the Census-year level.

Tabela 3.5.2 – Agricultural price shocks and out-migration

	TWFE				PPML		
	(1)	(2)	(3)		(4)	(5)	(6)
Price shock	-0.006***	-0.006***			-0.319***	-0.251***	
	(0.001)	(0.001)			(0.056)	(0.052)	
Negative shock			-0.014***				-0.615***
			(0.002)				(0.109)
Positive shock			0.002				0.112
			(0.002)				(0.100)
Observations	87,538	87,538	87,538	87,538	87,538	87,538	87,538
Municipalities	3,979	3,979	3,979	3,979	3,979	3,979	3,979
Year dummies	✓	✓	✓		✓	✓	✓
Municipality dummies	✓	✓	✓		✓	✓	✓
Baseline × year		✓	✓			✓	✓

Notes: This table shows the relationship between migration outflows and the agricultural price shocks defined in equation 3.3.1. In Columns (1)-(3) the dependent variable is the number of migrants leaving the municipality of origin divided by the total population, while in Columns (4)-(6) the dependent variable is the absolute number of migrants leaving the origin. All specifications control for year and origin-municipality fixed effects. Columns (3) and (6) also control for origin-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection). Standard errors clustered at the origin municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 3.5.3 – Dependent variable: Δ Share of slum area (Stacked differences)

	(1)	(2)	(3)	(4)	(5)
	Panel A: All municipalities				
In-migration (IHS)	0.01776	0.02009	-0.02561*	0.00203	0.00319
	(0.01735)	(0.01967)	(0.01339)	(0.02579)	(0.02303)
Share of slum area 1985			1.24089***	1.23362***	1.23361***
			(0.45881)	(0.44451)	(0.44456)
Observations	7835	7835	7745	7745	7745
Municipalities	3927	3927	3877	3877	3877
	Panel B: Only metropolitan areas				
In-migration (IHS)	-0.01296	-0.00554	-0.02713	0.02308	0.02317
	(0.01954)	(0.01693)	(0.02104)	(0.06966)	(0.06890)
Share of slum area 1985			1.24420***	1.23532***	1.23525***
			(0.45969)	(0.43757)	(0.43728)
Observations	1044	1044	1016	1016	1016
Municipalities	525	525	510	510	510
	Panel C: Only non-metropolitan areas				
In-migration (IHS)	0.00008***	0.00009***	0.00002	0.00006**	0.00005**
	(0.00003)	(0.00003)	(0.00001)	(0.00003)	(0.00003)
Share of slum area 1985			0.30759**	0.31609**	0.31547**
			(0.13919)	(0.13342)	(0.13338)
Observations	6791	6791	6729	6729	6729
Municipalities	3402	3402	3367	3367	3367
Time dummies		✓	✓	✓	✓
Baseline covariates				✓	✓
State × year					✓

Notes: This table shows the relationship between migration outflows and the expansion of slums in destination municipalities. The dependent variable is the stacked first difference (1991-2000 and 2000-2010) in the area occupied by slums divided by the available area. In-migration is measured as the inverse hyperbolic sine of the number of migrants entering the destination. Columns (3)-(5) control for the predetermined share of the area covered by slums in 1985. Columns (4)-(5) also control for destination-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection). All regressions are weighted by the total population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 3.5.4 – Dependent variable: Δ IHS of the population in slum areas (Stacked differences: 1991-2010)

	(1)	(2)	(3)	(4)	(5)
Panel A: All municipalities					
In-migration (IHS)	0.17790***	0.20304***	0.44050***	0.42864***	0.44411***
	(0.04573)	(0.04960)	(0.11864)	(0.11595)	(0.14367)
Slum population 1991 (IHS)			-0.11538***	-0.11064**	-0.10878***
			(0.04437)	(0.04564)	(0.04076)
Observations	7980	7980	7980	7980	7980
Municipalities	3990	3990	3990	3990	3990
Panel B: Only metropolitan areas					
In-migration (IHS)	0.10292	0.17792**	0.63885***	0.52118***	0.52537***
	(0.07548)	(0.08864)	(0.14712)	(0.11045)	(0.11287)
Slum population 1991 (IHS)			-0.21230***	-0.20190***	-0.19648***
			(0.05710)	(0.05022)	(0.04312)
Observations	1070	1070	1070	1070	1070
Municipalities	535	535	535	535	535
Panel C: Only non-metropolitan areas					
In-migration (IHS)	0.14230***	0.17109***	0.17454***	0.19466**	0.13109*
	(0.03818)	(0.04499)	(0.05842)	(0.08629)	(0.07744)
Slum population 1991 (IHS)			-0.00447	-0.00652	-0.00691
			(0.01900)	(0.02047)	(0.02031)
Observations	6910	6910	6910	6910	6910
Municipalities	3455	3455	3455	3455	3455
Time dummies		✓	✓	✓	✓
Baseline covariates				✓	✓
State × year					✓

