Internal Migration and Local Labor Markets in Brazil

Raphael Corbi[*], Tiago Ferraz[†] and Renata Narita[‡]

Preliminar version: November 3, 2020

Click here for most recent version

Abstract

We investigate the labor market impacts of weather-induced internal migration in Brazil between 1996 and 2010 using a shif-share instrument which combines variation from the Brazilian Semiarid migrant flows with the preexisting support network in each destination based on the migrant's region of origin. Our results indicate that increasing in-migration rate by $1p.p.$ reduces formal employment by $0.2p.p.$, increases the number of informal workers by $0.1p.p.$, while reduces the earnings of native workers by 1.5% in the informal sector and by 1.4% for those who are self-employed. We find no effect in the earnings of workers in formal jobs, but there is a reduction in the proportion of those who receive fringe benefits.

Keywords: Internal migration, labor supply, shift-share

JEL Codes: J21, J22, J61, R23

[*]University of São Paulo, Department of Economics, email: rcorbi@usp.br

[†]University of São Paulo, Department of Economics, email: tiago.ferraz@usp.br

[‡]University of São Paulo, Department of Economics, email: rnarita@usp.br

---

1 Introduction

Internal migration can be a powerful tool to fight poverty and an important coping mechanism against climate change. Although it is hard to precisely estimate its economic impacts, there is a growing concern that large weather shocks could harm vulnerable populations in rural areas of developing countries (Skoufias et al., 2011). In such scenario, migration could help to attenuate the longer term welfare impacts of changes in environmental conditions (Chein and Assunção, 2008).

On the other hand, the impacts in the destination areas are hard to anticipate. While some individuals may suffer because of increasing competition in the labor market, others could benefit from a growing demand for goods and services, rendering the net effect a question to be empirically answered. Not only that, in the context of a developing country where regulated labor markets coexist with a less frictional informal sector one could expect the effects of labor reallocation across destination areas to be quite distinct depending on the sector analyzed.

We investigate what are the impacts of internal migration on labor market outcomes for native workers in the destination municipalities where people move into. We study the flow of migrants coming from the Brazilian Semiarid and settling in destinations outside the region, during the period 1996-2010. Over 3 million people left their hometowns in the Semiarid region along this period.

To conduct our empirical analysis we construct a shift-share instrument in two steps. First, we exploit exogenous rainfall and temperature shocks at the origin region to predict the number of individuals leaving the origin municipalities. Then, we use the past settlement network to allocate them in the destination areas. Thus, we can use this predicted flow as instrumental variable for the observed in-migration.[1]

We find that increasing the share of migrants by one percentage point reduces formal employment among native workers by $0.3p.p.$, while increases the number of informal jobs by $0.2p.p.$ and self-employment by $0.1p.p.$. How can we interpret these results? Because natives have to face competition from an increase in the supply of workers, they are less likely to maintain a formal job. And, because the informal sector and self-employment are less frictional markets they are more likely to absorb those individuals who lost their formal position. We also estimate that the unemployment rate and labor force participation increased by $0.2p.p.$ and $0.1p.p.$, respectively.

There is no effect on the earnings of workers employed in the formal sector, but they are less likely to receive fringe benefits. The null effects on earnings occur not only on average, but across the entire distribution. For individuals employed in the informal sector or self-employed we show a 1.4% decrease on earnings. These effects are stronger for those at the bottom of the distribution.

[1] For a similar approach, see Boustan et al. (2010); Imbert et al. (2016).

Finally, we highlight that all those effects are more striking for native workers who are less educated, which is consistent with the fact that they are more likely to compete with Semiarid's migrants.

This article relates to several strands of the literature on the economics of migration and labor markets. First, there are several papers showing the relationship between climatic conditions and migration, either focusing on weather related shocks (Bastos et al., 2013; Peri and Sasahara, 2019; Kubik and Maurel, 2016; Gray and Mueller, 2012) or more broadly on natural disasters (Gröger and Zylberberg, 2016; Hornbeck, 2012; Mahajan and Yang, 2020).

Second, there is an extensive branch of the literature analyzing the labor market effects of internal migration, like Boustan et al. (2010) who focus on the Great Depression period to study the effects of internal migration in the US; Imbert and Papp (2020) studying the impacts of a rural public works program on seasonal migration and spatial spillovers; and Imbert et al. (2016) who use Chinese data to assess the effects of internal migration on manufacturing growth. More closely related to our work is Kleemans and Magruder (2017) who analyze the Indonesian labor market and show that internal migration reduces employment in the formal sector and earnings in the informal sector. We contribute to this literature by showing that when firms operate in a more frictional labor market they may also adjust to supply shocks via non-wage compensation.

Finally, our paper is also related to several others that use shift-share instruments to identify the effects of immigration on a wide range of outcomes (Card, 2001, 2009; Ottaviano and Peri, 2012; Ottaviano et al., 2013; Cattaneo et al., 2015; Foged and Peri, 2016). Despite being a very traditional empirical approach, there is more recent literature examining potential issues with shift-share instruments (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2018; Jaeger et al., 2018; Adao et al., 2019). We will briefly discuss some of these problems and implement some robustness checks using the procedure developed by Borusyak et al. (2018).

This paper is organized as follows. In the next section, we first present some background information on the Brazilian Semiarid region and labor markets, then we discuss the data used in our empirical analysis. In section 3, we explain the empirical strategy and the identifying assumptions we make. In section 4, we present and analyze the main results. Finally, in Section 5 we discuss the main threats to our identification strategy and show some robustness checks.

2 Background and Data

In this section, we first describe the economic background and weather conditions at the Semiarid region and the functioning of local labor markets in Brazil, 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.

2.1 Brazilian Semiarid

According to the official definition provided by the Ministry of National Integration, the Brazilian Semiarid encompasses 1,133 municipalities distributed by 9 states, covering an area of around 976,000km^2 (roughly 11 percent of the country's territory).[2] A municipality officially belongs to the Semiarid region 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[3]; (iii) risk of drought above 60%[4].

Average historical precipitation in the Semiarid is about 780mm, as opposed to around 1,500 mm for the rest of the country[5], 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 typically starts.

The Semiarid displays some very bad development indicators: about 80% of the children lived in households below the poverty line and infant mortality reached 31 per 1000 births in 1996, compared to a national average of 25% and 15 per 1000 births, respectively.[6] Municipalities are typically small (median population is around 20,000) and their economies are susceptible to suffer from climate shocks. Average human capital level is relatively low: more than 80% of the adult population had less than 8 years of schooling in 1991. Even though the situation improved along the years, this proportion was still above 60% in 2010.

2.2 Local labor markets in Brazil

A common feature of labor markets in developing countries is the existence of a two-sector economy, where firms operating in the formal sector coexist with those in the more unregulated informal sector. When firms hire workers formally they have to comply with several legal obligations, like paying minimum wages or observing safety regulations, among others.

According to the Brazilian law, firms have to register every employee but the enforcement is costly and the penalties for those who are caught can be discussed in courts. Therefore, some firms can choose to take the risk of hiring workers informally.

2.3 Migration, Labor Market and Weather Data

We collect data from three rounds of the Brazilian Census (1991, 2000 and 2010), provided by the *Instituto Brasileiro de Geografia e Estatística (IBGE)*, to construct a municipality-level panel of migration. For each census-year, they asked individuals who were not

[2]It includes almost all Northeastern states, except for *Maranhão*, plus the northern area of *Minas Gerais*, but it does not includes any state capital.

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

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

[5]See Figure 9.

[6]See Rocha and Soares (2015).

born in the municipality (i) how many years they lived in the current municipality; (ii) what is the municipality where they lived before. This allowed us to construct a measure of yearly out-migration from each origin municipality in the Semiarid and a measure of in-migration to each destination during 1996-2010. The 2000 Census only asked respondents where they lived 5 years before, so for this particular round, we can only track the individuals during the period 1996-2000. Since 1988 several municipalities were split into new ones. In order to avoid potential miscoding regarding migration status or municipality of origin, we aggregate our data using the original municipal boundaries as they were in 1991.

We also built a measure of pre-existing networks by associating the share of migrants from each Semiarid origin municipality in each destination, using the 1991 Census. This is especially relevant for our identification strategy, discussed in more detail in the next section, to resolve endogeneity problems that could arise when migrants choose the place where they move into.

Weather data were retrieved from the Climatic Research Unit at 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.[7]

Then we define the rainfall shocks as deviations from the historical average. More specifically

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

where $r_{o\tau t}$ is the rainfall in 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).[8] Historical averages are calculated over the period from 1901 to 2010. Temperature shocks are computed in a similar way, using average temperatures instead of the summation.

Table 1 describes municipality-level data for origin (Panel A) and destination (Panel B) municipalities. Semiarid's areas show lower levels of rainfall, slightly higher temperatures and are less populated than destination municipalities. On average, 1.0% of Semiarid's population move to larger cities within the country every year.

We gather labor market outcomes from the *Pesquisa Nacional por Amostra de Domicílios (PNAD)* , a survey also conducted by the *IBGE*, covering 851 municipalities in all 27 states. Although the number of municipalities can appear to be small, these are the destination choice of about half of the migrants from the Semiarid region and concentrate around 65% of the employed population in Brazil. The survey is realized

[7]This methodology follows that used by Rocha and Soares (2015).

[8]We also calculated this measure using the 12 months of the year. Results are virtually the same.

every year, except for those when the Census is conducted. Therefore, we have data from 1996 to 1999 and from 2001 until 2009.

We restricted the sample to individuals between 18 and 65 years old, living in the municipality for 10 years or more and we refer to them as natives. To avoid concerns about spatial correlation in the weather shocks we drop from the sample all the observations in municipalities that belong to the Semiarid region. The final sample has 2,153,328 individuals at 684 municipalities.

We use as outcomes the log of earnings and several dummies indicating whether the individual is employed; whether the job is in the formal sector, whether is informal or self-employment; whether she is unemployed and whether is out of the labor force; and some combinations of these dummies with indicators for low/high education level. We also create indicator variables for some forms of non-wage compensation. The survey asks specifically whether the individual received any kind of payment to help cover expenses with food, transport or health insurance. We create a dummy for each type, for those individuals who are employed in the formal sector. Finally, we pooled the 13 years of the survey and took the average of these variables at the municipality-by-year level.

Table 3 provides descriptive statistics for destination municipalities. In our sample, 63% of individuals are employed, 32% have a formal job, 16% have an informal job and 16% are self-employed. Unemployment rate is 13% and 24% of individuals are not in the labor force. The average monthly earning[9] is R$ 637.87, but is substantially higher in the formal sector (R$ 788.18) than in the informal sector (R$ 382.31) and a little higher than that for self-employed natives (R$ 600.83).

Among worker employed in the formal sector, 39% receive some help to cover expenses with food, 36% for transport and 20% for health expenditures.

3 Empirical Strategy and Identifying Assumptions

In this section, we first describe the empirical framework that allows us to (i) isolate the observed variation in out-migration induced by exogenous weather shocks, and (ii) the in-migration flows into destination municipalities predicted by the pre-existing migrant networks. We then discuss and present supportive evidence on the validity of this procedure that is key to isolate the causal effect of in-migration on labor market outcomes for native workers.

3.1 Empirical Framework. Weather-induced Migration.

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

$$\Delta y_{dt} = \alpha + \beta m_{dt} + \gamma \Delta X_{dt} + \psi_t + \epsilon_{dt} \tag{2}$$

[9] Earnings are measured in R$ (2012).

where y_{dt} is a vector of labor outcomes at destination municipality d in year t, m_{dt} is the destination migrant inflow, X_{dt} are municipality-level controls, ψ_t absorb year fixed effects and ϵ_{dt} is the error term. By differencing the outcome variables we can account for time-invariant unobserved characteristics that could be correlated with in-migration flows, but the error term may include unobserved time-varying confounders which would potentially bias OLS estimates. In particular, migrants could choose a specific destination municipality due to demand shocks leading to higher wages or job prospects.

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 out-migration rate[10] from origin municipality o in year t, using weather shocks in the previous year:

$$m_{ot} = \alpha + \beta' Z_{ot-1} + \phi_o + \delta_t + \varepsilon_{ot} \tag{3}$$

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:

$$\widehat{M}_{ot} = \widehat{m}_{ot} \times P_o \tag{4}$$

In the second step we use the pre-existent network of migrants from the origin o to municipality d in order to distribute them throughout the destination areas.

$$\widetilde{m_{dt}} = \sum_{o=1}^{O} \frac{s_{od} \times \widehat{M}_{ot}}{P_d} \tag{5}$$

Where s_{od} is the share of migrants from origin municipality o who settled in the destination area d in 1991.[11] We normalize the instrument dividing each destination migrant shock by its total population in 1991, P_d.

Hence the specification

$$\Delta y_{dt} = \alpha + \beta \widetilde{m_{dt}} + \gamma \Delta X_{dt} + \psi_t + \epsilon_{dt} \tag{6}$$

is the reduced-form relationship that associates labor market outcomes and the predicted migrant flow at the destination.

