Dynamic Migration:

From Local Effects to Aggregate Implications *

Alejandro Parraguez–Tala †

October 31, 2024

Click here for the latest version.

Abstract

What are the local and aggregate dynamic effects of regional migration flows? To address this question, I combine a panel SVAR framework with a shift-share instrument based on lagged birth rates to analyze the short-, medium-, and long-term impacts of a 1% increase in migration to U.S. commuting zones. My estimates indicate that these inflows raise wages and output per worker, with effects peaking between 0.2% and 0.3% over time. However, the most substantial impact is observed in housing prices, which reach a long-run level 0.9% higher than pre-shock levels. To assess whether these local effects translate into significant aggregate outcomes, I examine labor "deportation" and "relocation" through the prism of a dynamic spatial model that I estimate through impulse response matching to the empirical findings. The deportation scenario removes workers unevenly across regions, resulting in a long-term reduction in aggregate output of 1%. In contrast, the relocation scenario redistributes these workers across other regions in the country. Although this redistribution does not alter the steady state, it can lead to a temporary output decline of up to 1%, with recovery times exceeding 50 years, depending on the skill composition of relocated workers.

Keywords: Migration, Labor Mobility, Economic Growth

*I am deeply grateful to Nitya Pandalai-Nayar, Andres Drenik, Olivier Coibion and Caitlin Gorback for their exceptional mentorship and support throughout this research. I'm also grateful to Christoph Boehm, Saroj Bhattarai, Oliver Pfäuti, Carola Binder, Edson Wu, Richard Faltings, Yusik Kim and all participants at the UT Austin Macroeconomics workshop and the Texas Macro Job Market Conference at the Federal Reserve Bank of Dallas.

†University of Texas at Austin. Email: aparragueztala@utexas.edu.

1 Introduction

What are the dynamic effects of local migration shocks? Recent fluctuations in internal migration trends, along with international immigration disproportionately affecting certain states, have brought greater attention to the effects of regional population inflows on the broader economy. In spatial models, a larger workforce leads to increases in productivity through agglomeration economies, all the while congestion forces such as higher rents make these areas less attractive. However, so far these frameworks have mostly remained static and therefore offer no insights into how long these productivity gains take to materialize, specially when compared with short-run increases in housing prices. Moreover, this literature has placed limited emphasis on aggregate outcomes, despite the fact that the macroeconomic impact of relocation shocks can vary depending on the regions receiving migrants.

This paper aims to fill this gap. First, I estimate the dynamic local effects of migration shocks using a panel SVAR structure at the commuting zone (CZ) level, coupled with an external instrument. The results consist of structural impulse response functions (IRFs) that illustrate how variables such as wages, output, and housing prices respond over time to higher population inflows. Second, I develop a quantitative dynamic spatial model, with migration frictions as well as productivity spillovers, and use the empirical IRFs to estimate the structural parameters. Using the calibrated model, I demonstrate how the aggregate impact of commonly discussed immigration policies varies across the short, medium, and long term. Specifically, I consider a "deportation" shock that removes a spatially heterogeneous share of regional population that corresponds to undocumented immigrants. I then simulate a "relocation" shock, where that same proportion of workers is moved to other areas within the US.

To identify the local causal effect of migration inflows, I estimate a panel SVAR-IV model using shift-share instruments derived from lagged birth rates. This vector autoregression structure enables me to capture the intertemporal relationships between population inflows and other key variables. As a result, I can account not only for the direct impact of migration inflows on economic outcomes but also for the potential feedback effect, where favorable economic conditions attract additional future migration. To address potential simultaneity bias arising from contemporaneous productivity shocks that could drive migration into the area, I employ an instrumental variable approach. After estimating the reduced-form panel VAR, I apply the methodology of Stock and Watson (2018) to identify the impact of structural market size shocks using an external instrument. Specifically, for each CZ, I predict population inflows using a Bartik-style instrument: the 'pull' factor are lagged migration shares at the destination and the 'push' factor is the lagged birth rate of each origin, following a similar approach to Karahan, Pugsley and Şahin (2019).

The exclusion restriction of this instrument relies on the assumption that birth rates in

the areas sending migrants are uncorrelated with destination-specific shocks 25 years later. Following the strategy used in the immigration literature, such as in Altonji and Card (1991) and Card (2001), the weights of the shift-share instrument correspond to the lagged share of migrants originating from each area. However, unlike most studies in this field, I use migration patterns from the previous year rather than a fixed pre-period. Although these shares may be endogenous, I follow Borusyak, Hull and Jaravel (2022) and estimate the IV at the 'shock level'—in this case, the level of the county sending migrants. This approach allows me to leverage the exogeneity of past birth rates while constructing a stronger instrument with inflow shares that more accurately predict current migration inflows. Additionally, it enables me to calculate the appropriate standard errors for inference, thereby avoiding the limitations of conventional shift-share methods discussed by Adao, Kolesár and Morales (2019).

At the CZ level, a 1% increase in migration inflows leads to higher future in-migration, raises wages and output, stimulates establishment entry, and drives up housing prices. The shock initially produces an immediate effect on all variables, which intensifies over time. Specifically, productivity proxies such as output and wages reach a peak between 0.15% and 0.25% above pre-shock levels after four years. Simultaneously, some workers begin to leave the area, likely in response to congestion forces, as indicated by a 1.1% rise in housing prices. However, these cumulative effects then decrease slightly from their peak and stabilize at a higher permanent level. For example, housing prices eventually settle at a new level that is 0.9% above pre-shock values but below the four-year peak.

I also provide further evidence to better understand the mechanisms driving the baseline results. First, I split the CZs in my sample according to population per square mile and re-estimate the SVAR for each group. The new IRFs show that, although the magnitude of the estimate increases with CZ density, its variance does so as well, which suggests the net effect from the agglomeration and congestion externalities becomes more ambiguous. Furthermore, since the instrument is positively correlated with income per capita among migrants, a simple composition effect could be behind the rise in wages. Nevertheless, when I look at the response of wages within narrowly defined sectors with homogeneous low-skill requirements, I find that all of them experience a sustained increase.

These results underscore the importance of estimating the dynamic effects of such a shock. First, the shape of the IRF reveals the relative significance of different forces as time passes. In this sense, the smaller long-term productivity response, compared to the medium term, suggests that congestion externalities outweigh agglomeration economies after four years, whereas the reverse is true in the short and medium term. Second, using a dynamic model that includes lagged variables allows me to control for persistent unobserved shocks from the past, which could otherwise bias the IV estimates. Finally, examining the effect of population flows over different time horizons through the lens of a theoretical model can

help identify distinct structural elasticities. For instance, housing prices in the short term are largely determined by the housing supply elasticity, while in the long term, they depend on the interaction between this and other factors, such as the elasticity of migration.

Given these results, how does aggregate productivity react to local migration shocks? Inherently, this question connects heterogeneous treatments at the local level to aggregate outcomes and answering it requires a theoretical framework. I therefore develop a continuous time spatial economic model in which forward looking workers choose where to live each period, subject to mobility costs on the intensive and extensive margin. Furthermore, the model includes static agglomeration externalities à la Romer (1990) as well as local dynamic spillovers similar to the ones in Peters (2022), where a larger local mass of firms in a region decreases the cost of entry in the future. On the other hand, endogenously congested amenities as well as a housing sector with fixed land make locations less attractive as population increases. Mobility frictions slow the spatial redistribution of labor, and, when combined with intertemporal productivity spillovers, create conditions in which migration shocks can have persistent impacts that vary over time on variables like wages and housing prices. This is what we observe in the empirical IRFs.