Notes: This table shows the relationship between migration outflows and the expansion of the population living in slums in destination municipalities. The dependent variable is the stacked first difference (1991-2000 and 2000-2010) in the inverse hyperbolic sine of the number of people living in slums. In-migration is measured as the inverse hyperbolic sine of the number of migrants entering the destination. Columns (3)-(5) control for the predetermined share of the area covered by slums in 1985. Columns (4)-(5) also control for destination-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection) All regressions are weighted by the total population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 3.5.5 – Dependent variable: Δ Share of slum area

	All		Metropolitan		Non-metropolitan	
	Lower free area	Higher free area	Lower free area	Higher free area	Lower free area	Higher free area
	(1)	(2)	(3)	(4)	(5)	(6)
In-migration (IHS)	0.00181	0.00000	0.03918	-0.00008*	0.00012*	0.00002
	(0.03715)	(0.00002)	(0.09856)	(0.00004)	(0.00006)	(0.00001)
Share of slum area 1985	1.23546***	0.91119***	1.24290***	0.36221***	0.31760**	0.53335***
	(0.44401)	(0.34077)	(0.43869)	(0.08834)	(0.13018)	(0.11183)
Observations	3758	3987	483	533	3275	3454
Municipalities	1883	1994	243	267	1640	1727
Time dummies	✓	✓	✓	✓	✓	✓
Baseline covariates	✓	✓	✓	✓	✓	✓

Notes: This table shows the relationship between migration outflows and the expansion of slums in destination municipalities. The dependent variable is the stacked first difference (1991-2000 and 2000-2010) in the area occupied by slums divided by the available area. In-migration is measured as the inverse hyperbolic sine of the number of migrants entering the destination. All specifications control for the inverse hyperbolic sine of the population living in slums in 1991 and destination-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection) All regressions are weighted by the total population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 3.5.6 – Δ IHS of the population in slum areas (Stacked differences: 1991-2010)

	All		Metropolitan		Non-metropolitan	
	Lower free area	Higher free area	Lower free area	Higher free area	Lower free area	Higher free area
	(1)	(2)	(3)	(4)	(5)	(6)
In-migration (IHS)	0.48601**	0.02879	0.68935***	-0.22236	0.35798*	-0.04520
	(0.22504)	(0.07035)	(0.21869)	(0.22595)	(0.21605)	(0.16230)
Slum population 1991 (IHS)	-0.13311***	-0.02509	-0.23478***	0.05685*	-0.00386	-0.05429*
	(0.05049)	(0.02412)	(0.04797)	(0.03015)	(0.03830)	(0.03200)
Observations	3992	3988	536	534	3456	3454
Municipalities	1996	1994	268	267	1728	1727
Time dummies	✓	✓	✓	✓	✓	✓
Baseline covariates	✓	✓	✓	✓	✓	✓

Notes: This table shows the relationship between migration outflows and the expansion of the population living in slums in destination municipalities. The dependent variable is the stacked first difference (1991-2000 and 2000-2010) in the inverse hyperbolic sine of the number of people living in slums. In-migration is measured as the inverse hyperbolic sine of the number of migrants entering the destination. All specifications control for the inverse hyperbolic sine of the population living in slums in 1991 and destination-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection) All regressions are weighted by the total population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 3.5.7 – Dependent variable: Δ Share of slum area (Stacked differences: 1985-2010)

	Panel A: All municipalities				
	(1)	(2)	(3)	(4)	(5)
In-migration (IHS)	0.01272	0.01399	-0.01728	0.00766	0.00742
	(0.01341)	(0.01473)	(0.01149)	(0.02360)	(0.02090)
Share of slum area 1985			0.85432**	0.84947**	0.84947**
			(0.43285)	(0.41916)	(0.41916)
Observations	11714	11714	11624	11624	11624
Municipalities	3934	3934	3884	3884	3884
	Panel B: Only metropolitan areas				
In-migration (IHS)	-0.00651	-0.00172	-0.01633	0.03218	0.03185
	(0.01420)	(0.01223)	(0.01766)	(0.06209)	(0.06133)
Share of slum area 1985			0.85701**	0.85078**	0.85104**
			(0.43384)	(0.41225)	(0.41181)
Observations	1553	1553	1525	1525	1525
Municipalities	525	525	510	510	510
	Panel C: Only non-metropolitan areas				
In-migration (IHS)	0.00008***	0.00009***	0.00001	0.00005**	0.00004**
	(0.00003)	(0.00003)	(0.00001)	(0.00003)	(0.00002)
Share of slum area 1985			0.32505***	0.33172***	0.33114***
			(0.11838)	(0.11447)	(0.11436)
Observations	10161	10161	10099	10099	10099
Municipalities	3409	3409	3374	3374	3374
Time dummies		✓	✓	✓	✓
Baseline covariates				✓	✓
State × year					✓