To be clear, our instrument is an adaptation of the shift-share instrument developed in Card (2001), using an estimated flow of migrants from the Semiarid region. We claim

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

[11] For every year we use the predetermined share of migrants in the 1991 Census. We also tested another specification with an updated network using the predetermined flow of migrants in the previous Census. There is no much difference in the results, but updating the network may raise concerns about the persistence in migrant flows as discussed by Jaeger et al. (2018).

that this strategy provides convincingly exogenous shocks exploiting two sources of variation that allows us to identify the parameters of interest. First, we have variation in the cross-section given by the predetermined network of migrants (share). Second, we exploit variation along the time using the migrant flows induced by rainfall and temperature shocks (shift). Leveraging variation from the origin location weather conditions may alleviate concerns that our results are driven by demand shocks in the destination areas.

The validity of this type of instrument has been scrutinized by recent literature which analyzes its formal structure, distinguishing sufficient conditions for identification according to each source of variation. Goldsmith-Pinkham et al. (2020) show that a sufficient condition for consistency of the estimator is the strict exogeneity of the shares, which can be a very stringent assumption. In contrast, Borusyak et al. (2018) discuss how one can use variation from the shocks instead of the shares and establish sufficient conditions for the estimator to be consistent. They propose to take destination-level outcomes and endogenous variables and calculate its origin-level weighted averages to estimate the parameter of interest using the shocks as instruments. They also demonstrate that even if the shares are not exogenous this method renders a consistent estimator. This is the ideal approach in our setting because we have a plausible exogenous shock to exploit in order to identify the causal effect of internal migration on labor outcomes.

One last aspect of the shift-share design which has been analyzed in further detail is inference. Adao et al. (2019) argue that in the standard setup, regression residuals are correlated across destinations with similar past settlement shares independently of their geographic location, potentially understating the true variability of OLS and leading to null hypothesis overrejection. Borusyak et al. (2018) show that by estimating shift-share coefficients with a shock-level IV regression one obtains standard errors that are asymptotically valid regardless of the correlation structure of the error term. In Appendix B we further discuss those issues and present some results to show that our findings remain robust.

3.2 Identifying Assumptions.

We use Semiarid municipality-level data to estimate variations of specification 3 and report the results in Table 4. All regressions control for temperature shocks and the log of 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. Standard errors are clustered at the grid level to account for the fact that municipalities in the same grid will have similar shocks[12].

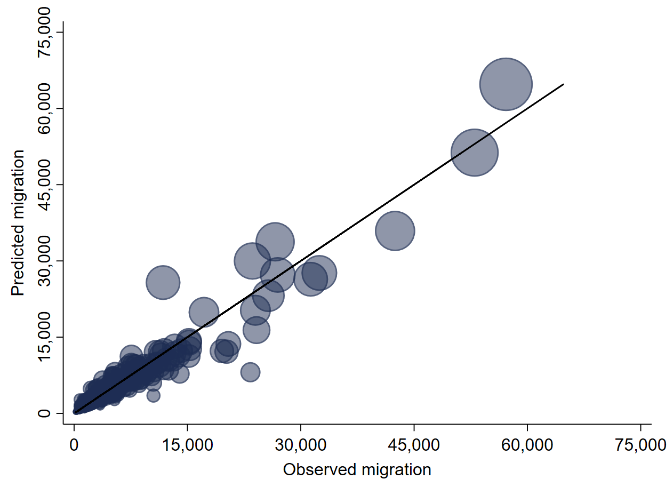
As expected, rainfall shocks in the previous year are negatively correlated with

[12] Similar but no equal, because the 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.

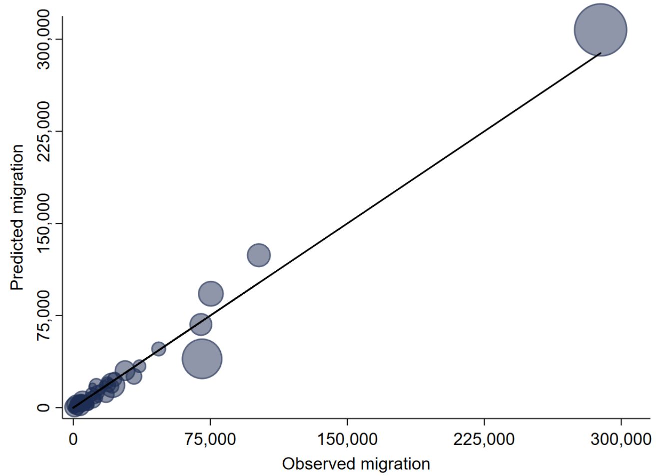
out-migration indicating that Semiarid's inhabitants are induced to leave the region during drought periods. Coefficient estimates are remarkably stable across specifications and adding more lags do not change the baseline results. More important to our identification, we include as control rainfall 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 in the Semiarid where annual rainfall is 10% below historical average will experience an increase of $1.0p.p.$ in the out-migration rate.

In the second step, we distribute the predicted out-migration shock using the pre-existent network of migrants from origin municipality o to destination d. A *sine qua non* requirement implicit in our empirical framework is that both predicted out- and in-migration rates, $\widetilde{m_{ot}}$ and $\widetilde{m_{dt}}$ respectively, should be strongly correlated with their observed counterparts. Figure 1 illustrates that our prediction provides a reasonable approximation of the observed migrant flows. Panel (a) shows the relationship between the predicted and observed number of migrants leaving the Semiarid region and entering all non-Semiarid municipalities, accumulated along the period 1996-2010. Panel (b) shows the predicted and observed numbers of incoming Semiarid migrants in all destinations along the same period. In Panel (c) we restrict the analysis only to destination municipalities with a maximum native population of 200,000 people and the results are very similar.

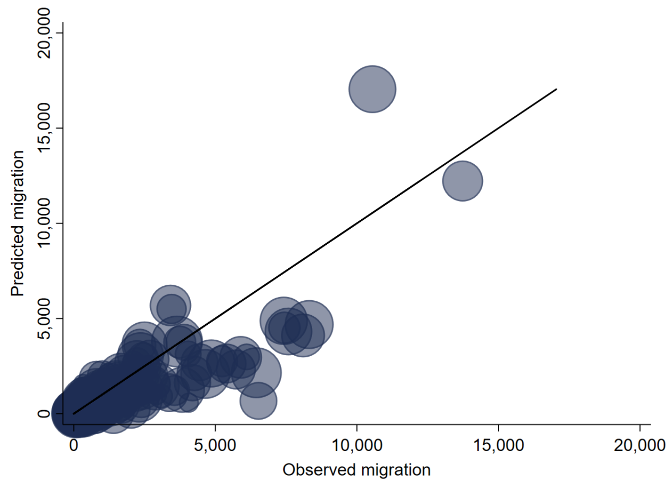
Figure 1: Observed vs predicted migration



(a) Out-migration



(b) In-migration



(c) In-migration

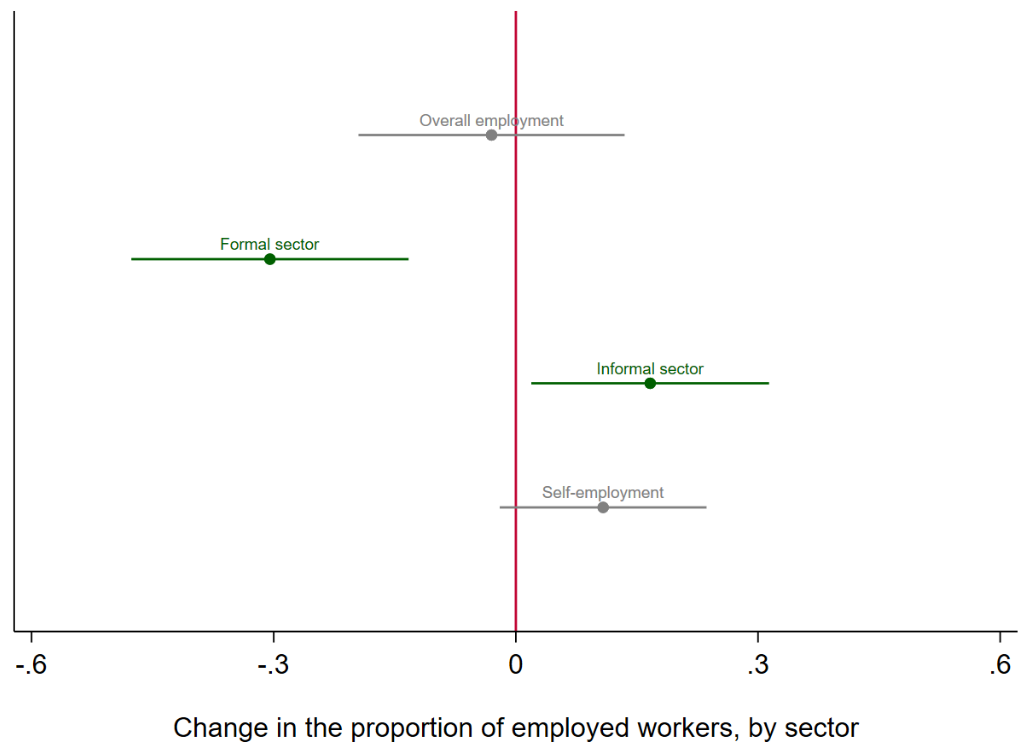
Notes: This figure presents the relationship between the predicted and observed migration flows among Brazilian municipalities, accumulated between 1996 and 2010. Panel (a) shows the number of migrants leaving the Semiarid region to all non-Semiarid municipalities. Panel (b) shows the number of migrants from the Semiarid entering each destination. Panel (c) shows the same measure as Panel (b) but only for municipalities with less than 200,000 natives in 1991. Circle size represents the municipality's total population in 1991. *Data sources*: Census microdata (IBGE).

4 Migration Flows and Labor Markets

4.1 Reduced form estimates

Now we turn our attention to the destination labor markets and focus on the question: how does internal migration affects earnings and employment prospects of native individuals? Our starting point is to show the estimates of the relationship between labor market outcomes and the predicted number of migrants from the Semiarid region. Figure 2 summarizes our main results, while in Tables 5-9 we present the estimates from several specifications for the reduced form equation (6).

Figure 2: **Reduced form: Effects of predicted in-migration on employment**



Notes: This figure shows the relationship between changes in employment rate, by sector, and 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. Controls include time dummies, the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Standard errors are clustered at the municipality level. Green markers are statistically significant at the 5% level.

Table 5 presents the relationship between changes in the employment rates, by sector, and the predicted number of migrants from the Semiarid region, measured as a fraction of the working-age population in 1991. Column (1) displays a flexible specification, without any control. In column (2) we control for the log of the native population in the previous year and for changes in a set of municipality-level demographics (the share of individuals with high school and college education; the share of female, white and black individuals; and a cubic polynomial of age). Column (3) also controls for these changes in demographics and includes time dummies. Finally, there are two important changes occurring during our period that could be a concern for our identification: population aging and an increase in education levels. To account for this, in Column (4) we include a linear trend of age and the shares of high school and college educated individuals. All regressions are weighted by the 1991 native population to account for the fact that heterogeneity in municipality size could lead to differential effects.

In Panel A we examine the effect of the predicted migrant inflow on the total employment, but we find no significant effect. Although the coefficient is always negative, the magnitude is virtually null in every specification and never statistically significant. But, this overall result is actually masked by the employment sectoral composition.

We focus on the employment rate in the formal sector in Panel B . Our estimate implies that one percentage point increase in the predicted in-migration reduces the formal employment rate by $0.3p.p.$ and is very precise. This estimate is very stable across specifications.

Panel C shows the effect of increasing the predicted number of migrants on the employment in the informal sector. The point estimates are slightly smaller and little less precise, but still significant. Increasing the predicted migrant inflow by $1p.p.$ raises the informal employment by $0.2p.p.$ This result is not altered after we add demographic controls, time dummies or linear trends in the covariates.

Finally, in Panel D we explore the impact on the self-employment rate. In this case, there is a smaller impact from increases in the in-migration rate ($0.1p.p.$), but the precision is reduced by adding controls. It is hard to tell whether this effect is statistically different from zero.

