Rural Migrants and Urban Informality: Evidence from Brazil

Clément Imbert Gabriel Ulyssea*

May 30, 2024

Abstract

This paper studies the long-run effects of rural-urban migration on Brazilian cities. Using a shift-share IV design, we show that immigration reduces informality and wages, has no effect on unemployment, and increases the number of formal firms and jobs over a decade. These results are in sharp contrast with the short-run, informality-increasing effects previously documented in the literature. To rationalize these surprising results, we develop and estimate a model of firm dynamics and informality that can quantitatively replicate the long-run IV results. Assuming sluggish formal wage adjustment in the transition between equilibria, immigration shocks increase firm and labor informality in the short-run, but they gradually decline to lower long-run levels, consistent with our IV results. A large share of new formal firms come from the informal sector, which serves as a "stepping-stone". However, the overall economic benefits of immigration are higher in a counterfactual with no informality.

Keywords: Informality, rural-urban migration, labor markets, firm dynamics

*Imbert: Sciences Po, clement.imbert@sciencespo.fr. Ulyssea: University College London, g.ulyssea@ucl.ac.uk. We thank Doug Gollin, Kirill Borusyak, Julieta Caunedo, Frederic Docquier, Nezih Guner, Jonas Hjort, Oriana Bandiera, Joan Monras, Giovanni Prarolo, Simon Quinn, Imran Rasul, Diego Restuccia, Selim Gulesci, Federico Rossi, Chris Woodruff, as well as participants of several seminars, conferences and workshops for insightful comments and discussions. We are grateful to David Crommelynck, Henrique Mota, Ana Beatriz Pousada, Viktor Veterinarov and Caio de Holanda for outstanding research assistance throughout different stages of this project. Ulyssea gratefully acknowledges the funding received from the European Research Council (ERC) Starting Grant under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 852744).

1 Introduction

Urban population in developing countries grew 12.5 percent between 2015-2020, and it is projected to grow 64.7 percent until 2050 (UNCTAD, 2021). Rural-urban migration flows account for a substantial fraction of this population growth (Jedwab et al., 2017), which are only likely to intensify due to climate change (Rigaud et al., 2018). Whether urban developing economies will be able to generate enough good jobs to accommodate this fast growing workforce is a fundamental question for economic development (The World Bank, 2013).

The available empirical evidence on the *short-run* effects of rural-urban migration provides a gloomy picture (El Badaoui et al., 2017; Kleemans and Magruder, 2018; Corbi et al., 2021). It largely confirms the traditional view that stems from the works of Harris and Todaro (1970) and Fields (1975) who argue that immigration leads to higher unemployment or informality in urban areas. More broadly, these results are consistent with the vast evidence that urban developing economies are characterized by labor market frictions (e.g. Abebe et al., 2021; Carranza et al., 2022; Donovan et al., 2023), low firm growth (e.g Hsieh and Klenow, 2014; Quinn and Woodruff, 2019), high informality (Ulyssea, 2020), and unemployment, especially among young workers (Alfonsi et al., 2020; Bandiera et al., 2021).

This paper investigates the long-run economic effects of rural-urban migration on local urban economies in Brazil. We start by combining detailed data on workers and firms with a shift-share IV design to identify the causal effects of immigration at urban destinations over a decade. We find evidence that immigration *reduces* informality with no effect on unemployment, reduces formal and informal wages, and increases the number of formal firms and jobs. We show that these surprising effects are due to the long time horizon of our analysis: as we move to the usual year-on-year specification, we find the opposite, informality-increasing effects documented in the literature. To further understand these results, we develop a new model of firm dynamics that features the intensive and extensive margins of informality. The model counterfactuals quantitatively replicate the long-run IV results. If we impose downward wage rigidity in the formal sector during the transition between steady states, informality increases in the short run, but gradually decreases as formal wages adjust. It eventually becomes lower than the baseline, which is consistent with our short- and long-run reduced-form estimates. Our counterfactuals also show that informality serves as "stepping-stone": a large fraction of formal firm creation following immigration is explained by informal firms formalizing. On aggregate, however, the presence of a large informal sector dampens the economic dividends from immigration.

In the first part of the paper, we combine matched employer-employee data on the universe of formal firms and workers in Brazil between 1995 and 2018, with individual level data from three waves of the Demographic Census (1991, 2000 and 2010). We use these data and a shift-share IV design to identify the causal effects of immigration on local labor market outcomes and formal firm dynamics at destination. We focus on decadal changes in outcomes between 2000 and 2010. To construct our shift-share instrument, we use detailed information on previous migration patterns across municipalities to construct the "shares". We combine those with granular data on agricultural suitability and land use, and with international price shocks to construct push-shocks at origin, the "shifts".

We find that internal immigration in Brazil between 2000 and 2010 had no effect on unemployment, and *increased* the share of formal employment: an increase in the immigration rate of one percentage point increases the share of workers in formal wage employment by 0.4 percentage points (a 1.7 percent increase). This effect is entirely driven by a shift of workers from informal to formal jobs, with no change in wage employment overall, unemployment, or self-employment. Wages in the formal and informal sectors fall by 2.2 and 1.9 percent, respectively. We find similar results if we look at migrants and non-migrants or high- and low-skilled workers separately. The latter is consistent with the fact that immigration increases the share of men and young in the workforce at destination, but does not change its composition by skill.

We then turn to the effects of internal immigration on formal firms' outcomes. We find that an increase in the immigration rate by one percentage point leads to a 2.3 percent increase in the number of firms and a 2.2 percent increase in the number of formal jobs. There is also sharp increase in the number of firms entering and exiting every year (7.5 and 5.6 percent, respectively). There are no effects on average firm size nor firm growth, and we show that these new firms are not being created by the migrants themselves. When we examine effects over time, we show that they remain constant or slightly increase throughout the 2010's (until 2018).

We carry out a number of robustness checks to alleviate concerns about identification. Reassuringly, we do not find evidence of differential pre-trends, and we assess various potential threats to the exclusion restriction. We show that our results are robust to controlling for past trends in outcomes, for GDP and population in 2000, for the share of manufacturing and construction in 2000, and for immigration in the previous decade (between 1995 and 2000). As we run our regressions in first difference, this effectively allows for differential trends along these dimensions. Another concern may be that price shocks in rural areas affect firms at destination through other channels than migration. We check that our results are robust to excluding agricultural firms or firms that process agricultural products, to controlling for local price shocks and in neighboring municipalities that could depress the demand for firms' products, and to controlling for exposure to capital reallocation from rural areas via the network of banks' branches across municipalities. Following Borusyak et al. (2022), we run regressions at the level of the municipality of origin and check that results are the same, and robust to controlling for lagged agricultural price shocks.[1]

We present three additional pieces of evidence that shed light on the effects of immigration on local economic development. First, the vast economic heterogeneity across regions and municipalities in Brazil allows us to investigate heterogeneous effects across terciles of GDP per capita at baseline (in 2000): average GDP p.c. in the top tercile is more than twice as large the average in the middle tercile and 4.5 times larger than the bottom one. We show that the effects of immigration on workers and firms we document are driven by the bottom and middle terciles, and not by the largest and richest cities like São Paulo and Rio de Janeiro. This suggests that our results may be relevant for other low-middle income urban settings.

We then show that the composition of formal firms and jobs shifts towards firms in services, retail and construction, while moving away from larger and manufacturing firms. These effects are intuitive, since construction, retail and services are more labor intensive and require lower and less specific skills than manufacturing. Non-tradables are also more likely to benefit from the positive demand shock brought about higher presence of immigrants. These findings resonate with Gollin et al. (2016)'s argument that the expected relationship between urbanization and industrialization (e.g. Gollin et al., 2002; Duarte and Restuccia, 2010) is absent in many developing countries.

Finally, we turn to the effects on the composition of formal firms and jobs by firm size and productivity, which we proxy by residual firm-level average wage. We show that immigration leads to a reallocation of firms and jobs towards the bottom of the

[1]This also eliminates potential concerns about inference in our baseline specification, as this specification is asymptotically equivalent to the the approach proposed by Adão et al. (2019).

firm size distribution and the bottom-middle of the firm productivity distribution, away from large and high-productivity firms. These effects could be explained by the formalization of existing informal firms, which are on average less productive than formal firms. We cannot directly test this in the data, which only includes formal firms, but we do so using the structural model.

The results discussed so far stand in stark contrast with those from the previous literature, which usually estimates year-on-year effects, while we consider changes over a decade. The differences in findings could be due to the presence of labor market frictions that constrain formal (but not informal) labor demand expansion in the short run, but are at least partially alleviated in the longer run. In a final empirical exercise, we use the National Household Survey (PNAD) to revisit the year-on-year specification typically used in the literature to directly assess this conjecture. One challenge is that we must rely on a higher frequency push shock to construct the shift-share instrument, as the price shocks are not predictive of migration in the immediate short run. We thus use an alternative identification strategy that employs droughts as push shocks (see Corbi et al., 2021; Albert et al., 2021), combined with pre-existing migration networks to instrument for immigration.

We first replicate our main, long difference specification using the drought shocks and find very similar effects on both workers and firms. This is an important result in its own right, because it sheds light on how developing country cities will cope with the expected surge in climate-driven migration (Rigaud et al., 2018). It also strengthens the credibility of our main empirical strategy, since the results are independent of the push shocks we use.[2] Moving to the year-on-year specification, we find the patterns documented in the previous literature: rural to urban migration leads to an increase in informality driven by a decline in formal wage employment, with no effect on wages.

In the second part of our analysis, we develop a new model of firm dynamics and informality that is able to rationalize these surprising findings. It extends the canonical model of firm dynamics of Hopenhayn (1992) in two ways. First, we include the two margins of informality considered in Ulyssea (2018): (i) whether to register their business or not, the extensive margin; and (ii) whether firms that are formally registered hire their workers with or without a formal contract, the intensive margin.

[2] The drought and the price shocks are completely independent across origins, the correlation between the two shift-shares at destination is low, and the composition of compliers is different: climatic migrants' demographics are not significantly different from natives'.

Modeling the intensive margin is key, because it links formal firms and migrants even if they only take up informal jobs when arriving in cities.

Second, we allow for heterogeneous growth profiles (similar in spirit to Sterk et al., 2021), which implies that informality can have opposing effects in the economy. On the one hand, informality can act as a "stepping-stone" for potentially high-growth firms. More broadly, it introduces greater *de facto* flexibility that could lead to higher firm entry and growth, which would otherwise not happen due to other frictions, such as burdensome regulations. However, it also allows less productive firms to survive and compete with more productive formal firms. This weakens the natural selection process in the economy, and shifts resources away from high-productivity firms, hindering their ability to grow. The net effect of these forces and their interaction with immigration shocks is an empirical question.

We calibrate the model and use it to perform counterfactual simulations of the effects of a once-and-for-all 10 percent labor supply increase, which corresponds to the 80th percentile of the migration shocks we see in the data. The counterfactual results confirm the main findings from the instrumental variable analysis: the supply shock leads to a small reduction in the share of informal workers, a reduction in wages and a sizable increase in the number of formal firms. Importantly, the *magnitude* of the effects predicted by the model are quite close to those we estimate using the IV method for an immigration shock of the same size.

Furthermore, the counterfactual results show that 40 percent of the increase in the number of formal firms comes from higher formalization of informal firms throughout their life cycle. This highlights the importance of accounting for firm dynamics, as a static model would not capture this mechanism and could therefore substantially underestimate the formalization effects. Output increases by less than the increase in labor supply, so there is a decline in output per worker. This is largely explained by compositional effects: the share of formal firms and workers at the bottom quartiles of the firm productivity distribution increases, which is consistent with the IV results.

In a second counterfactual, we examine the transition dynamics between steady states assuming that there is downward wage rigidity in the formal sector. This is a friction that constraints formal – but not informal – labor demand expansion in the short-run, and there is a large literature suggesting that it is pervasive in developed and developing countries (e.g. Schmitt-Grohé and Uribe, 2016; Grigsby et al., 2021). This is also broadly consistent with the fact that formal jobs are more stable, employment contracts are legally binding in the formal sector, and nominal wage cuts are not allowed by law in Brazil. We show that immediately after the labor supply shock, both labor and firm informality increase substantially. The latter starts declining almost immediately after the shock, while the share of informal labor hovers around its initial level for a few years before converging to its lower level in the new steady state. This is due to the intensive margin of informality, as the new formal firms tend to be small and hire a substantial fraction of their labor force informally.

Finally, our third counterfactual combines the labor supply shock with higher intensity of enforcement on informal firms (the extensive margin of informality), nearly shutting down the informal sector. We do not see this as an actual policy counterfactual but rather as a thought experiment that sheds light on the role of the informal sector in absorbing this shock. Our results show that migration has stronger positive effects on output and firm productivity, at the cost of a substantial displacement of the least productive informal firms. This suggests that despite its role as a "stepping-stone", the informal sector overall dampens the economic benefits from immigration by allowing the least productive firms in the economy to survive.

Our paper is relevant to four strands of the literature. First, we bring new evidence on the effects of rural-urban migration on urban labor markets in developing countries. The literature has grown around the idea of wage rigidity in the formal sector (Harris and Todaro, 1970; Fields, 1975), which implies that rural-urban migration only adds workers to the pool of unemployed or informal workers queuing for a limited number of formal jobs. Consistent with this view, the empirical literature tends to find negative immediate employment effects of immigration for resident workers, especially in the formal sector (Kleemans and Magruder, 2018; El Badaoui et al., 2017).[3] In the same context as our paper, Corbi et al. (2021) show that in the year following an inflow of migrants due to droughts in the semi-arid regions of Brazil, the share of workers in the informal sector increases. We have similar findings than theirs when we estimate the short-run effects of drought-induced migration in our sample, which covers all of Brazil. In contrast with these short-run results, we document positive effects on formal employment in the longer run and negative effects on both informal and formal sector wages. Using a model-based counterfactual, we show that sluggish wage adjustment

[3]These effects are larger than in the literature on international migration in developed countries (Card and DiNardo, 2000; Card, 2001; Borjas, 2003; Clemens and Hunt, 2019). In the context of international migration, there is also a direct link between undocumented migration and informality, which is not relevant in our context (Bahar et al., 2021; Elias et al., 2018).

in the formal sector can explain these differences between short- and long-run effects.

Second, our findings provide new insights to the broader discussion on job creation in developing countries. Most of the literature takes labor demand in the formal sector as given, and documents labor market frictions that make it hard for job-seekers to enter the formal sector (Franklin, 2018; Alfonsi et al., 2020; Donovan et al., 2023; Abebe et al., 2021; Carranza et al., 2022). In contrast, our paper focuses on the demand side, and shows that a labor supply shock can lead to formal employment creation. For this, we build on the large literature that studies firms and informality (see Ulyssea, 2020, for a review) and on the recent but growing literature of firm dynamics and informality (e.g. D'Erasmo and Boedo, 2012; Dix-Carneiro et al., 2021; Erosa et al., 2022). Our contribution is to introduce a richer entry structure and heterogeneous growth profiles, which allows the informal sector to act as a "stepping-stone" for potentially productive firms. This proves to be key in our context, where 40% of the formalization effects of immigration come from greater entry of informal firms that eventually grow to become formal firms.

Third, our paper relates to a handful of papers that causally estimate the economic effects of internal migration in developing countries beyond labor markets. Imbert et al. (2022) study large-scale manufacturing firms in China and show that internal immigration linked to agricultural price shocks leads to employment growth and a shift towards labor-intensive production patterns. Albert et al. (2021) study climate adaptation via capital and labor reallocation in Brazil, and show that local economies are insured against droughts thanks to financial integration. Importantly, this capital reallocation does not explain our results: they are robust to controlling for the exposure to shocks via the network of banks' branches. Albert et al. (2021) also document that climate migrants go to smaller non-manufacturing firms, which is consistent with our own results on firm composition.[4] Our contribution is to combine empirical and model-based approaches to investigate the relationship between internal migration and the urban economy. We show that immigration generates formal sector expansion but decreases productivity per worker. This is consistent with the view that the reallocation of workers across sectors increases aggregate output and narrows the rural-urban productivity gap (Gollin et al., 2014).

[4] Innovation is another channel through which immigration may affect employment growth. In the US context, the literature has documented the positive contribution of high-skilled immigration to innovation and science (Moser et al., 2014; Akcigit et al., 2016). In Brazil, (Bustos et al., 2018) find that research and innovation decline when agriculture releases low-skilled workers.

Finally, we contribute to the literature on population growth and firm dynamics. Our findings that higher labor supply due to internal migration spurs formal firm and job creation in Brazil are a mirror image of the recent literature on demographic decline and the "start-up deficit" in the US (Hopenhayn et al., 2018; Karahan et al., 2019; Peters and Walsh, 2022). The US literature argues that the demographic slowdown had negative effects on firm entry, with detrimental consequences for labor reallocation, employment and firm growth (e.g. Pugsley and Sahin, 2019). Our findings also relate to the growing literature on immigration and firms in developed economies (Lewis, 2011; Peri, 2012; Olney, 2013; Kerr et al., 2015; Mitaritonna et al., 2017). In particular, Dustmann and Glitz (2015) highlight the role of firm entry in the absorption of immigrant labor supply. We bring to this literature some of the first empirical evidence on the effect of exogenous increases in internal migrant labor supply on formal firms dynamics and worker allocation between the informal and formal sectors in a developing country context.

The paper is structured as follows. Section 2 presents the data and Section 3 the shift-share instrumental variable empirical design and its results. Section 4 develops the model, while Section 5 discusses the calibration and counterfactual results.

2 Data

This section describes the data sources we use to estimate the causal effects of immigration on the local economy (Section 3), and to calibrate our model (Section 5).

Migration and Labor Market Outcomes The first dataset we use is the Decennial Population Census, which contains information on individuals' socioeconomic characteristics and labour market outcomes. Crucially, the Census also contains detailed information about individuals' migration patterns. In our analysis, we focus on the last two waves of the Census – 2000 and 2010 – and restrict the sample to working-age adults (15 to 64 years old).

Our unit of analysis is the Minimum Comparable Area (MCA), which combines municipalities whose borders have changed during the study period: there are 3,545 such MCAs in our final sample. We will call them "municipalities" for simplicity. Following the census data structure, we define as migrant a person who came to their current municipality of residence in the last ten years, and compute the cumulative immigration flows between 2000 and 2010 between each municipality pair. We focus on flows to urban locations, as defined in the census (88% of all migration between 2000 and 2010), and across state borders (40% of migration to urban areas between 2000 and 2010). We then compute the immigration rate in each urban destination as the sum of immigration flows over the decade divided by the population in 2000. This is our main endogenous regressor in the instrumental variable analysis (Section 3). Internal immigration in the average urban destination is large: 17.6% overall, 7% for state-to-state migration (Panel D in Table 1). Figure 1 shows the geographical variation in immigration rates across municipalities, which is one of the main sources of variation used in our empirical analysis.

Additionally, we use the census to compute socio-demographic characteristics and labor market outcomes for each destination municipality, which are shown in Table 1. The main socio-demographics (Panel A) are the share of female, of young (below 18) and high-skilled (completed secondary education and above) in the working age population. Our main labor market outcomes are the share of working-age adults employed in the private sector as their main occupation, which we split between formal and informal wage employment. In Brazil, formal employees are required to have a "signed work booklet" (*carteira de trabalho assinada*), and the Census directly asks wage workers whether they have it. We categorize formal workers as those who have this booklet, and informal otherwise.[5] We compute the average real monthly wage for all private sector workers, and separately for formal and informal workers (we compute hourly wages as well for some robustness checks). As Panel B in Table 1 shows, formal private wage work expanded in Brazil between 2000 and 2010, while informal wage work stagnated. Wages increased by 11% (6% in the formal sector).