Notes: This table shows the relationship between migration outflows and the expansion of slums in destination municipalities. The dependent variable is the stacked first difference (1985-1991, 1991-2000 and 2000-2010) in the area occupied by slums divided by the available area. In-migration is measured as the inverse hyperbolic sine of the number of migrants entering the destination. Columns (3)-(5) control for the predetermined share of the area covered by slums in 1985. Columns (4)-(5) also control for destination-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection). All regressions are weighted by the total population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Tabela 3.5.8 – Dependent variable: Δ Share of slum area (Stacked differences: 1985-2010)

	All		Metropolitan		Non-metropolitan	
	Lower free area	Higher free area	Lower free area	Higher free area	Lower free area	Higher free area
	(1)	(2)	(3)	(4)	(5)	(6)
In-migration (IHS)	0.01084	0.00001	0.04964	-0.00010	0.00009*	0.00002
	(0.03453)	(0.00001)	(0.08741)	(0.00008)	(0.00005)	(0.00002)
Share of slum area 1985	0.85079**	0.95044***	0.85734**	0.35019***	0.33269***	0.72989***
	(0.41861)	(0.24549)	(0.41223)	(0.09738)	(0.11214)	(0.16051)
Observations	5643	5981	725	800	4918	5181
Municipalities	1890	1994	243	267	1647	1727
Time dummies	✓	✓	✓	✓	✓	✓
Baseline covariates	✓	✓	✓	✓	✓	✓

Notes: This table shows the relationship between migration outflows and the expansion of slums in destination municipalities. The dependent variable is the stacked first difference (1985-1991, 1991-2000 and 2000-2010) in the area occupied by slums divided by the available area. In-migration is measured as the inverse hyperbolic sine of the number of migrants entering the destination. All specifications control for the inverse hyperbolic sine of the population living in slums in 1991 and destination-level characteristics in 1980 (age; the total population; the share of the non-white population; the share of the female population, the shares of the population with an elementary, high school, and college education; the shares of agricultural and manufacturing employment; familiar income per capita; and the shares of the households with access to water, electricity and sewage collection) All regressions are weighted by the total population in 1980. Standard errors clustered at the destination municipality level in parenthesis. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

References

ADAO, R.; KOLESÁR, M.; MORALES, E. Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, Oxford University Press, v. 134, n. 4, p. 1949–2010, 2019.

AGÉNOR, P.-R.; CANUTO, O. Gender equality and economic growth in brazil. World Bank, Washington, DC, 2013.

ALMEIDA, R.; CARNEIRO, P. Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics*, v. 4, n. 3, p. 64–89, July 2012. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/app.4.3.64>.

ALTONJI, J. G.; CARD, D. The effects of immigration on the labor market outcomes of less-skilled natives. In: *Immigration, trade, and the labor market*. [S.l.]: University of Chicago Press, 1991. p. 201–234.

AMIOR, M.; MANNING, A. *Monopsony and the Wage Effects of Migration*. [S.l.], 2020.

AMIOR, M.; STUHLER, J. *Immigration and Monopsony: Evidence Across the Distribution of Firms*. [S.l.], 2022.

ANS. *Painel de precificação: Planos de Saúde*. 2018. Agência Nacional de Saúde Suplementar, <http://www.ans.gov.br/materiais-publicados/periodicos>.

ANS. *Dados e Indicadores do Setor*. 2019. Agência Nacional de Saúde Suplementar, <http://www.ans.gov.br/perfil-do-setor/dados-e-indicadores-do-setor>. Last accessed on 28/10/2020.

ARBACHE, J. S. *Uma Introdução aos Salários Indiretos no Brasil*. [S.l.], 1995.

ARBACHE, J. S.; FERREIRA, D. M. Salários indiretos e previdência social no brasil. *Estudos Empresariais*, v. 6, n. 1, p. 37–44, 2001.