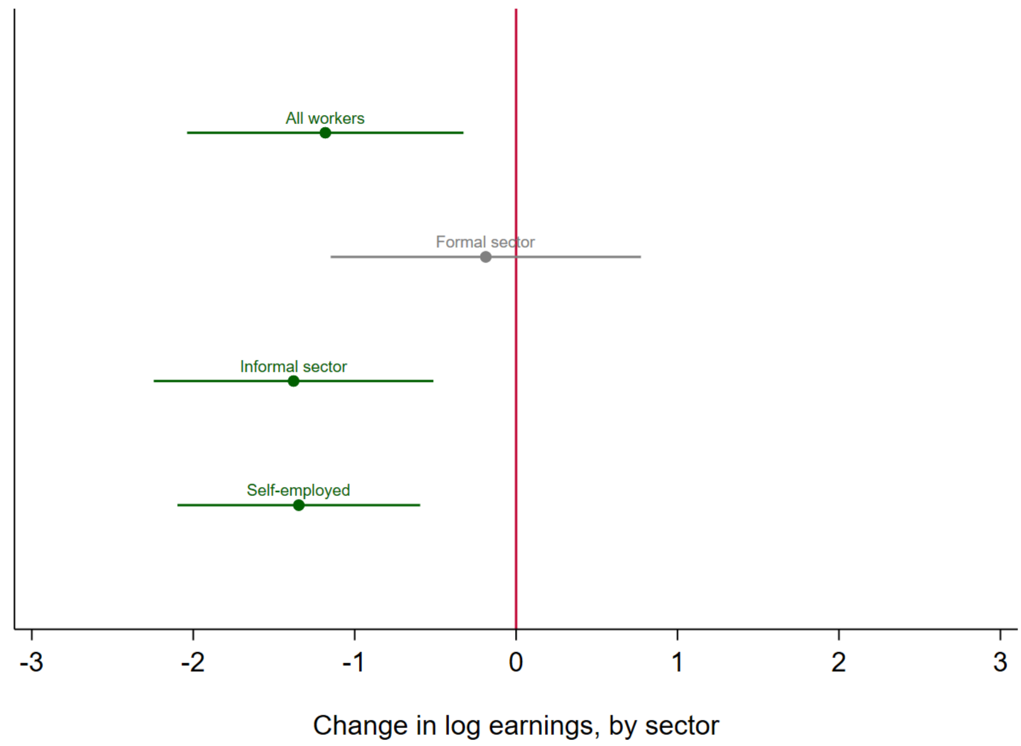
We should highlight that these results are consistent with our intuition about the expected effects. When the labor market in a destination municipality is hit by a supply shock, the native workers will have to face increased competition. Therefore, the negative impact on formal employment. The fact that we see a raise in the proportion of native workers in the informal sector and self-employed may reflect that these sectors are less frictional and more able to absorb those workers who lost the formal jobs.

In Table 6 we present the reduced form estimates on log earnings. Panel A reveals a strong negative effect of the predicted inflow of migrants on log earnings for native workers. The inclusion of controls diminishes a little the magnitude of our coefficient of interest, but that does not change substantially our conclusions. One percentage point increase in the predicted number of migrants reduces earnings by 1.3%.

In Panel B we restricted our analysis to native workers holding a formal job, while in Panels C and D we focus on those in the informal sector and self-employed, respectively. There is no significant effect on the earnings of formal workers, but for those individuals in the informal sector and self-employed earnings are reduced by 1.5% and 1.4%, respectively. We illustrate all these findings in Figure 3, plotting the coefficients from the specification presented in column (3).

Once again, the estimated coefficients are aligned with our expectations about the

Figure 3: **Reduced form: Effects of predicted in-migration on earnings**

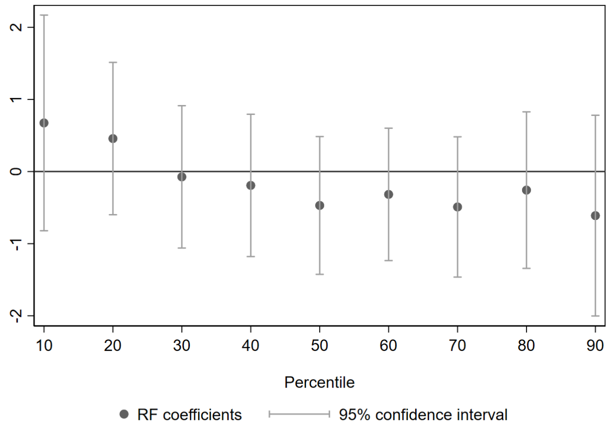


Notes: This figure shows the relationship between changes in log earnings, by sector, and the predicted number of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991. Controls include time dummies, the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Standard errors are clustered at the municipality level. Green markers are statistically significant at the 5% level.

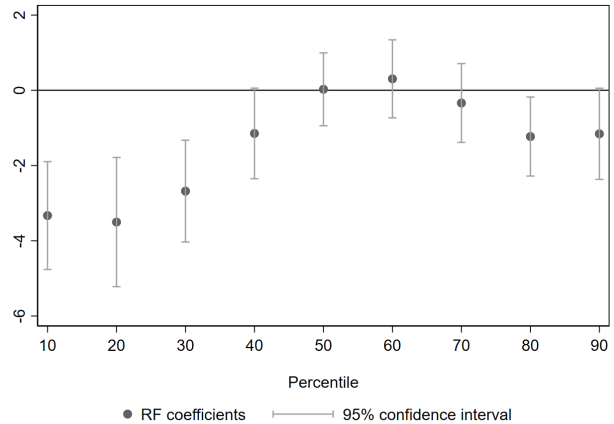
effects. The fact that earnings are decreasing more for workers employed in the informal sector and self-employment compared to those who have a formal job reflects the presence of frictions in this latter market. It should be more difficult for firms operating in the formal sector to reduce wages in order to adjust to the new supply of labor. On the other hand, informal firms face no restrictions like minimum wages or collective bargain with unions, for example.

We also investigate the existence of differential effects according to the native worker's position in the earnings distribution and present the estimates in Figure 4. We estimate the reduced form on each decile of earnings, by sector. For those native workers employed in the formal sector, there is no statistically significant effect along the entire distribution. But, for informal workers and self-employed individuals the impact is substantially stronger for those at the bottom of the distribution.

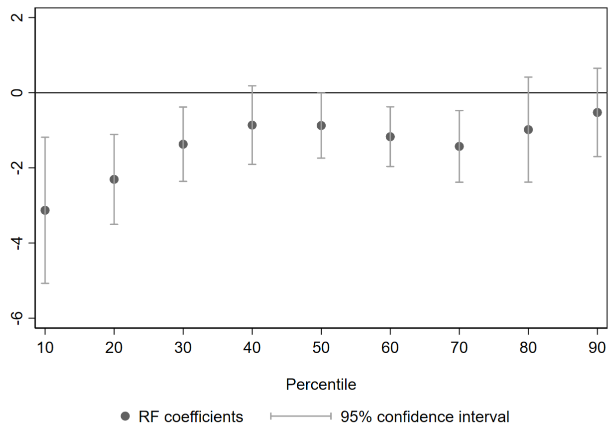
Figure 4: Effects of predicted internal migration along the earnings distribution



(a) Formal sector



(b) Informal sector



(c) Self-employment

Notes: This figure presents the impacts of predicted in-migration along the earnings distribution, by sector. Panel (a) uses only formal workers, while in Panels (b) and (c) we restrict the sample to individuals holding an informal job and self-employed, respectively. Each point is the coefficient from a regression of the change in the average of log earnings, in each decile, on the predicted number of migrants from the Semiarid, measured as a fraction of the native working-age population in 1991. Controls include time dummies, the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. The capped lines show the 95% confidence interval. Standard errors are clustered at the municipality level.

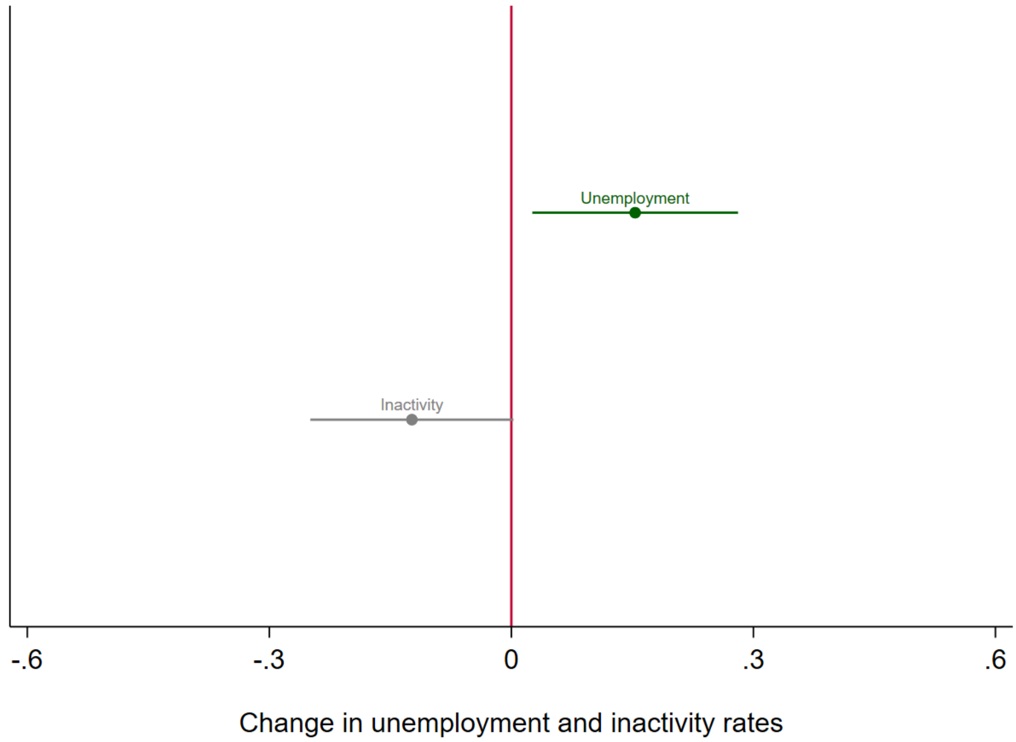
To draw a more complete picture we also look at the impacts on unemployment and labor force participation. In Panel A from Table 7 the dependent variable is the change in the municipality-level proportion of unemployed native workers. As one should expect, raising in-migration leads to a small increase in the unemployment rate ($0.2p.p.$). In Panel B, the outcome used is the change in the proportion of individuals out of the labor force and the point estimate is almost identical, even though with opposite sign. But, in this case is hard to tell ex-ante what should be the most likely effect. On one hand, increased competition in the labor market could discourage native individuals to even look for a job. On the other hand, if the main earner in the household looses his job because of the increased competition , then it is possible that other members of the household would enter the market. We test this second mechanism running the same regressions separately for individuals identified as head or non-head of the household. As we can see in Table 8, almost all the impact on the employment margin comes from those native workers who are head of the household, while the change in unemployment and inactivity rates are led by those identified as non-head. This confirms our intuition that the second channel prevails. Also, the symmetry between the effects on unemployment and inactivity suggests that once those individuals enter the market it takes some time for them to actually find a job.

Finally, there is another margin of adjustment for in-migration shocks. Because firms operating in the formal sector cannot reduce wages below the minimum required by law they may adjust to labor supply shocks reducing the fringe benefits they offer. We focus on individuals who are currently holding a formal job because these benefits are not mandatory for those firms operating in the informal sector, neither by force of law or collective bargaining with unions. Figure 6 present the main results for this mechanism. We find a decrease in the proportions of workers receiving some assistance from the employer to pay for food, transport or health expenses, although the coefficient for food is estimated with less precision. Table 9 shows all different specifications. In Panels A, B and C the dependent variables are the proportions of individuals who reported that they received some help to cover expenses with food, transport and health insurance, respectively. Our estimates suggest a negative relationship between these benefits and the predicted inflow of migrants from the Semiarid region, although in this case the results are more sensitive to the specification. For example, in Panel A the magnitude and the precision of the coefficients increase as we include controls. The same is true in Panels B and C, except that adding a linear trend for the covariates (age and the shares of high school and college educated) actually reduces the magnitude and makes the estimate more noisy.