Finally, to capture year-on-year variation in urban labor market outcomes we use the National Household Survey (PNAD), which is conducted annually by the National Bureau of Statistics (IBGE) also responsible for the Demographic Census. The PNAD is a repeated cross section that is representative at the state, not at the municipality level, but in which it is possible to identify 700 municipalities (see Corbi et al., 2021).[6] We use data from 2001 to 2009 for these 700 municipalities to compute the same labor market outcomes computed from the Census.

[5]See for example, Ponczek and Ulyssea (2022) for a description of labor regulations in Brazil.

[6]We are very grateful to Corbi et al. (2021) for providing us with the municipality codes that allow us to identify these municipalities in the PNAD data.

Firms The third data set we use is the *Relação Anual de Informações Sociais* (RAIS), which is a matched employer-employee administrative dataset from the Ministry of Labour in Brazil that contains the universe of formal firms and workers. We use these data for the period of 1997-2018. For the instrumental variable analysis, we use the RAIS data to compute moments related to firm dynamics at the municipality level (Panel C in Table 1). In particular, we compute the logarithm of the total number of formal establishments and formal jobs, and of the number of establishments that entered and exited in the last year. Entry in the data can be due to firm creation or formalization of an establishment that was previously operating informally. Exit is characterized when a given establishment is no longer found in the administrative data. We also calculate the real average formal wage in each municipality by dividing the monthly wage by the number of hours worked and an inflation index and averaging it over all formal employees. As Panel C in Table 1 shows, the 2000s have seen tremendous growth in the number of formal firms (40%) and workers (60%).

We use additional data on informal firms and informal workers within formal firms to compute the relevant moments we need for the model calibration. We thus complement the RAIS data with the ECINF survey (*Pesquisa de Economia Informal Urbana*), a cross-section representative of all Brazilian firms with up to five employees, which was also collected by IBGE in 2003.[7] This is a matched employer-employee data set that contains information on entrepreneurs, their businesses and employees. In particular, it includes information about the businesses' age (months/years in operation), their formalization status and of their employees. Thus, the data allow us to capture both the extensive and the intensive margins of informality. The ECINF is not representative at finer levels of geographic disaggregation (only at the state level), which prevents us from using it in the instrumental variable analysis.

Push Shocks To identify the causal effect of migration on destination outcomes, we need variation in migration flows that are exogenous to conditions at destination. In our main specification, we consider variations in agricultural income due to changes in international crop prices. In Section 3.6, we investigate the short-run effects of migration using a year-on-year specification, for which we need a higher frequency shock. We thus use the occurrence of droughts as second push shock.

[7] The data include firms with up to 10 employees, but the information for firms with more than five employees is not representative (de Paula and Scheinkman, 2010).

We combine three additional data sources to construct these shocks. For the price shock, we collect monthly information on international prices for 12 crops (including bananas, cocoa, coffee, cotton, maize, orange, rice, soybeans, sugar, tobacco, wheat and wood) between 1970 and 2011. For each crop c and month m, we compute a price shock ε_{cm} as the residual of an AR(1) process. We then aggregate these crop- and month-level shocks into a (origin) municipality level shock s_o^p using the share of each crop π_{oc} in the value of agricultural production in the municipality of origin o. We use the 1980 Agricultural Census to compute these shares for each municipality and crop.[8] Figure 2 shows the geographic distribution of the price shock. Formally, the price shock is given by:

$$s_o^{prices} = \sum_m \sum_c \left(\pi_{oc} \times \varepsilon_{cm}\right) \tag{1}$$

For the drought shock, we combine a measure of dryness in each municipality in each month, the SPEI (Standardized Precipitation-Evapotranspiration Index),[9] with information on the growing season of each crop in a given region, and the value of each crop harvested in each municipality from the 1980 Agricultural Census (see Appendix E.1 for more details on the drought shock). Interestingly, the price and the drought shocks are completely independent across origins (correlation of 0.007).

3 Instrumental Variable Analysis

3.1 Empirical Design

We estimate the impact of immigration on labor markets outcomes by regressing the change in outcome y in municipality d between 2000 and 2010, on the immigration rate over the decade Mig_d:

$$\Delta y_d = \beta_0 + \beta_1 Mig_d + \beta_2' \mathbf{X_d} + \varepsilon_d \tag{2}$$

[8] The construction of the shock is similar to Imbert et al. (2022) for China.

[9] The SPEI has been built by climate scientists Vicente-Serrano et al. (2010) and can be freely downloaded here `https://spei.csic.es/home.html`. It has been used in economics by Bertoli et al. (2020) and Albert et al. (2021).

where $\Delta y_d = y_{d,2010} - y_{d,2000}$ is the long difference in labor market outcomes: employment rate in the private wage sector (formal and informal) and log wages in the private sector (formal and informal). The vector $\mathbf{X_d}$ includes socio-demographic controls: the share of female, young and high-skilled workers in municipality d in 2000. We report robust standard errors, and use 2000 population as regression weight.

We then turn to firm outcomes, for which we have annual data between 1997 and 2018. We use a specification similar to Equation 2 and regress changes in the number of formal firms, entrants and exiting firms, in the number of formal jobs and in the firm wage on the migration rate, controlling for socio-demographic controls. To reduce noise and increase stability of the entry and exit measures (especially in very small municipalities), we compute outcomes as two-year averages. For comparability with the labor market results, we first focus on the decadal changes between 1999-00 and 2011-12. We then exploit the full extent of the panel data and estimate dynamic effects of migration on changes in outcomes between 1999-00 and 2011-12, 2013-14, 2015-16 and 2017-18. We also investigate the presence of pre-trends, and use as a dependent variable the change in outcomes between 1997-98 and 1999-00.

Identification

Since we estimate all regressions in first differences, our specification implicitly accounts for municipality fixed effects. By including controls, we also allow for differential trends across municipalities with different initial socio-demographic conditions. However, the regressions above may not identify the causal effect of immigration on labor markets and firms due to reverse causation and omitted variable bias. For example, cities with thriving labor markets are likely to attract more migrants.

To overcome these threats to identification, we rely on a shift-share instrumental variable design, in which the shifts are assumed to be exogenous (Adão et al., 2019; Borusyak et al., 2022). More specifically, we combine cross-sectional variation across destination municipalities in their pre-existing migration networks with different origins (the "share"), and time variation in exogenous push shocks that affect migration incentives at origin (the "shift"). Formally, the instrument writes:

$$Z_d = \sum_o \lambda_{o,d} s_o \tag{3}$$

where Z_d denotes the instrument for immigration into the municipality of destination d, based on the price shock s_o to agricultural productivity in the municipality of origin o described in section 2. $\lambda_{o,d}$ denotes the share of migrants from origin o among migrants who had come at destination d between 1995 and 2000.[10] We then use Z_d as an instrument for Mig_d in a 2SLS estimator.

There are a number of potential threats to our empirical strategy. Regarding identification, one may worry that shifts are not randomly assigned, but correlated with potential outcomes at destination. To alleviate this concern, we carry out placebo checks in which we regress changes in firm outcomes between 1997-98 and 1999-00 on immigration between 2000 and 2010. We also estimate our main specification including as controls lagged changes in outcomes, log GDP at baseline, and shares of the different industries. Another concern is that immigration between 1995 and 2000, which we use to compute the shares, may have long lasting effects on firm outcomes between 2000 and 2010. To alleviate this concern, we add 2000 log population and the 1995-2000 migration rate as controls to our main specification. We also follow Borusyak et al. (2022), and transform our estimation into an origin-level regression and check that we obtain similar results when we control for lagged price shocks.[11]

There may also be concerns about the exclusion restriction. First, shocks to rural incomes may affect the demand for goods produced by urban firms (Santangelo, 2016). To control for this potential demand channel, we control for agricultural price shocks in the municipality of destination and for the sum of agricultural price shocks in all other municipalities weighted by the inverse of distance. Second, negative shocks to rural productivity may lead to a reallocation of capital towards other parts of the country, including migrants' destinations. To account for this channel, we control for the exposure of each destination via the bank network, measured as the share of loans that come from banks that draw more deposit from municipalities hit by price shocks (as in Albert et al., 2021). Third, international price shocks for agricultural commodities may have indirect effects on urban firms that process agricultural goods. To alleviate this concern, we exclude from the sample agricultural

[10] The 2000 Census only contains retrospective information on migration going 5 years back, hence we compute the migration shares from 1995 to 2000.

[11] Adão et al. (2019) show that conventional standard errors can be invalid in frameworks such as ours, as observations which similar exposure shares can have correlated residuals. The inference procedure in Borusyak et al. (2022) is asymptotically equivalent to that of Adão et al. (2019), and therefore it also eliminates any concerns about inference in our main specification.

firms, and manufacturing firms that process food and beverages or tobacco.

Finally, we use an alternative shift-share instrument in which the shifts are droughts at origin and check that we obtain similar results. Given that the shifts are independent across origins and the two shift-share instruments are only partly correlated across destinations (correlation of -0.3), this tests that our results are not driven by specific pairs of sending and receiving municipalities.

3.2 Effects on Labor Markets

We start by discussing the effects of internal immigration on urban labor markets. Table 2 presents the results. The OLS estimates suggest that immigration is associated with higher formal employment, lower informal employment and (insignificantly) higher wages. One would expect these results to be biased since migrants would be attracted by destinations with better labor markets outcomes. We next turn to the IV estimates, which have a causal interpretation.[12] Surprisingly, the effect on formal employment is even more positive, and the effect on informal employment more negative than what the OLS estimates suggest (Columns 2 and 3). This could be due to the fact that large and rich cities attract a large share of immigration, but may not be the most dynamic labor markets. We show below that our results are driven by municipalities in the bottom and middle terciles of baseline GDP per capita.

Table 2 shows that a one percentage point increase in the migration rate – which corresponds to 18.5 percent of a standard deviation – increases formal employment by 0.4 percentage points (a 1.7 percent increase from the mean of 23 percent) and decreases informal employment by 0.29 percentage point (a 2.9 percent decrease from the mean of 10 percent). Overall, wage employment increases by an insignificant 0.10 percentage points: most of the increase in formal sector work is driven by a shift away from informal employment. In Appendix Table A.2 we show that there is no effect on non-employment, self-employment, or domestic work, but a small decline in the fraction of workers who report being employers and working in the public sector. Turning to Columns 4 to 6 (Table 2), we find evidence that immigration decreases wages. A one percentage point increase in the immigration rate reduces wages in the formal sector by 2.2 percent, and by 1.9 percent in the informal sector. The overall effect on wages is smaller, 1.6 percent, which is due to a compositional effect, as

[12] Table A.1 in the Appendix shows first stage estimation results.

workers shift from the low paying informal sector to the formal sector.[13]

The literature has also emphasized the importance of looking at the labor market response to immigration separately by skill, especially when migrants have different skill levels than natives (Dustmann and Glitz, 2015). Table A.4 in the Appendix shows that the immigration flows do not change the skill composition at destination, but they lead to an increase in the proportion of young males in the population. Consistently, Appendix Table A.5 shows very similar effects on employment and wages for high- and low-skilled workers, although wage effects are more negative for low-skilled workers. Finally, we check in Appendix Table A.6 that our results are the same among migrants and non-migrants.

3.3 Effects on Firm Dynamics

Our results from the population census suggest that internal immigration decreases labor costs and increases the share of workers in the formal sector. We now turn to the firm-level data, and examine the effects of immigration on decadal changes in the number of formal firms, entry, exit, number of jobs and firm average wages. Panel A in Table 3 present the OLS estimates, while Panel B presents the IV results. Comparing the OLS and IV results, the effects on number of firms, entry, exit and number of jobs remain positive and increase in magnitude. Our estimates indicate that a one percentage point increase in the immigration rate leads to an increase of 2.3 percent in the number of firms and 2.2 percent increase in the number of formal jobs. There is also higher churn, with a strong increase in the number of firms entering and exiting every year, 7.5 and 5.6 percent for a one percentage point increase in the immigration rate, respectively. The negative effect on average firm wage (-3.4 percent) confirms the worker-level findings.

To put these effects in perspective, the start-up deficit (i.e. decline in entry) documented in the U.S. corresponds to a decline of 5 percentage points between 1980 and 2012 (Pugsley and Sahin, 2019). Hence, the labor supply shock induced by rural to urban migration has a first order effect on firm dynamics. Interestingly, this effect does not seem to be driven by migrants creating firms at higher rates: Table A.7 in the Appendix shows that migrants are not more likely to be business owners in

[13] Although we use the main occupation in the last month and monthly wages as our main outcomes, we check in Appendix Table A.3 that the results are similar when we use the share of hours spent in each occupation and hourly wages instead.

municipalities that receive more immigration. This is true for the three business size categories in the census (self-employed, less than 5 employees, at least 6 employees).

We can also use the time dimension of the RAIS firm data to analyze the dynamics of these migration effects. Specifically, we estimate the effects of migration on changes in outcomes for different time windows: between 1999-00 and 2011-12 (our main results), 2013-14, 2015-16, and 2017-18. As Figure 3 shows, the effects tend to be somewhat stable or slightly increasing over time. Reassuringly, the figures do not show any clear evidence of pre-trends.

3.4 Robustness

The main threat to our identification strategy would be a failure of the common trend assumption, i.e. if firms or workers located in destinations that receive immigration shocks between 2000 and 2010 would have experienced differential changes in outcomes even in absence of the shocks. The shift-share instrument partly alleviates this concern: in our case, identification relies on the assumption that although the baseline migration shares may be endogenous to future outcomes, the push-shocks (shifts) are as good as randomly assigned (Adão et al., 2019; Borusyak et al., 2022).

Figure 3 already suggests that pre-trends may not be a concern. To check that pre-trends do not drive our results, we run the benchmark specification controlling for lagged differenced outcomes and show that our basic results remain essentially unchanged (Panel A of Appendix Tables B.1 and B.2). We perform a number of additional robustness checks, which we report in Panels B–E in Appendix Tables B.1 and B.2. Another potential threat to our instrumental variable strategy would arise if migration rates are very persistent overtime, so that we are capturing the effects of previous migration waves, a concern which has been raised in particular by Jaeger et al. (2018). To alleviate this concern, we control for log population (Panel B) at baseline and for migration rates between 1995 and 2000, in Panel C. Finally, we allow for differential trends by baseline GDP and industry composition (both computed in 2000) in Panels D and E, respectively. Our results are robust to these controls.

Although we interpret our IV results as the effects of immigration, in principle agricultural price shocks could affect firm outcomes via other channels. First, they could change the demand for goods produced at destination. In Panel A in Appendix Tables B.3 and B.4, we show that our results do not change after we control for agricultural price shocks in the municipality of destination and the sum of shocks in other municipalities weighted by the inverse of distance. Second, price shocks could induce a capital reallocation towards migration destinations. To capture capital reallocation, we construct alternative shift-share inspired by Albert et al. (2021) that combines price shocks at origin with bank links (deposits and loans) between all municipalities. Our results remain unchanged when add it as a control (Panel B in Appendix Tables B.3 and B.4). Finally, we check that our results are not driven by firms that produce or process agricultural goods: our results are unchanged when we exclude them (Panel C in Appendix Table B.4).

In Tables B.5 and B.6 we implement a transformation from destinations to origins and run the regressions at the origin level as suggested by Borusyak et al. (2022). In addition to the benchmark OLS and IV regressions transformed at the origin level (Panels A and B, respectively), we also control for the lagged price shock in Panel C. All of our results are robust to these exercises.

3.5 Implications for local economic development

The results discussed so far show that rural-urban immigration induced a considerable shift away from informal employment and into formal employment without any negative effects on total employment in receiving municipalities. In this Section we examine what are the additional implications for local economies.

We start by exploring changes in firm composition at destination (all results are reported in Appendix Section C). Table C.1 shows that there is a shift of firms, towards retail, services and construction, and away from manufacturing, with insignificant but similar effects on jobs. This is intuitive, as retail and services are more labor intensive and require lower and less specific skills than manufacturing, while construction is also a sector that traditionally employs many migrants. Additionally, non-tradables are also more likely to benefit from the positive demand shock caused by the arrival of immigrants. These new entrants are quite small in size: Table C.2 shows that there is an increase in the share of firms with up to 5 employees. Hence, the urbanization process that we are capturing shares similarities with the "urbanization without industrialization" discussed in Gollin et al. (2016).

Given these compositional effects, one might expect the results to be driven by the entry or survival of low-productivity firms. To test this, we classify firms in four productivity quartiles and estimate the effects of migration on the composition of firms. We do not directly observe firm productivity in the data, so we rely on a proxy: average wage paid by the firm (over all workers and years the firm operates). We take into account inflation and spatial differences in costs of living by regressing firms' wage on year and micro-region fixed effects, and use the residual as our measure of productivity. One potential drawback from using wages as proxy for productivity is that since migration reduces wages, one may expect to find mechanically more firms in lower quartiles. Interestingly, however, the results presented in Appendix Table C.3 suggest that firms that expand in response to a migrant labor supply are in the middle of the productivity distribution rather than the very bottom. We also observe a relative decline of top-productivity firms. The effects on the composition of jobs mirror those of the composition of firms. Our results therefore strongly suggest that immigration favors the entry of firms in the middle of the productivity distribution, which strive in the long-run, and make the formal sector grow.

Finally, we examine which municipalities benefit the most from these immigration flows, in terms of labor force formalization and firm creation in the formal sector. To do so, in Appendix Section D we investigate the heterogeneous effects across terciles of GDP per capita at baseline (in 2000). There are huge differences in economic development across regions at baseline: the average GDP p.c. in the top tercile is more than twice as large the average in the middle tercile and more than 4.5 times larger than the bottom one. To put these in international perspective, the poorest State in Brazil has a development level comparable to Egypt, while the richest is close to Czechia. We show that the effects on formal employment from the worker and firm side are mostly driven by poor municipalities from the bottom and, to a lesser extent, middle terciles. This has important implications for how one can interpret our findings and their external validity. Broadly, the results indicate that urban economies in the lower end of the development spectrum can benefit from the inflows of migrants.

3.6 Long vs. Short Time Differences

To a large extent, the rural-urban migration literature has grown around the idea that internal migrants join the ranks of casually employed or unemployed workers in urban areas, a view that dates back at least to Harris and Todaro (1970) and Fields (1975), and which has been supported by recent empirical evidence (Kleemans and Magruder, 2018; El Badaoui et al., 2017). Our results run counter this view, as they show that internal migration increases formal employment, reduces informality, reduces both formal and informal wages, with no effect on unemployment.[14]

One potential explanation for our contrasting findings is that the previous literature estimates short-run effects, while we consider changes over a decade.[15] Such differences between short and long run results could be rationalized by the presence of frictions that constrain formal labor demand (but not informal) in the short run, and which would be at least partially alleviated in the longer run. Downward wage rigidity in the formal sector is one such friction, which we directly analyze in the counterfactuals discussed in Section 5.2.