ASSUNçãO, J.; CHEIN, F. Climate change and agricultural productivity in brazil: future perspectives. *Environment and Development Economics*, Cambridge University Press, v. 21, n. 5, p. 581–602, 2016.

AZOULAY, P. et al. Immigration and entrepreneurship in the united states. *American Economic Review: Insights*, v. 4, n. 1, p. 71–88, 2022.

BABECKÝ, J. et al. Non-base wage components as a source of wage adaptability to shocks: evidence from european firms, 2010–2013. *IZA Journal of Labor Policy*, v. 8, n. 1, p. 1, 2019. Disponível em: <https://doi.org/10.1186/s40173-018-0106-8>.

BABECKý, J. et al. How do european firms adjust their labour costs when nominal wages are rigid? *Labour Economics*, v. 19, n. 5, p. 792–801, 2012. ISSN 0927-5371. Special Section on: Price, Wage and Employment Adjustments in 2007-2008 and Some Inferences for the Current European Crisis. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0927537112000334>.

BADAOUI, E. E.; STROBL, E.; WALSH, F. Impact of internal migration on labor market outcomes of native males in thailand. *Economic Development and Cultural Change*, The University of Chicago Press, v. 66, n. 1, p. pp. 147–177, 2017.

BAKER, M.; GRUBER, J.; MILLIGAN, K. Universal child care, maternal labor supply, and family well-being. *Journal of political Economy*, The University of Chicago Press, v. 116, n. 4, p. 709–745, 2008.

BARBIERI, A. F. et al. Climate change and population migration in brazil's northeast: scenarios for 2025–205. *Population and Environment*, v. 31, p. 344–370, 2010.

BARONE, G.; MOCETTI, S. With a little help from abroad: the effect of low-skilled immigration on the female labour supply. *Labour Economics*, Elsevier, v. 18, n. 5, p. 664–675, 2011.

BAUERNSCHUSTER, S.; SCHLOTTER, M. Public child care and mothers' labor supply—evidence from two quasi-experiments. *Journal of Public Economics*, Elsevier, v. 123, p. 1–16, 2015.

BAUM-SNOW, N.; HAN, L. The microgeography of housing supply. *Work in progress, University of Toronto*, 2019.

BAZZI, S. Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425, v. 9, n. 2, p. 219–255, 2017.

BEENSTOCK, M.; FELSENSTEIN, D. et al. *The econometric analysis of non-stationary spatial panel data.* [S.l.]: Springer, 2019.

BELCHIOR, C.; GONZAGA, G.; ULYSSEA, G. Unpacking neighborhood effects: Experimental evidence from a large-scale housing program in brazil. IZA Discussion Paper, 2023.

BEZERRA, V. M. d. C. B.; CRUZ, J. M. Censo demográfico de 1960: Favelas do estado da guanabara. *Revista Brasileira de Geografia*, IBGE Rio de Janeiro, n. 2, 1982.

BLUNDELL, R. et al. The labour market impact of the working families' tax credit. *Fiscal studies*, Wiley Online Library, v. 21, n. 1, p. 75–104, 2000.

BLUNDELL, R.; HOYNES, H. W. Has' in-work'benefit reform helped the labor market? In: *Seeking a Premier Economy: The Economic Effects of British Economic Reforms, 1980-2000*. [S.l.]: University of Chicago Press, 2004. p. 411–460.

BORJAS, G. J. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics*, MIT Press, v. 118, n. 4, p. 1335–1374, 2003.

BORJAS, G. J. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics*, Oxford University Press, v. 118, n. 4, p. 1335–1374, 2003. ISSN 00335533, 15314650. Disponível em: <http://www.jstor.org/stable/25053941>.

BORJAS, G. J. *Immigration Economics*. [S.l.]: Cambridge: Harvard University Press, 2014.

BORJAS, G. J.; MONRAS, J. The labour market consequences of refugee supply shocks. *Economic Policy*, v. 32, p. 361—-413, 2017.

BORUSYAK, K.; HULL, P.; JARAVEL, X. Quasi-experimental shift-share research designs. *The Review of Economic Studies*, p. Forthcoming, 2021.

BOSCH, M.; ESTEBAN-PRETEL, J. Job creation and job destruction in the presence of informal markets. *Journal of Development Economics*, v. 98, n. 2, p. 270–286, 2012.

BOUSTAN, L. P.; FISHBACK, P. V.; KANTOR, S. The effect of internal migration on local labor markets: American cities during the great depression. *Journal of Labor Economics*, University of Chicago Press Chicago, IL, v. 28, n. 4, p. 719–746, 2010.