4.2 Differential effects by education level

In the analysis presented so far we implicitly assumed that native workers and migrants are perfectly substitutable. We relax this assumption now and consider the possibility that individuals with different education levels are not affected the same

Figure 5: **Reduced form: Effects of predicted in-migration on labor force participation**



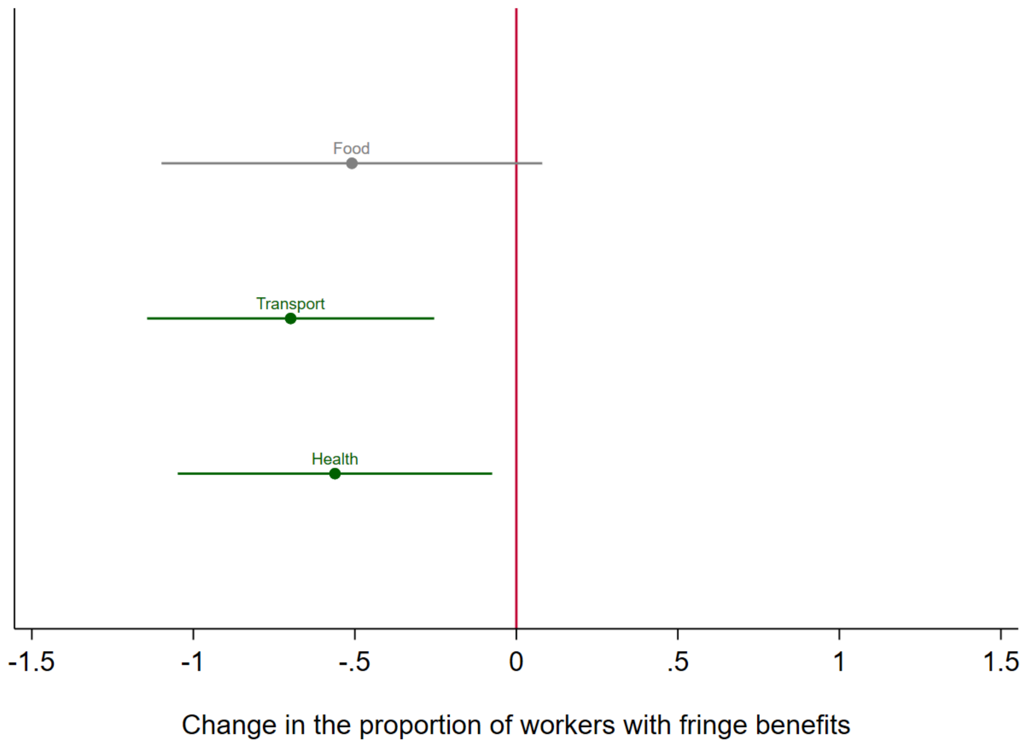
Notes: This figure shows the relationship between changes in unemployment and inactivity rates and the predicted number of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991. Controls include time dummies, the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Standard errors are clustered at the municipality level. Green markers are statistically significant at the 5% level.

way.

We define as low educated someone who have up to 7 years of schooling, which is equivalent to an incomplete elementary education. In our sample, 59% of natives are low educated.

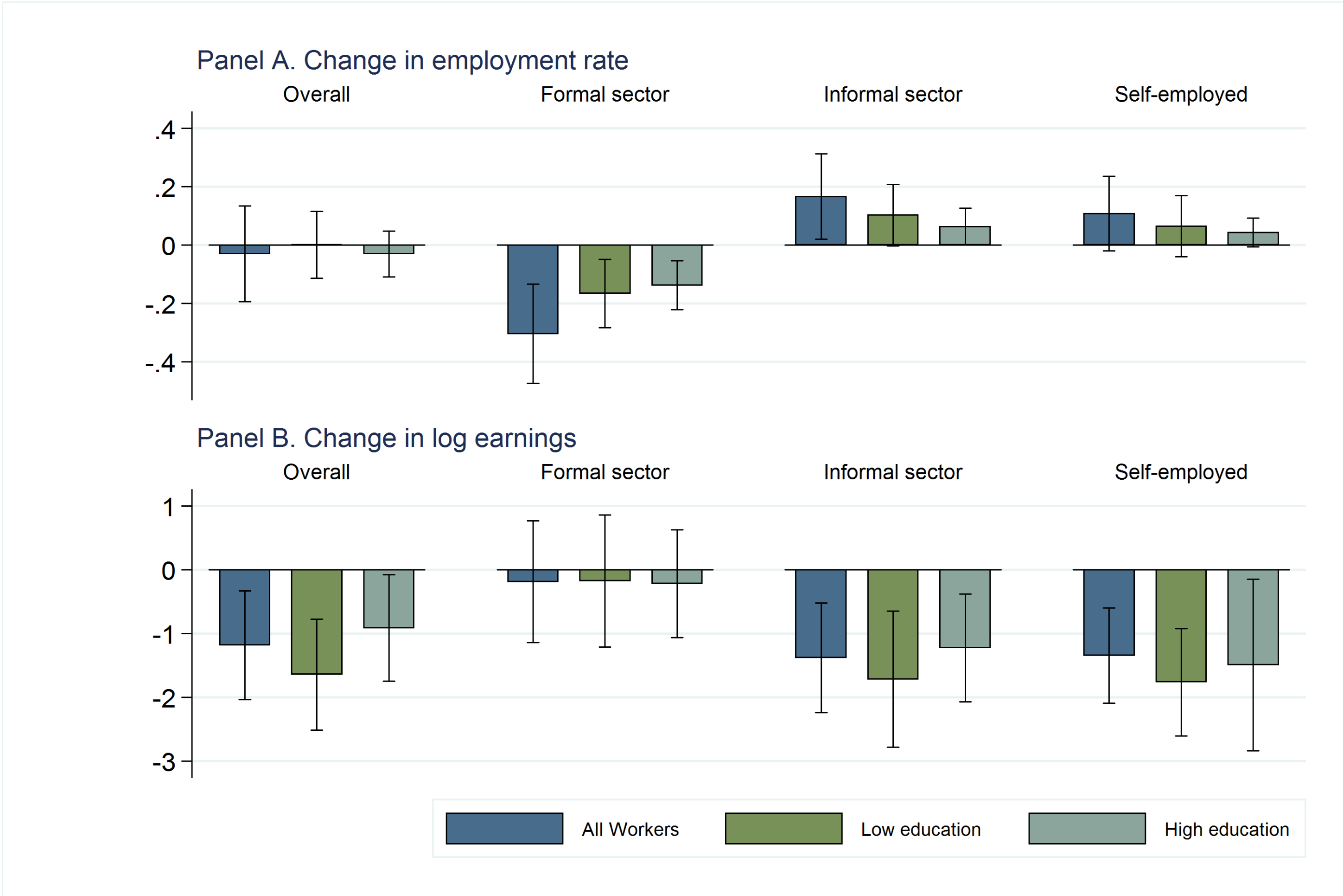
One should expect that native workers who are more alike Semiarid's migrants would be more affected by competition in the labor market. Our findings confirm this intuition. Figure 7 illustrate the reduced form relationship by education group. Panel A shows the effect of predicted in-migration on the changes in employment rates, by sector and group. Low-educated native individuals are more likely to lose a formal job and to be employed in the informal sector compared to those who have higher education. In Panel B we analyze the differential effects on log earnings. There is no significant impact on native workers employed in the formal sector, while in the less frictional markets (informal and self-employed) individuals with less education are those who suffer the most from the increased competition with Semiarid's migrants.

Figure 6: **Reduced form: Effects of predicted in-migration on fringe benefits**



Notes: This table shows the relationship between changes in the proportions of working-age native population, employed in the formal sector, who receive some assistance from the employer to pay for food, transport or health expenses and 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. Controls include time dummies, the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Standard errors are clustered at the municipality level. Green markers are statistically significant at the 5% level.

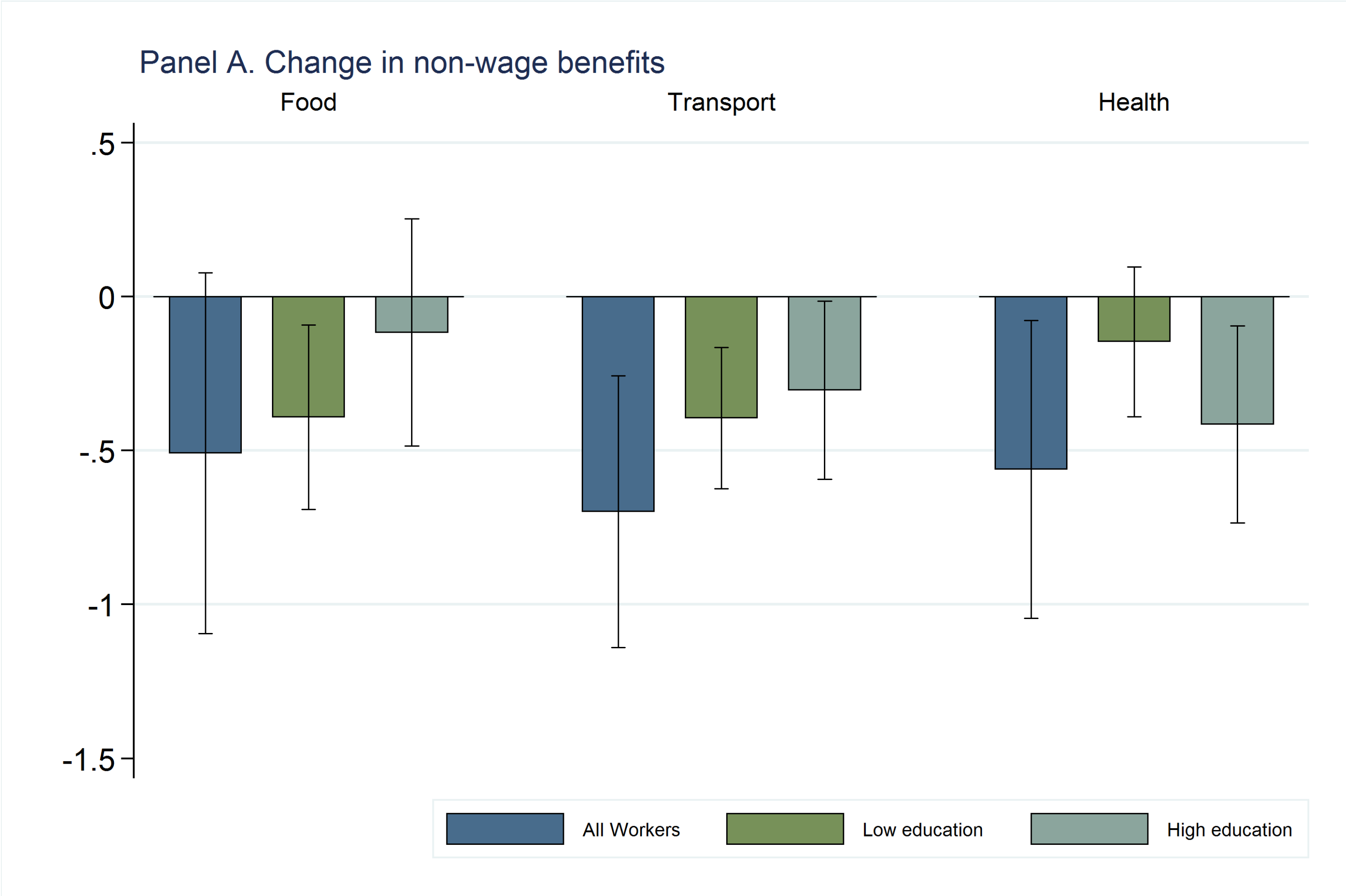
Figure 7: **Labor market effects of predicted internal migration, by education level**



Notes: This figure shows the relationship between changes in labor market outcomes, by education level, and 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. 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 reduced form coefficient by education level. All regressions are weighted by total native population in 1991, include time dummies and control for the log of native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. The capped lines show the 95% confidence intervals.

We also consider the possibility that adjustments in the non-wage benefits could differ by education group. That would be the case if each type of worker attributed different values for such benefits. In Figure 8 we show that the negative impact on food and transport benefits come from low educated workers, but the more educated workers are more likely to lose health benefits.

Figure 8: **Effects of predicted internal migration non-wage benefits, by education level**



Notes: This figure shows the relationship between changes in non-wage benefits, by education level, and 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. All regressions are weighted by total native population in 1991, include time dummies and control for the log of native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. The capped lines show the 95% confidence intervals.

4.3 IV Estimates

In this section we present the two-stage least squares (2SLS) estimates of the parameters of interest using the predicted in-migration rate as an instrument for the observed inflow.

In Table 10 we report the IV estimates including the first-stage coefficients and F-statistics. Columns (1), (3), (5) and (7) report OLS estimates while in columns (2), (4), (6) and (8) we present the IV/2SLS coefficients. All regressions are weighted by the total native population in 1991 and include time and state dummies, as well as demographic controls. Standard errors were calculated via bootstrapping to account for the fact that our instrumental variable was created with an estimated shock.

In a more simple framework one should expect that individuals would move into municipalities where they could find better work conditions, leading OLS estimates to be upward biased. We compare OLS and IV coefficients in Tables 10 and show that OLS estimates are biased toward zero. In Panel A we present the impacts on changes in the employment rate, for each sector. One percentage point increase in the number of migrants from the Semiarid region reduces formal employment by $0.3p.p.$, raises the proportion of informal native workers by $0.2p.p.$ and of self-employed by $0.1p.p.$.

Panel B shows a negative relation between the change in log earnings and the inflow of migrants and these estimates are much more precise. One percentage point increase in the number of incoming Semiarid migrants reduces overall earnings by 1.3%. Such decrease occurs both in the informal sector and among self-employed individuals (1.5%). The effect on the earnings of those employed in the formal sector is not significant, neither economically or statistically.

Table 11 compares OLS and IV estimates on labor force participation. Raising the inflow of migrants by 1p.p. increases the unemployment rate by $0.2p.p.$ and reduces the inactivity rate by $0.1p.p.$.

We already presented an illustration of our first-stage in Figure 1 and have shown that the predicted and observed numbers of migrants leaving the Semiarid and entering the destination municipalities are strongly correlated. As we can see in both tables the F-statistic is sufficiently high to avoid concerns about weakness of the instrument, even in light of the recent discussion brought by Lee et al. (2020) who show that a 5 percent test requires a F statistic of 104.7, much higher than the broadly accepted threshold of 10.