To discipline the main parameters of the model I use impulse response matching. In particular, I simulate the transition path of the economy implied by the empirical birth rates in the shift-share instrument. I then replicate the estimation described above using the simulated data to obtain model generated IRFs. The parameter estimates I obtain are the ones that minimize the distance between these simulated moments and the empirical values. This guarantees the model generates the correct causal effect of migration inflows on variables such as housing prices and output per worker. Other parameters are estimated using calibration by targeting cross-sectional evidence with moments generated in the initial steady-state.

I then use the calibrated model to assess the aggregate impact of local migration shocks. To do so, I construct two policy counterfactuals. The first is a "deportation" shock: using the empirical geographical distribution of undocumented migrants in the U.S., this scenario removes a segment of the workforce from the economy. It is essential to emphasize that these are region-specific shocks, with deportation shares varying across regions, representing between 3% and 5% of local populations. The areas directly affected by the shock experience higher amenities, lower housing prices, and real incomes that are, on average, 7.5% higher. As a result, these locations attract workers from other regions, with increased inflows and reduced outflows. However, the overall reduction in the workforce leads to a long-term decline in the number of firms and total output by 6% and 2%, respectively, across all regions. These dynamics ultimately cause aggregate output per capita to decline to a new steady state that is 1% below the initial level. The magnitude of these effects depends on the skill composition of the deported workers, with potentially greater impacts if higher-skilled

workers are removed.

In contrast, the second policy scenario involves a "relocation" shock, in which the same population is redistributed uniformly across other regions. A similar policy has been implemented recently at a local level, with some states relocating undocumented immigrants to different regions within the country (Goodman et al. (2024)). The results in this paper indicate that, if applied on a national scale, such a policy would produce no long-term effects, as the economy would eventually converge back to the original steady state through migration flows. However, in the short term, the shock would create significant impacts, including a 1% increase in housing prices across all locations and an immediate 9% rise in real incomes in areas losing workers. Moreover, the reduction in the mass of firms by more than 2% in these regions would lead to a regional output decline of between 2% and 8%, depending on the skill composition of the relocated households. At the aggregate level, relocating low-skilled households results in a slight increase in output per capita. Conversely, redistributing high-skilled workers leads to a drop in output per worker of over 1%, with recovery taking more than 50 years.

Related literature This paper follows previous research on market size, labor mobility, and economic growth. From a theoretical perspective, Duranton and Puga (2004) summarize the main mechanisms driving agglomeration economies. The first of these is *sharing*: producers in larger cities can divide the fixed cost of indivisible inputs as well as the gains of specialization. Another channel is *matching*, as a larger labor force increases both the quantity and quality of firm-worker matches, especially in the presence of mismatching costs due to skill heterogeneity. Finally *learning* from a larger and more diverse market raises productivity through knowledge creation, diffusion, and accumulation (Lucas, 1988; Moretti, 2004; Crews, 2023). The relevance of each mechanism, and ultimately the effect of immigration, depends on the time frame. For example, a fixed capital stock in the short run causes wages to drop in the immediate years following worker inflows (Borjas, 2013). By estimating a dynamic model over several years, the following analysis determines how the relative importance of each mechanism evolves.

Numerous empirical studies estimate the response to labor inflows and obtain ambiguous results. Using previous immigration shares as an instrument, some papers find a negative effect on native employment and wages (Altonji and Card, 1991; Card, 2001). Similarly, the wage elasticities estimated by Borjas (2003) imply immigration to the US in the 1980s and 1990s reduced wages by 3.2%. Nevertheless, as the survey by Edo (2019) points out, the estimated impact on wages, employment, and other outcomes depends on the methodology, period, and country under study. Some papers have found evidence of agglomeration economies with positive spillovers on manufacturing productivity (Greenstone, Hornbeck and Moretti, 2010; Kline and Moretti, 2014) as well as employment opportunities (Moretti, 2010). Most of these studies estimate a single elasticity for multiple periods and adopt

reduced-form strategies that do not account for the intertemporal correlation between migration and other economic measures.

To incorporate these interactions, this paper models local population inflows and outflows, along with other outcome variables, using a panel VAR structure. Barcellos (2010) adopted this approach in her analysis of migration between US states whereas Boubtane, Coulibaly and Rault (2013) did so for 22 OECD countries. To identify the dynamic response to population inflow shocks, these studies rely on ordering assumptions and Cholesky decompositions of the variance-covariance matrix. I avoid making such assumptions by estimating the coefficients of interest using an IV strategy as described by Stock and Watson (2018) and applied in the fiscal multipliers (Mertens and Ravn, 2013) as well as the monetary policy literature (Gertler and Karadi, 2015).

The use of exogenous variation resembles the analysis of previous authors leveraging historical episodes to estimate the effect of population movements. Notable examples include Peters (2022) and Burchardi and Hassan (2013), both of whom examine the impact of post-WWII German migrants who resettled in West Germany. However, their estimation relies on a single event, whereas the SVAR-IV exploits changes across time. To accomplish this, I use a shift-share instrument (estimated at the shock level as stated in Borusyak, Hull and Jaravel (2022)) based on migration shares and lagged birthrates. This follows a similar identification strategy as the one proposed by previous work that estimates the effect of increasing labor supply (Shimer, 2001; Karahan, Pugsley and Şahin, 2019).

The methodology in this paper identifies exogenous shocks that influence local inmigration flows to estimate the latter's effect on economic conditions. In this sense, it closely resembles the Local Projection IV (LP-IV) used in Howard (2020), which uses a dynamic model along with an Altonji and Card (1991)-style instrument to determine the effect of migration on unemployment. The paper finds that unemployment falls whereas the employment-to-population ratio rises. However, it differs from the analysis I carry out in several ways. First, its econometric approach does not account for the interaction between migration, unemployment, and other relevant variables such as wages, output, and establishments, which is an element the VAR structure allows me to consider. Furthermore, it only estimates the effect on labor market outcomes, whereas this paper determines the response on a set of comprehensive economic measures. Finally, I estimate the impact on CZs which cover a broader geographical sample than MSAs, which are only representative of urban economies.

This paper also follows previous work on the implications of local dynamics for aggregate outcomes. Davis, Fisher and Whited (2014) study the contribution of agglomeration in cities to aggregate productivity growth through the lens of a structurally estimated spatial model. In their framework, local output per worker depends positively on output density. Crews (2023) further endogenizes these externalities in a heterogeneous agents model: as the

aggregate stock of skill in a city increases, the rate of individual human capital accumulation rises as well[1]. On the other hand, Peters (2022) proposes a different mechanism by relating current fixed costs of entry to past market size, creating dynamic spillovers. He embeds this in a model with variety gains, as in Romer (1990), and uses it to examine the scale effects on productivity in the context of postwar Germany.

The rest of the paper is structured as follows. The next section explains the SVAR-IV methodology, as well as the shock-level shift-share instrument I use to estimate the effects of migration. Then, in Section 3, I analyze the structural impulse responses to a 1% increase in inflows and discuss further supporting evidence. Section 4 develops the formal framework relating the local evidence to aggregate outcomes, whereas Section 5 implements the structural estimation of the model parameters. Finally Section 6 implements the policy counterfactuals and Section 7 concludes.

2 Econometric Model

2.1 Panel SVAR-IV Model

The first part of this paper seeks to determine the dynamic effect of population inflows on local economies. Specifically, for any CZ i and year t, I determine the impact total in-migration from within the US m_{it} has on the population flow dynamics in the years following the initial shock. In other words, I estimate the effect on future in- and out-migration, m_{it} and o_{it} respectively, since the arrival of new residents could drive native workers to leave as well as more immigrants to relocate in the future. This could be due, in part, to changes in employment opportunities. Consequently, I also study the effect on the employment-to-population ratio l_{it} to establish if new workers compete for the same jobs as previous employees or if their arrival leads to additional opportunities.