To investigate whether differences in time horizon can explain the discrepancy between our results and those found in the previous literature, we lean on the work of Corbi et al. (2021), who show that immigration from the semi-arid regions of Brazil increases informality and unemployment in urban destinations. We use the same data as theirs for destination outcomes, but we expand the sample to all sending municipalities (See Section 2).[16] We estimate the following year-on-year specification, which is similar to that used in previous studies (including Corbi et al., 2021):

$$\Delta y_{dt} = \beta_0 + \gamma_1 Mig_{dt} + \gamma_2' \mathbf{X_d} + \eta_t + \varepsilon_d \tag{4}$$

where d again denotes destination municipalities, η_t denotes year fixed effects, and $\Delta y_{dt} = y_{d,t} - y_{d,t-1}$ for $t = 2001, ..., 2010$. The vector $\mathbf{X_d}$ is the same as before and includes the share of female, young and high-skilled workers in municipality d in 2000. Mig_{dt} measures the inflow of migrants in year t to destination municipality d, which we compute using the Census.

This specification requires a higher frequency instrument, as prices do not vary so much year-on-year, and are not predictive of outmigration in the very short-run. We thus turn to the droughts push shock discussed in Section 2 (see also Appendix E.1 for more details), which have high yearly variation and predict well outmigration in the short-run (e.g. Corbi et al., 2021; Albert et al., 2021).

[14] Imbert et al. (2022) also find negative wage effects and positive employment effects of internal immigration focusing on large manufacturing firms in China over five years.

[15] Strictly speaking, our specification captures a mix of short and long-run effects, since we estimate the effect of migration waves that happened throughout the decade.

[16] We are grateful to the authors for providing us with the codes that identify municipalities in the National Household Survey, without which it would have not been possible to conduct this analysis.

We start by re-estimating our main long-term specification (equation 2) using droughts instead of price shocks to construct the shift-share instrument. The Appendix Section E.2 provides the details. Tables E.1 and E.2 in the appendix show that we find very similar results, with the same reallocation from informal to formal employment, and positive effects on the number of formal firms and jobs. Effects on wages are negative but smaller in magnitude and statistically insignificant. These results are important, as they suggest that the expected surge in climate-induced rural-urban migration could have positive effects on receiving urban labor markets. They also show that our main results are not driven by specific pairs of sending and receiving regions, as the two shocks are independent across origins (correlation of 0.007), and the two shift-shares are only weakly correlated across destinations (around 0.3). Table A.4 also shows that the drought shocks have no discernible effects on workforce composition at destination, in contrast to the price shocks that induce young men to migrate. Yet, both shocks have the same effects on urban outcomes.

Finally, we estimate the year-on-year specification 4, using droughts as push shocks. Table E.5 shows that municipalities that receive more migrants experience a reduction in overall wage employment that is driven by a fall in formal wage employment, while informal employment is unaffected. The share of informal jobs is thus increasing in regions that receive more migrants. We find no statistically significant effects on formal and informal wages, although point estimates are sizable and go in the expected direction. The effects on firms are consistent, with no effect on the number of formal firms but a large negative effect on the number of formal jobs.[17] Hence, in contrast to the long-run increase in formal employment we find that in the short-run immigration leads to an increase in informality in urban destinations (as in Kleemans and Magruder, 2018; El Badaoui et al., 2017; Corbi et al., 2021).

4 Model

This section develops a simple equilibrium model of firms dynamics and informality to rationalize the results discussed so far. We build on Hopenhayn (1992), with key innovations in the entry structure, firms' productivity processes and, crucially,

[17]We check that our long-term results are unaffected if we restrict the sample to the 700 municipalities in the PNAD data (Appendix Tables E.3 and E.4). Thus, the difference between short- and long-term results is not due to sample restrictions.

the two margins of informality (Ulyssea, 2018). Importantly, this is not a spatial model that aims at capturing the equilibrium effects of migration for the country as a whole. Instead, we model a single urban economy, calibrate it to the average urban destination in Brazil, and use it to assess the combined equilibrium effects of the labor supply and demand shocks represented by immigration flows in the average destination economy, as well as to gain a deeper understanding of the mechanisms behind our empirical results.

4.1 Set up

Every period, there is a continuum of firms that are indexed by their idiosyncratic productivity, θ. All firms have the same technology and use labor as their only input: $f(\ell) = \theta q(\ell)$, where $q(\cdot)$ is increasing and concave. Formal and informal firms operate in the same industry, produce a homogeneous good and face the same prices in a competitive market. For simplicity, we assume that labour is homogeneous.[18] The informal sector is composed by informal firms (the extensive margin), while informal workers can be found in both sectors (because of the intensive margin).

These assumptions regarding market structure imply complete integration between the formal and informal sectors, which is in sharp contrast with the "dual economy" view that goes back to Lewis (1954), Harris and Todaro (1970) and Fields (1975). This view, however, is at odds with the more recent evidence available (e.g. Hsieh and Olken, 2014), in particular that formal and informal firms coexist within narrowly defined industries, there is a large degree of overlap between formal and informal firms' productivity distributions, and a large fraction of informal employment is in fact located within formal firms (Ulyssea, 2020).

4.2 Profits

The static, within-period problem follows very closely Ulyssea (2018). Informal firms are able to avoid taxes, but face an "informality cost" captured by $1 \leq \tau_i(\cdot) < \infty$, where $\tau_i', \tau_i'' > 0$. This function is a general formulation for the different costs asso-

[18]The model can accommodate different skill levels and imperfect substitution between migrants and non-migrants. Nevertheless, our empirical results show that: immigration was neutral in terms of the skill composition at destination; conditional on skills, internal migrants tend to be close substitutes to non-migrants; and effects are the same for skilled and non-skilled workers. Thus, we believe that the worker homogeneity assumption is a reasonable simplification in our context.

ciated to informality, such as the probability of being inspected by the government. The latter is likely to be increasing in firms' size, as larger firms are more visible to the government and inspected with higher probability, for example. Informal firms' profit function is given by

$$\Pi_i(\theta, w) = \max_{\ell} \{\theta q(\ell) - \tau_i(\ell) w\} \tag{5}$$

Formal firms must pay revenue and payroll taxes, but can evade the latter by hiring informal workers. However, formal firms also face a cost of hiring informal workers that can be rationalized along similar lines as above – e.g. a large formal firm is more likely to be audited by the government. We assume that formal firms' cost of hiring informal workers is given by $1 \leq \tau_f(\ell_i) < \infty$, where also $\tau_f', \tau_f'' > 0$. Hence, formal and informal workers have different marginal costs due to regulations that are imperfectly enforced. As both types of workers are perfect substitutes, at the margin firms hire the cheaper one and there is a unique threshold, $\tilde{\ell}$, above which all additional workers are hired formally.[19] Formal firms' profit function can thus be written as follows:

$$\Pi_f(\theta, w) = \max_{\ell} \{(1 - \tau_y)\theta q(\ell) - C(\ell)\} \tag{6}$$

where τ_y denotes the revenue tax and

$$C(\ell) = \begin{cases} \tau_f(\ell) w, & \text{for } \ell \leq \tilde{\ell} \\ \tau_f\left(\tilde{\ell}\right) w + (1 + \tau_w)\left(\ell - \tilde{\ell}\right) w, & \text{for } \ell > \tilde{\ell} \end{cases} \tag{7}$$

Additionally, every period firms must pay a fixed cost of operation and the net profit is given by: $\pi_s(\theta, w) = \Pi_s(\theta, w) - \bar{c}_s,\ s = i, f$.

4.3 Entry and Dynamics

Every period, there is a large mass of potential entrants that decide whether to enter the informal or formal sectors, or not to enter at all. Before entry, they observe their

[19] The marginal cost of hiring informal workers is increasing and given by $\tau_f'(\ell)w$, while the marginal cost of hiring formal workers is constant and given by $(1 + \tau_w) w$, where τ_w is the payroll tax. The threshold $\tilde{\ell}$ is given by the point where these two marginal costs are equated.

potential long-run (fixed) productivity parameter, ν, which is drawn from the CDF $H(\nu)$. These are assumed to be i.i.d., such that entry in one period does not affect the composition of entrants in the following period. Entrants in both sectors must pay a fixed entry cost, c_s^e. These parameters will be estimated, but we expect that $c_f^e > c_i^e$ due to regulatory costs associated with formal firm creation.

Post-entry dynamics are driven by the evolution of firms' idiosyncratic productivity, θ, which is the only source of uncertainty to the firm (there are no aggregate shocks). The productivity process is characterized by the following expressions:

$$\ln \theta_{j,1} = \ln \nu_j + \ln \epsilon_{j,1} \tag{8}$$

$$\ln \theta_{j,t} = \rho_s \ln \theta_{j,t-1} + (1-\rho_s) \ln \nu_j + \ln \epsilon_{j,t}, \;\; t \geq 2 \tag{9}$$

where j indexes firms, the unexpected shock is i.i.d. and $\ln \epsilon \sim \ln \mathcal{N}(0, \sigma_s^2)$, $s = i, f$.

Equation 9 is very close to standard formulations in the dynamic panel literature (e.g. Blundell and Bond, 1998), where firms' long-run productivity level is given by $\ln \theta_\infty = \ln \nu$. Crucially, Equations 8 and 9 imply that firms' current productivity can differ from their long-run productivity level, and that there is heterogeneity across firms in terms of life-cycle growth profiles.

We assume that formal firms cannot become informal,[20] and therefore face a simple stopping-time problem (to remain active or exit). Informal incumbents have the additional option of transiting to the formal sector. To formalize, an informal firm must pay the difference between entry costs, $\tilde{c}^e = c_f^e - c_i^e$, and after that it will face the regulatory costs implied by formality. In addition to endogenous exit, we assume that formal and informal firms face exogenous death shocks denoted by δ_s.

The value functions of formal and informal *incumbents*, respectively, can be written in recursive form as follows:

$$V_f(\theta, w) = \pi_f(\theta, w) + (1-\delta_f)\beta \max\left\{0, E_\nu\left[V_f(\theta', w)\,\middle|\theta\right]\right\} \tag{10}$$

$$\begin{aligned} V_i(\theta, w) &= \pi_i(\theta, w) \\ &+ \beta \max\left\{0, (1-\delta_i) E_\nu\left[V_i(\theta', w)\,\middle|\theta\right], (1-\delta_f) E_\nu\left[V_f(\theta', w)\,\middle|\theta\right] - \tilde{c}^e\right\} \end{aligned} \tag{11}$$

[20]Once the firm is visible to the government, it is very hard to become informal and therefore "invisible" again, as the government has formal firms' complete information. Firms could change locations to become informal, which in our model would correspond to exiting the formal sector and rejoining the pool of potential entrants (but with no memory of previous experiences).

where β denotes the discount factor and the subindex ν in the continuation value indicates that the expectation depends on the firm-specific long-run productivity parameter, as described in expressions 8 and 9.

The exit decision is made before future productivity is revealed and it follows a cut-off rule. If $\theta < \underline{\theta}_s$ firms choose to exit, where $\underline{\theta}_s$ is given by

$$E_\nu \left[V_s \left(\theta', w \right) \middle| \underline{\theta}_s \right] = 0, \;\; s = i, f \tag{12}$$

Entry is also characterized by a threshold rule, as follows:

$$\begin{aligned} E_\nu \left[V_i \left(\theta, w \right) \middle| \nu = \underline{\nu}_i \right] &= c_i^e & (13) \\ E_\nu \left[V_f \left(\theta, w \right) - V_i \left(\theta, w \right) \middle| \nu = \underline{\nu}_f \right] &= c_f^e - c_i^e & (14) \end{aligned}$$

where $\underline{\nu}_s$ characterizes the last firm to enter sector $s = i, f$.

Finally, informal firms formalize if $\theta \geq \overline{\theta}_i$, where $\overline{\theta}_i$ is given by:

$$E_\nu \left[V_f \left(\theta', w \right) - V_i \left(\theta', w \right) \middle| \overline{\theta}_i \right] = \tilde{c}^e \tag{15}$$

The *timing* of informal firms' decisions is the following. At the beginning of the period, firms draw their productivity shock, θ, and decide how much to produce (i.e. how much labor to hire). If $\theta \in \left[\underline{\theta}_i, \overline{\theta}_i \right)$, they will start next period in the informal sector again; if $\theta \geq \overline{\theta}_i$, they will pay $\tilde{c}^e$ and start next period in the formal sector; and if $\theta < \underline{\theta}_i$, they exit. The timing for formal firms is the same, except that they only choose whether to stay or exit.

4.4 Steady state equilibrium

To close the model, we assume that there is a representative household that inelastically supplies $\overline{L}$ units of labor and that derives utility from consuming the final good. We model migration as a once-and-for-all increase in $\overline{L}$.[21] Households consume all of their income, which is given by $w\overline{L} + \Pi + T$, where Π denotes the mass of profits in the economy, $w\overline{L}$ is total wage income and T denotes total tax revenues. Therefore,

[21] Our empirical results show that immigration has no effect on unemployment nor on the share of individuals out of the labor force. Hence, we do not model unemployment nor labor force participation decisions.

migration simultaneously increases labor supply and demand for the final good.

We focus on stationary competitive equilibria, which correspond to a set of allocations, wage, cutoffs and measures of firms such that they remain constant over time and the following conditions hold in every period: (i) product and labor markets clear;[22] (ii) the cutoffs $\left(\underline{\theta}_s, \overline{\theta}_i, \underline{\nu}_s\right)$, $s = i, f$, are determined according to (12)–(15) and entry conditions hold. In Appendix F we show that there exists a unique stationary competitive equilibrium where the formal and informal sectors exist and have positive entry and exit (Proposition 1).

Appendix F provides other characterization results that shed light on the model's main mechanisms. In particular, we show that the productivity distribution among active firms within a sector and cohort is increasing in the cohort's age (Proposition 2). Put differently, the productivity distribution of older cohorts first-order stochastically dominates that of younger cohorts. From this result, it follows that any function that is increasing in firms' productivity, θ, will also be increasing in cohort's age. In particular, this is true for the survival rate, firms' average size and average revenue in *both the formal and informal sectors* (Corollary 1). These results also imply that the average share of informal workers within formal firms (intensive margin) and the average share of informal firms (extensive margin) within a given cohort decline with the cohort's age (Corollaries 2 and 3, respectively). Importantly, these results constitute falsifiable predictions and show that the model has empirical content.

4.5 Discussion

In response to an immigration shock that increases both labor supply and demand for the final good, one would expect places that receive more migrants to experience a decline in real wages, which would cause incumbents' labor demand to expand. Incumbents are eventually limited by scale, so the increase in labor supply is likely to be also met by greater firm entry. In equilibrium, not only entry rates are affected, but also exit rates, and firm size distribution. The literature that explains the "startup deficit" in the US by a slowdown in labor supply growth highlights the same mechanisms working in reverse (Hopenhayn et al., 2018; Karahan et al., 2019).

It is unclear, however, what the net effects should be in terms of overall firm and labor informality, or output per worker. If migrants are largely absorbed by informal

[22] We normalize the price of the final good to one throughout the analysis.

jobs in informal firms or low-productivity formal firms, the net result will most likely be an increase in informality among firms and workers. However, if the decline in wages induces greater entry also of firms with higher growth potential – either in the informal or formal sectors – then this can lead to more firm growth in the formal sector, and to a new steady state equilibrium with lower informality. The results discussed in Section 3 suggest that the latter force dominates, but these are relative effects between municipalities more and less exposed to migration. We now move to the model calibration and counterfactuals to quantify these effects in equilibrium.

5 Counterfactuals

5.1 Model Calibration

To calibrate the model discussed in Section 4, and use it to perform counterfactual simulations, we must parameterize the model's remaining objects that were left unspecified. Starting by the production function, we assume a simple span-of-control formulation: $y(\theta, \ell) = \theta\ell^{\alpha}$, $0 < \alpha < 1$. The cost functions are defined as in Ulyssea (2018): $\tau_s(\ell) = \left(1 + \frac{\ell}{\varphi_s}\right)\ell$, where $\varphi_s > 0$ and $\tau_s', \tau_s'' > 0$. The larger φ_i the more informal firms can grow before it becomes too costly to be informal, and the larger φ_f the easier it is for formal firms to hire informal workers. We assume that the potential long-run productivity parameter, ν, follows a Pareto distribution. Hence, firms' first productivity draw, $\theta_1 = \nu\epsilon_1$, has a Pareto-Lognormal distribution. This distribution is very well-suited to characterize firm size distributions in developing countries (e.g. Ulyssea, 2018).

We use a two-step minimum distance calibration. In the first step, we use the RAIS establishment-level panel to estimate the persistence parameter of formal firms' productivity process, which gives $\rho_f = 0.921$ (see Appendix G for details). We set the tax parameters to their statutory values: $\tau_w = 0.375$ and $\tau_y = 0.293$. The value of τ_w corresponds to the main regulatory costs that are proportional to firms' payroll (social security contribution, direct payroll tax and severance contributions). The τ_y corresponds to the federal VAT taxes (IPI and PIS/COFINS), and excludes state-level value-added taxes, which vary across state borders and create an intricate system of tax substitution along the production chain. As in Ulyssea (2018), the Pareto distribution scale parameter (ν_0) is set so that the firms' minimum size is one employee, while formal sector's fixed cost of operation parameter is set to be 70 percent of the baseline wage. Table 4 describes all parameter values.

The second step takes the first-step parameters as given to calibrate the remaining 12 parameters of the model, which are the following:

$$\Omega = \left\{\varphi_f, \varphi_i, \delta_i, \delta_f, \bar{c}_i, \xi, c_f^e, c_i^e, \alpha, \sigma_i, \sigma_f, \rho_i\right\}$$

where φ_i and φ_f are the parameters of extensive and intensive margins' cost functions, respectively; δ_i and δ_f are the exogenous death shocks in the informal and formal sectors, respectively; $\bar{c}_i$ determines the per-period fixed cost of operation; ξ is the Pareto shape parameter; c_f^e and c_i^e are the formal and informal entry costs, respectively; α is the span-of-control; σ_i^2 and σ_f^2 are the informal and formal variances of the productivity shock; and ρ_i is the persistence of the productivity process in the informal sector. The vector Ω is estimated by minimizing the distance between moments from the data, $\hat{m}_N$, and moments computed from simulated data generated by the structural model. Appendix Section G provides more details about the implementation.

Moments and parameters

We use two main data sources to estimate the moments used in the calibration: the RAIS, which covers all formal firms, and the ECINF survey which is representative of all formal and informal firms with up to five employees (See Section 2). Since the ECINF survey is only available in 2003, we restrict the RAIS to the same year when computing static moments. For moments relating to firm growth, we use a 2000-2011 panel from the RAIS, the same period used in the IV analysis. We compute the following moments: overall share of informal firms and by size brackets (less than 2, and 3 to 5 employees); average share of informal workers within formal firms; formal and informal firm growth at ages 5 and 10 relative to age 1; formal firms size distribution by size brackets (less than 5, 5 to 10, 11 to 20, 20 to 50, and more than 50 employees); informal firms size distribution by size brackets (less than 2, and 3 to 5 employees). We compute the share of informal workers in the economy using the 2003 National Household Survey (PNAD).

Table 4 shows the results. The estimated cost of entry in the formal sector is more than twice as large as for the informal sector. The values denominated in 2003 *Reais* are not negligible: formal sector's entry costs correspond to more than 30 times the monthly national minimum wage at the time. Informal firms face a relatively high exogenous death shock of nearly 15 percent ($\delta_i = 0.148$), which is more than twice as large as in the formal sector. Apart from the exogenous exit shock, informal firms do not face higher uncertainty, as the variances of the productivity innovation in both sectors, σ_s^2, are very similar. The same is true for persistence of the productivity process in both sectors.

Model fit

Table 5 shows the main targeted data moments and their model counterparts. To further assess how the model fits the data, we examine non-targeted moments in Figure 4. Panels (a) and (b) show that the model reproduces well the behavior observed for the extensive and intensive margins of informality, respectively. The model predicts a steeper gradient in the share of informal firms by firm size than what is observed in the data, but the fit for the intensive margin is very good and the model is able to match well the behavior of the average share of informal workers within formal firms relative to firm size.