BRUECKNER, J. K.; MATION, L.; NADALIN, V. G. Slums in brazil: Where are they located, who lives in them, and do they 'squeeze'the formal housing market? *Journal of Housing Economics*, Elsevier, v. 44, p. 48–60, 2019.

BUSSO, M.; CHAUVIN, J. P.; Herrera L, N. Rural-urban migration at high urbanization levels. *Regional Science and Urban Economics*, p. forthcoming, 2021. ISSN 0166-0462. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166046221000181>.

CARABETTA, J.; PEREIRA, R. H. M.; GONCALVES, C. N. geobr: Loads shapefiles of official spatial data sets of brazil. python package version 0.1.5. *The Python Package Index (PyPI)*, 2020. Disponível em: <https://pypi.org/project/geobr/0.1.5/>.

CARD, D. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, The University of Chicago Press, v. 19, n. 1, p. 22–64, 2001.

CARD, D. Immigration and inequality. *American Economic Review*, v. 99, n. 2, p. 1–21, May 2009. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/aer.99.2.1>.

CARD, D.; LEMIEUX, T. Can falling supply explain the rising return to college for younger men? a cohort-based analysis. *The Quarterly Journal of Economics*, Oxford University Press, v. 116, n. 2, p. 705–746, 2001. ISSN 00335533, 15314650. Disponível em: <http://www.jstor.org/stable/2696477>.

CARD, D.; PERI, G. Immigration Economics by George J. Borjas: A Review Essay. *Journal of Economic Literature*, v. 54, n. 4, p. 1333–49, 2016.

CARDOSO, A.; LAGE, T. Labor inspection in Brazil. *Dados*, scieloss, v. 2, p. Rio de Janeiro, 2006. ISSN 0011-5258. Disponível em: <http://socialsciences.scielo.org/scielo.php?script=sci_arttext&pid=S0011-52582006000200004&nrm=iso>.

CARDOSO, A.; PORTUGAL, P. Contractual wages and the wage cushion under different bargaining settings. *Journal of Labor Economics*, v. 23, n. 4, p. 875–902, 2005. Disponível em: <https://doi.org/10.1086/491608>.

CARVALHO, J. A.; GARCIA, R. A. *Estimativas decenais e quinquenais de saldos migratorios e taxas líquidas de migração do Brasil*. [S.l.], 2002.

CAVALCANTI, T.; MATA, D. D.; SANTOS, M. On the determinants of slum formation. *The Economic Journal*, Oxford University Press, v. 129, n. 621, p. 1971–1991, 2019.

CLEMENS, J. How do firms respond to minimum wage increases? understanding the relevance of non-employment margins. *Journal of Economic Perspectives*, v. 35, n. 1, p. 51–72, February 2021. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.35.1.51>.

CLEMENS, J.; KAHN, L. B.; MEER, J. *The Minimum Wage, Fringe Benefits, and Worker Welfare*. [S.l.], 2018. Disponível em: <http://www.nber.org/papers/w24635>.

CORBI, R.; FERRAZ, T.; NARITA, R. Internal migration and labor market adjustments in the presence of nonwage compensation. 2021.

CORBI, R. et al. *Labor Courts, Job Search and Employment: Evidence from a Labor Reform in Brazil.* [S.l.], 2022.

CORBI, R. B.; SANCHES, F. M. et al. *Church competition, religious subsidies and the rise of Evangelicalism: A dynamic structural analysis.* [S.l.]: FEA/USP, 2022.

CORTES, P.; PAN, J. Outsourcing household production: Foreign domestic workers and native labor supply in hong kong. *Journal of Labor Economics*, University of Chicago Press Chicago, IL, v. 31, n. 2, p. 327–371, 2013.

CORTÉS, P.; PAN, J. When time binds: Substitutes for household production, returns to working long hours, and the skilled gender wage gap. *Journal of Labor Economics*, The University of Chicago Press Chicago, IL, v. 37, n. 2, p. 351–398, 2019.

CORTES, P.; TESSADA, J. Low-skilled immigration and the labor supply of highly skilled women. *American Economic Journal: Applied Economics*, v. 3, n. 3, p. 88–123, 2011.

COSTA, V. G.; NASCIMENTO, J. A. S. d. O conceito de favelas e assemelhados sob o olhar do ibge, das prefeituras do brasil e da onu. *Anais do X Encontro de Geógrafos da América Latina*, USP São Paulo, p. 3794–3808, 2005.

DUSTMANN, C.; FRATTINI, T.; PRESTON, I. P. The effect of immigration along the distribution of wages. *Review of Economic Studies*, Oxford University Press, v. 80, n. 1, p. 145–173, 2012.