Another relevant labor market outcome is the average wage w_{it} within a CZ, as an in-migration shock can increase labor supply and could therefore lead to lower wages. On the contrary, the presence of agglomeration economies can make labor productivity increase with larger market size. These two opposing effects can lead to different net impacts across different timelines. In addition to wages, I also estimate the effect on output per worker y_{it}, another measure of aggregate productivity, as well as on the number of establishments per worker k_{it}. Through the lens of an expanding variety model, a larger workforce leads to higher entry, which in turn increases productivity. Thus, accounting for establishments will allow me to test the predictions of such frameworks.

Finally, I determine the response of housing prices hp_{it}, a major component of households' local expenditures. As the population of an area increases, these values should

[1]Previous papers in the growth literature suggested this mechanism (Uzawa, 1965; Lucas, 1988), albeit at an aggregate level in a representative agent economy.

increase in the short run, when the good is in fixed supply. This would in turn act as a congestion force, counteracting the agglomeration externalities and therefore leading to different effects across timelines. Moreover, recent trends in real estate prices, especially in markets experiencing a high inflow of migrants, have prompted a policy discussion aimed at increasing the housing supply. Consequently, to study these counterfactual policies we first need to estimate the effect of population increases on real estate values.

To estimate these effects, the first step is to model these variables using a panel SVAR structure, similar to the one in Boubtane, Coulibaly and Rault (2013), which accounts for the reciprocal relation between population flows and economic performance in a given geographical area. In this sense, define a vector of stationary observables $\boldsymbol{Y}_{it}$ and assume it follows a linear process[2]:

$$\boldsymbol{A} \cdot \boldsymbol{Y}_{it} = \sum_{l=1}^{p} \boldsymbol{\alpha}_l \cdot \boldsymbol{Y}_{it-l} + \boldsymbol{\varepsilon}_{it} \tag{1}$$

where $\boldsymbol{A}$ is a non-singular matrix and $\boldsymbol{\alpha}_l$ is a $7 \times 7p$ matrix which, along with $\boldsymbol{A}$, describes the relation between present and past values of $\boldsymbol{Y}_{it}$. Furthermore, $\boldsymbol{\varepsilon}_{it}$ is a vector of structural innovations uncorrelated across localities and time, with $E[\boldsymbol{\varepsilon}_{it}] = 0$ and $E[\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}'_{it}] = I$. The reduced form of the model above is given by:

$$\boldsymbol{Y}_{it} = \sum_{l=1}^{p} \boldsymbol{\delta}_l \cdot \boldsymbol{Y}_{it-l} + \boldsymbol{B}\boldsymbol{\varepsilon}_{it} \tag{2}$$

where $\boldsymbol{B} = \boldsymbol{A}^{-1}$ and $\boldsymbol{\delta}_1 = \boldsymbol{A}^{-1}\boldsymbol{\alpha}_1$.

Let $\boldsymbol{u}_{it} = \boldsymbol{B}\boldsymbol{\varepsilon}_{it}$ be the reduced-form residuals. After estimating equation (2) we need to identify $\boldsymbol{B}$ to construct impulse responses to structural migration shocks ε^m_{it}. Typically, when making timing assumptions, we would rely on the Cholesky decomposition of Σ, the covariance matrix of $\boldsymbol{e}_{it}$, to achieve this. This is the approach Boubtane, Coulibaly and Rault (2013) adopt in their analysis of OECD immigration. However, since I am only interested in the response to immigration shocks, and given the ordering of variables in $\boldsymbol{Y}_{it}$, I only need to identify the first column $\boldsymbol{b}^m$. By instead using instrumental variables as detailed in Stock and Watson (2018), I avoid imposing any timing or sign restrictions.

To see this, rewrite the reduced-form residuals of the j^{th} VAR equation as a function of the structural migration ε^m_{it} and non-migration $\boldsymbol{\varepsilon}^{-m}_{it}$ shocks:

$$u^j_{it} = b^m_j \cdot \varepsilon^m_{it} + \boldsymbol{b}'_{j,2:5} \cdot \boldsymbol{\varepsilon}^{-m}_{it} \tag{3}$$

where b^m_j corresponds to the first coefficient in the j^{th} row of matrix $\boldsymbol{B}$. If we use a unit-effect normalization, so that $b^m_m = 1$, we can substitute ε^m_{it} with the residuals from the

[2]This includes all the variables discussed above so that $\boldsymbol{Y}_{it} = [\Delta m_{it}, \Delta o_{it}, \Delta y_{it}, \Delta w_{it}, \Delta l_{it}, \Delta k_{it}, \Delta hp_{it}]'$.

migration equation and obtain:

$$u_{it}^{j} = b_{j}^{m} \cdot u_{it}^{m} + e_{it}^{j} \tag{4}$$

where e_{it}^{j} is a linear combination of the non-migration structural shocks $\boldsymbol{\varepsilon}_{it}^{-m}$.

As Stock and Watson (2018) note, although the residuals u_{it}^{j} are unobserved, to estimate b_{j}^{m} we can use each equation in the reduced-form VAR in (2) to rewrite equation (4) as[3]:

$$Y_{it}^{j} = b_{j}^{m} \cdot m_{it} + \sum_{l=1}^{p} \boldsymbol{\Gamma}_{l}^{j} \cdot \boldsymbol{Y}_{it-l} + e_{it}^{j} \tag{5}$$

where Y_{it}^{j} is the j^{th} entry in $\boldsymbol{Y}_{it}$. Since in-migration rates m_{it} are correlated with non-migration shocks we have $E[m_{it}^{m} e_{it}^{j}] \neq 0$. Consequently, estimating the coefficients of interest requires a vector of instruments $\boldsymbol{z}_{it}$ that satisfies the relevance and exclusion conditions, $E[\varepsilon_{it}^{m} \boldsymbol{z}_{it}] \neq 0$ & $E[\varepsilon_{it}^{-m} \boldsymbol{z}_{it}] = 0$ respectively. With these instruments, I can estimate equation (5) using two-stage least squares, recover column $\boldsymbol{b}^{m}$ and finally compute the dynamic response of all the local economic variables in $\boldsymbol{Y}_{it}$ to a structural population inflow shock ε_{it}^{m}.

2.2 Shift-Share Instruments

To identify the effect of structural market size shocks using the methodology above, I construct a set of Altonji and Card (1991)-style instruments z_{it}^{g}, based on migration shares s_{int} and demographic shocks g_{nt}:

$$z_{it}^{g} = \sum_{n} s_{int} \cdot g_{nt} \tag{6}$$

where n is the county from which migrants originate. To see how this instrument captures variation in market size, note that we can decompose population growth $\Delta \ln N_{it}$ as:

$$\Delta \ln N_{it} \approx \frac{\Delta N_{it-1}}{N_{it-1}} + \frac{\Delta b_{it}^{net}}{N_{it-1}} + \underbrace{\sum_{n} \left(\frac{\tilde{m}_{int-1}}{N_{it-1}} \cdot \frac{\Delta \tilde{m}_{int}}{\tilde{m}_{int-1}} \right)}_{\text{In-migration rate } \Delta m_{it}} - \frac{\Delta \tilde{o}_{it}}{N_{it-1}} \tag{7}$$

where b_{it}^{net} is births net of deaths, $\tilde{m}_{int}$ is the total flow of migrants coming from n to i and $\tilde{o}_{it}$ is the total outflow of individuals from i.