We also use the panel structure of RAIS to examine how the model fits moments directly related to formal firms' dynamics. As panel (c) shows, the model fits well the growth profile for formal firms up to age 11, which is the maximum age that we are able to measure in our panel (see Appendix). The simulated data shows a more concave profile, however, which implies that the model slightly overestimates the accumulated growth for ages 3 to 9, but then converges to the data at age 11. Similarly, panel (d) in Figure 4 shows that the model reproduces very well the behavior of the autocorrelations of log-employment relative to log-employment at age one. This is reassuring, as these moments are never used in the estimation, and are quite informative about the structure of the firms' productivity process, in particular the importance of permanent components (Sterk et al., 2021).

In the Appendix G, we provide a careful discussion of which dimensions of the data provide the variation that allows us to pin down the different model parameters and investigate how informative the moments used in the calibration are.

5.2 Results

We start our quantitative analysis by investigating the equilibrium effects of a large immigration shock. We simulate a once-and-for-all increase in the aggregate labor supply of 10 percent, which is around the 80th percentile of the distribution of immigration rates we observe in our data.

To compare our simulation and empirical results, we run a modified version of the IV specification described in Section 3.1 (Expression 2). We restrict the sample to municipalities in the bottom and top quartiles of the immigration rate distribution, for which the average immigration rate is 1.8 and 13.4 percent, respectively. We then replace the immigration rate, Mig_d, with a dummy that equals one if the municipality belongs to the top quartile and zero otherwise, and we continue to use the same shift-share instrument using price shocks. This modification allows us to estimate the average effect of receiving a large immigration shock relative to receiving very few migrants (only a small number of municipalities have a zero immigration rate). This more closely corresponds to our our first counterfactual simulation (a 10 percent increase in the labor force).

Table 6 presents the results. The model fits the empirical results remarkably well for the three outcomes that are directly comparable. The counterfactual simulation shows a 4.1 percent decline in the share of informal workers (1.3 percentage point reduction), while the IV estimation shows a 3.9 percent decline (1.9 percentage point reduction). The effects on the number of formal firms are also comparable, with model simulation and IV estimation implying an increase of 16.3 and 14.7 percent, respectively. The model shows a 3.4 percent wage reduction compared to 5.7 in the IV estimation, thus underestimating the overall wage decline. Given that the model does not have worker heterogeneity, it is not surprising that it underestimates wage variation, albeit it remains close to the reduced-form effects.

The model allows us to go further and assess the source of this increase in the number of formal firms. As Table 6 shows, around 40 percent comes from the formalization of previously existing informal firms, while the remaining 60 percent comes from the creation of new formal businesses. This result highlights the importance of accounting for firm dynamics and the linkages between the informal and formal sectors. A static framework would largely underestimate the increase in the number of formal firms, as it would not be able to account for the transition of informal firms to the formal sector after entry occurs. Similarly, the dual view of informality which rules out such transitions by assumption would miss this important dimension of the economy's adjustments to shocks. This result also suggests that the informal sector might play a positive role in the adjustment process of economies exposed to labor supply shocks. We directly examine this in our third counterfactual below.

Table 6 also shows the equilibrium effects on additional key aggregate outcomes. First, the simulation shows a 5.3 percent decline in the share of informal firms in the economy (from 69.6 to 65.9 percent), which confirms that the increase in the number of active firms in the economy (of 3.7 percent) is driven by the increase in the number of formal firms. Second, even though informality is declining, average firm productivity in the economy declines by 1.4 percent. As Panel (a) in Figure 5 shows, this is the result of worsening firm composition within the formal sector: the share of firms in the lower productivity quartile increases by more than 4 percentage points, while the share in the upper quartile declines by around 3 percentage points. Employment composition changes in a similar but less pronounced way. Panel (b) in Figure 5 also documents a very small positive effect on formal firm growth.

Finally, output and taxes increase by 7.1 and 8.7 percent respectively, and therefore decline in per capita terms. This is not surprising given the results on firm composition. Nevertheless, it is worth highlighting that these results are likely to be lower bounds of the effects on productivity and output, as we assume that there are no technological differences between the formal and informal sectors, or differential access to capital. If firms that formalize gain access to cheaper credit or better technologies, then one could expect effects on output and average firm productivity to become more positive.

From Short- to Long-Run: Wage Rigidity in the Formal Sector

The results discussed so far show that the long-run results stand in sharp contrast with the short-run labor market effects of immigration. As discussed in Section 3.6, this discrepancy could be due to frictions in the labor market that constrain formal – but not informal – labor demand expansion in the short-run. This point is important, as whatever friction one considers it cannot constraint informal firms, otherwise there would be no increase in informality in the short-run, only in unemployment. This potentially excludes many frictions that affect firms in general, such as informational frictions. In this subsection, we focus on downward wage rigidity in the formal sector, which plays a central role in the "Harris-Todaro-Fields" framework. More broadly, there is a large literature suggesting that downward wage rigidity is pervasive in developed and developing countries (e.g. Schmitt-Grohé and Uribe, 2016; Grigsby et al., 2021).

We introduce downward wage rigidity in the formal sector, but assume that wages in the informal sector are fully flexible in every period. This could be a reasonable approximation if, for example, there is a binding minimum wage in the formal sector and low inflation, such that real wages take time to adjust, which was the case in Brazil in the 2000's. More generally, formal jobs are more stable, employment contracts are legally binding in the formal sector, and perfect downward nominal wage rigidity is imposed by law in Brazil (nominal wage cuts are not allowed). By definition, informal contracts are not subject to these constraints.

In period $t = 0$ the economy is in the initial steady state, and the immigration shock hits at $t = 1$. We treat this as a "MIT shock", which is an unpredictable shock to the steady state equilibrium of an economy without aggregate shocks (Boppart et al., 2018). We focus on the equilibrium along a perfect-foresight path, and under the assumption that no shock will ever hit this economy again. In the final period, $t = T$, the economy reaches the new steady state equilibrium discussed in the previous section (see Table 6), wages in the formal sector have fully adjusted and wages in both sectors are equalized. We follow Schmitt-Grohé and Uribe (2016) and assume that during the transition between steady states wages in the formal sector evolve as follows: $w_{f,t} = \gamma w_{t-1}$, where we set $\gamma = 0.996$ following their estimates for Argentina. This implies that it takes 10 years for formal wages to reach the new steady state value (see Appendix Section H for more details).

Figure 6 shows the results for the evolution of the share of informal workers and firms. In the immediate short run, both labor and firm informality increase substantially, as the formal sector wage cannot adjust and the labor supply shock in entirely absorbed by informal jobs. Interestingly, the number of informal firms shows limited increase in the first two years (around 1.8 percent), and the employment adjustment largely occurs via the intensive margin of informality. This also explains why the share of informal labor hovers around its initial level for a few years before falling, while firm informality starts declining almost immediately after the shock: the new formal firms tend to be small and lower productivity, and therefore hire a substantial fraction of their labor force informally. Finally, the economy performs worse during the transition when compared to the final steady state when wages have fully adjusted: output, average firm productivity, and tax revenues are substantially lower.

Immigration shock without the informal sector

The previous result that part of the increase in the number of formal firms comes from informal firms formalizing seems to suggest that the informal sector serves as a "stepping stone" for formal firm creation. We examine this hypothesis by combining the labor supply shock with nearly shutting down the extensive margin of informality. This corresponds to making the cost function of operating in the informal sector very steep by substantially reducing the parameter φ_i. Conceptually, this simulates a scenario in which the government devotes a lot of resources to increasing enforcement on informal firms, making it prohibitively costly for firms to operate unregistered, unless they are very small (at most one employee, say). Given the costs of such a policy, we see this counterfactual as a way to analyze the role of the informal sector rather than as a feasible policy experiment.

As Table 7 (column 4) shows, nearly shutting down the extensive margin of informality leads to a substantial reduction in the share of informal firms and workers. However, this comes at the cost of a sizable reduction in the total number of firms in the economy, as many low productivity informal firms do not survive. As a result, the effect on firm composition is positive and average firm productivity increases by 2.5 percent. Consistently, the effect on total output is higher than in the basic labor supply shock scenario (8.3 and 7.1 percent, respectively), and the effect on taxes is much higher, as the economy is now much more formal than in the baseline.

Overall, the economy reaps larger dividends from the labor supply shock when the informal sector is nearly shut down. Thus, the informal sector is dampening the economy's ability to reap the dividends of a growing labor force. This echoes the results from Dix-Carneiro et al. (2021), who show that even though the informal sector acts as an "employment buffer" when the economy faces a negative shock, it is not a "welfare buffer" since the counterfactual with stricter enforcement still leads to a higher welfare.

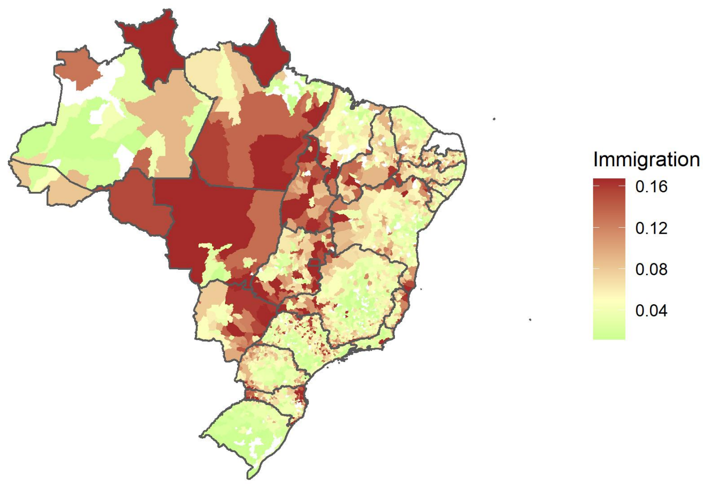
6 Conclusion

In this paper, we study the economic effects of rural-urban migration on Brazilian cities between 2000 and 2010. To identify the causal effect of immigration on firms and workers, we use a shift-share IV design to show that immigration fosters formal employment and lowers wages in the informal and the formal sectors, with no effect on unemployment. We also find large positive effects on the number of formal firms and formal jobs. These results run counter to the view that rural-urban migration only increases the number of informal or unemployed workers in developing country cities (Harris and Todaro, 1970; Fields, 1975), which has been confirmed by the previous literature on the short-run effects of migration (Kleemans and Magruder, 2018; Corbi et al., 2021). We confirm these short-run effects in our context using a year-on-year specification, which suggests that these differences stem from the longer time horizon in our analysis, which examines decadal changes in labor market and firms' outcomes.

To rationalize these surprising results, we develop a structural model of firm dynamics and informality that we calibrate using the same data. The model is able to quantitatively replicate the long-difference, IV results: large labor supply shocks generate wage reductions, and substantial gains in formal employment and formal firm creation. If we impose sluggish wage adjustment in the formal sector during the transition between steady states, we show that labor and firm informality increase in the short-run, but converge to the new lower levels in the longer run. Thus, our results suggest that urban developing economies might experience demographic dividends from internal migration in the long-run even in the presence of frictions, which is reassuring given the expected increase in climatic migration (Rigaud et al., 2018). The extent of these dividends may however depend on the skill level of internal migrants, which in our context is similar to natives'.

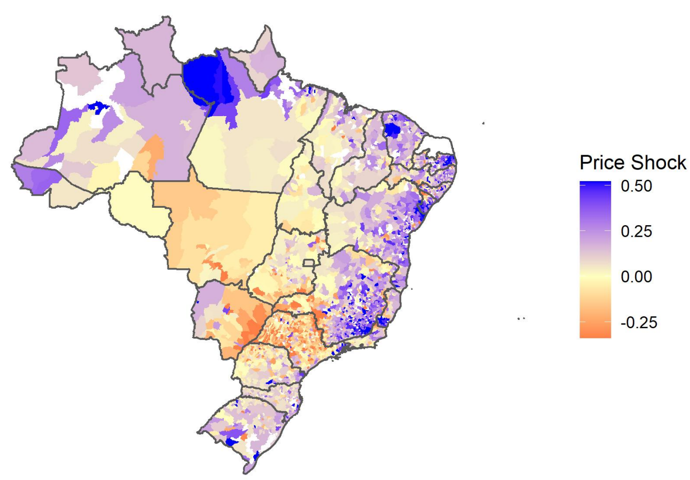
Tables and Figures

Figure 1: In-Migration, 2000-2010



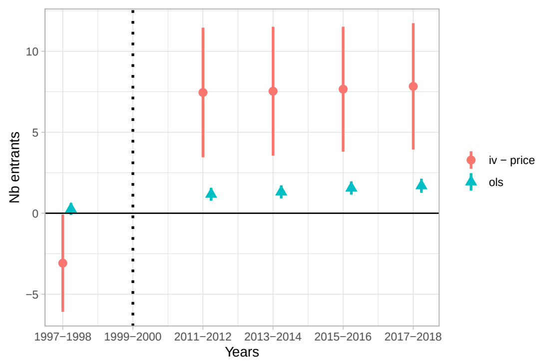
Notes: Computed using the Decennial Population Census. Darker areas denote higher in-migration rates.

Figure 2: Push Shocks: Crop Prices

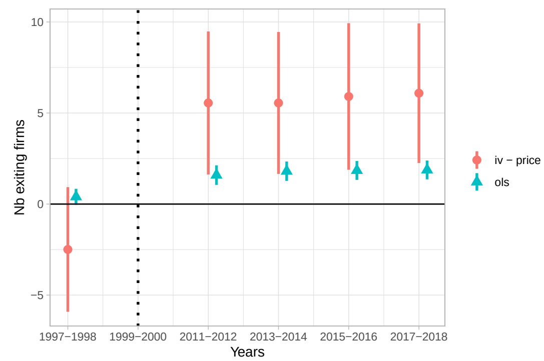


Notes: The Price shocks are constructed according to expression 1 in the text.

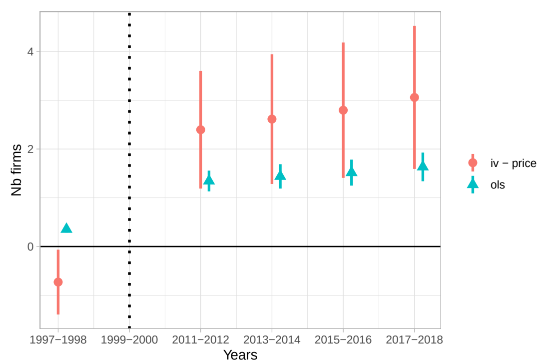
Figure 3: Dynamic Effects



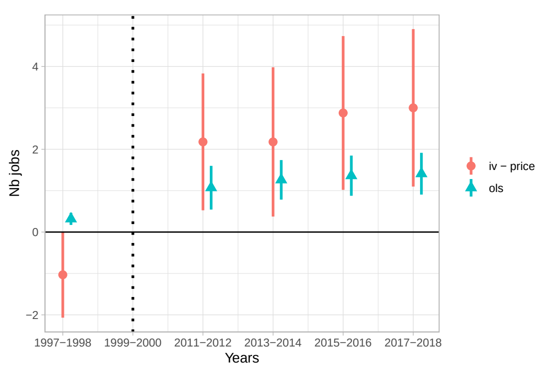
(a) Number of entrants



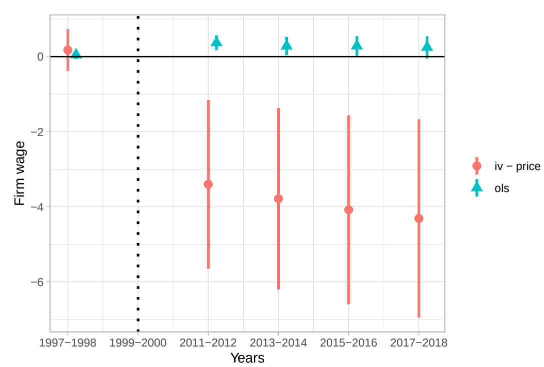
(b) Number of exiting firms



(c) Number of firms

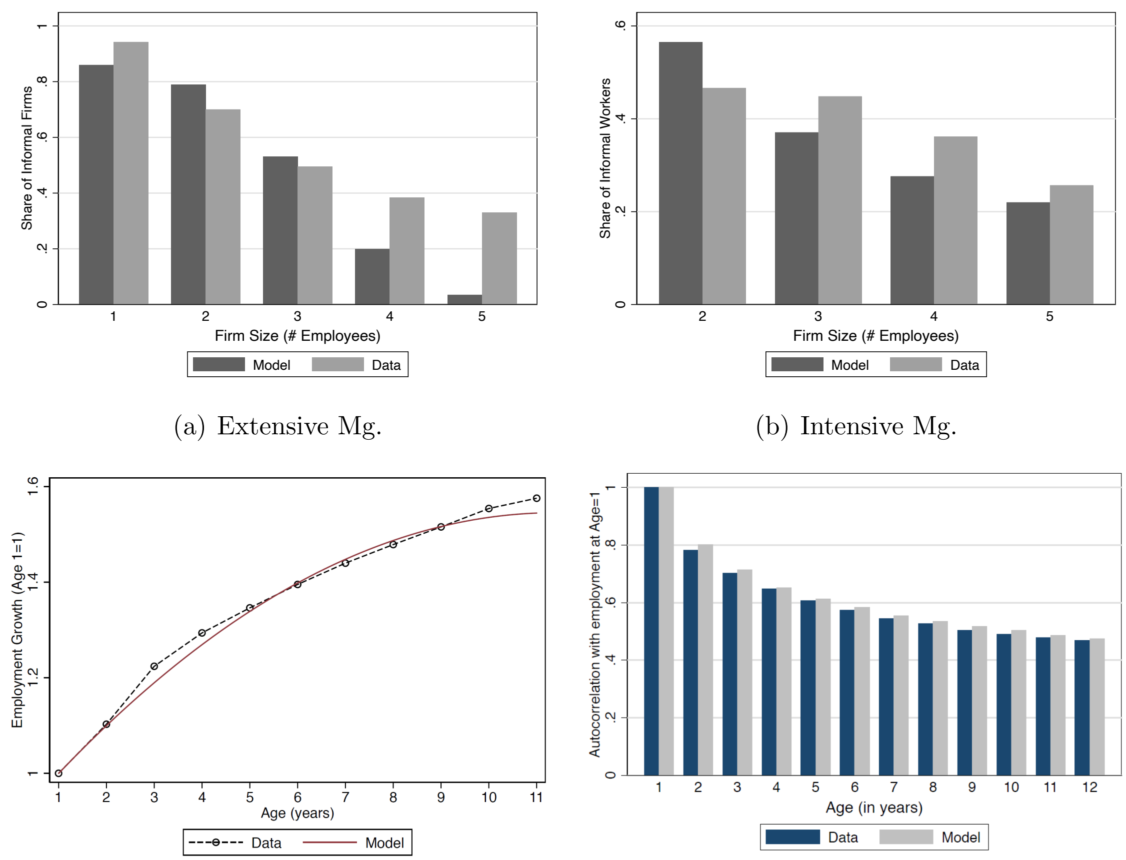


(d) Number of jobs



(e) Firm wage

Figure 4: Model fit: non-targeted moments



(a) Extensive Mg.