DUSTMANN, C.; GLITZ, A. How do industries and firms respond to changes in local labor supply? *Journal of Labor Economics*, v. 33, n. 3, p. 711–50, 2015.

DUSTMANN, C.; SCHÖNBERG, U.; STUHLER, J. The impact of immigration: Why do studies reach such different results? *The Journal of Economic Perspectives*, American Economic Association, v. 30, n. 4, p. 31–56, 2016.

DUTTA, A.; GANDHI, S.; GREEN, R. K. Distant shocks, migration, and housing supply in india. *Migration, and Housing Supply in India (April 5, 2021)*, 2021.

EAST, C. N.; VELÁSQUEZ, A. Unintended consequences of immigration enforcement: Household services and high-educated mothers' work. *Journal of Human Resources*, University of Wisconsin Press, p. 0920–11197R1, 2022.

FARRÉ, L.; GONZÁLEZ, L.; ORTEGA, F. Immigration, family responsibilities and the labor supply of skilled native women. *The BE Journal of Economic Analysis & Policy*, De Gruyter, v. 11, n. 1, 2011.

FERREIRA, P. C.; MONGE-NARANJO, A.; PEREIRA, L. Torres de M. Of cities and slums. *FRB St. Louis Working Paper*, n. 2016-22, 2016.

FIELD, E. Property rights and investment in urban slums. *Journal of the European Economic Association*, Oxford University Press, v. 3, n. 2-3, p. 279–290, 2005.

FOGED, M.; PERI, G. Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, American Economic Association, v. 8, n. 2, p. 1–34, 2016.

FRANKLIN, S. Slums to failed estates? evidence from a lottery for low-cost housing in ethiopia. *Job Market Paper*, v. 51, 2018.

FURTADO, D. Immigrant labor and work-family decisions of native-born women. *IZA World of Labor*, 2015.

FURTADO, D. Fertility responses of high-skilled native women to immigrant inflows. *Demography*, Duke University Press, v. 53, n. 1, p. 27–53, 2016.

GALIANI, S.; SCHARGRODSKY, E. Property rights for the poor: Effects of land titling. *Journal of Public Economics*, Elsevier, v. 94, n. 9-10, p. 700–729, 2010.

GECHTER, M.; TSIVANIDIS, N. Spatial spillovers from urban renewal: evidence from the mumbai mills redevelopment. *Penn State University (mimeo)*, 2020.

GLAESER, E. *Triumph of the city: How urban spaces make us human.* [S.l.]: Pan Macmillan, 2011.

GOLDSMITH-PINKHAM, P.; SORKIN, I.; SWIFT, H. Bartik instruments: What, when, why, and how. *American Economic Review*, v. 110, n. 8, p. 2586–2624, August 2020. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/aer.20181047>.

GONZAGA, G. Labor turnover and labor legislation in Brazil. *Economía (Journal of the Latin American and Caribbean Economic Association)*, v. 4, n. 1, p. 165–222, 2003.

GONZAGA, G. M.; FILHO, N. A. M.; CAMARGO, J. M. Os efeitos da redução da jornada de trabalho de 48 para 44 horas semanais em 1988. *Revista Brasileira de Economia*, SciELO Brasil, v. 57, p. 369–400, 2003.

GOURIEROUX, C.; MONFORT, A.; TROGNON, A. Pseudo maximum likelihood methods: Theory. *Econometrica: journal of the Econometric Society*, JSTOR, p. 681–700, 1984.

GREEN, R. K.; MALPEZZI, S.; MAYO, S. K. Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources. *American Economic Review*, American Economic Association, v. 95, n. 2, p. 334–339, 2005.

HARRIS, I. et al. Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. *Scientific data*, Nature Publishing Group, v. 7, n. 1, p. 1–18, 2020.

HARRIS, J. R.; TODARO, M. P. Migration, unemployment and development: A two-sector analysis. *American Economic Review*, American Economic Association, v. 60, n. 1, p. 126–142, 1970. ISSN 00028282. Disponível em: <http://www.jstor.org/stable/1807860>.

HENDERSON, J. V.; REGAN, T.; VENABLES, A. J. Building the city: from slums to a modern metropolis. *The Review of Economic Studies*, Oxford University Press, v. 88, n. 3, p. 1157–1192, 2021.

HORNBECK, R. The enduring impact of the american dust bowl: Short-and long-run adjustments to environmental catastrophe. *American Economic Review*, v. 102, n. 4, p. 1477–1507, 2012.

IBGE. As favelas do distrito federal eo censo demográfico de 1950. *Documentos censitários*, IBGE Rio de Janeiro, n. 9, 1953.