We can see that the third term in equation (7) is the in-migration rate, which is decomposed into inflow shares and growth rates. I use the former to construct the instrument so that $s_{int} = \tilde{m}_{int-1}/N_{it-1}$. This is the "pull" factor of migration: since individuals tend to move to places that previously received migrants coming from the same area, the share s_{int}

[3] From (2) we know $u_{it}^{j} = Y_{it}^{j} - \sum_{l=1}^{p} \boldsymbol{\delta}_{l}^{j} \cdot \boldsymbol{Y}_{it-l} - \nu_{i}^{j} - \nu_{t}^{j}$. Use this in (4) and define $\boldsymbol{\Gamma}_{l}^{j} = \boldsymbol{\delta}_{l}^{j} - b_{j}^{m} \boldsymbol{\delta}_{l}^{m}$ to obtain (5).

"pulls" households from n to i. On the other hand, the "push" factor will be the demographic shocks that produce an outflow from n and proxy for the growth rates within the sum in equation (7). To this end, and following a similar strategy to Karahan, Pugsley and Şahin (2019) and Shimer (2001), I set g_{nt} equal to the 25-year lags of birth rates in county n.

Note that, unlike previous migration studies that use this type of instrument, the shares are lagged by one year instead of multiple periods. Although this produces a stronger instrument it also introduces endogeneity since the predictable element of current productivity could influence past migration shares. Nevertheless, I follow Borusyak, Hull and Jaravel (2022) and estimate the IV regression at the *shock-level*, which in this case is at the county of origin level. Thus for each dependent variable Y_{it}^{j} in the VAR, instead of estimating equation (5), I use the following specification:

$$\bar{Y}_{nt}^{j\perp} = b_{j}^{m} \cdot \bar{m}_{nt}^{\perp} + \bar{e}_{nt}^{j\perp} \tag{8}$$

where, for every variable x, we have $x^{\perp}$ is the residual from a regression of x_{it} on the original controls, including the lagged values of Y_{it}^{j}, and $\bar{x}_{nt}^{\perp} = \sum_{i} \frac{s_{int}}{s_{nt}} x_{it}^{\perp}$.

In addition to recovering the same coefficient of interest b_{m}^{j}, this specification has other benefits. On the one hand, it recovers the appropriate standard errors and therefore avoids the issues with classic shift-share designs shown in Adao, Kolesár and Morales (2019), such as the potential correlation between residuals and instruments across observations. Furthermore, this method relies on the exogeneity of the shock g_{nt} and not that of s_{int}, which is why I can use migration shares from the previous period to strengthen the instrument.

To satisfy this exclusion restriction, birth rates 25 years ago in counties sending migrants need to be exogenous to unobservable variables affecting the current dependent variables in destinations. Specifically, following Borusyak, Hull and Jaravel (2022), $\bar{e}_{nt}^{j}$ is the average unobserved shock to variable j (for example wages) across CZs that are mostly receiving migrants from n. Lagged birth rates need to be uncorrelated with these unobservable shocks. This would be violated if, for example, individuals making fertility decisions could predict economic conditions 25 years in the future in the counties receiving the most migrants after that time interval. Aggregate shocks, such as booms and recessions, as well as permanent productivity differences, are controlled for by time fixed effects and using differenced variables respectively.

2.3 Data

The IRS provides annual Statistics of Income (SOI), which include county-to-county population flows from 1990 to 2018. Specifically, for any origin i and destination j in year t, we observe the number of individual income tax returns that changed addresses from i in $t-1$

to j in t. This dataset also includes the number of exemptions as well as the total adjusted gross income associated with those returns. Following IRS guidance, the first variable proxies for the number of households that move from one county to another whereas the second represents the number of individuals that do so. Consequently, I define migration inflows m_{it} as the total exemption inflow from counties within the US. I also define outflows o_{it} in a similar manner.

The BEA also provides GDP, wages, and employment at the county level, which are some of the variables in $\boldsymbol{Y}_{it}$[4]. This data is available from 2001 to 2019. Furthermore, I obtain the number of establishments for each county from the Business Dynamics Statistics (BDS) produced by the Census Bureau. To aggregate all these measures at the CZ level, I use the USDA ERS definitions from 2000, which map each county to one of 709 CZs. As described in Autor and Dorn (2013), these geographic areas group counties with strong commuting ties, and therefore constitute a better unit of analysis for studying the local effects of migration. Furthermore, unlike metropolitan areas that only cover certain locations, CZs encompass the entire continental US. Finally, to find housing price growth Δhp_{it}, I use the FHFA house price index, which is available at the county level. Therefore, I use the same delineation of CZs to compute an average housing index growth rate, weighted by the population of each county within a commuting zone. Table 1 shows the descriptive statistics for these variables.

Table 1 Descriptive Statistics

Commuting Zones	Mean	St. Dev.	10%	50%	90%
Inflow Rate (%)	1.61	1.25	0.60	1.37	2.77
Outflow Rate (%)	1.63	1.39	0.75	1.48	2.56
Net Immigration Rate (%)	-0.02	1.27	-0.49	-0.07	0.48
Employment-to-population	0.42	0.07	0.33	0.42	0.52
GDP per worker (000's)	98.85	36.02	75.44	91.68	124.06
Average Wage (000's)	37.55	6.89	30.70	36.15	45.85

Note: Sample consists of 660 commuting zones within the continental US (excluding Alaska) and years 2002-2018.

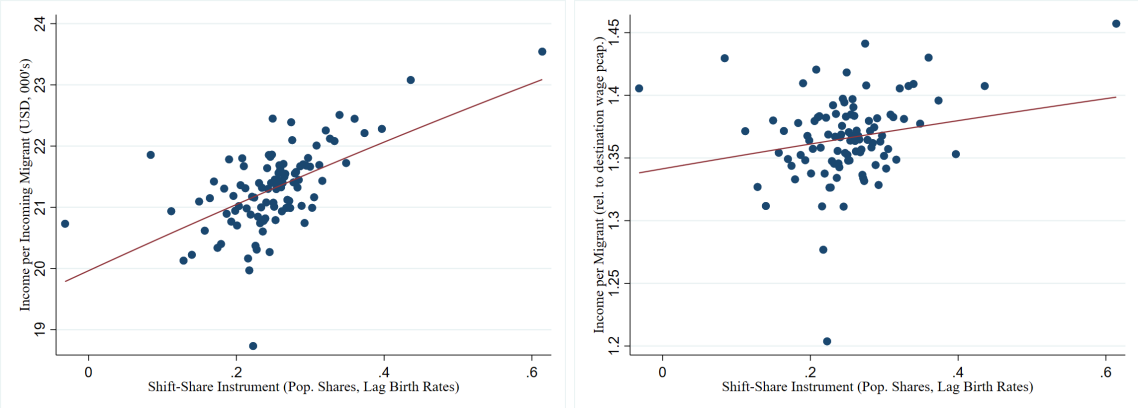
To construct the instrument I need birth rates at the county level. I obtain this from the CDC's National Vital Statistics microdata, which provides details on each birth during the calendar year[5]. Once I build the instrument I observe how it correlates with migrant income. The left panel of Figure 1 shows how the instrument is positively correlated with the absolute income per capita among households arriving to CZs. In other words, this

[4]GDP figures correspond to CAGDP2 whereas the remaining BEA variables come from the Economic Profile for counties (CAINC30)

[5]This data is publicly available for all counties until 1988. From 1989 onwards, I only have the number of births for counties with a population over 100,000 (which corresponds to 500 counties). Consequently, I impute the missing births using the average share (out of all state births) for each county for previous years.

suggest the "push" and "pull" factors are associated with high income migration. However, when we control for the average wage at the destination this relationship becomes flatter, as evidenced by the right panel of the same figure.