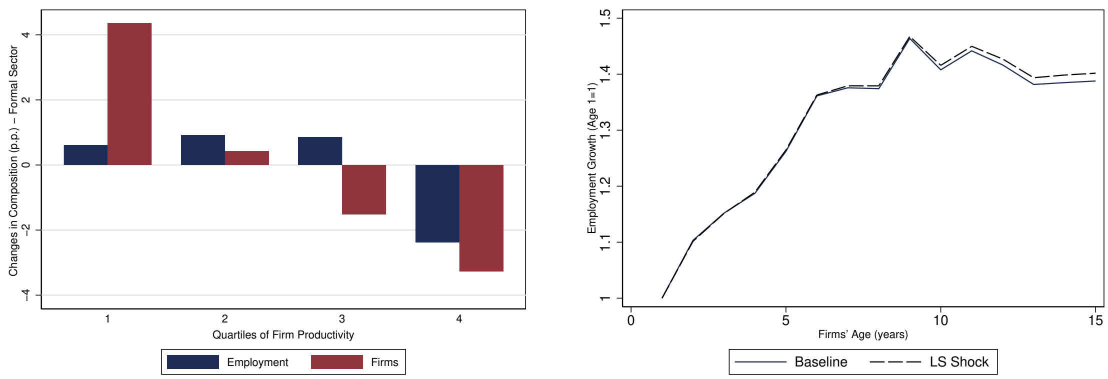
(b) Intensive Mg.

(c) Formal Firms' growth

(d) Log-Employment Autocorrelations

Notes: Size is measured as number of employees. For panels (c) and (d), both simulated and empirical moments use only formal firms and their formal workers. The empirical moments are computed using the panel data from RAIS.

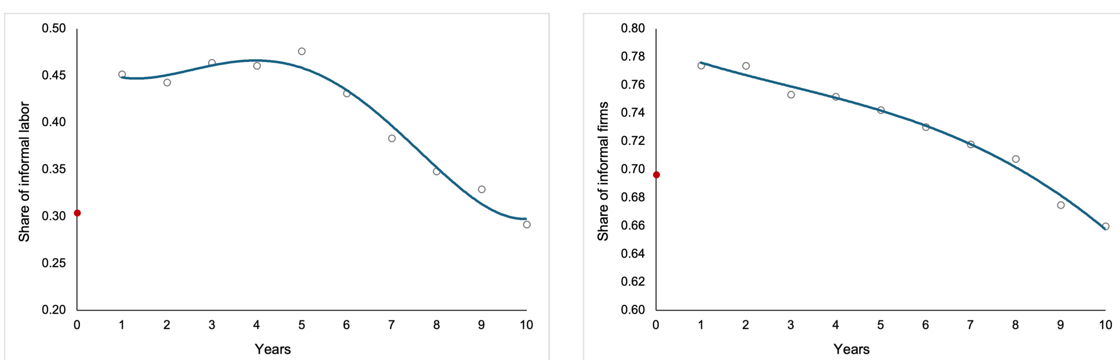
Figure 5: Counterfactuals: Formal Sector Firms



(a) Changes in Composition

(b) Growth Profiles

Figure 6: Transition Dynamics – Informality



(a) Labor Informality

(b) Firm Informality

Notes: Red dot indicates the initial steady state of the baseline economy.

Table 1: Descriptive Statistics

	2000		2010	
	Mean	SD	Mean	SD
Panel A: Socio-Demographics				
% Female	0.483	0.014	0.482	0.013
% Young	0.128	0.021	0.102	0.016
% Low-skilled	0.72	0.096	0.579	0.094
Panel B: Labor Market Outcomes				
% Overall wage employment	0.332	0.061	0.4	0.071
% Formal wage employment	0.229	0.076	0.309	0.096
% Informal wage employment	0.103	0.036	0.09	0.037
Log overall monthly wage	6.886	0.415	6.992	0.321
Log formal monthly wage	7.017	0.361	7.074	0.28
Log informal monthly wage	6.578	0.391	6.712	0.296
Panel C: Firm Outcomes				
Log number of firms	8.728	2.341	9.286	2.195
Log entry	7.169	2.273	7.58	2.117
Log exit	6.918	2.345	7.306	2.14
Log number of jobs	11.303	2.514	11.955	2.389
Log firm monthly wage	7.091	0.455	7.376	0.322
Panel D: Immigration				
Immigration rate			0.176	0.092
Immigration rate (State to state)			0.07	0.054
Panel E: Population				
Population	25135	155267	31436	183082

Notes: For all variables there are 3,545 observations in both 2000 and 2010. Mean and Standard Deviation (SD) in panels A-D are weighted by population in 2000.

Table 2: Effects of Immigration on Workers

	Wage employment			Log monthly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
Panel A: OLS						
Immigration	0.037	0.105	−0.068	0.062	0.032	0.035
	(0.019)	(0.023)	(0.014)	(0.076)	(0.068)	(0.092)
Panel B: IV-Price						
Immigration	0.100	0.394	−0.294	−1.599	−2.168	−1.909
	(0.103)	(0.149)	(0.100)	(0.580)	(0.680)	(0.743)
F Statistic (IV)	16.5	16.5	16.5	16.5	16.5	16.5
Baseline Mean	0.332	0.229	0.103	–	–	–
Observations	3,545	3,545	3,545	3,545	3,545	3,545

Notes: Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table 3: Effects of Immigration on Firms

	Nb firms (1)	Entry (2)	Exit (3)	Nb jobs (4)	Firm wage (5)
Panel A: OLS					
Immigration	1.342	1.167	1.588	1.071	0.372
	(0.108)	(0.203)	(0.274)	(0.270)	(0.101)
Panel B: IV - Price					
Immigration	2.354	7.526	5.583	2.176	−3.436
	(0.622)	(2.085)	(2.045)	(0.855)	(1.169)
F Statistic (IV)	16.5	16.5	16.5	16.5	16.5
Observations	3,545	3,545	3,545	3,545	3,545

Notes: Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. Columns 1-3 refer to the log of total number of firms, entrants and exiting firms, respectively.

Table 4: Parameters of structural model

Parameter	Description	Source	Value
First Step			
τ_w	Payroll Tax	Statutory values	0.375
τ_y	Revenue Tax	Statutory values	0.293
ρ	Productivity Process: Persistence Parameter	GMM Estimation	0.92
ν_0	Pareto's Location Parameter	Calibrated	7.3
γ_f	Per-period fixed cost (Formal)	Calibrated	0.7
Second Step: Min. Dist. Calibration			
φ_f	Intensive margin: $\tau_f = \left(1 + \frac{\ell}{\varphi_f}\right)\ell$	–	5.830
φ_i	Extensive margin: $\tau_i = \left(1 + \frac{\ell}{\varphi_i}\right)\ell$	–	5.427
δ_i	Informal death shock	–	0.148
δ_f	Formal death shock	–	0.066
γ_i	Per-period fixed cost (Informal)	–	0.340
ξ	Pareto shape parameter	–	3.801
$c_f^{e\,\dagger}$	Formal sector's entry cost	–	6,205
$c_i^{e\,\dagger}$	Informal sector's entry cost	–	2,800
α	Span-of-control	–	0.643
σ_i	Informal productivity process: SD	–	0.144
σ_f	Formal productivity process: SD	–	0.145
ρ_i	Informal productivity process: persistence	–	0.935

† Estimates and SD expressed in R$ of 2003.

Table 5: Model Fit – Targeted moments

	Model	Data
Share Informal workers	0.304	0.298
Share Informal Firms	0.696	0.696
Informal Firms Size Distribution		
≤ 2 employees	0.933	0.957
≤ 5 employees	0.999	0.998
Formal Firms Size Distribution		
≤ 5 employees	0.658	0.697
6 to 10	0.136	0.144
11 to 20	0.092	0.083
21 to 50	0.053	0.048
> 50	0.023	0.028

Notes: Data moments computed using the RAIS, ECINF and PNAD data sets.

Table 6: Impacts of Immigration – Model vs. IV Estimation

	IV Estimation	Model
Share Informal Workers (Δ%)	-3.9	-4.1
Wages (Δ%)	-5.7	-3.4
Number Formal Firms (Δ%)	14.7	16.3
Newly created firms	–	9.9
Previously informal firms	–	6.4
Share Informal Firms (Δ%)	–	-5.3
Average Firm Productivity (Δ%)	–	-1.4
Output (Δ%)	–	7.1
Taxes (Δ%)	–	8.7

Notes: IV estimation results from a regression contrasting municipalities at the top and bottom quartiles of immigration rates (see text). Model results from simulating a permanent increase of 10% in population, and effects are measured as percentage change relative to baseline values.

Table 7: Counterfactuals

	Baseline	Labor Supply Shock	LS Shock + Wage Rigidity	LS Shock + Enforcement
Share Informal Labor	0.304	0.291	0.525	0.188
Share Informal Firms	0.696	0.659	0.687	0.221
Wages	1.000	0.966	–	0.979
Formal	1.000	–	1.000	–
Informal	1.000	–	0.664	–
# of Formal Firms	1.000	1.163	1.027	2.280
Avg. Firm Productivity	1.000	0.986	0.900	1.025
Output	1.000	1.071	1.036	1.083
Taxes	1.000	1.087	0.905	1.309

References

Abebe, G., A. S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2021). Anonymity or distance? job search and labour market exclusion in a growing african city. *The Review of Economic Studies 88*(3), 1279–1310.

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

Adda, J., C. Dustmann, and K. Stevens (2017). The career costs of children. *Journal of Political Economy 125*(2), 293–337.

Akcigit, U., S. Baslandze, and S. Stantcheva (2016). Taxation and the international mobility of inventors. *American Economic Review 106*(10), 2930–81.

Albert, C., P. Bustos, and J. Ponticelli (2021, June). The Effects of Climate Change on Labor and Capital Reallocation. CEPR Discussion Papers 16312, C.E.P.R. Discussion Papers.

Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2020). Tackling youth unemployment: Evidence from a labor market experiment in uganda. *Econometrica 88*(6), 2369–2414.

Altonji, J. G., A. A. Smith, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica 81*(4), 1395–1454.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), pp. 277–297.

Bahar, D., A. M. Ibanez, and S. V. Rozo (2021). Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees. *Journal of Development Economics 151*(C).

Bandiera, O., V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2021). The search for good jobs: evidence from a six-year field experiment in uganda. *Available at SSRN 3910330*.

Bertoli, S., F. Docquier, H. Rapoport, and I. Ruyssen (2020). Weather Shocks and Migration Intentions in Western Africa: Insights from a Multilevel Analysis. CESifo Working Paper Series 8064.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115–143.

Boppart, T., P. Krusell, and K. Mitman (2018). Exploiting mit shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control 89*, 68–92.

Borjas, G. J. (2003, November). The Labor Demand Curve Is Downward Sloping: Reexamining The Impact Of Immigration On The Labor Market. *The Quarterly Journal of Economics 118*(4), 1335–1374.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

Bruins, M., J. A. Duffy, M. P. Keane, and A. A. Smith Jr (2015). Generalized indirect inference for discrete choice models. Yale University, mimeo.

Bustos, P., J. M. Castro Vincenzi, J. Monras, and J. Ponticelli (2018, December). Structural Transformation, Industrial Specialization, and Endogenous Growth. CEPR Discussion Papers 13379, C.E.P.R. Discussion Papers.

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. and J. DiNardo (2000). Do immigrant inflows lead to native outflows? *American Economic Review 90*(2), 360–367.

Carranza, E., R. Garlick, K. Orkin, and N. Rankin (2022, November). Job search and hiring with limited information about workseekers' skills. *American Economic Review 112*(11), 3547–83.

Clemens, M. A. and J. Hunt (2019, August). The Labor Market Effects of Refugee Waves: Reconciling Conflicting Results. *ILR Review 72*(4), 818–857.

Corbi, R., T. Ferraz, and R. Narita (2021). Internal migration and labor market adjustments in the presence of nonwage compensation. Manuscript.

de Paula, A. and J. A. Scheinkman (2010). Value-added taxes, chain effects, and informality. *American Economic Journal: Macroeconomics 2*(4), 195–221.

D'Erasmo, P. and H. Boedo (2012). Financial structure, informality and development. *Journal of Monetary Economics 59*(3), pp. 286–302.

Dix-Carneiro, R., P. K. Goldberg, C. Meghir, and G. Ulyssea (2021). Trade and informality in the presence of labor market frictions and regulations. Technical report, National Bureau of Economic Research.

Donovan, K., W. J. Lu, and T. Schoellman (2023). Labor market dynamics and development. Technical report, Unpublished Manuscript.

Duarte, M. and D. Restuccia (2010). The role of the structural transformation in aggregate productivity. *The Quarterly Journal of Economics 125*(1), 129–173.

Dustmann, C. and A. Glitz (2015). How Do Industries and Firms Respond to Changes in Local Labor Supply? *Journal of Labor Economics 33*(3), 711 – 750.

El Badaoui, E., E. Strobl, and F. Walsh (2017). Impact of Internal Migration on Labor Market Outcomes of Native Males in Thailand. *Economic Development and Cultural Change 66*, 147–177.

Elias, F., J. Monras, and J. VÃ¡zquez Grenno (2018, February). Understanding the Effects of Legalizing Undocumented Immigrants. CEPR Discussion Papers 12726, C.E.P.R. Discussion Papers.

Erosa, A., L. Fuster, and T. R. Martinez (2022). Public financing with financial frictions and underground economy. *Journal of Monetary Economics*.

Fields, G. S. (1975). Rural-urban migration, urban unemployment and underemployment, and job-search activity in ldcs. *Journal of development economics 2*(2), 165–187.

Franklin, S. (2018, September). Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies. *Economic Journal 128*(614), 2353–2379.

Gollin, D., R. Jedwab, and D. Vollrath (2016). Urbanization with and without industrialization. *Journal of Economic Growth 21*, 35–70.

Gollin, D., D. Lagakos, and M. E. Waugh (2014). The agricultural productivity gap. *The Quarterly Journal of Economics 129*(2), 939–993.

Gollin, D., S. Parente, and R. Rogerson (2002). The role of agriculture in development. *American economic review 92*(2), 160–164.

Grigsby, J., E. Hurst, and A. Yildirmaz (2021, February). Aggregate nominal wage adjustments: New evidence from administrative payroll data. *American Economic Review 111*(2).

Harris, J. R. and M. P. Todaro (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review 60*(1), pp. 126–142.

Hopenhayn, H., J. Neira, and R. Singhania (2018). From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. Technical report, National Bureau of Economic Research.

Hopenhayn, H. A. (1992, September). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

Hsieh, C.-T. and P. J. Klenow (2014). The life cycle of plants in india and mexico. *The Quarterly Journal of Economics 129*(3), 1035–1084.

Hsieh, C.-T. and B. A. Olken (2014). The missing" missing middle". *Journal of Economic Perspectives 28*(3), 89–108.

Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg (2022, June). Migrants and firms: Evidence from china. *American Economic Review 112*(6), 1885–1914.

Jaeger, D. A., J. Ruist, and J. Stuhler (2018, February). Shift-Share Instruments and the Impact of Immigration. CEPR Discussion Papers 12701, C.E.P.R. Discussion Papers.

Jedwab, R., L. Christiaensen, and M. Gindelsky (2017). Demography, urbanization and development: Rural push, urban pull and...urban push? *Journal of Urban Economics 98*(C), 6–16.

Karahan, F., B. Pugsley, and A. Şahin (2019). Demographic origins of the startup deficit. Technical report, National Bureau of Economic Research.

Kerr, S. P., W. R. Kerr, and W. F. Lincoln (2015). Skilled Immigration and the Employment Structures of US Firms. *Journal of Labor Economics 33*(S1), S147 – S186.

Kleemans, M. and J. Magruder (2018, August). Labour Market Responses To Immigration: Evidence From Internal Migration Driven By Weather Shocks. *Economic Journal 128*(613), 2032–2065.

Lewis, A. (1954). Economic development with unlimited supplies of labour. *The Manchester School 22*(2), 139–191.

Lewis, E. (2011). Immigration, Skill Mix, and Capital Skill Complementarity. *The Quarterly Journal of Economics 126*(2), 1029–1069.

Mitaritonna, C., G. Orefice, and G. Peri (2017). Immigrants and firms' outcomes: Evidence from france. *European Economic Review 96*, 62–82.

Moser, P., A. Voena, and F. Waldinger (2014). German jewish émigrés and us invention. *American Economic Review 104*(10), 3222–55.

Olney, W. W. (2013). Immigration and firm expansion. *Journal of Regional Science 53*(1), 142–157.

Peri, G. (2012, February). The Effect Of Immigration On Productivity: Evidence From U.S. States. *The Review of Economics and Statistics 94*(1), 348–358.

Peters, M. and C. Walsh (2022). Population growth and firm-product dynamics. Technical report, Yale.

Ponczek, V. and G. Ulyssea (2022). Enforcement of labour regulation and the labour market effects of trade: Evidence from brazil. *The Economic Journal 132*.

Pugsley, B. and A. Sahin (2019). Grown-up business cycles. *The Review of Financial Studies 32*(3), 1102–1147.

Quinn, S. and C. Woodruff (2019). Experiments and entrepreneurship in developing countries. *Annual Review of Economics 11*, 225–248.

Rigaud, K. K., A. de Sherbinin, B. Jones, J. Bergmann, V. Clement, K. Ober, J. Schewe, S. Adamo, B. McCusker, S. Heuser, and A. Midgley (2018, March). Groundswell. World Bank Publications - Reports 29461.

Santangelo, G. (2016). Firms and farms: The impact of agricultural productivity on the local indian economy. Manuscript.

Schmitt-Grohé, S. and M. Uribe (2016). Downward nominal wage rigidity, currency pegs, and involuntary unemployment. *Journal of Political Economy 124*(5), 1466–1514.

Sterk, V., P. Sedláček, and B. Pugsley (2021). The nature of firm growth. *American Economic Review 111*(2), 547–79.

Tauchen, G. (1986). Finite state markov-chain approximation to univariate and vector autoregressions. *Economic Letters 20*, 177–81.

The World Bank (2013). World development report 2013: Jobs. Technical report.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from Brazil. *American Economic Review 108*(8), 2015–47.

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics 12*, 525–546.

UNCTAD (2021). Handbook of statistics. Technical report, United Nations Conference on Trade and Development.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate 23*(7), 1696–1718.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics 126*(1), 25 – 51.

ONLINE APPENDIX

A Additional results

We start this Section by reporting First stage results in Table A.1. As the table shows, both instruments are very predictive of immigration flows.

Table A.1: First stage

	Immigration	
	(1)	(2)
Price	−0.053	
	(0.006)	
Drought		0.077
		(0.004)
Observations	3545	3545

Notes: Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

A.1 Additional Results for Workers

In Table A.2 we use the benchmark IV specification to examine the impacts of immigration on occupational structure by looking at all possible employment statuses. Table A.3 reproduces the main results in Table 2, but computing labor market shares using total hours worked (columns 1–3), and log hourly wages for all, formal and informal workers (columns 4-6).

Table A.4 investigates the effects of immigration flows on labor force composition at destination using both IV-price and IV-droughts. We estimate the IV regression described in 2 without any controls, and having as dependent variables the share of females, low skill and young individuals in the population (columns 1–3), population change and out-migration rate, all computed at destination. Table A.5 shows the results estimated separately for high- and low-skilled workers, and Table A.6 separately for migrants and non-migrants.

Finally, Table A.7 investigates whether migrants are more likely to be business owners in municipalities that receive higher inflows of migrants. We do the same looking at non-migrants as well, and by different firm sizes (self-employed, firms with less than 5 employees and firms with at least 6 employees).