5 Robustness

Once we established the main results we need to assess the validity of our findings by performing some robustness checks.

The first issue we want to address is whether a shift in local labor demand may be confounding our identification. If that was the case, then we should expect that migrants from other regions outside the Semiarid would be attracted for 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 12 we show the coefficients from a regression of the in-migration rate of migrants from other regions on our instrument. Column (1) includes time and municipality fixed effects. In Column (2) we add the same set of controls from our main results and Column (3) also includes the same linear trends as before. There is no statistically significant impacts in any specification, which alleviates our concerns about a demand shock confounding the identification.

The second issue is that our strategy relies on the assumption that rainfall at origin municipalities affects destination labor markets only through internal 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. We believe that this problem was somehow addressed because we use only destinations outside the Semiarid region. Nevertheless, our data allow us to identify the industry where the individual is currently employed. Therefore, we are able to evaluate the impacts separately on natives working in agriculture, manufacturing or other industries. In Table 13 we report the coefficients from a regression of the predicted in-migration rate on changes in log earnings by industry. Panel A displays the effects on native workers employed in the agricultural sector, while in Panel B we focus on individuals holding a manufacturing job. In both cases, although the coefficients show a negative sign, the magnitudes are not very large and the estimates are not precise. In contrast, in Panel C we show that the effect on those employed in other industries is economically and statistically significant. Taken together, these estimates suggest that our identification remains valid.

6 Conclusion

In this paper we investigated the labor market impacts of weather-induced internal migration in Brazil. We used a shift-share instrument 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.

We find that increasing in-migration rate by $1p.p.$ reduces formal employment by $0.3p.p.$ and raises the proportion of informal workers by $0.2p.p.$. We also find a reduction of 1.4% for those natives in the informal sector or self-employed, but no effect for those in formal jobs. For these workers, the adjustment occurs at non-wage benefits. All these effects are stronger for low educated individuals, which are more likely to be substituted by migrants from the Semiarid region.

References

Adao, R., M. Kolesár, and E. Morales (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics 134*(4), 1949–2010.

Bastos, P., M. Busso, and S. Miller (2013). Adapting to climate change: Long-term effects of drought on local labor markets.

Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-experimental shift-share research designs. Technical report, National Bureau of Economic Research.

Boustan, L. P., P. V. Fishback, and S. Kantor (2010). The effect of internal migration on local labor markets: American cities during the great depression. *Journal of Labor Economics 28*(4), 719–746.

Carabetta, J., R. H. M. Pereira, and C. N. Gocalves (2020). geobr: Loads shapefiles of official spatial data sets of brazil. python package version 0.1.5. *The Python Package Index (PyPI)*.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

Card, D. (2009, May). Immigration and inequality. *American Economic Review 99*(2), 1–21.

Cattaneo, C., C. V. Fiorio, and G. Peri (2015). What happens to the careers of european workers when immigrants "take their jobs"? *Journal of Human Resources 50*(3), 655–693.

Cavalcanti, F. (2018). Voters sometimes provide the wrong incentives. the lesson of the brazilian drought industry.

Chein, F. and J. Assunção (2008). Climate change, agricultural productivity and poverty. In *XXX Meeting of the Brazilian Econometric Society*.

Correia, S. (2016). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Technical report. Working Paper.

Foged, M. and G. Peri (2016). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics 8*(2), 1–34.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020, August). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Gray, C. and V. Mueller (2012). Drought and population mobility in rural ethiopia. *World development 40*(1), 134–145.

Gröger, A. and Y. Zylberberg (2016). Internal labor migration as a shock coping strategy: Evidence from a typhoon. *American Economic Journal: Applied Economics 8*(2), 123–53.

Harris, I., T. J. Osborn, P. Jones, and D. Lister (2020). Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. *Scientific data 7*(1), 1–18.

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

Imbert, C. and J. Papp (2020). Short-term migration, rural public works, and urban labor markets: Evidence from india. *Journal of the European Economic Association 18*(2), 927–963.

Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg (2016). Migrants and firms: Evidence from china.

Jaeger, D. A., J. Ruist, and J. Stuhler (2018). Shift-share instruments and the impact of immigration.

Kleemans, M. and J. Magruder (2017). Labour market responses to immigration: Evidence from internal migration driven by weather shocks. *The Economic Journal*.

Kubik, Z. and M. Maurel (2016). Weather shocks, agricultural production and migration: Evidence from tanzania. *The Journal of Development Studies 52*(5), 665–680.

Lee, D. S., J. McCrary, M. J. Moreira, and J. Porter (2020). Valid t-ratio inference for iv. Papers, arXiv.org.

Mahajan, P. and D. Yang (2020). Taken by storm: Hurricanes, migrant networks, and us immigration. *American Economic Journal: Applied Economics 12*(2), 250–77.

Ottaviano, G. I. and G. Peri (2012). Rethinking the effect of immigration on wages. *Journal of the European economic association 10*(1), 152–197.

Ottaviano, G. I., G. Peri, and G. C. Wright (2013). Immigration, offshoring, and american jobs. *American Economic Review 103*(5), 1925–59.

Peri, G. and A. Sasahara (2019). The impact of global warming on rural-urban migrations: evidence from global big data. Technical report, National Bureau of Economic Research.

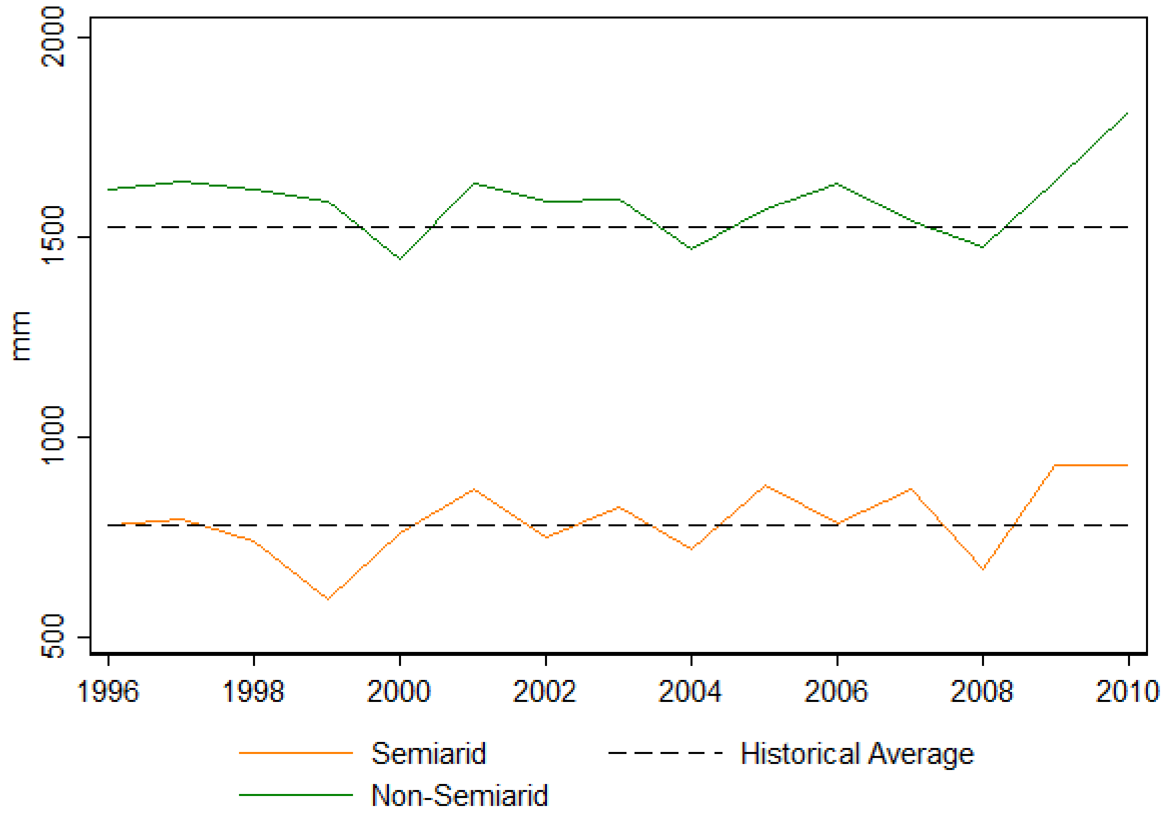
Rocha, R. and R. R. Soares (2015). Water scarcity and birth outcomes in the brazilian semiarid. *Journal of Development Economics 112*, 72–91.

Skoufias, E., K. Vinha, and H. V. Conroy (2011). *The impacts of climate variability on welfare in rural Mexico*. The World Bank.

Xavier, A. C., C. W. King, and B. R. Scanlon (2016). Daily gridded meteorological variables in brazil (1980–2013). *International Journal of Climatology 36*(6), 2644–2659.

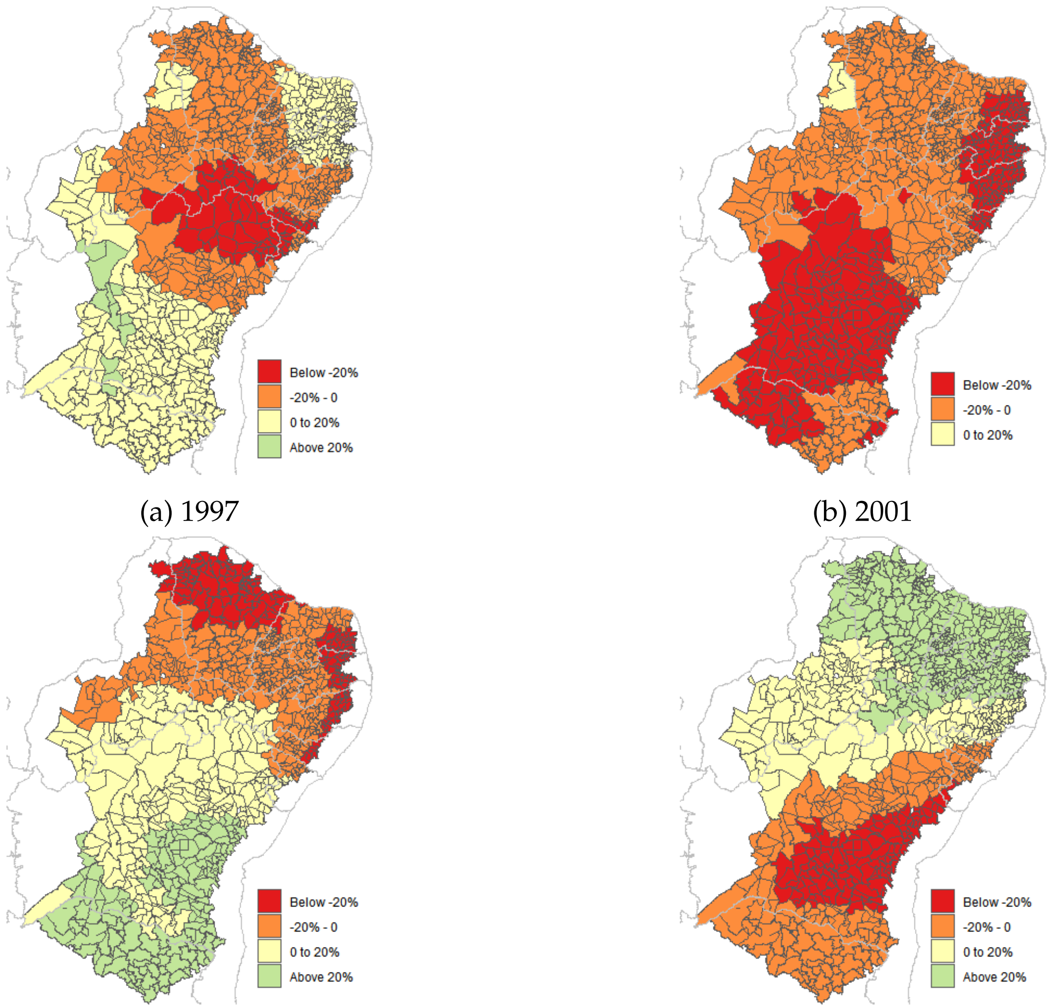
Figures and tables

Figure 9: **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).

Figure 10: 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).