FIGURE 1. MIGRANT INCOME & SHIFT-SHARE INSTRUMENT



(A) Migrant income (absolute) & instrument

(B) Migrant income (relative) & instrument

Note: Each figure is a binned scatter plot with 100 bins. Controls for year and commuting zone fixed effects.

3 Empirical Results

The first subsection estimates the SVAR-IV model described above on the entire sample. The result is a set of structural impulse responses that show the evolution of local economies after a 1% increase in the migration rate. The remaining parts of this section delve deeper into these effects by studying the heterogeneity across different types of commuting zones as well as the impact on wages in several service industries.

3.1 Baseline SVAR-IV

The shift-share structure of the instrument allows any type of shock at the origin county level g_{nt} that is exogenous with current economic conditions. Consequently, I compare lagged birth rates with other demographic "push" factors in Table 2. The first is the total outflow from county n: note from equation (7), this effectively replaces the specific migration flow from n to i with the total number of people leaving n. This explains why this county-level shock produces a stronger instrument than the other two, with an F-statistic of 81.96[6] when combined with the migration-to-population shares found in equation (7). Nevertheless, this variable is the least likely to be uncorrelated with current unobservable economic shocks affecting destinations.

[6]All the first stage results, including the F-statistic come from the *shock-level* regression developed by Borusyak, Hull and Jaravel (2022).

Table 2 Shock-level IV First Stage

	Δm_{it} (1)	Δm_{it} (2)	Δm_{it} (3)	Δm_{it} (4)	Δm_{it} (5)	Δm_{it} (6)
z_{it}^{outflow}	74.90***	190.21***				
	(8.27)	(39.97)				
$z_{it}^{\text{pop. growth}}$			269.63***	12484.44*		
			(67.39)	(4974.16)		
$z_{it}^{\text{birth rate (25y ago)}}$					0.31*	26.12***
					(0.12)	(6.58)
# of destination-years	8,232	8,232	8,232	8,232	8,232	8,232
# of origin-years	11,241	11,241	11,357	11,357	11,357	11,357
F-stat	81.960	22.642	16.006	6.299	6.193	15.759
Shares relative to total migration	✓		✓		✓	
Shares relative to total population		✓		✓		✓

Note: Robust standard errors clustered at the originating county level in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; All regressions control for the lagged values of the variables in $\boldsymbol{Y}_{it}$ and contain time fixed effects. Year range is 2002-2018. For each shock (total outflow, population growth and lagged birth rate), this table reports two sets of first stage results as a robustness check. The first uses an instrument based on the share of migration from n to i relative to total migration $\tilde{m}_{int-1}/m_{it-1}$. The second is the migration-to-population share $\tilde{m}_{int-1}/N_{it-1}$.

Another county level "push" variable we can use is the total population growth rate in origin n. As column (4) in Table 2 shows, the resulting instrument is highly correlated with in-migration growth, with the highest coefficient among all three shocks. The increasing population in origin counties produces a higher outflow that gets allocated according to the migration shares. Nevertheless, its F-statistic of approximately 6.3 makes this a weak instrument. Furthermore, although total population growth is more likely to satisfy the exclusion restriction than total outflow, it can still suffer from endogeneity as it is not a lagged variable.

Finally, Table 2 also reports the first stage for the instrument that uses lagged birth rates, which is the one I use to compute the baseline IRFs later in this paper. As column (6) shows, as z_{it}^{birth} increases so does the growth rate of migration inflows. Furthermore, this coefficient is statistically significant and possesses an F-statistic of 15.76. These results, along with the arguments provided in the previous section, make this third instrument the preferred variable for estimating the response to a structural migration shock.

Indeed, Table 3 shows the results from the two-stage least squares estimation using the birth rate instruments. Recall these estimates correspond to the first column of matrix $\boldsymbol{B}$ in equation (2). As such, they identify the initial response of each variable to a migration inflow shock $\boldsymbol{\varepsilon}_{it}^{m}$. In this sense, we can see that outflows initially rise after an increase in inflows. However, when comparing them with the OLS estimates, we can see the true effect is lower and statistically not significant. The upward bias could be due to unaccounted shocks (in the OLS case) that are positively correlated with both inflows and outflows,

TABLE 3 SHOCK-LEVEL SSIV

	Δo_{it}		Δy_{it}		Δw_{it}		Δl_{it}		Δk_{it}		Δhp_{it}	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Δm_{it}	0.516***	0.123	0.001	0.141*	0.005***	0.091**	0.005***	0.125***	-0.002	-0.117**	0.004**	0.535***
	(0.038)	(0.174)	(0.003)	(0.064)	(0.001)	(0.029)	(0.001)	(0.036)	(0.001)	(0.036)	(0.001)	(0.144)
# of destination-years	10,258	8,232	10,258	8,232	10,258	8,232	10,258	8,232	10,258	8,232	10,258	8,232
# of origin-years		11,357		11,357		11,357		11,357		11,357		11,357

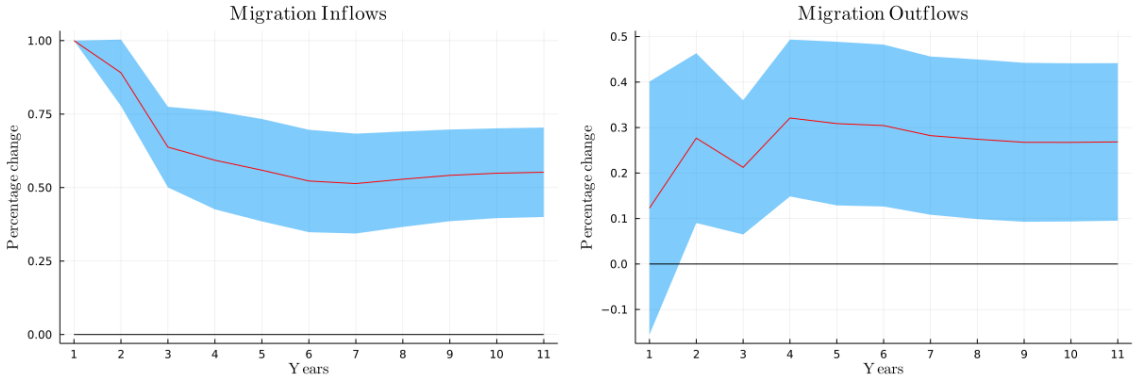
Note: Robust standard errors clustered at the originating county level in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; All regressions control for the lagged values of the variables in $\boldsymbol{Y}_{it}$ and contain time fixed effects. Year range is 2002-2018.

such as shifts in the local production process that benefit workers from other counties who replace established residents.

On the other hand, the contemporaneous response of output and wages is statistically significant and higher when using the instrument as seen in columns (4) and (6). While GDP per worker increases by 0.14%, average salaries do so by 0.09%, which suggests the 1% rise of in-migration mostly represents a demand shock that is passed through from the goods to the labor market. Additionally, the agglomeration externalities discussed in the introduction could be present upon impact. These mechanisms would also explain the rise in employment of 0.13% in column (8). Unobserved variables, such as changes in industrial composition that benefit native residents over incoming migrants while still increasing productivity, could be driving the downward bias in the OLS measures.

Finally, the number of establishments per worker drops by -0.12% while housing prices experience a rise of 0.54%, the highest increase of all 7 variables. Thus existing establishments absorb the arrival of new workers by creating new positions since the employment-to-population rises as we saw above. On the other hand, the direction and magnitude of the real estate response reveal a potential congestion effect that could counteract the demand shock driving wage and productivity growth in the medium to long run. To evaluate this, we now study the structural impulse responses.