Table A.2: Effects on Occupational Composition

	Formal (1)	Informal (2)	Non-emp (3)	Self-emp (4)
Immigration	0.394	−0.294	0.092	−0.041
	(0.149)	(0.100)	(0.129)	(0.047)
Baseline Mean	0.229	0.103	0.435	0.119
Observations	3545	3545	3545	3545
	Employer (5)	Domestic (6)	Public (7)	Non-remun (8)
Immigration	−0.040	0.020	−0.102	−0.029
	(0.020)	(0.031)	(0.061)	(0.022)
Baseline Mean	0.019	0.046	0.038	0.012
Observations	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table A.3: Labor Market Effects using Hours

	Share of hours			Log hourly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
Immigration	−0.015	0.393	−0.408	−1.227	−1.866	−1.529
	(0.109)	(0.145)	(0.127)	(0.515)	(0.614)	(0.713)
Observations	3,545	3,545	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table A.4: Effects on labor force composition at destination

	Female (1)	Low Skill (2)	Young (3)	Delta Population (4)	Out-Migration (5)
Panel A: IV-Price					
Immigration	−0.085	0.035	0.261	0.085	−0.868
	(0.030)	(0.214)	(0.059)	(0.514)	(0.328)
Observations	3,545	3,545	3,545	3,545	3,545
Panel B: IV-Drought					
Immigration	−0.019	−0.194	−0.010	0.478	0.125
	(0.013)	(0.119)	(0.020)	(0.246)	(0.111)
Observations	3,545	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions have only one regressor: immigration. All regressions are weighted by municipality's population in 2000.

Table A.5: Labor Market Effects by Skill Level

	Wage employment			Log monthly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
Panel A: High-Skilled Workers						
Immigration	0.130	0.364	−0.235	−1.313	−1.610	−1.076
	(0.125)	(0.172)	(0.104)	(0.455)	(0.524)	(0.670)
Observations	3,545	3,545	3,545	3,545	3,524	3,511
Panel B: Low-Skilled Workers						
Immigration	0.039	0.325	−0.286	−1.876	−2.268	−2.462
	(0.107)	(0.109)	(0.096)	(0.746)	(0.791)	(0.994)
Observations	3,545	3,545	3,545	3,545	3,543	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table A.6: Labor Market Effects by Migration status

	Wage employment			Log monthly wage		
	Overall	Formal	Informal	Overall	Formal	Informal
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Non-migrants						
Immigration	0.092	0.369	−0.276	−2.088	−2.632	−2.310
	(0.103)	(0.145)	(0.100)	(0.696)	(0.801)	(0.841)
Observations	3,545	3,545	3,545	3,545	3,543	3,545
Panel B: Migrants						
Immigration	0.159	0.475	−0.316	0.366	−0.239	−0.266
	(0.140)	(0.183)	(0.100)	(0.525)	(0.569)	(0.797)
Observations	3,545	3,545	3,545	3,540	3,435	3,524

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table A.7: Effects on Entrepreneurship by firm size

	All	Non-migrants	Migrants
	(1)	(2)	(3)
Panel A: Self-employed			
Immigration	−0.041	−0.064	0.136
	(0.047)	(0.055)	(0.097)
Baseline Mean	0.119	0.121	0.111
Panel B: Firms with ≤ 5 employees			
Immigration	−0.026	−0.040	0.025
	(0.013)	(0.016)	(0.024)
Baseline Mean	0.013	0.013	0.011
Panel C: Firms with ≥ 6 employees			
Immigration	−0.014	−0.020	0.011
	(0.010)	(0.012)	(0.012)
Baseline Mean	0.006	0.006	0.005
Observations	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

B Robustness

Table B.1: Effects of Immigration on Workers - Additional Controls

	Wage employment			Log monthly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
Panel A: Controlling for Lagged Outcome						
Immigration	0.040	0.297	−0.236	−1.474	−2.388	−2.184
	(0.102)	(0.117)	(0.084)	(0.510)	(0.788)	(0.825)
Observations	3,545	3,545	3,545	3,539	3,483	3,524
Panel B: Controlling for Lagged Population						
Immigration	0.188	0.471	−0.283	−1.219	−1.793	−1.321
	(0.106)	(0.169)	(0.101)	(0.417)	(0.533)	(0.522)
Observations	3,545	3,545	3,545	3,545	3,545	3,545
Panel C: Controlling for Lagged log(Migration)						
Immigration	0.102	0.402	−0.300	−1.286	−2.013	−1.504
	(0.113)	(0.168)	(0.114)	(0.471)	(0.633)	(0.610)
Observations	3,545	3,545	3,545	3,545	3,545	3,545
Panel D: Controlling for Lagged log(GDP)						
Immigration	0.117	0.421	−0.304	−1.245	−1.982	−1.404
	(0.111)	(0.168)	(0.113)	(0.438)	(0.592)	(0.565)
Observations	3,545	3,545	3,545	3,545	3,545	3,545
Panel E: Controlling for Lagged Industrial Composition						
Immigration	0.020	0.296	−0.276	−1.172	−1.820	−1.505
	(0.093)	(0.124)	(0.091)	(0.491)	(0.598)	(0.651)
Observations	3,545	3,545	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. Panel A has fewer units for wages because data on wages is missing for some municipalities in the census 1991.

Table B.2: Effects of Immigration on Firms - Additional Controls

	Nb firms (1)	Entry (2)	Exit (3)	Nb jobs (4)	Firm wage (5)
Panel A: Controlling for Lagged Outcome					
Immigration	2.505	5.423	3.725	1.954	−3.412
	(0.579)	(1.197)	(0.923)	(0.838)	(1.160)
Observations	3,545	3,545	3,545	3,545	3,545
Panel B: Controlling for Lagged Population					
Immigration	2.392	7.414	5.962	2.085	−2.997
	(0.652)	(2.134)	(2.116)	(0.895)	(1.033)
Observations	3,545	3,545	3,545	3,545	3,545
Panel C: Controlling for Lagged log(Migration)					
Immigration	2.357	8.522	6.842	2.188	−3.010
	(0.698)	(2.593)	(2.573)	(0.947)	(1.050)
Observations	3,545	3,545	3,545	3,545	3,545
Panel D: Controlling for Lagged log(GDP)					
Immigration	2.439	8.567	6.969	2.262	−2.923
	(0.692)	(2.508)	(2.541)	(0.922)	(0.991)
Observations	3,545	3,545	3,545	3,545	3,545
Panel E: Controlling for Lagged Industrial Composition					
Immigration	1.899	6.634	5.163	2.153	−2.578
	(0.537)	(1.886)	(1.898)	(0.820)	(0.945)
Observations	3,545	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. Columns 1-3 refer to the log of total number of firms, entrants and exiting firms, respectively.

Table B.3: Effects of Immigration on Workers - Alternative Channels

	Wage employment			Log monthly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
Panel A: Controlling for local and neighboring municipalities' price shocks						
Immigration	0.065	0.297	−0.232	−1.921	−2.399	−2.321
	(0.099)	(0.122)	(0.074)	(0.601)	(0.678)	(0.755)
Observations	3,545	3,545	3,545	3,545	3,545	3,545
Panel B: Controlling for capital reallocation channel						
Immigration	0.026	0.303	−0.276	−1.812	−2.325	−2.243
	(0.115)	(0.151)	(0.105)	(0.693)	(0.791)	(0.879)
Observations	2,627	2,627	2,627	2,627	2,627	2,627

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. Panel A has 3538 units because data on local prices is missing for 10 municipalities. Panel B has 2630 units because data on banks is missing for 918 municipalities.

Table B.4: Effects of Immigration on Firms - Alternative Channels

	Nb firms (1)	Entry (2)	Exit (3)	Nb jobs (4)	Firm wage (5)
Panel A: Controlling for local and neighboring municipalities' price shocks					
Immigration	2.158	4.697	2.667	2.219	−3.920
	(0.530)	(1.370)	(1.411)	(0.739)	(1.181)
Observations	3,545	3,545	3,545	3,545	3,545
Panel B: Controlling for capital reallocation channel					
Immigration	2.415	7.391	4.979	2.527	−3.494
	(0.692)	(2.266)	(2.223)	(0.941)	(1.336)
Observations	2,627	2,627	2,627	2,627	2,627
Panel C: Excluding firms that produce agricultural goods					
Immigration	2.785	7.513	5.542	2.395	−3.823
	(0.648)	(2.054)	(2.039)	(0.885)	(1.247)
Observations	3,545	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. Panel A has 3538 units because data on local prices is missing for 10 municipalities. Panel B has 2630 units because data on banks is missing for 918 municipalities. Columns 1-3 refer to the log of total number of firms, entrants and exiting firms, respectively.

Table B.5: Effects on Workers – Transformation from destinations to origins following (Borusyak et al., 2022)

	Wage employment			Log monthly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
			Panel A: OLS			
Transformed Immigration	0.030	0.146	−0.116	0.105	−0.049	0.100
	(0.014)	(0.014)	(0.009)	(0.103)	(0.098)	(0.125)
Observations	3,545	3,545	3,545	3,545	3,545	3,545
			Panel B: IV-Price			
Transformed Immigration	0.102	0.395	−0.293	−1.616	−2.163	−1.933
	(0.066)	(0.078)	(0.073)	(0.605)	(0.615)	(0.766)
F Statistic (IV)	11.82	11.82	11.82	11.82	11.82	11.82
Observations	3,545	3,545	3,545	3,545	3,545	3,545
			Panel C: IV-Price – Controlling for Lagged Price shock			
Transformed Immigration	0.070	0.372	−0.303	−1.603	−2.163	−1.961
	(0.068)	(0.080)	(0.075)	(0.625)	(0.633)	(0.787)
Observations	3,545	3,545	3,545	3,545	3,545	3,545
			Panel D: Reduced form			
Price shock	−0.0005	−0.002	0.001	0.008	0.010	0.009
	(0.0004)	(0.0005)	(0.0003)	(0.003)	(0.003)	(0.004)
Observations	3,545	3,545	3,545	3,545	3,545	3,545

Notes: Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured at the origin municipality in 2000. The dependent variables and Transformed Immigration are variables at destination (outcomes and the immigration rate) transformed into variables across origins, through a combination of (i) migration patterns between origins and destinations, and (ii) population within destinations at baseline. All regressions are weighted in line with (Borusyak et al., 2022) in the origin municipality.

Table B.6: Effects on Firms – Transformation from destinations to origins following (Borusyak et al., 2022)

	Nb firms (1)	Entry (2)	Exit (3)	Nb jobs (4)	Firm wage (5)
	Panel A: OLS				
Transformed Immigration	1.583	1.429	1.962	1.213	0.295
	(0.067)	(0.238)	(0.378)	(0.129)	(0.186)
Observations	3,545	3,545	3,545	3,545	3,545
	Panel B: IV-Price				
Transformed Immigration	2.343	7.485	5.540	2.171	−3.407
	(0.524)	(2.197)	(2.566)	(0.849)	(1.016)
F Statistic (IV)	11.82	11.82	11.82	11.82	11.82
Observations	3,545	3,545	3,545	3,545	3,545
	Panel C: IV-Price – Controlling for Lagged Price shock				
Transformed Immigration	2.321	7.575	5.630	2.107	−3.359
	(0.530)	(2.242)	(2.620)	(0.856)	(1.048)
Observations	3,545	3,545	3,545	3,545	3,545
	Panel D: Reduced form				
Price shock	−0.011	−0.036	−0.027	−0.011	0.016
	(0.005)	(0.006)	(0.009)	(0.006)	(0.006)
Observations	3,545	3,545	3,545	3,545	3,545

Notes: Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured at the origin municipality in 2000. The dependent variables and Transformed Immigration are variables at destination (outcomes and the immigration rate) transformed into variables across origins, through a combination of (i) migration patterns between origins and destinations, and (ii) population within destinations at baseline. All regressions are weighted in line with (Borusyak et al., 2022) in the origin municipality. Columns 1-3 refer to the log of total number of firms, entrants and exiting firms, respectively.

C Changes in the Composition of Firms

In all tables, we report the effect in terms of the fraction of firms (Panel A) and workers (Panel B). Table C.1 shows the results relative to the sectoral composition of firms and jobs, while Table C.2 shows the effects across firm size bins (measured as number of employees). Finally, Table C.3 reports the effects on the share of firms and jobs across quartiles of firm quality (see text for the measure of firm quality).

Table C.1: Effects on Industry Composition (2011-2012)

Industries:	Retail and Services (1)	Construction (2)	Manufacturing (3)	Other Sectors (4)
	Panel A: Shares of Firms			
Immigration	0.226	0.176	−0.303	−0.099
	(0.150)	(0.053)	(0.136)	(0.136)
Baseline Mean	0.738	0.033	0.111	0.118
	Panel B: Shares of Jobs			
Immigration	0.373	−0.114	−0.341	0.082
	(0.402)	(0.109)	(0.257)	(0.408)
Baseline Mean	0.465	0.041	0.185	0.309
Observations	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table C.2: Effects on Firm Size Composition (2011-2012)

Sizes of firms:	≤5 (1)	6 to 10 (2)	11 to 20 (3)	21 to 50 (4)	>50 (5)
			Panel A: Shares of Firms		
Immigration	0.367	−0.161	−0.122	−0.079	−0.005
	(0.138)	(0.068)	(0.048)	(0.041)	(0.033)
Baseline Mean	0.706	0.131	0.078	0.048	0.036
			Panel B: Shares of Jobs		
Immigration	0.050	−0.067	−0.136	−0.087	0.240
	(0.095)	(0.062)	(0.072)	(0.103)	(0.258)
Baseline share	0.129	0.079	0.086	0.112	0.594
Observations	3,545	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table C.3: Effects on Firm Quality Composition (2011-2012)

Quartiles:	Bottom (1)	Mid-bottom (2)	Mid-top (3)	Top (4)
		Panel A: Shares of Firms		
Immigration	−0.393	2.686	1.250	−3.543
	(0.626)	(0.976)	(1.239)	(1.487)
		Panel B: Shares of Jobs		
Immigration	−0.165	0.951	1.083	−1.869
	(0.310)	(0.440)	(0.733)	(1.072)
Observations	3,545	3,545	3,545	3,545

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. We proxy firm quality by the residual of firms' average wage regressed on year and micro-region fixed effects.

D Heterogeneity

Table D.1: Effect of Immigration on Workers by Baseline GDP p.c. Terciles

	Wage employment			Log monthly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
			Panel A: Bottom			
Immigration	0.293	0.380	−0.087	−0.265	−0.906	−0.399
	(0.112)	(0.116)	(0.087)	(0.360)	(0.457)	(0.389)
F Statistic (IV)	10.95	10.95	10.95	10.95	10.95	10.95
Observations	1,170	1,170	1,170	1,170	1,170	1,170
			Panel B: Middle			
Immigration	0.330	0.637	−0.307	−0.644	−0.916	−0.777
	(0.126)	(0.173)	(0.113)	(0.541)	(0.552)	(0.739)
F Statistic (IV)	19.82	19.82	19.82	19.82	19.82	19.82
Observations	1,170	1,170	1,170	1,170	1,170	1,170
			Panel C: Top			
Immigration	−0.402	0.318	−0.720	−3.222	−4.017	−4.035
	(0.308)	(0.325)	(0.474)	(2.333)	(2.814)	(2.901)
F Statistic (IV)	2.19	2.19	2.19	2.19	2.19	2.19
Observations	1,205	1,205	1,205	1,205	1,205	1,205

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. GDP per capita is non-agricultural GDP per urban population measured in 1999-2000.

Table D.2: Effect of Immigration on Firms by Baseline GDP p.c. Terciles

	Nb firms (1)	Entry (2)	Exit (3)	Nb jobs (4)	Firm wage (5)
Panel A: Bottom					
Immigration	3.058	4.538	4.045	2.967	-1.969
	(0.913)	(1.729)	(1.936)	(1.196)	(0.803)
F Statistic (IV)	10.95	10.95	10.95	10.95	10.95
Observations	1,170	1,170	1,170	1,170	1,170
Panel B: Middle					
Immigration	0.611	4.617	3.166	1.415	-2.185
	(0.782)	(2.518)	(3.040)	(1.136)	(1.049)
F Statistic (IV)	19.82	19.82	19.82	19.82	19.82
Observations	1,170	1,170	1,170	1,170	1,170
Panel C: Top					
Immigration	1.611	9.249	5.330	0.939	-6.799
	(1.216)	(5.817)	(3.248)	(1.538)	(5.141)
F Statistic (IV)	2.19	2.19	2.19	2.19	2.19
Observations	1,205	1,205	1,205	1,205	1,205

Notes: IV-price estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. GDP per capita measured in 1999-2000. Columns 1-3 refer to the log of total number of firms, entrants and exiting firms, respectively.

E Short- vs. Long-Run Effects

E.1 Construction of the drought shocks

We follow Bertoli et al. (2020); Albert et al. (2021) and use the SPEI (Standardized Precipitation-Evapotranspiration Index), which measures the water balance based on precipitation and evapo-transpiration due to temperature (Vicente-Serrano et al., 2010).[23] The data are available on a geo-referenced grid which we match to Brazilian municipalities.

We construct an indicator for drought which corresponds to negative values of the SPEI for each month m in each municipality of origin o (D_{om}). From the agricultural census, we also build an indicator g_{ocm} equal to one if the crop c is growing in municipality o in month m. Finally, we create the drought shock as the combination of the drought indicator in the growing season of each group, with weights equal to the agricultural value the crop for the municipality.

For the long run (decadal change) analysis in Section E.2, we proceed in an analogous way to the price shock used in our main specification, and we accumulate all shocks throughout the decade. The drought shock is thus given by:

$$s_o^{drought} = \sum_m \sum_c (\pi_{oc} \times g_{ocm} \times D_{om}) \tag{16}$$

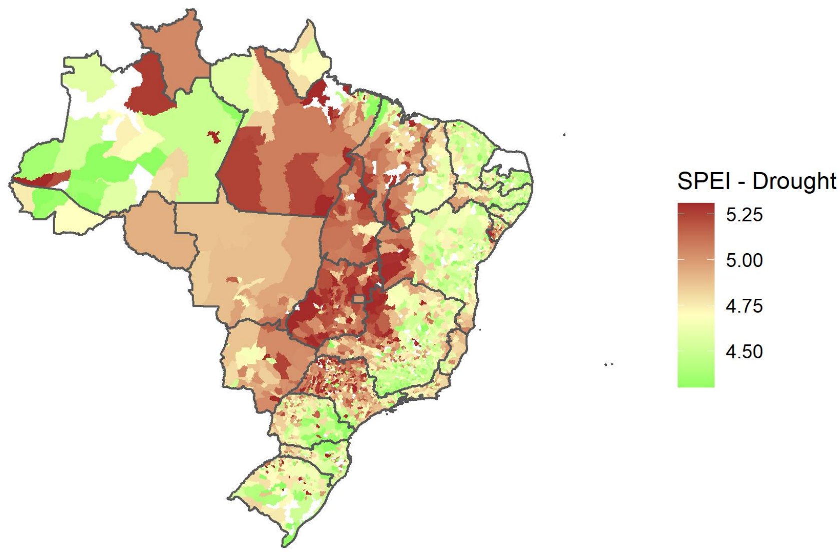
For the short run (year-on-year) specification, we explore the higher frequency of the shock and aggregate up to year t, so that the instrument is now given by:

$$s_{o,t}^{drought} = \sum_{m=t_0}^{t} \sum_c (\pi_{oc} \times g_{ocm} \times D_{om}) \tag{17}$$

Figure E.1 shows the spatial distribution of the decadal drought shock given by expression 16. As mentioned in the main text, the price and drought shocks are independent across origins, with a correlation of 0.007.