Table 1: **Summary statistics: Weather and migration data**

Table 2: Municipality characteristics

	Mean	Std. Dev.	Min	Max	Obs
Annual Rainfall	791.77	213.94	266.31	2,052.87	14,400
Historical Rainfall	779.62	163.06	355.00	1,431.38	14,400
Annual Temperature	25.55	1.39	21.42	28.93	14,400
Historical Temperature	25.20	1.48	21.22	27.93	14,400
Out-migration	214.16	323.66	0.00	5,773.00	14,400
Predicted out-migration	211.66	284.16	12.66	4,551.83	14,400
Out-migration rate (p.p.)	1.05	0.62	0.00	7.22	14,400
Predicted out-migration rate (p.p.)	1.05	0.03	0.99	1.14	14,400
Population	21,260.91	30,092.29	1,265.00	480,949.00	14,400
Panel B: Non-Semiarid region	Mean	Std. Dev.	Min	Max	Obs
In-migration	146.69	896.95	0.00	25,423.00	8,190
Predicted in-migration	135.47	929.32	0.00	21,543.50	8,190
In-migration rate (p.p.)	0.27	0.90	0.00	24.30	8,190
Predicted in-migration rate (p.p.)	0.20	0.41	0.00	4.77	8,190
Native population	56,382.74	248,912.77	312.42	5,113,798.35	8,190

Notes: Rainfall is measured in mm. Temperature is measured in degrees Celsius. Population observed in the previous Census.

Table 3: **Summary statistics: Native individuals in destination municipalities**

	Mean	Std. Dev.	Min	Max	Obs
	Individual Characteristics				
	Mean	Std. Dev.	Min	Max	Obs
Female	51.08	3.64	0	72.72	8,190
Black	6.23	5.98	0	53.85	8,190
Age	37.45	1.96	30.15	55	8,190
Schooling	6.58	1.78	0	13.52	8,190
Low educated	65.32	15.74	4.71	100	8,190
	Employment				
	Mean	Std. Dev.	Min	Max	Obs
Overall	62.72	7.95	10	100	8,190
Formal sector	31.34	11.85	0	100	8,190
Informal sector	15.84	6.49	0	64.7	8,190
Self-employed	15.54	6.50	0	71.93	8,190
Unemployed	13.05	7.73	0	80	8,190
Outlabor	24.23	7.08	0	58.14	8,190
	Earnings				
	Mean	Std. Dev.	Min	Max	Obs
Overall	637.87	349.00	60.88	3,582.08	8,190
Formal sector	788.18	439.50	58.67	15,167.10	8,174
Informal sector	382.31	238.34	20	6,000.00	8,172
Self-employed	600.83	385.11	20	6,384.64	8,155
	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 municipality-year cell. Low-educated individuals are those with incomplete elementary schooling. Earnings are measured in R$ of 2012. Non-wage benefits are calculated only for native workers employed in the formal sector.

Table 4: **Out-migration induced by weather shocks**

	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall$_{t-1}$	-0.099***	-0.086***	-0.087***	-0.086***	-0.088***	-0.094***
	(0.028)	(0.030)	(0.030)	(0.030)	(0.030)	(0.031)
Rainfall$_{t-2}$			0.042	0.053		
			(0.032)	(0.033)		
Rainfall$_{t-3}$				0.056*		
				(0.031)		
Rainfall$_{t}$					-0.041	
					(0.032)	
Rainfall$_{t+1}$						-0.018
						(0.041)
Observations	14,400	14,400	14,400	14,400	14,400	14,400
Municipalities	960	960	960	960	960	960
R-Squared	0.461	0.472	0.472	0.473	0.473	0.472
F Stat	8.372	5.987	3.134	2.291	4.362	3.049
Time dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Temperature shocks	✓	✓	✓	✓	✓	✓
Covariates trends		✓	✓	✓	✓	✓

Notes: Each observation is a municipality-year. 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 historical average. All specifications include controls for temperature shocks, municipality and year fixed effects. Columns (2)-(6) also control for the log of population in the previous census and include interactions between time dummies and 1991 municipality-level characteristics (age and the share of high school and college educated individuals). Standard errors are clustered at the grid level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 5: **Effects of predicted in-migration on employment**

	(1)	(2)	(3)	(4)
	A. Change in employment rate			
Predicted inflow	-0.014	-0.033	-0.030	-0.017
	(0.091)	(0.084)	(0.084)	(0.087)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
	B. Change in formal employment rate			
Predicted inflow	-0.316***	-0.323***	-0.305***	-0.289***
	(0.104)	(0.088)	(0.087)	(0.096)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
	C. Change in informal employment rate			
Predicted inflow	0.160**	0.181**	0.166**	0.153**
	(0.079)	(0.075)	(0.075)	(0.076)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
	D. Change in self-employment rate			
Predicted inflow	0.143**	0.109*	0.108*	0.119
	(0.064)	(0.066)	(0.065)	(0.077)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in employment rate and 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. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Standard errors clustered at the municipality level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 6: **Effects of predicted in-migration on earnings**

	(1)	(2)	(3)	(4)
	A. Change in log earnings			
Predicted inflow	-1.655***	-1.323***	-1.182***	-1.265***
	(0.576)	(0.438)	(0.436)	(0.411)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
	B. Change in log earnings, formal sector			
Predicted inflow	-0.848	-0.300	-0.188	-0.462
	(0.674)	(0.488)	(0.489)	(0.444)
Observations	8,162	8,162	8,162	8,162
Municipalities	684	684	684	684
	C. Change in log earnings, informal sector			
Predicted inflow	-1.929***	-1.517***	-1.379***	-1.490***
	(0.607)	(0.439)	(0.441)	(0.390)
Observations	8,156	8,156	8,156	8,156
Municipalities	684	684	684	684
	D. Change in log earnings, self-employed			
Predicted inflow	-1.557***	-1.516***	-1.346***	-1.343***
	(0.450)	(0.391)	(0.383)	(0.405)
Observations	8,133	8,133	8,133	8,133
Municipalities	683	683	683	683
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in log earnings and 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. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Standard errors clustered at the municipality level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 7: **Effects of predicted in-migration on labor force participation**

	(1)	(2)	(3)	(4)
	A. Change in unemployment rate			
Predicted inflow	0.177**	0.161**	0.153**	0.156**
	(0.084)	(0.063)	(0.065)	(0.064)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
	B. Change in inactivity rate			
Predicted inflow	-0.163**	-0.128**	-0.123*	-0.139**
	(0.079)	(0.064)	(0.064)	(0.067)
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in unemployment and inactivity rates and 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. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Standard errors clustered at the municipality level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 8: **Effects of predicted in-migration on labor market outcomes, by condition in the household**

	(1) Employment	(2) Formal	(3) Informal	(4) Self-Employment	(5) Unemployment	(6) Inactivity
			A. Head			
Predicted inflow	-0.067	-0.220***	0.088**	0.065	0.009	0.021
	(0.054)	(0.067)	(0.043)	(0.041)	(0.032)	(0.035)
Observations	8,190	8,190	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684	684	684
			B. Non-head			
Predicted inflow	0.036	-0.085*	0.079**	0.043	0.144***	-0.144**
	(0.064)	(0.050)	(0.040)	(0.033)	(0.043)	(0.059)
Observations	8,190	8,190	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684	684	684
Demographics	✓	✓	✓	✓	✓	✓
Time dummies	✓	✓	✓	✓	✓	✓

Notes: This table shows the relationship between changes in employment (by sector), unemployment and inactivity rates and 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. Each observation is a municipality-year cell. 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 control for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; a cubic polynomial of age, include time dummies and are weighted by the 1991 native population. Standard errors clustered at the municipality level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 9: **Effects of predicted in-migration on fringe benefits**

	(1)	(2)	(3)	(4)
	A. Food			
Predicted inflow	-0.342	-0.472	-0.509*	-0.701**
	(0.313)	(0.299)	(0.300)	(0.301)
Observations	8,147	8,147	8,147	8,147
Municipalities	684	684	684	684
	B. Transport			
Predicted inflow	-0.540*	-0.655***	-0.698***	-0.542*
	(0.306)	(0.228)	(0.226)	(0.278)
Observations	8,147	8,147	8,147	8,147
Municipalities	684	684	684	684
	C. Health			
Predicted inflow	-0.443**	-0.546**	-0.562**	-0.226
	(0.212)	(0.239)	(0.248)	(0.218)
Observations	8,147	8,147	8,147	8,147
Municipalities	684	684	684	684
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in the proportions of working-age native population, employed in the formal sector, who receive non-wage benefits and 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. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Standard errors clustered at the municipality level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 10: **Impacts of internal migration on labor markets: IV estimates**

	Overall		Formal		Informal		Self-employment	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Change in employment rates							
Migrant inflow	-0.114	-0.033	-0.218***	-0.330***	0.087	0.180**	0.017	0.117*
	(0.078)	(0.095)	(0.075)	(0.093)	(0.063)	(0.086)	(0.056)	(0.071)
First-stage coefficient		.92		.92		.92		.92
First-stage F-statistic		220.6		220.6		220.6		220.6
Observations	8,190	8,190	8,190	8,190	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684	684	684	684	684
	B. Change in log earnings							
Migrant inflow	-0.695**	-1.281***	-0.265	-0.203	-1.147***	-1.494***	-0.693	-1.459***
	(0.295)	(0.465)	(0.309)	(0.527)	(0.392)	(0.485)	(0.465)	(0.441)
First-stage coefficient		.92		.92		.92		.92
First-stage F-statistic		220.6		220.2		220.8		220.4
Observations	8,190	8,190	8,162	8,162	8,156	8,156	8,133	8,133
Municipalities	684	684	684	684	684	684	683	683
Time dummies	✓	✓	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents OLS and IV estimates of the relationship between the number of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991, and labor market outcomes. Columns (1), (3), (5) and (7) show OLS estimates and columns (2),(4),(6) and (8) present the IV coefficients. In Panel A the dependent variables are the changes in employment rates while in Panel B are the log earnings, for each sector. All specifications are weighted by the total native population in 1991, include time dummies and control for the log of total native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Bootstrapped standard errors in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 11: **Impacts of internal migration on labor force participation: IV estimates**

	Unemployment		Inactivity	
	OLS	IV/2SLS	OLS	IV/2SLS
	(1)	(2)	(3)	(4)
Migrant inflow	0.173**	0.166**	-0.059	-0.133*
	(0.072)	(0.072)	(0.067)	(0.075)
First-stage coefficient		.92		.92
First-stage F-statistic		220.6		220.6
Observations	8,190	8,190	8,190	8,190
Municipalities	684	684	684	684
Time dummies	✓	✓	✓	✓
Demographics	✓	✓	✓	✓

Notes: This table presents OLS and IV estimates of the relationship between changes in unemployment and inactivity rates and the number of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991. Columns (1) and (3) show OLS estimates and columns (2) and (4) present the IV coefficients. All specifications are weighted by the total native population in 1991, include time dummies and control for the log of total native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Bootstrapped standard errors in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 12: **Correlation between predicted migration from the Semiarid and other regions**

	(1)	(2)	(3)
	Migrants from other regions		
Predicted inflow	0.217	0.006	0.198
	(2.854)	(2.823)	(2.846)
Observations	8190	8190	8190
Municipalities	684	684	684
Time dummies	✓	✓	✓
Municipality dummies	✓	✓	✓
Demographics		✓	✓
Covariates Trends			✓

Notes: This table shows the relationship between the predicted number of migrants from the Semiarid region in each destination municipality and the inflow of migrants coming from other regions, both measured as a fraction of the native working-age population in 1991. All specifications are weighted by total native population in 1991 and include municipality and year fixed effects. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. In column (3) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). Standard errors clustered at the municipality level in parenthesis.*** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 13: **Effects of predicted migration on changes in log earnings, by industry**

	(1)	(2)	(3)	(4)
	A. Agriculture			
Predicted inflow	-2.476	-0.709	-0.693	-0.391
	(2.222)	(1.863)	(1.821)	(1.866)
Observations	6,251	6,251	6,251	6,251
Municipalities	661	661	661	661
	B. Manufacturing			
Predicted inflow	-1.003*	-0.866	-0.667	-0.467
	(0.584)	(0.578)	(0.578)	(0.579)
Observations	7,902	7,902	7,902	7,902
Municipalities	683	683	683	683
	C. Other industries			
Predicted inflow	-1.319**	-0.971**	-0.830**	-1.042***
	(0.570)	(0.386)	(0.386)	(0.357)
Observations	8,175	8,175	8,175	8,175
Municipalities	684	684	684	684
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

*Notes:*This table shows the relationship between changes in log earnings by industry and 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. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Standard errors clustered at the municipality level in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Appendix A Shift-share instrument

In this section we present the details about how we constructed our instrument using two sources of exogenous variation.

A.1 Weather data

Our main data source for weather shocks comes from the CRUTS v4, a gridded dataset produced by the Climatic Research Unit at the University of East Anglia (Harris et al., 2020). This dataset provides information on monthly precipitation and temperature covering the whole globe[13], from 1901 to 2018. The grid resolutions is 0.25° $\times$ 0.25° (around 56km^2) and is created by interpolation from ground-based weather stations around the world.

We use the R package 'geobr' (Carabetta et al., 2020) to download the shapefile of Brazilian municipalities and georreference the coordinates from each municipality's centroid and keep only municipalities that belong to the Semiarid region. Then, for each municipality, we find the grid's four points which are closest to it's centroid and calculate the average level of precipitation and temperature from this 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 in the Semiarid, from 1901 to 2018, which we aggregate in yearly measures. Precipitation is defined as the sum of monthly levels and temperature as the average. For each municipality we calculate the historical mean from both variables and take log of the levels and long term averages.