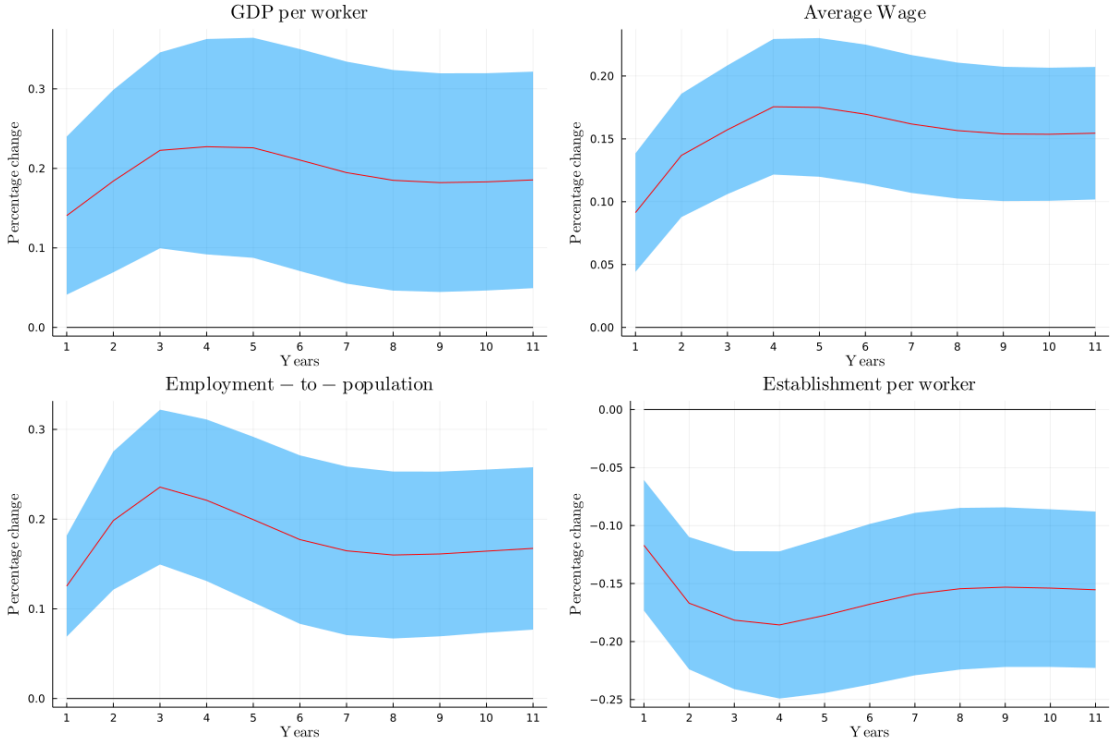
FIGURE 2. CUMULATIVE IRFS FOR MIGRATION FLOWS



Note: The red line represents the cumulative impulse response. The blue shade indicates a 90% confidence level. Standard errors are generated by Monte-Carlo with 200 repetitions.

After a 1% increase in the migration rate, the net inflow of commuting zones increases and remains high after several years. Figure 2 decomposes this effect into inflows and outflows following the initial shock. On the one hand, more people leave the area even after 8 years. However, this is only 0.3% higher than it was before the first arrival. On the other hand, inflows permanently increase by 0.5%. Thus, although some workers decide to move from the CZ, maybe because the shock negatively impacts their salaries and employment prospects, enough individuals decide to move into the area to maintain a positive net inflow. It is important to note that the cumulative effect of in-migration is lower than the initial rise. As the population suddenly increases, this could hurt certain amenities (such as long-term residential costs as we will see later), which makes the destination less attractive to future migrants so that part of the initial shock fades out.

FIGURE 3. CUMULATIVE IRFS FOR ECONOMIC VARIABLES



Note: The red line represents the cumulative impulse response. The blue shade indicates a 90% confidence level. Standard errors are generated by Monte-Carlo with 200 repetitions.

Since net inflows remain positive, the local market size continues to rise and therefore output and wages continue to increase as a response to surging demand. This occurs until they reach their peak between 3 and 4 years after the shock, as the top two panels of Figure 3 show. Notice all variables in the figure experience a medium run effect within the same time frame, as employment-to-population grows by 0.23% and the number of establishments per worker drops by 0.2%. During this time firms hire an increasing number of workers to

FIGURE 4. CUMULATIVE IRFS FOR ECONOMIC VARIABLES

Note: The red line represents the cumulative impulse response. The blue shade indicates a 90% confidence level. Standard errors are generated by Monte-Carlo with 200 repetitions.

meet demand, but this is also the period when longer-term agglomeration externalities, such as those described in Duranton and Puga (2004), can start affecting productivity.

Nevertheless, housing prices also grow in the medium term, exerting pressure on the budget constraint of residents. The cumulative impulse response in Figure 4 shows how these values are 1% higher once they reach their peak. This is a much larger effect than that of GDP per worker, which only grew by 0.2% in the same period. Consequently, newer migrants are discouraged from moving into the CZ and inflows fall as we saw before. The response of real estate values can also be a proxy for other unobservable negative externalities that also increase with population, such as traffic and pollution.

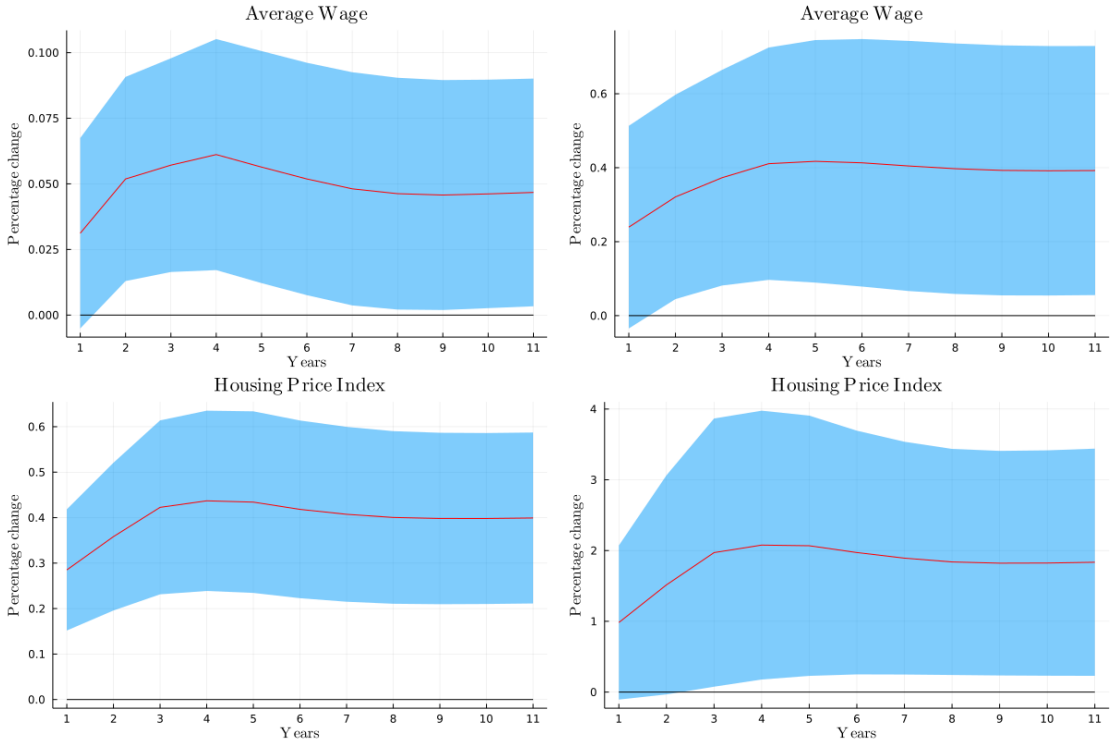
These are examples of the congestion forces that avoid a sustained rise in the in-migration rate. However, as time passes several factors can alleviate their negative effects. In the case of housing, growth in supply and new constructions can reduce the pressure on prices. This could be why we see a slight decline in Figure 4, after which values stabilize. Furthermore, since net inflows decreased from their initial level, at this point, the population is growing at a smaller rate as well as residential demand.