[23]The data can be freely downloaded here https://spei.csic.es/home.html

Figure E.1: Droughts (SPEI measure)



Notes: The Drought shocks are constructed according to expression 16 in the text.

E.2 Long Differences (Decadal Changes) using Droughts as Push Shocks

We start by re-estimating our main specification described in expression 2, but instead of using price shocks as shifts in the shift-share instrument, we use the drought shock described in the previous section. Thus, we combine the same variation across destination municipalities in their pre-existing migration networks with different origins (the "share"), and time variation in drought shocks that affect migration incentives at origin (the "shift"). Formally, the instrument writes:

$$Z_d = \sum_o \lambda_{o,d} s_o^{drought} \tag{18}$$

where $\lambda_{o,d}$ denotes the share of migrants from origin o among migrants who had come at destination d between 1995 and 2000; and $s_o^{drought}$ is given by expression 16. We then use Z_d as an instrument for Mig_d in a 2SLS estimator.

Table E.1: Long-Run Effects on Workers – Drought Shocks

	Wage employment			Log monthly wage		
	Overall (1)	Formal (2)	Informal (3)	Overall (4)	Formal (5)	Informal (6)
IV-Drought Immigration	−0.014	0.271	−0.284	−0.126	−0.671	−0.200
	(0.060)	(0.089)	(0.072)	(0.284)	(0.336)	(0.352)
F Statistic (IV)	18.11	18.11	18.11	18.11	18.11	18.11
Observations	3,545	3,545	3,545	3,545	3,545	3,545

Notes: IV-drought estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table E.2: Long-Run Effects on Firms – Drought Shocks

	Nb firms (1)	Entry (2)	Exit (3)	Nb jobs (4)	Firm wage (5)
IV - Drought Immigration	1.625	2.555	2.758	2.031	−0.747
	(0.306)	(0.877)	(1.123)	(0.624)	(0.554)
F Statistic (IV)	18.11	18.11	18.11	18.11	18.11
Observations	3,545	3,545	3,545	3,545	3,545

Notes: IV-drought estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000. Columns 1-3 refer to the log of total number of firms, entrants and exiting firms, respectively.

E.3 Long Differences (Decadal Changes) using Droughts as Push Shocks and Municipalities Available in PNAD

In this subsection, we re-estimate the regression described in Section E.2 above, but restricting the sample to the 700 municipalities identifiable in the National Household Survey (PNAD).

Table E.3: Long-Run Effects on Workers – Drought Shocks with PNAD Municipalities

	Wage employment			Log monthly wage		
	Overall	Formal	Informal	Overall	Formal	Informal
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Immigration	0.004	0.067	−0.062	0.059	0.024	0.044
	(0.022)	(0.023)	(0.018)	(0.099)	(0.086)	(0.124)
Panel B: IV-Price						
Immigration	−0.091	0.148	−0.239	−1.663	−2.102	−1.906
	(0.103)	(0.119)	(0.101)	(0.682)	(0.792)	(0.831)
F Statistic (IV)	11.16	11.16	11.16	11.16	11.16	11.16
Baseline Mean	0.338	0.243	0.095	-	-	-
Observations	700	700	700	700	700	700

Notes: IV-drought estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table E.4: Long-Run Effects on Firms – Drought Shocks with PNAD Municipalities

	Nb firms	Entry	Exit	Nb jobs	Firm wage
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Immigration	1.215	1.296	1.870	0.922	0.498
	(0.119)	(0.283)	(0.365)	(0.319)	(0.134)
Panel B: IV-Price					
Immigration	2.310	6.632	5.126	2.089	−2.609
	(0.644)	(2.234)	(2.223)	(0.910)	(1.203)
F Statistic (IV)	11.16	11.16	11.16	11.16	11.16
Observations	700	700	700	700	700

Notes: IV-drought estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

E.4 Short-Run Results

We now turn to the short run, year-on-year specification discussed in Section 3.6. We construct the instrument in a similar way to the one described in Section E.2 (expression 18), but we fully explore the annual variation in the drought shocks:

$$Z_{d,t} = \sum_{o} \lambda_{o,d} s_{o,t}^{drought}$$

where $s_{o,t}^{drought}$ is given by expression 17.

Table E.5 presents the results for the short-run effects of immigration on workers, while Table E.6 presents the results for firm outcomes.

Table E.5: Short-Run Effects on Workers

	Wage employment			Log monthly wage		
	Overall	Formal	Informal	Overall	Formal	Informal
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Immigration	0.099	0.087	0.011	0.162	0.285	−0.176
	(0.040)	(0.040)	(0.022)	(0.175)	(0.188)	(0.343)
Panel B: IV-Drought						
Immigration	−1.233	−1.199	−0.034	0.512	1.226	−1.890
	(0.624)	(0.576)	(0.324)	(2.080)	(2.507)	(3.256)
F Statistic (IV)	21.53	21.53	21.53	21.53	21.53	21.61
Baseline Mean	0.335	0.242	0.092	-	-	-
Observations	6,407	6,407	6,407	6,407	6,381	6,377

Notes: IV-drought estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

Table E.6: Short-Run Effects on Firms

	Nb firms	Entry	Exit	Nb jobs	Firm wage
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Immigration	0.106	−0.081	−0.246	−0.011	0.077
	(0.024)	(0.105)	(0.063)	(0.076)	(0.042)
Panel B: IV-Drought					
Immigration	0.704	2.921	3.079	−14.425	−0.609
	(0.332)	(4.115)	(0.996)	(5.211)	(0.528)
F Statistic (IV)	21.52	21.52	21.52	21.52	21.52
Observations	6,382	6,382	6,382	6,382	6,382

Notes: IV-drought estimation. Robust standard errors in parenthesis. All regressions control for the share of women, youth (less than 18 years old), and high skill individuals (at least completed high-school) measured in 2000. All regressions are weighted by municipality's population in 2000.

F Model Appendix

This section contains the propositions and corollaries discussed in Section 4 and their respective proofs. For notational simplicity, we discuss the case without heterogeneous growth profiles, that is, without a firm-specific intercept in the productivity process.

Proposition 1 ***(existence and uniqueness of the stationary equilibrium):*** *Under the assumptions that (i) the general law of motion of the productivity process in both sectors, denoted here by $G(\theta'|\theta)$, is continuous in both arguments and strictly decreasing in θ; and (ii) the profit functions are such that q_s and ℓ_s are continuous, single-valued and strictly increasing in θ (as in 5 and 6). Then, there exists a unique stationary competitive equilibrium where both the formal and informal sector exist, and both sectors have positive entry and exit.*

Proof: The proof of this proposition is organized in several steps.

Step 1: Properties of the profit and value functions

The assumption about the production and cost functions in both sectors implies that the profit functions are continuous and that q_j^* and l_j^* are continuous, single valued, and strictly increasing in θ. Write the corresponding equilibrium prices as

$$w^*(\mu) \;=\; W\Big(L^*(\mu, w(\mu))\Big) \tag{19}$$

It follows directly from Hopenhayn (1992), Lemma 3, that the function $w^*(\mu)$ is well defined and continuous. Using 19, one can re-write the value functions in each sector as follows:

$$V_f(\theta,\mu) \;=\; \tilde{\pi}_f(\theta,\mu) + (1-\delta_f)\beta \max\left\{0, E\left[V_f(\theta',\mu)\left|\theta\right.\right]\right\} \tag{20}$$

$$\begin{aligned} V_i(\theta,\mu) \;=\;& \tilde{\pi}_i(\theta,\mu) \\ &+\beta \max\left\{0, (1-\delta_i)E\left[V_i(\theta',\mu)\left|\theta\right.\right], (1-\delta_f)E\left[V_f(\theta',\mu)\left|\theta\right.\right] - \tilde{c}^e\right\} \end{aligned} \tag{21}$$

where $\tilde{\pi_s}(\theta,\mu) \equiv \pi_s(\theta, w^*(\mu))$.

Lemma 1: *The functions $\tilde{\pi}_s$ are continuous in both arguments, strictly increasing in θ and decreasing in μ.*

Proof: The proof follows directly from the proof in Hopenhayn (1992). Continuity follows from the continuity of the profit functions in θ and w, and from continuity of

$w^*(\mu)$. Since $p^*(\mu) > 0$, $\forall\mu$, it also follows from the properties of the profit function that $\tilde{\pi}_s$ is increasing in θ.

It remains to show that $\tilde{\pi}_j$ is decreasing in μ. Let $\mu_2 > \mu_1$ with the corresponding equilibrium prices w_1^* and w_2^*. Suppose by way of contradiction that $w_2^* < w_1^*$. This implies that $q_s(\theta, w_2^*) > q_s(\theta, w_1^*)$ $\forall\theta$ and therefore $Q_s^*(\mu_2) = \int q_s(\theta, w_2^*)\, d\mu_2^j(\theta) > \int q_s(\theta, w_1^*)\, d\mu_1^j(\theta) = Q_s^*(\mu_1)$ (the last inequality follows because q_j is strictly increasing in θ), $s = i, f$. But if $Q^*(\mu_2) > Q^*(\mu_1)$, then $L^*(\mu_2) > L^*(\mu_1)$ and therefore $w_2^* > w_1^*$ (as labor supply is fixed), which contradicts the initial assumption that $w_2^* < w_1^*$. Hence, if $\mu_2 > \mu_1$ then $w_2^* > w_1^*$. Hence, one can easily verify that $\tilde{\pi}_s(\theta, \mu_2) \equiv \tilde{\pi}_s(\theta, w_2^*) < \tilde{\pi}_s(\theta, w_1^*) \equiv \tilde{\pi}_s(\theta, \mu_1)$, which establishes the desired result. $\square$

Given these properties of the profit functions, we can go ahead and establish the properties of the value functions:

LEMMA 2: *The functions, V_s, that solve 20 and 21 are unique and have the following properties: (i) they are continuous functions; (ii) strictly increasing in θ and decreasing in μ; and (iii) the option value in both sectors (the integral terms in 20 and 21) is strictly increasing in θ.*

Proof: Given the assumptions about the productivity process and the properties of $\tilde{\pi}_j$ established in Lemma 1, the properties (i) and (ii) follow directly from standard dynamic programming arguments. The same can be said about (iii), except that it also relies on (ii). Note that the presence of V_f in the option value of the informal firm does not alter the argument, as the value functions in both sectors share the same properties. $\square$

Step 3: Uniqueness of the cut-offs for entry, exit and informal to formal transitions.

Due to the properties of the value functions established in Lemma 2, it follows that $\underline{\theta}_s, \underline{\nu}_s$, and $\overline{\theta}_i$, are uniquely determined by the following expressions:

$$\int V_s(\theta', \mu)\, dF\left(\theta' \middle| \underline{\theta}_s\right) = 0, \qquad s = i, f \tag{22}$$

$$\int V_i(\theta, z)\, dF\left(\theta \middle| \underline{\nu}_i\right) = c_i^e \tag{23}$$

$$\int V_f(\theta, z)\, dF\left(\theta \middle| \underline{\nu}_f\right) = V_i^e\left(\underline{\nu}_f, w\right) - \left(c_i^e - c_f^e\right) \tag{24}$$

$$\int \left[V_f(\theta', \mu) - V_i(\theta', \mu)\right] dF\left(\theta' \middle| \overline{\theta}_i\right) = \tilde{c}^e \tag{25}$$

Using assumption (A.2.ii) and the functional forms assumed for the production and cost functions in each sector, one can verify that $\underline{\theta}_f > \underline{\theta}_i$. This will be the case even if at $\theta = \underline{\theta}_f$ the formal firm hires all of its employees informally. Using a similar reasoning, $c_f^e > c_i^e$ implies that $\underline{\nu}_f > \underline{\nu}_i$.

Finally, condition 25 requires a more careful explanation. There is a productivity threshold $\overline{\theta}$ above which $\pi_f(\theta) > \pi_i(\theta)$, for any $\theta > \overline{\theta}$. Hence, from the results proved in Lemmas 1 and 2 it follows that, if there is informal to formal transition, then the threshold $\overline{\theta}_i$ defined by 25 is unique.

Step 4: The industry measures

Following Hopenhayn (1992), we define the operator $\hat{P}_k$ as follows:

$$\hat{P}_k \equiv \hat{P}_k(\theta, B) = \begin{cases} \int_B dF\left(x \middle| \theta\right) & \text{if} \quad \theta \in I_k \\ 0 & \text{otherwise} \end{cases}$$

for all Borel sets $B \subset \Theta$; where $I_1 = \left[\underline{\theta}_i, \overline{\theta}_i\right)$, $I_2 = \left[\overline{\theta}_i, \infty\right)$, and $I_3 = \left[\underline{\theta}_f, \infty\right)$. This is a bounded $\left(\left\|\hat{P}_k\right\| \leq 1\right)$, linear operator on the space of positive bounded measures: $\hat{P}_k\mu(B) = \int \hat{P}_k(\theta, B)\, d\mu(\theta)$, for all Borel sets $B \subset \Theta$ (see Hopenhayn, 1992). Using this notation, one can rewrite the law of motion of both sectors' measure as follows:

$$\mu_i' = \hat{P}_1\mu_i + \Lambda_i'\tilde{G}_i' \tag{26}$$
$$\mu_f' = \hat{P}_3\mu_f + \Lambda_f'\tilde{G}_f' + \hat{P}_2\mu_i \tag{27}$$

The properties of the productivity process and the definition of the sets I_k imply that $\left\|\hat{P}_k\right\| < 1$. Hence, the argument presented in Hopenhayn (1992, Lemma 4) follows directly and the operator $\left(\mathrm{I} - \hat{P}_k\right)$ has an inverse. Hence, the invariant measures μ_i and μ_f are well-defined and can be written as

$$\mu_i = \left(\mathrm{I} - \hat{P}_1\right)^{-1} \Lambda_i\tilde{G}_i \tag{28}$$
$$\mu_f = \left(\mathrm{I} - \hat{P}_3\right)^{-1} \left(\Lambda_f\tilde{G}_f + \hat{P}_2\mu_i\right) \tag{29}$$

Write the above measures as $\mu_i = m_i\left(\underline{\theta}_i, \overline{\theta}_i, \Lambda_i\right)$ and $\mu_f = m_f\left(\underline{\theta}_f\overline{\theta}_i, \Lambda_f\right)$, where we use the fact that the operators $\hat{P}_k$ are functions of the cut-offs that define the sets I_k. It is then useful to establish the following result:

LEMMA 3: *The functions* $m_i\left(\underline{\theta}_i, \overline{\theta}_i, \Lambda_i\right)$ and $m_f\left(\underline{\theta}_f, \overline{\theta}_i, \Lambda_f\right)$ *are continuous in all of their arguments. Moreover,* $m_i\left(\underline{\theta}_i, \overline{\theta}_i, \Lambda_i\right)$ is *strictly increasing in* Λ_i *and* $m_f\left(\underline{\theta}_f, \overline{\theta}_i, \Lambda_f\right)$ is *strictly increasing in* Λ_f.

Proof: From expressions 28 and 29, it is clear that the functions $m_s(\cdot)$ are continuous and strictly increasing in Λ_s. The continuity in the cut-offs is a consequence of the fact that the operator $\hat{P}_k$ is continuous in the cut-offs that define the corresponding set I_k (see Hopenhayn, 1992, Lemma 5 for a proof).$\square$

Step 5: Existence and uniqueness

The starting point is to use the following results established in Hopenhayn (1992): (i) there exists a stationary equilibrium with invariant measure μ associated to an unique aggregate input-output pair, (L, Q), and unique prices (Theorem 2); (ii) if the entry cost is low enough, a stationary equilibrium with positive entry exists (Theorem 3); (iii) if the profit function is multiplicatively separable between productivity and prices, $\pi(\theta, p, w) = h(\theta) g(p, w)$, then if there exists an equilibrium with entry and exit, it is unique (Theorem 4).[24] It remains to show that there exists an unique stationary equilibrium with an unique formal-informal partition and entry into both sectors.

First, write $\mu = m\left(\underline{\theta}_i, \overline{\theta}_i, \underline{\theta}_f, \Lambda_f, \Lambda_i\right) = m_i\left(\underline{\theta}_i, \overline{\theta}_i, \Lambda_i\right) + m_f\left(\underline{\theta}_f, \overline{\theta}_i, \Lambda_f\right)$. Fix μ; from Step 3 (equations 22-25) we know that there are unique cut-off points for entry, exit and transition between sectors. In particular, because $c_f^e > c_i^e$ the following holds: $\underline{\nu}_f > \underline{\nu}_i$. Thus, for any $\nu \in [\underline{\nu}_i, \underline{\nu}_f)$, entry occurs into the informal sector and for any $\nu \geq \underline{\nu}_f$, entry occurs into the formal sector. Hence, the unique thresholds $\left(\underline{\nu}_i, \underline{\nu}_f\right)$ pin down the mass of entrants into the informal sector, Λ_i. Additionally, fixing μ uniquely determines thresholds $\left(\underline{\theta}_i, \overline{\theta}_i\right)$, which therefore pins down a unique informal sector size, $\mu_i = m_i\left(\underline{\theta}_i, \overline{\theta}_i, \Lambda_i\right)$.

Finally, the industry size μ also uniquely determines the formal sector's threshold, $\underline{\theta}_f$. But $\mu = \mu_i + \mu_f$ and $\mu_f = m_f\left(\underline{\theta}_f, \overline{\theta}_i, \Lambda_f\right)$ is strictly increasing in Λ_f. Thus, once the informal sector size is determined and the thresholds $\left(\underline{\theta}_f, \overline{\theta}_i\right)$ are fixed, there is an unique value of Λ^f that satisfies the identity $\mu = \mu_i + \mu_f$.

Thus, there is an unique stationary equilibrium with invariant μ_s, $\left(\underline{\nu}_s, \underline{\theta}_s, \overline{\theta}_i\right)$, $s = i, f$, aggregate prices and quantities, (Q, L, w), and entry levels in both sectors, (Λ_f, Λ_i).

$\square$

The life cycle of firms: size, productivity and informality

Let the productivity distribution among firms of age n in sector s be denoted by the probability measure λ_s^n. This measure is defined over the set of active firms with a given age, A_s^n. The evolution of the productivity distribution in the informal and

[24] The functional form assumed for the production function belongs to this class and thus the theorem applies directly.

formal sectors within a given cohort is given by:

$$\begin{aligned}
\lambda_i^{n+1}(\theta') &= \int_{X_1} G\left(\theta' \middle| \theta\right) d\lambda_i^n(\theta) \\
\lambda_f^{n+1}(\theta') &= \int_{X_2} G\left(\theta' \middle| \theta\right) d\lambda_f^n(\theta) + \int_{X_3} G\left(\theta' \middle| \theta\right) d\lambda_i^n(\theta)
\end{aligned}$$

for all $\theta' \in \Theta$; $X_1 \equiv A_i^n \cap \left[\underline{\theta}_i, \overline{\theta}_i\right)$, $X_2 \equiv A_f^n \cap \left[\underline{\theta}_f, \infty\right)$, and , $X_3 \equiv A_i^n \cap \left[\overline{\theta}_i, \infty\right)$.