IBGE. Censo demográfico de 1960: Favelas do estado da guanabara. *Documentos censitários*, IBGE Rio de Janeiro, n. 9, 1953.

IBGE. *Divisão do Brasil em mesorregiões e microrregiões geográficas*. 1990. Instituto Brasileiro de Geografia e Estatística, <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15778-divisoes-regionais-do-brasil.html?edicao=16163&t=downloads>. Last accessed on 13/08/2021.

IMBERT, C. et al. Migrants and firms: Evidence from china. *American Economic Review*, v. 112, n. 6, p. 1885–1914, 2022.

IMBERT, C.; ULYSSEA, G. *Firm Dynamics and the Labor Market Effects of Rural-Urban Migration*. [S.l.], 2022.

INGLEHART, R. et al. World values survey: Round six-country-pooled datafile version. *Madrid: JD systems institute*, v. 12, 2014.

JAEGER, D. A.; RUIST, J.; STUHLER, J. *Shift-Share Instruments and the Impact of Immigration*. [S.l.], 2018. Disponível em: <https://ideas.repec.org/p/nbr/nberwo/24285.html>.

JAYACHANDRAN, S. The roots of gender inequality in developing countries. *economics*, Annual Reviews, v. 7, n. 1, p. 63–88, 2015.

JONES, K.; WILCHER, B. Reducing maternal labor market detachment: A role for paid family leavefl. 2019.

JR, C. M. S. et al. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. *Remote Sensing*, MDPI, v. 12, n. 17, p. 2735, 2020.

KAHN, A. S.; CAMPUS, R. T. *Effects of drought on agricultural sector of Northeast Brazil*. [S.l.], 1992.

KLEEMANS, M.; MAGRUDER, J. Labour market responses to immigration: Evidence from internal migration driven by weather shocks. *The Economic Journal*, Wiley Online Library, v. 128, n. 613, p. 2032–2065, 2018.

KLEVEN, H.; LANDAIS, C.; SØGAARD, J. E. Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, v. 11, n. 4, p. 181–209, 2019.

LAGAKOS, D. Urban-rural gaps in the developing world: Does internal migration offer opportunities? *Journal of Economic Perspectives*, v. 34, n. 3, p. 174–92, August 2020. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.34.3.174>.

LAGOS, L. *Labor Market Institutions and the Composition of Firm Compensation: Evidence from Brazilian Collective Bargaining*. [S.l.], 2020.

LALL, S. V.; LEBRAND, M. S. M.; SOPPELSA, M. E. *The Evolution of City Form : Evidence from Satellite Data*. [S.l.], 2021. Disponível em: <https://ideas.repec.org/p/wbk/wbrwps/9618.html>.

LALL, S. V. et al. Do urban land regulations influence slum formation? evidence from brazilian cities. *Brazilian Association of Graduate Programs in Economics*, 2007.

LEE, D. S. et al. *Valid t-ratio Inference for IV*. [S.l.], 2020. Disponível em: <https://EconPapers.repec.org/RePEc:arx:papers:2010.05058>.

LEFEBVRE, P.; MERRIGAN, P. Child-care policy and the labor supply of mothers with young children: A natural experiment from canada. *Journal of Labor Economics*, The University of Chicago Press, v. 26, n. 3, p. 519–548, 2008.

LEFEBVRE, P.; MERRIGAN, P.; VERSTRAETE, M. Dynamic labour supply effects of childcare subsidies: Evidence from a canadian natural experiment on low-fee universal child care. *Labour Economics*, Elsevier, v. 16, n. 5, p. 490–502, 2009.

LEWIS, E. Immigration, skill mix, and capital skill complementarity. *Quarterly Journal of Economics*, v. 126, n. 2, p. 1029–69, 2011.

LUNDBERG, S. The added worker effect. *Journal of Labor Economics*, [University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago], v. 3, n. 1, p. 11–37, 1985. ISSN 0734306X, 15375307. Disponível em: <http://www.jstor.org/stable/2535048>.

MARINHO, C. R. V. *Essays on Health Economics*. [S.l.], 2020.

MARX, B.; STOKER, T.; SURI, T. The economics of slums in the developing world. *Journal of Economic perspectives*, American Economic Association, v. 27, n. 4, p. 187–210, 2013.

MATION, L. F.; NADALIN, V. G.; KRAUSE, C. *Favelização no Brasil entre 2000 e 2010: Resultados de uma classificação comparável*. [S.l.], 2014.

MCKENZIE, R. The labor market effects of minimum wage laws: A new perspective. *Journal of Labor Research*, v. 1, p. 255–264, 1980.