We can see the same pattern for the other variables in the VAR, with a long-run response that is smaller in magnitude than the one after 4 years. Output per worker, average wages, and the employment-to-population ratio all fall after their peak in Figure 3. Furthermore, the number of establishments rises slightly from its trough. The reason behind these responses is the same as the one we mentioned above: although the population is increasing, it is doing so at a lower rate. Presumably, this is the growth rate that balances agglomeration and congestion externalities. For example, if we consider wages, this is the market size growth rate at which the increase in demand and supply for labor offset one another.

3.2 Further Evidence

The previous results show the importance of studying the dynamic response of jointly distributed variables, as the short-, medium-, and long-run effects all differ. Now I explore whether these effects are heterogeneous across different types of CZs and industries, and if so what this reveals regarding the underlying mechanisms.

FIGURE 5. CIRFS FOR LOW AND MEDIUM DENSITY CZS



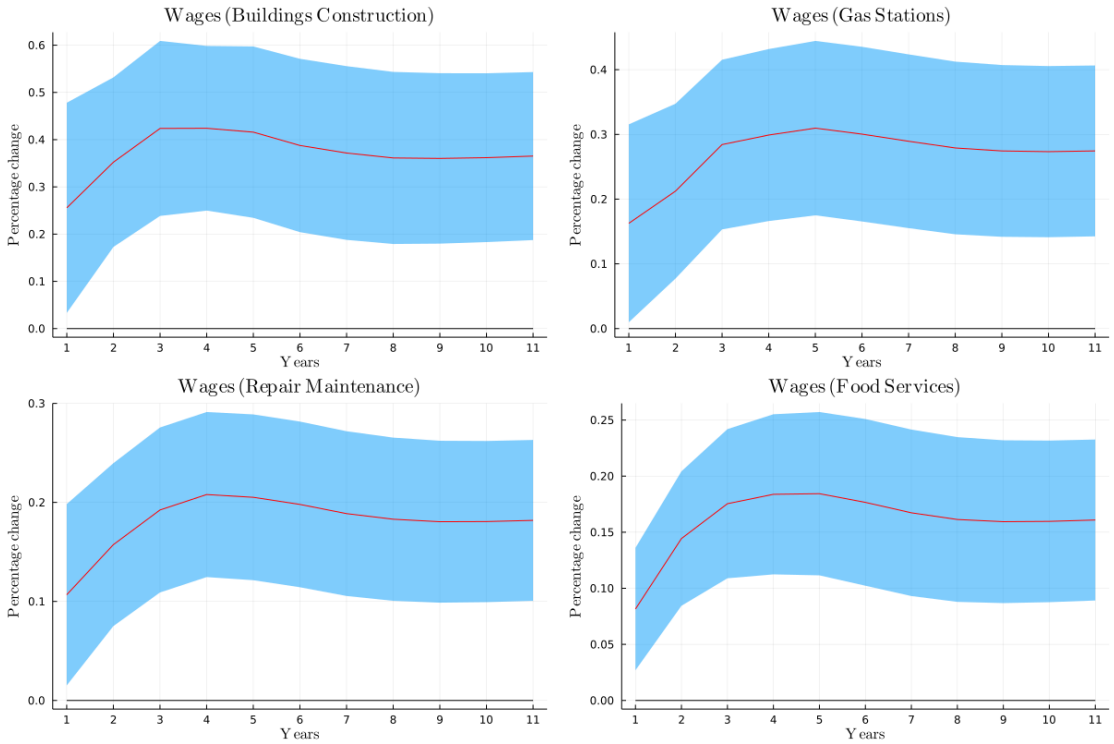
Note: The red line represents the cumulative impulse response. The blue shade indicates a 90% confidence level. Standard errors are generated by Monte-Carlo with 200 repetitions.

First, we can find the density of each CZ and determine whether the local effect changes depending on this factor. Specifically, I compute the average population for each zone throughout the years and classify each on whether they are in the bottom, medium or top third of the density distribution. I then re-estimate the SVAR-IV model specified above for each of these groups separately. The reason different densities could produce heterogeneous responses is the differential relevance of congestion and agglomeration forces. Workers in New York City interact with each other more frequently than those in Phoenix, where land is less densely populated, which facilitates the human capital externalities that Lucas (1988) described. Nevertheless, these areas also experience higher congestion, and therefore the net effect of migration shocks is ambiguous.

When we look at the plots by density in Figure 5 we can see the response in medium-density CZs is higher in magnitude than that in areas with a low population by square

mile. The average wage, for example, reaches a peak that is 0.4% higher in the former whereas in the latter it only grows by 0.05%. This difference in magnitude suggests the agglomeration externalities that drive productivity increases are more prevalent in denser areas. One such mechanism is the acceleration of human capital accumulation in cities explored in Crews (2023). Another is misallocation: in sparse markets, it is less likely for workers with particular skills to supply their labor in the industry where they possess a comparative advantage. However, as the right plots in the figure show, the standard errors become wider with density. This could be due to congestion effects that mitigate the agglomeration externalities.

FIGURE 6. CIRFS FOR WAGES ACROSS NAICS SECTORS



Note: The red line represents the cumulative impulse response. The blue shade indicates a 90% confidence level. Standard errors are generated by Monte-Carlo with 200 repetitions.

Recall from Figure 1 that migrant income is positively correlated with the instrument. Since the exogenous variation is coming from high-income migrants, the baseline results in Figure 3, and especially the response of wages, could be due to a composition effect. To test this I estimate the path of wages after the shock within industries with homogeneous low-skill requirements. If we observe a change in these salaries, we can conclude the impact of in-migration is broad-based. To carry this out I select a set of 3-digit NAIC sectors and add each of their average wages to the VAR to re-estimate the model with 8 variables

instead of 7[7]. I then plot the impulse response of each wage in Figure 6.

The results indeed suggest this is more than a composition effect. For reference, recall from the baseline responses that wages increased between 0.1% to 0.2%. As Figure 6 shows, this increase in salaries affects workers in a diverse set of industries. These include building construction and repair & maintenance services, where the impact is 0.4% and 0.2% at its peak respectively. Besides raising the average wage due to their higher income, incoming migrants increase the market size of CZs. This in turn produces a positive demand shock as well as triggering agglomeration externalities that affect services and non-tradable production. The following section explores these mechanisms quantitatively through the lens of a formal framework.

4 Theoretical Model

Increasing migration inflows by 1% produces a positive and sustained change in local economic outcomes such as wages, output and housing prices. However, the shape of the effect across time suggests the interaction between agglomeration and congestion externalities results in different effects at different intervals. To rationalize these results, this section develops a theoretical model that will later be disciplined with the structural impulse responses.