Now consider the ordering of measures given by the the first order stochastic dominance criterion, so that $\lambda_s^{n+1} \succeq \lambda_s^n$ means that the productivity distribution at age $n+1$ first order stochastically dominates the one at age n. Then the following result holds:

Proposition 2: *Assume that the assumptions of Proposition 1 hold and that the pre-entry signal parameter has a continuous distribution, $H(\nu)$. Then the productivity distribution among active firms within a sector and a cohort is increasing in the cohort's age. That is, $\lambda_s^{n+1} \succeq \lambda_s^n$ for all n and $s = i, f$.*

Proof:

Define the following operators:

$$T_k(B) = \begin{cases} \int_B dF\left(x \middle| \theta\right), & \text{if } \theta \in X_k \\ 0 & \text{otherwise} \end{cases}$$

for $k = 1, 2, 3$ and all Borel sets $B \subseteq \Theta$. As before, $X_1 \equiv A_i^n \cap \left[\underline{\theta}_i, \overline{\theta}_i\right)$, $X_2 \equiv A_f^n \cap \left[\underline{\theta}_f, \infty\right)$, and , $X_3 \equiv A_i^n \cap \left[\overline{\theta}_i, \infty\right)$. One can then rewrite the expressions for the λ_s^n as follows:

$$\begin{aligned}
\lambda_i^{n+1} &= T_1 \lambda_i^n \\
\lambda_f^{n+1} &= T_2 \lambda_f^n + T_3 \lambda_i^n
\end{aligned}$$

The productivity distribution of newborn firms, however, is simply given by $\lambda_i^1([\theta_l, \infty)) = \int_{\nu > \underline{\nu}_i} F\left(\theta \middle| \nu\right) dG(\nu)$ and $\lambda_f^1([\theta_l, \theta)) = \int_{\nu > \underline{\nu}_f} F\left(\theta \middle| \nu\right) dG(\nu)$.

From the second period on, the productivity distribution is obtained by first applying the truncation implied by the conditions $\theta \in X_k$ in the operators T_k. This always implies a truncation on the lower tail of the productivity distribution, which is increasing in the FOSD ordering. Finally, assumption (A.2) implies that after the truncation, the operator T_k is also monotone and hence the operator T_k as defined is increasing in the FOSD criterion. Thus, the following holds almost by definition: $\lambda_i^2 \equiv T_1 \lambda_i^1 \succeq \lambda_i^1$. By induction, $\lambda_n^{n+1} \succeq \lambda_i^n$ for any n.

The analysis of λ_f^n is not so straightforward because of the presence of the term

$T_3\lambda_i^n$, so a more careful argument is needed. First, note that because of the argument just made for λ_i^n, in the absence of the term $T_3\lambda_i^n$ one would observe $\lambda_f^{n+1} \succeq \lambda_f^n$ for any n. Second, the conditioning embedded in operator T_3 is stronger than the one in T_2, as $\overline{\theta}_i > \underline{\theta}_f$. Thus, $\lambda_f^2 = T_2\lambda_f^1 + T_3\lambda_i^1 \succeq \lambda_f^1$. But given the result that λ_i^n is increasing in n, and that the T_k operators are increasing, then λ_f^n will be also increasing in n. $\square$

Corollary 1: *As the productivity distribution among active firms within a sector and a cohort is increasing in the cohort's age, so is the integral of any function that is increasing in θ. In particular, this is true for the survival rate, average size and average revenue.*

Proof: Given the result in Proposition 2, this corollary follows mechanically and no proof is provided.

Corollary 2 (intensive margin of informality): *As a consequence of Proposition 2 and Corollary 1, the average informality rate within formal firms in a given cohort is decreasing in cohort's age.*

Proof: There is an unique threshold $\tilde{\ell}$ above which the formal firm only hires formal workers (on the margin). For a formal firm that has its optimal level of labor below the threshold, $\ell_f^*(\theta) < \tilde{\ell}$, $s_i(\theta) = 1$. In this case, the within informality rate will stay constant at one for some period while the firm is growing, but as soon as $\ell_f^*(\theta) = \tilde{\ell}$ the informal share will start declining monotonically with firm's size. Similarly, for any initial value of $s_i < 1$, as the firm grows the s_i will decline monotonically. Hence, the s_i is a constant function of θ if $\ell(\theta) < \tilde{\ell}(\theta)$ and it is strictly decreasing in θ over the range where $\ell(\theta) \geq \tilde{\ell}(\theta)$. Combined with the result in Proposition 2, this implies that the average within firm informality for a given cohort will be monotonically decreasing with cohort's age almost everywhere in the relevant range $\left[\underline{\theta}_f, \infty\right]$. As long as there is always at least one firm within the cohort that has a $s_i(\theta) < 1$, the average within firm informality will be strictly decreasing in the cohort's age. $\square$

It is straightforward to see that the share of informal workers is decreasing in firm's size (and therefore productivity) in the within-period (static) problem of the firm. Thus, by Proposition 1 and Corollary 1 one gets Corollary 2. Finally, let the informality rate among firms of a given cohort of age n be expressed as $B^n = \frac{\mu\left(A_i^n\right)}{\mu\left(A_i^n\right)+\mu\left(A_f^n\right)}$, where $\mu\left(A_s^n\right)$ denotes the measure of active firms in sector s with age n. With this notation in hand, the last result of this section can be stated as follows:

Corollary 3 (extensive margin of informality) *Under the conditions of Proposition 2, the informality rate B^n within a given cohort is weakly decreasing in cohort's age. That is, $B^{n+1} \leq B^n$ for all n.*

Proof: Proposition 2 established that the distribution of productivity is increasing with firms' age in both sectors. At the same time, the informal to formal transition threshold, $\overline{\theta}_i$ remains constant. Hence, as the cohort gets older a smaller number of firms (among the survivors) will remain informal as the most productive ones keep making the transition into the formal sector. This implies that $\mu_i^{n+1} \leq \mu_i^n$. The same is not true for μ_f^n, as there is no upper limit for formal stayers. This means that formal firms do not face the additional exit margin that informal firms do, as there is no upper limit for their growth. Since the productivity shock in both sectors is the same, even if the set of active firms in the formal sector reduces in size as the cohort ages, it does so at a lower rate than informal firms in the same cohort. Hence, the ratio $\gamma_b^n = \frac{\mu_i^n}{\mu_i^n + \mu_f^n}$ is weakly decreasing in the cohort's age. □

G Model Calibration Appendix

G.1 Estimating formal sector's persistence parameter

One can use the first order condition of formal firms that hire at least one formal worker and the law of motion of formal firms' log-productivity to show that the establishment-level employment process can be represented as a simple AR(1) process with the same persistence parameter, ρ_f. This procedure assumes that there are no adjustment costs, such as hiring costs, which could lead to an overestimation of ρ_f. However, Dix-Carneiro et al. (2021) estimate this persistence parameter in a richer model with adjustment costs and find slightly higher (and not lower) values.

Let the log of firm j's employment at time t be denoted by $n_{j,t}^f = \ln(\ell_{j,t})$; using formal firms' first order condition and the law of motion for the log-productivity (8), one can write

$$n_{j,t}^f = \gamma_0 + \gamma_1 \log(\theta_{j,t}) - \gamma_2 w_t + m_{j,t} \tag{30}$$

where w_t denotes wages, $\gamma_0 = \frac{1}{1-\alpha}\left[\log(\alpha) + \log\left(\frac{1-\tau_y}{1+\tau_w}\right)\right]$, $\gamma_1 = \frac{1}{1-\alpha}$, and $m_{j,t}$ is measurement error.

One can then use (9) to write (30) in its dynamic representation:

$$n_{j,t}^f = b_0 - \gamma_1 w_t + b_2 w_{t-1} + \rho_f n_{j,t-1}^f + \eta_j + e_{j,t} \tag{31}$$

where $\eta_j \equiv \gamma_1(1-\rho_f)\ln\nu_j$, $b_0 \equiv (1-\rho_f)\gamma_0$, $b_2 \equiv \gamma_1\rho_f$, and the error term is given by $e_{j,t} = \gamma_1\epsilon_{j,t} + m_{j,t} - \rho_f m_{j,t-1}$.

The employment process can be represented as a simple AR(1) process with an MA(1) error, where the MA(1) component arises if one allows for measurement error. The final regression estimated is the following:

$$n_{j,t}^s = \rho_f n_{j,t-1}^f + \Gamma \mathbf{X}_{j,t} + \eta_j + e_{j,t}, \tag{32}$$

where $\mathbf{X}_{j,t}$ denotes a vector of controls in addition of $n_{j,t-1}^f$ and η_j denotes firm's fixed effect. The vector $\mathbf{X}_{j,t}$ includes a set of year dummies and the current and lagged log-average wage rate calculated at the 4-digit industry level. This specification, which includes current and lagged wages, is standard in the empirical literature (e.g. Blundell and Bond, 1998).

We start by estimating (32) using both a standard OLS estimator and a within-groups estimator. The former is known to be upward biased while the latter is downward biased, and they can thus be used as upper and lower bounds to any consistent estimator. The third model used is the standard first-differenced GMM estimator, which can be subject to finite sample biases (towards zero) when the lagged levels are weak instruments for the first-differenced equation. The forth model is the system GMM estimator (Blundell and Bond, 1998), which uses lagged differences as instruments for equations in levels.

For the first-differenced and system GMM models, we consider two scenarios for the error term. The first is no measurement error, which allows for the use of lagged levels dated $t-2$ and earlier as instruments for the differenced equations, and lagged differences dated $t-1$ and earlier for the level equations. The second allows for measurement error, which implies that the error term in (32) will have a MA(1) structure. I this case one can only use lagged levels dated $t-3$ (and earlier) and lagged differences dated $t-2$ and earlier as instruments. All GMM regressions consider the log-wage as a predetermined variable and they are estimated using the two-step estimator with the correction for the variance-covariance suggested by Windmeijer (2005). Table G.7 shows the results.

Table G.7: Productivity process estimation

Dep. Var.: Log-Employment ($n_{j,t}$)						
	OLS	FE	DIFF1	DIFF2	SYS1	SYS2
$n_{j,t-1}$	0.944**	0.497**	0.594**	0.728**	0.713**	0.921**
	(0.000)	(0.002)	(0.007)	(0.011)	(0.009)	(0.005)
$\log(w_t)$	0.0030	0.0100	-0.0920	-0.158*	-0.210**	-0.339**
	(0.006)	(0.008)	(0.066)	(0.062)	(0.072)	(0.069)
$\log(w_{t-1})$	0.006	0.005	0.069	0.054	0.110	0.075
	(0.006)	(0.007)	(0.095)	(0.048)	(0.069)	(0.053)
Obs.	741,268	741,268	741,268	741,268	741,268	741,268

Notes: DIFF1 and DIFF2 are the difference GMM models without and with measurement error; SYS1 and SYS2 are the system GMM models without and with measurement error, respectively. Significant at ***1%, **5% and *10% levels.

The results shown in Table G.7 follow the pattern expected from the standard results in the literature. The OLS result already points to a very persistent series, which is confirmed by the downward bias apparent in the first-differenced GMM model (DIFF) as compared to the system GMM estimator (SYS). Blundell and Bond

(1998) find similar results when comparing these two estimators using a small sample of British firms. There is also evidence that measurement error is indeed present, as the estimates under the assumption of no measurement error (DIFF1 and SYS1) seem to be substantially downward biased. The Sargan tests for the additional instruments available in the DIFF1 and SYS1 models (relatively to the DIFF2 and SYS2, respectively) strongly reject the validity of these additional instruments, reinforcing the evidence of measurement error (results not reported). That said, the preferred model is the system GMM under the assumption of measurement error (SYS2), which provides a reasonable value for the persistence parameter.[25]

G.2 Details of calibration implementation

Let the vector of simulated moments is denoted by $m_S(\Omega;\zeta)$, where ζ denotes the vector of parameters determined in the first step. The calibrated parameters vector is given by

$$\hat{\Omega} = \arg\min_{\Omega} \left(\hat{m}_N - m_S(\Omega;\zeta)\right)' \left(\hat{m}_N - m_S(\Omega;\zeta)\right) \tag{33}$$

We take the observed wages as given. In order to obtain a wage measure that is more consistent with the model, we estimate a log-wage regression controlling for schooling, gender (male dummy), 4-digit industry dummies, state of residence, a dummy for whether the worker holds a formal contract, dummy for whites, age and age squared, tenure in current job and tenure squared. We restrict the sample to employees only (formal or informal), who are 18 to 69 years old, and who have worked at least 20 hours but at most 84 hours (the 99th percentile) in a given week. We use the estimated coefficients to compute the adjusted wage evaluated at the mean of the vector of observables.

We simulate a cohort of $500,000$ potential entrants, which we follow until the age of 50 (within this class of models, Sterk et al. (2021) use a cohort of size $100,000$, which they follow for 20 years). For each potential entrant, we draw a pre-entry productivity parameter (ν_j) and a sequence of post entry productivity shocks, $\epsilon_{j,t}$ for $j = 1, ..., 50$. The stochastic components of the model are drawn only once in the beginning of the procedure and are kept fixed throughout. Each potential entrant has an individual pre-entry productivity parameter, ν_j, which is a firm-specific intercept in the AR(1) process that the firm faces after entry. We use a fine grid of 101 equally spaced points for the productivity space. We compute a separate transition probability matrix for each point in the grid and also for the formal and informal sectors (which have different persistence, ρ_s, and shock variance, σ_s^2). We do so using the method proposed by Tauchen (1986).

[25] The Sargan test of overidentifying restrictions rejects instruments' validity in all models, with p-values of zero (not reported). However, Arellano and Bond (1991) show in their simulations exercise that the Sargan test rejects too often in the presence of heteroskedasticity, which is confirmed in their empirical application (they estimate a dynamic employment equation very similar to (32).

Smooth Policy Functions

Some of the main decisions firms make in this model are discrete, which implies that some of the policy functions will be step functions. This implies that is not possible to use derivative-based methods, which are faster and more accurate than derivative-free methods or random search algorithms. We therefore use the smoothing function proposed by Bruins et al. (2015) to correct for the choppiness of the policy functions:

$$h\left(\tilde{V}(\beta), c, \lambda\right) = \frac{\tilde{V}_c(\beta)/\lambda}{1 + \sum_k \tilde{V}_k(\beta)/\lambda}$$

where $\tilde{V}(\beta)$ is the set of payoffs associated to firms' choices, such as whether to transit to the formal sector or not (if the firm is informal). $\tilde{V}_c(\beta)$ denotes the net payoff of a given choice c and λ denotes the smoothing parameter. As $\lambda \to 0$, $h(\cdot)$ goes to one if the alternative c provides the highest payoff and zero otherwise. There is a trade-off between bias and smoothness in the choice of the smoothing parameter: a large value for λ provides a smoother objective function, but can lead to biased estimates; a small value reduces bias but increases choppiness. We follow Altonji et al. (2013) and choose $\lambda = 0.05$.

G.3 Relationship between data moments and parameters

We start by discussing what variation in the data allows us to pin down different parameters. The cost function parameters for both margins of informality are identified by the variation in the share of informal firms by firm size (for the parameter φ_i) and by the behavior of the share of informal workers within formal firms by firm size (for the parameter φ_f). In the model, the cost functions of both margins are increasing in firms' size, which implies that the average share of informal firms and the average share of informal workers within formal firms will be decreasing in firm size, as observed in the data. The intensity of this negative relationship in the model is governed by φ_i and φ_f and therefore the behavior of these moments in the data provides the relevant information to estimate these parameters.

The death shocks in both sectors determine the relative disadvantage of being informal regardless of firms' productivity, size or age: the higher the informal death shock, δ_i (relative to the formal sector's, δ_f), the greater the relative disadvantage of being informal. They are disciplined by the moments that speak to the relative size of the informal sector. The estimated exogenous exit probability in the formal sector is consistent with the exit probability observed in the data among older formal firms.[26] This is reassuring, as older firms (both in the model and in the data) are more stable and less likely to endogenously exit due to a very negative productivity

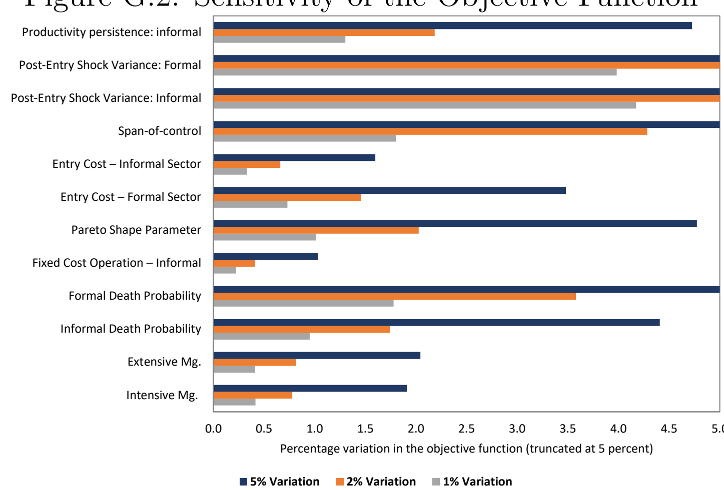
[26] Using a simple linear probability regression conditioning on year and industry dummies, the predicted exit rate at ages 6 to 10 monotonically declines from 0.074 to 0.038.

shock. The entry costs into both sectors are also directly linked to the relative size of the informal sector, but more importantly they directly affect the left tail of the size distribution in both sectors.

The shape parameter of the Pareto distribution, ξ, is largely determined by the moments of firm size distribution in the formal and informal sectors, and in particular the skewness in formal sector's distribution. The informal persistence and the variance of the shock of the productivity processes in both sectors – ρ_i and σ_s^2, $s = i, f$ – are linked to the moments related to firm growth in both sectors. The σ_s are directly connected to the degree of overlap between formal and informal firm size distributions, as they determine how much productivity dispersion there is conditional to sector choice. These sources of variation in the data contribute to separately estimate the variances and exogenous death shocks.

Finally, we follow Adda et al. (2017) and evaluate if the objective function has curvature around the estimated vector of parameters. If the objective function is flat, then it would suggest that the moments used do not provide relevant information to estimate the parameters. We re-compute the objective function varying each parameter by 1, 2 and 5 percent from its estimated value and compute the corresponding percentage change in the objective function. Figure G.2 shows that there is substantial variation in the objective function in response to small perturbations of the different parameters.

Figure G.2: Sensitivity of the Objective Function



Note: The horizontal bars show the percentage change in the objective function with respect to one, two and five percent changes in the given parameter of the model. The figure is truncated on the right at 5 percent.

H Computing the transition dynamics

This section presents the details of the computation of the deterministic transition dynamics of the model after the unexpected shock of a permanent increase in labor supply hits the economy. The stationary equilibrium before the shock, in $t = 0$, is simply our baseline economy (Brazil in 2003), while the new steady state after the shock is given by the first counterfactual described in Section 5.2 and Table 6. Importantly, the transition dynamics induced by this shock is deterministic (as it is the case with all MIT shocks). The computation algorithm used is the following:

1. Compute the initial and final steady states.

2. Define the formal wage path as $w_{f,t} = \gamma w_{t-1}$. Following Schmitt-Grohé and Uribe (2016), our main results use $\gamma = 0.996$, which implies that the final steady state is reached in 10 periods. As a robustness check, we also use $\gamma = 0.997$, which implies that the final steady state is reached in 15 periods (results below).

3. We assume that informal wages adjust freely at all times. Given that we have the value functions at the new steady state in $t = T$, we can solve the model backwards from $t = T-1, ..., 1$. For each period, we solve for the informal wage, $w_{i,t}$, that clears the labor market.

4. Simulate the model forward from $t = 1, ..., T$ using the sequences of formal and informal wages, and value functions computed in the previous step.

5. Compute the maximum difference between total labor demand and (fixed) labor supply in all periods, $\Delta = max\, |(LD_t - LS)/LS|$. If Δ is small enough, stop.

6. Otherwise, update the guess for informal wages and go to step 3.