Finally, our weather shock variables are defined as

$$Rainfall_{ot} = \ln\left(\sum_{\tau \in \{GS\}} r_{o\tau t}\right) - \ln(\bar{r}_o) \tag{A1}$$

where $r_{o\tau t}$ is the rainfall in 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. Our main specification focus on the growing season, but results are very similar when we use the full year.[14]

[13] Except for the Antartica.

[14] The growing season in the Semiarid region occurs from November to April.

Table A1: **Out-migration induced by weather shocks (12 months)**

	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall$_{t-1}$	-0.126***	-0.120***	-0.118***	-0.128***	-0.121***	-0.124***
	(0.033)	(0.035)	(0.036)	(0.036)	(0.035)	(0.035)
Rainfall$_{t}$			-0.032	-0.055	-0.029	
			(0.041)	(0.042)	(0.040)	
Rainfall$_{t-2}$			0.053	0.069*		
			(0.040)	(0.041)		
Rainfall$_{t-3}$				0.006		
				(0.035)		
Rainfall$_{t+1}$						-0.040
						(0.042)
Observations	14,400	14,400	14,400	14,400	14,400	14,400
Municipalities	960	960	960	960	960	960
R-Squared	0.461	0.469	0.469	0.470	0.469	0.469
F Stat	8.033	7.232	3.626	3.491	4.398	4.687
Time dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Temperature shocks	✓	✓	✓	✓	✓	✓
Covariates trends		✓	✓	✓	✓	✓

Notes: Each observation is a municipality-year. 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 historical average. All specifications include controls for temperature shocks, municipality and year fixed effects. Columns (2)-(6) also control for the log of population in the previous census and include interactions between time dummies and 1991 municipality-level characteristics (age and the share of high school and college educated individuals). Standard errors are clustered at the grid level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

A.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 gather new data from Xavier et al. (2016), who provides a gridded dataset with daily averages of precipitation and potential evaporation, from 1980 to 2013, based on ground data from weather stations interpolated to create high-resolution grids (0.25° $\times$ 0.25°) across the Brazilian territory. They calculate potential evaporation using maximum and minimum temperatures, solar radiation, relative humidity and wind speed. We aggregate the daily precipitation and evaporation data into monthly measures and follow Cavalcanti (2018) to construct a measure of drought severity, the aridity index, as follows:

$$AI_{mt} = \frac{\sum_{\tau\in\{GS\}} PE_{m\tau t}}{\sum_{\tau\in\{GS\}} Pr_{m\tau t}} \tag{A2}$$

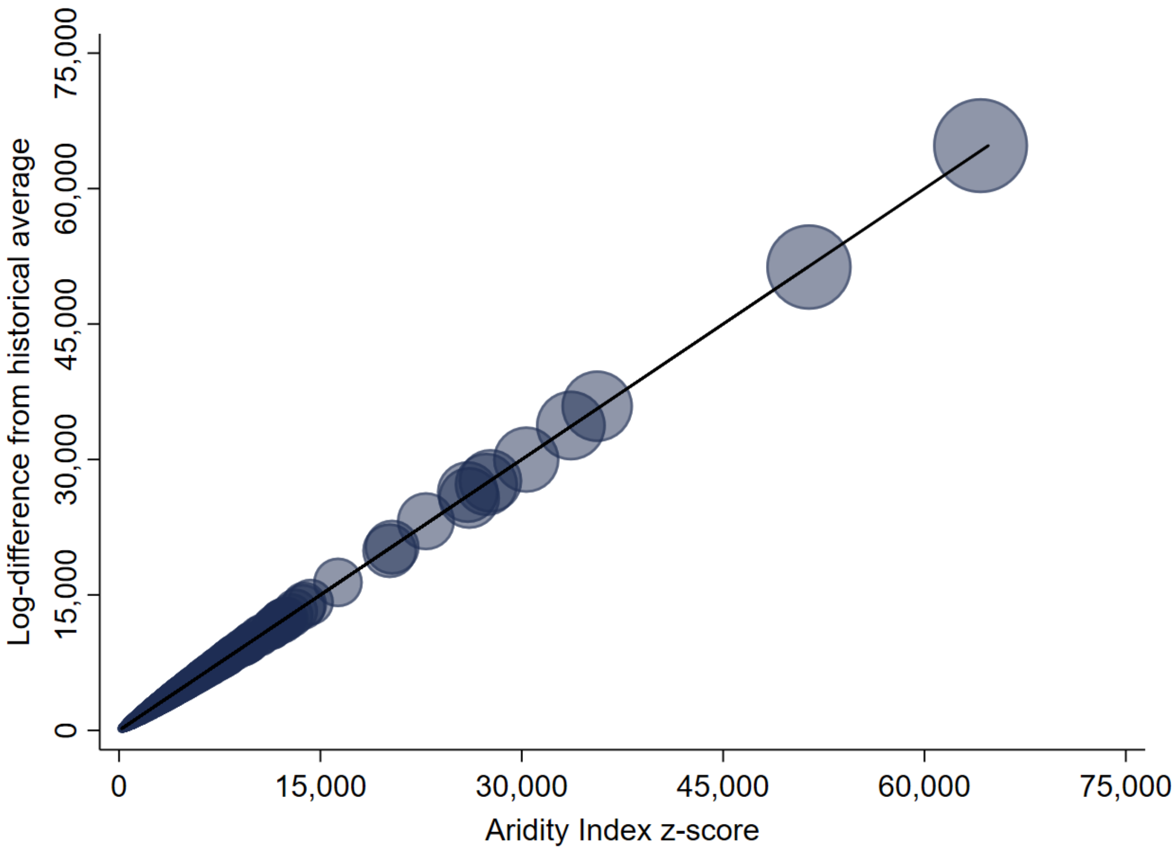
where $PE_{m\tau t}$ is the potential evaporation in the municipality m, at the month τ of the growing season(GS) in year t. Then we standardize this measure to simplify interpretation and calculate de aridity index z-score as

$$Z^{AI}_{mt} = \frac{\left(AI_{mt} - \overline{AI}\right)}{AI_{sd}} \tag{A3}$$

We show in Table A2 that this alternative measure is also strongly correlated with the out-migration rate. Including up to three lags and one lead does not affect the main coefficient, neither does the inclusion of controls. In Panel B we regress the out-migration rate on a categorical variable indicating the quartile of the Aridity Index z-score. Our estimates show that extreme events of drought increase out-migration even further.

More importantly, in Figure A1 we compare the predicted number of migrants using this alternative measure to the prediction we had before and show they are virtually identical, which means that our instrument is robust to measurement errors in water availability.

Figure A1: **Predicted out-migration using alternative measures of weather**



Notes: Y-axis shows the number of migrants predicted by rainfall and temperature log-differences from historical average. X-axis presents the number of migrants predicted by the aridity index z-score. Circle size represents the municipality's total population in 1991. *Data sources*: CRU TS v4 (Harris et al., 2020) and Xavier et al. (2016).

Table A2: **Out-migration induced by weather shocks: Aridity Index**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: *Continuous Z-score*						
Aridity Index$_t$	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Aridity Index$_{t-1}$		0.003***	0.003***	0.003***			
		(0.001)	(0.001)	(0.001)			
Aridity Index$_{t-2}$			0.001	0.001			
			(0.001)	(0.001)			
Aridity Index$_{t-3}$				0.000			
				(0.001)			
Aridity Index$_{t+1}$					0.001		
					(0.001)		
Constant	1.020***	0.995***	0.987***	0.985***	1.015***	-2.022**	
	(0.007)	(0.012)	(0.015)	(0.018)	(0.011)	(0.851)	
	Panel B: *Drought severity*						
Second quartile							0.028**
							(0.014)
Third quartile							0.010
							(0.016)
Fourth quartile							0.076***
							(0.019)
Constant							1.024***
							(0.011)
Observations	14,400	14,400	14,400	14,400	14,400	14,400	14,400
Municipalities	960	960	960	960	960	960	960
R-Squared	0.461	0.462	0.462	0.462	0.461	0.470	0.462
Time dummies	✓	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓	✓
Demographics						✓	

Notes: Each observation is a municipality-year. Dependent variable is the number of individuals who left the origin municipality divided by the total population in the 1991 Census. Aridity Index is measured as the municipality z-score of the ratio between evaporation and precipitation accumulated from November to April. All specifications include municipality and year fixed effects. Column (6) also control for the log of population in the previous census and include interactions between time dummies and 1991 municipality-level characteristics (age and the share of high school and college educated individuals). Drought severity measures are the quartiles of the Aridity Index z-score. Standard errors are clustered at the municipality level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

A.3 Out-migration prediction

We use the weather shocks to predict the number of migrants leaving the Semiarid municipalities and settling in other places across the country. We estimate the following two-way fixed effects regression:

$$m_{ot} = \alpha + \beta' Z_{ot-1} + \phi_o + \delta_t + \varepsilon_{ot} \tag{A4}$$

where m_{ot} is the migration rate in origin municipality o and year t, Z is a vector of rainfall and temperature shocks in the previous year, ϕ_o and δ_t are municipality and year fixed effects, respectively, and ε_{ot} is a random error term. Then we project the migration rate and multiply by the 1991 population to obtain the predicted number of migrants who left the Semiarid region because of weather shocks.

$$\widehat{M}_{ot} = \widehat{m}_{ot} \times P_o \tag{A5}$$

As shown in Figure 1a this procedure results in a good approximation of the observed migration.

From 1996 to 2010 over 1.5 million people left the Semiarid region to settle elsewhere. Figure A2 indicates the main origin municipalities, responsible for around 7% of the out-migrants in our sample.

A.4 The pre-existent migrant network (1991)

We use the 1991 Census' microdata to construct the predetermined network of migrants from the Semiarid region settling in the destination municipalities. Individuals were asked how many years they lived in the current municipality and what was the municipality where they lived before[15]. With this information we construct a yearly panel of migration for every pair of origin-destination municipalities in Brazil.

For each destination d we define the share of Semiarid's migrants from origin o as

$$s_{od} = \frac{m_{od}}{\sum\limits_{o} m_{od}} \tag{A6}$$

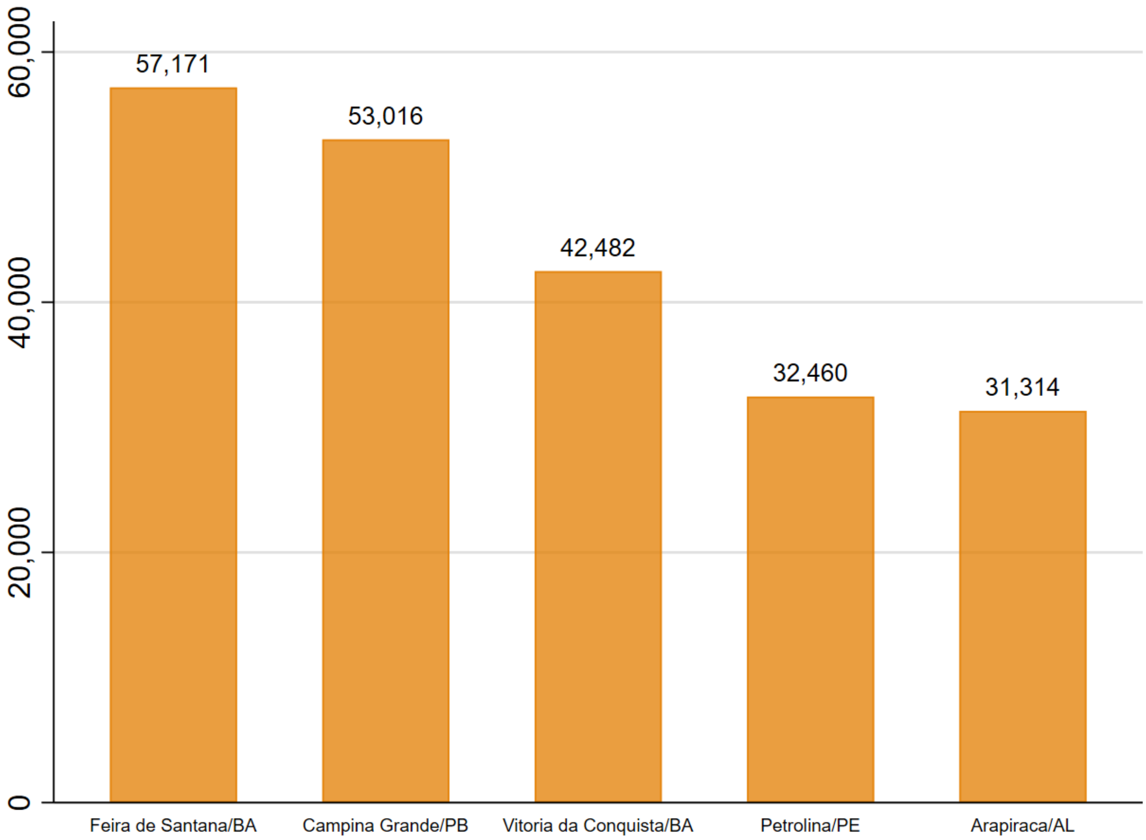
By construction, we have $\sum\limits_{o} s_{od} = 1$, meaning that every individual predicted to migrate will be allocated in some destination. The final step is to use these shares to allocate the predicted number of migrants from the Semiarid region into the destination municipalities

$$\widetilde{m_{dt}} = \sum_{o=1}^{O} \frac{s_{od} \times \widehat{M}_{ot}}{P_d} \tag{A7}$$

where $\widehat{M}_{ot}$ is obtained from equation A5 and P_d is the 1991 population in the destination municipality, which we use as normalization.