MEGHIR, C.; NARITA, R.; ROBIN, J.-M. Wages and informality in developing countries. *American Economic Review*, v. 105, n. 4, p. 1509–46, April 2015. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/aer.20121110>.

MOLLOY, R.; SMITH, C. L.; WOZNIAK, A. Internal migration in the united states. *Journal of Economic Perspectives*, v. 25, n. 3, p. 173–96, September 2011. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.25.3.173>.

MUNSHI, K. Networks in the modern economy: Mexican migrants in the us labor market. *The Quarterly Journal of Economics*, MIT Press, v. 118, n. 2, p. 549–599, 2003.

NARITA, R. Self-employment in developing countries: A search-equilibrium approach. *Review of Economic Dynamics*, v. 35, p. 1–34, 2020. ISSN 1094-2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1094202518302874>.

NEUMARK, D.; WASCHER, W. *Minimum Wages and Employment: A Review of Evidence from the New Minimum Wage Research*. [S.l.], 2006. (Working Paper Series, 12663). Disponível em: <http://www.nber.org/papers/w12663>.

NIU, D.; SUN, W.; ZHENG, S. The role of informal housing in lowering china's urbanization costs. *Regional Science and Urban Economics*, Elsevier, v. 91, p. 103638, 2021.

OLIVETTI, C.; PETRONGOLO, B. The economic consequences of family policies: lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418, v. 31, n. 1, p. 205–230, 2017.

OTTAVIANO, G. I.; PERI, G. Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, Oxford University Press, v. 10, n. 1, p. 152–197, 2012.

OTTAVIANO, G. I.; PERI, G.; WRIGHT, G. C. Immigration, offshoring, and american jobs. *American Economic Review*, v. 103, n. 5, p. 1925–59, 2013.

PERRY, G. et al. *Informality: Exit or Exclusion*. [S.l.]: Washington, DC: World Bank, 2007.

PINHEIRO, L. S. et al. Mulheres e trabalho: breve análise do período 2004-2014. Instituto de Pesquisa Econômica Aplicada (Ipea), 2016.

REY, E. D.; KYRIACOU, A.; SILVA, J. I. Maternity leave and female labor force participation: evidence from 159 countries. *Journal of Population Economics*, Springer, v. 34, p. 803–824, 2021.

ROCHA, R.; SOARES, R. R. Water scarcity and birth outcomes in the brazilian semiarid. *Journal of Development Economics*, Elsevier, v. 112, p. 72–91, 2015.

ROSER, M.; ORTIZ-OSPINA, E. Income inequality. *Our World in Data*, 2013. Https://ourworldindata.org/income-inequality.

SKOUFIAS, E.; VINHA, K.; CONROY, H. The impacts of climate variability on household welfare in rural mexico. *Population and Environment*, v. 34, n. 3, p. 370–399, 2013.

SOTO, H. D.; DIAZ, H. P. The mystery of capital. why capitalism triumphs in the west and fails everywhere else. *Canadian Journal of Latin American & Caribbean Studies*, Taylor & Francis Ltd., v. 27, n. 53, p. 172, 2002.

TASCHNER, S. P. Favelas em são paulo–censos, consensos e contra-sensos. *Cadernos Metrópole*, n. 05, p. 09–27, 2001.

ULYSSEA, G. Regulation of entry, labor market institutions and the informal sector. *Journal of Development Economics*, v. 91, p. 87–99, 2010.

ULYSSEA, G. Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, v. 108, n. 8, p. 2015–47, 2018.

ULYSSEA, G. Informality: Causes and consequences for development. *Annual Review of Economics*, v. 12, n. 1, p. 525–546, 2020.

UN DESA. *International Migration Report 2017*. 2017. United Nations Department of Economic and Social Affairs, Population Division.

UNDP. *Overcoming barriers: Human mobility and development*. 2009. United Nations Development Program, Human Development Report 2009. <http://hdr.undp.org/en/content/human-development-report-2009>. Last accessed on 03/09/2021.

VELÁSQUEZ, A. The economic burden of crime evidence from mexico. *Journal of Human Resources*, University of Wisconsin Press, v. 55, n. 4, p. 1287–1318, 2020.

VICENTE-SERRANO, S. M.; BEGUERÍA, S.; LÓPEZ-MORENO, J. I. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, v. 23, n. 7, p. 1696–1718, 2010.

WANG, X. et al. Wet periods in northeastern brazil over the past 210 kyr linked to distant climate anomalies. *Nature*, v. 432, p. 740–743, 2004.