[15]Microdata only records the previous municipality for those individuals who moved during the decade covered by the 1991 Census. Therefore, we can only track individuals as far as 1982.

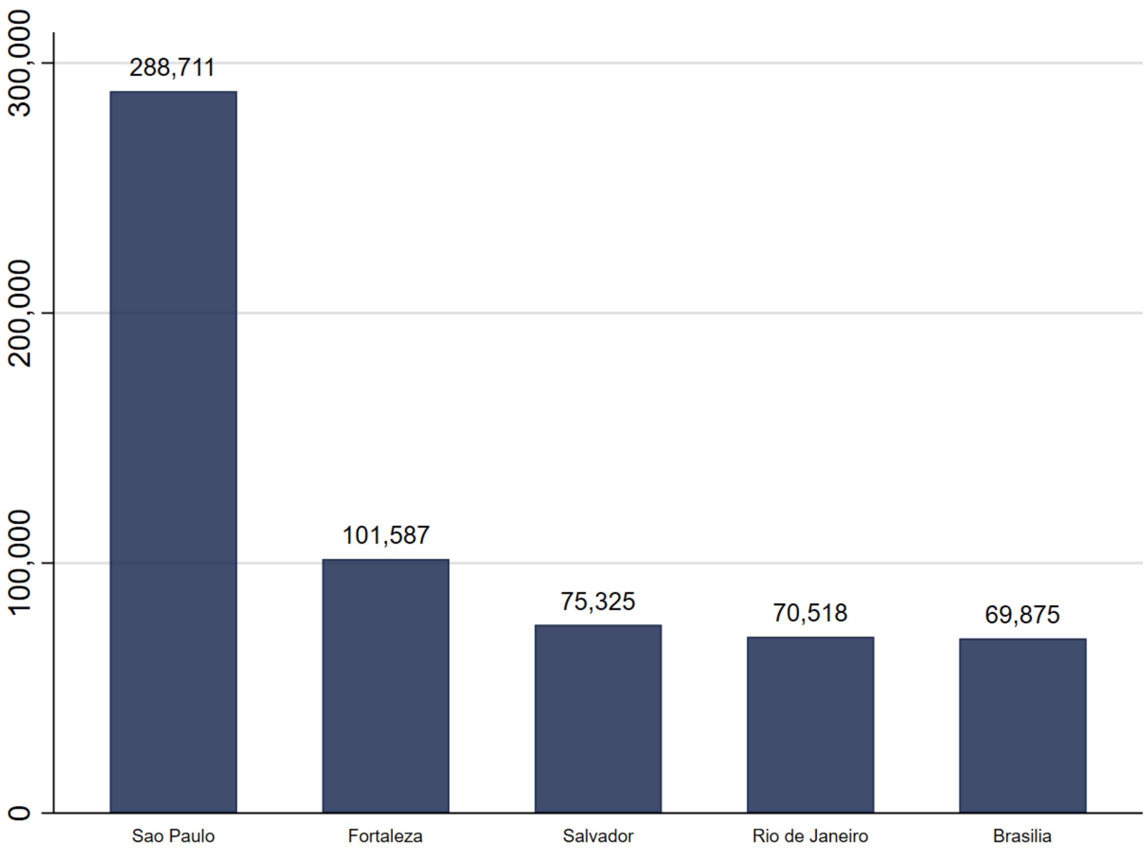
Figure A2: **Top 5 sending municipalities in the Semiarid region**



Notes: This figure shows the top five sending municipalities. *Data source*: Census microdata.

Figure 1b already shown that we obtained a good prediction of in-migration. We emphasize that destination areas are very concentrated. Figure A3 depicts the top five settling municipalities for Semiarid migrants during our period, which concentrate around 40% of the total migration in our sample. All of them are state capitals, but only two (Fortaleza and Salvador) are in the Northeast region.

Figure A3: **Top 5 receiving municipalities**



Notes: This figure shows the top five receiving municipalities. *Data source*: Census microdata.

Appendix B Inference correction

In this appendix we address the issue raised by Adao et al. (2019) who demonstrated that in the typical shift-share design standard errors may be underestimated because the regression residuals are correlated across units with similar shares.

B.1 Shock-level IV regressions

In this section we apply the method developed by Borusyak et al. (2018) to correct the standard errors in our estimates and show that our main results remain valid. Applying their procedure to our setting, we estimate the following IV regression

$$\bar{y}_{ot} = \alpha + \beta \bar{m}_{ot} + \bar{\varepsilon}_{ot} \tag{B1}$$

where $\bar{v}_{ot} = \frac{\sum\limits_{d} s_{od} v_{dt}}{\sum\limits_{d} s_{od}}$ is a share-weighted average of the outcomes y or the instrument m at the destination level and the regression is weighted by the average exposure to the shocks $s_o = \sum\limits_{d} s_{od}$. Put in a more simplistic way, in the traditional shift-share design we take a shock at the origin level and calculate a share-weighted average to use as instrumental variable at the destination. What Borusyak et al. (2018) propose is to take the destination-level outcomes and endogenous variables, calculate its weighted averages for each origin municipality and use the shocks as instruments to estimate the parameter β. They demonstrate that even if the shares are not exogenous their approach renders a consistent estimator. More importantly for this section, they also show that this procedure allow to obtain standard errors asymptotically valid in the framework of Adao et al. (2019).

We calculate share-weighted averages of our outcomes and instrument and estimate all the same specifications at the origin-level and present the results in Tables B1-B8.

Table B1: **Inference a la Borusyak et al. (2018): Reduced form estimates on employment**

	(1)	(2)	(3)	(4)
	A. Change in employment rate			
Predicted inflow	-0.014	-0.033	-0.030	-0.017
	(0.062)	(0.064)	(0.046)	(0.058)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Change in formal employment rate			
Predicted inflow	-0.316***	-0.323***	-0.305***	-0.289***
	(0.064)	(0.060)	(0.041)	(0.052)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	C. Change in informal employment rate			
Predicted inflow	0.160**	0.181**	0.166**	0.153**
	(0.079)	(0.075)	(0.075)	(0.076)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	D. Change in self-employment rate			
Predicted inflow	0.143***	0.109***	0.108***	0.119***
	(0.037)	(0.037)	(0.035)	(0.044)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in employment rate and 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. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table B2: **Inference a la Borusyak et al. (2018): Reduced form estimates on earnings**

	(1)	(2)	(3)	(4)
	A. Change in log earnings			
Predicted inflow	-1.655***	-1.323***	-1.182***	-1.265***
	(0.242)	(0.224)	(0.150)	(0.190)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Change in log earnings, formal sector			
Predicted inflow	-0.848***	-0.300*	-0.188	-0.462***
	(0.193)	(0.178)	(0.125)	(0.157)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	C. Change in log earnings, informal sector			
Predicted inflow	-1.929***	-1.516***	-1.378***	-1.489***
	(0.241)	(0.243)	(0.188)	(0.237)
Observations	8,156	8,156	8,156	8,156
Municipalities	684	684	684	684
	D. Change in log earnings, self-employed			
Predicted inflow	-1.557***	-1.516***	-1.345***	-1.342***
	(0.416)	(0.425)	(0.326)	(0.412)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in log earnings and 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. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table B3: **Inference a la Borusyak et al. (2018): Reduced form estimates on labor force participation**

	(1)	(2)	(3)	(4)
	A. Change in unemployment rate			
Predicted inflow	0.177***	0.161***	0.153***	0.156***
	(0.055)	(0.057)	(0.040)	(0.051)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Change in inactivity rate			
Predicted inflow	-0.163***	-0.128**	-0.123***	-0.139**
	(0.053)	(0.053)	(0.046)	(0.058)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in unemployment and inactivity rates and 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. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table B4: **Inference a la Borusyak et al. (2018): Reduced form estimates on non-wage benefits**

	(1)	(2)	(3)	(4)
	A. Food			
Predicted inflow	-0.342**	-0.472***	-0.508***	-0.700***
	(0.141)	(0.144)	(0.107)	(0.134)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Transport			
Predicted inflow	-0.540***	-0.655***	-0.698***	-0.541***
	(0.137)	(0.135)	(0.099)	(0.124)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	C. Health			
Predicted inflow	-0.443**	-0.546***	-0.561***	-0.226
	(0.176)	(0.184)	(0.125)	(0.157)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in the proportions of working-age native population, employed in the formal sector, who receive non-wage benefits and 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. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table B5: **Inference a la Borusyak et al. (2018): IV/2SLS estimates on employment**

	(1)	(2)	(3)	(4)
	A. Change in employment rate			
Migrant inflow	-0.015	-0.035	-0.033	-0.018
	(0.068)	(0.069)	(0.050)	(0.063)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Change in formal employment rate			
Migrant inflow	-0.347***	-0.350***	-0.330***	-0.310***
	(0.071)	(0.065)	(0.045)	(0.056)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	C. Change in informal employment rate			
Migrant inflow	0.175***	0.196***	0.180***	0.164***
	(0.048)	(0.049)	(0.035)	(0.044)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	D. Change in self-employment rate			
Migrant inflow	0.156***	0.119***	0.117***	0.128***
	(0.041)	(0.041)	(0.038)	(0.047)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in employment rate and the inflow of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table B6: **Effects of predicted in-migration on earnings**

	(1)	(2)	(3)	(4)
	A. Change in log earnings			
Migrant inflow	-1.815***	-1.433***	-1.281***	-1.359***
	(0.268)	(0.244)	(0.164)	(0.205)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Change in log earnings, formal sector			
Migrant inflow	-0.930***	-0.325*	-0.203	-0.496***
	(0.210)	(0.192)	(0.135)	(0.168)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	C. Change in log earnings, informal sector			
Migrant inflow	-2.116***	-1.642***	-1.494***	-1.599***
	(0.265)	(0.263)	(0.204)	(0.255)
Observations	8,156	8,156	8,156	8,156
Municipalities	684	684	684	684
	D. Change in log earnings, self-employed			
Migrant inflow	-1.708***	-1.642***	-1.458***	-1.441***
	(0.458)	(0.463)	(0.354)	(0.443)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in log earnings and the inflow of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table B7: **Effects of predicted in-migration on labor force participation**

	(1)	(2)	(3)	(4)
	A. Change in unemployment rate			
Migrant inflow	0.194***	0.174***	0.166***	0.168***
	(0.060)	(0.061)	(0.044)	(0.055)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Change in inactivity rate			
Migrant inflow	-0.179***	-0.139**	-0.133***	-0.150**
	(0.058)	(0.058)	(0.050)	(0.062)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in unemployment and inactivity rates and the inflow of migrants from the Semiarid region in each destination municipality, measured as a fraction of the native working-age population in 1991. Each observation is a municipality-year cell. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table B8: **Effects of predicted in-migration on non-wage benefits**

	(1)	(2)	(3)	(4)
	A. Food			
Migrant inflow	-0.375**	-0.511***	-0.551***	-0.751***
	(0.155)	(0.158)	(0.116)	(0.145)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	B. Transport			
Migrant inflow	-0.592***	-0.709***	-0.756***	-0.581***
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
	C. Health			
Migrant inflow	-0.592***	-0.709***	-0.756***	-0.581***
	(0.152)	(0.149)	(0.108)	(0.135)
Observations	11,460	11,460	11,460	11,460
Municipalities	955	955	955	955
Demographics		✓	✓	✓
Time dummies			✓	✓
Covariates trends				✓

Notes: This table shows the relationship between changes in the proportions of working-age native population, employed in the formal sector, who receive non-wage benefits and the inflow of migrants from the Semi-arid region in each destination municipality, measured as a fraction of the native working-age population in 1991. Column (2) controls for the log of the native population in the previous year and changes in the share of individuals with high school and college education; the shares of female, white and black individuals; and a cubic polynomial of age. Column (3) also include time dummies. In column (4) we add a linear trend of 1991 municipality-level characteristics (age and the shares of high school and college educated individuals). All regressions are weighted by the 1991 native population. Robust standard errors calculated by the procedure in Borusyak et al. (2018) in parentheses. *** Significant at 1%. ** Significant at 5%. * Significant at 10%.