4.1 Environment

Households The economy consists of $g \in G$ regions in continuous time $t \in \mathbb{R}^+$ and a mass $N(t)$ of workers who discount the future at rate ρ and exit the labor force at regional rate δ. Furthermore, each period, workers enter the regional markets at an exogenous rate $b_g(t)$ are born. When they are born workers are endowed with a permanent productivity level $z \in \Omega_z$, which they draw from an initial distribution $z \sim F(z)$. They derive utility from consumption $C_g(t)$ and housing $H_g(t)$, as well as an amenity $D_g(t)$ so that:

$$U_g(C_g(t), H_g(t)) = D_g(t) \cdot C_g(t)^{\alpha} \cdot H_g(t)^{1-\alpha}$$

where $D_g(t) = D_g \cdot N_g(t)^{-\phi}$ and $\phi > 0$. Note that amenities are decreasing with local population. This is to account for congestion externalities other than housing such traffic or the public goods. The consumption good is a CES aggregate from differentiated regional varieties $C_g = (\sum_r c_{rg}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$. To buy variety r in g, households need to pay $p_{rg}(t)$, whereas $P_g^h(t)$ is the price for each unit of housing they consume.

[7] Each of these industries will have a VAR model associated with it. The Quarterly Census of Employment and Wages (QCEW) provides average wages at the county-industry level. To aggregate them at the CZ level, I use the same delineation detailed in Section 2

Each period thereafter they receive a migration opportunity, whose arrival follows a Poisson process with rate λ[8]. This parameter captures the fixed costs associated with moving to new markets. This is one of the two mobility frictions the model includes. After receiving the migration opportunity, workers draw a preference shock for each location $\xi_g(t)$, which are distributed i.i.d. Frechet with shape parameter ν. The second friction is a flow-utility cost κ_{gd} that the household pays for moving from g to d, where $\kappa_{gd} < 1$ for $g \neq d$ and $\kappa_{gg} = 1$. Thus workers consider the value of living in each location, given these mobility costs and preference shocks that are both multiplicative and permanent as in Crews (2023).

The mobility decisions of workers, along with their entry and exit from the economy, will determine the spatial distribution of z at time t dennoted by $\psi_g(z,t)$. Thus the total labor available in region g will be given by

$$L_g(t) = N(t) \cdot \int_{z \in \Omega_z} z \cdot \psi_g(z,t) dz$$

since workers supply their efficiency units z inelastically in the region g they choose to live in, in exchange for the local wage $W_g(t)$. Additionally, the population of region g is determined by $N_g(t) = N(t) \cdot \int_{z \in \Omega_z} \psi_g(z,t) dz$.

Production Perfectly competitive local producers supply housing services $H_g(t)$ by combining land $T_g(t)$ and labor $L_g^h(t)$ as inputs using the following function:

$$H_g(t) = \Gamma_h T_g(t)^{\theta} (L_g^h(t))^{1-\theta}.$$

where $\Gamma_h = \theta^{-\theta}(1-\theta)^{-(1-\theta)}$. I assume land is a fixed factor owned by immobile landlords who collect rents $R_g(t)$.

Following Romer (1990), production of regional varieties is subject to variety gains. Each period, under perfect competition, a local firm sources $y_{gd}(j,t)$ at price $q_g(j,t)$ from firm j to produce y_{gd}, the amount of variety $g \in G$ it ships to destination d. These inputs are aggregated following:

$$y_{gd}(t) = \frac{1}{\tau_{gd}} \left(\int_0^{M_g(t)} y_{gd}(j,t)^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}}$$

where $M_g(t)$ is the mass of intermediate input producers in g at time t, and τ_{gd} represents the usual iceberg trade cost. Local intermediate producers on the other hand face monopolistic

[8] I assume that newborns do not receive this migration opportunity and are not subject to the exit probability.

competition as well as labor requirements given by

$$l_g(j,t) = \frac{y_g(j,t)}{A_g} + F_g(t)$$

where A_g is the fixed productivity in region g and $F_g(t)$ is the fixed cost in terms of labor units that firms have to pay. Similar to the discrete time setting in Peters (2022), this related to the growth rate of the mass of firms as well as the *inter-temporal knowledge elasticity* η:

$$F_g(t) = \gamma_M(t) \cdot M_g(t)^{-\eta} \tag{9}$$

where $\gamma_M(t) = \dot{M}_g(t)/M_g(t)$.

4.2 Static Equilibrium

At any point in time t, the state variables for each geography $g \in G$ consist of the mass of producers $M_g(t)$ and the distribution of efficiency units $\psi_g(z,t)$. We can therefore define a static equilibrium as follows:

Definition 1. *For each $g \in G$, given $M_g(t)$ and $\psi_g(z,t)$, a **static equilibrium** in period $t \in \mathbb{R}_+$ is a set of prices and allocations for skill $\{W_g(t), L_g(t)\}$, consumption imported from all regions $\{c_{rg}(t), p_{rg}(t)\}_{r\in G}$, intermediate inputs $\{q_g(j,t), y_{gd}(j,t)\}_{j\in M_g(t), d\in G}$, housing $\{H_g(t), P_g^h(t)\}$ as well as land $\{T_g(t), R_g(t)\}$ such that given prices i) households and firms behave optimally, ii) allocations clear their respective markets.*

Given their place of residence, households decide how much to consume of each region's variety with a local price index given by $P_g(t) = (\sum_r p_{rg}(t)^{1-\sigma})^{1-\sigma}$. Worker z will spend a constant share of her income $v_g(z,t) = z \cdot W_g(t)$ on consumption goods and housing services. Since land is fixed the supply of housing services under perfect competition will be:

$$H_g(t) = \left(\frac{P_g^h(t)}{W_g(t)}\right)^{\frac{1-\theta}{\theta}} \cdot \frac{T_g}{\theta} \tag{10}$$

On the other hand, local producers decide how much of their output they allocate to each region (including their own) as well as the demand for local varieties, which will be

$$y_g(j,t) = \left(\frac{Q_g(t)}{q_g(j,t)}\right)^{\varepsilon} Y_g(t),$$

where $Y_g(t)$ is the total output of region g and $Q_g(t)$ is the price index these firms face. Intermediate producers take this demand and choose prices and labor so as to charge a markup over marginal cost $q_g(j,t) = \frac{\varepsilon}{\varepsilon-1} \cdot \frac{W_g(t)}{A_g}$, which implies $Q_g(t) = M_g(t)^{\frac{1}{1-\varepsilon}} (\frac{\varepsilon}{\varepsilon-1} \cdot \frac{W_g(t)}{A_g})$.

Furthermore, free entry determines total labor demand for production in region g:

$$L_g^y(t) = \int_0^{M_g(t)} l_g(j,t)dj = M_g(t)F_g(t)\varepsilon \tag{11}$$

Market clearing for labor states demand from the goods and housing sector has to equal total supply $L_g(t)$. This, along with zero profits and housing market clearing, sets the total efficiency units in housing production as a constant share of regional labor: $L_g^h(t) = (1-\theta)(1-\alpha)L_g(t)$. Total income in region g is $W_g(t)L_g(t)$ and balanced trade implies:

$$W_g(t)L_g(t) = \sum_d \pi_{gd}(t)W_d(t)L_d(t) \tag{12}$$

where $\pi_{gd}(t)$ is the share of consumption expenditures in d spent in goods from g. This in turn is given by:

$$\pi_{gd}(t) = \frac{\Big(\tau_{gd}\cdot Q_g(t)\Big)^{1-\sigma}}{\sum_{i\in G}\Big(\tau_{id}\cdot Q_i(t)\Big)^{1-\sigma}} \tag{13}$$

Using equilibrium input prices I obtain the price index $P_g(t) = Q_g(t)\cdot\pi_{gg}(t)^{\frac{1}{\sigma-1}}$ as well as the price of housing units in region g given by:

$$P_g^h(t) = \left(\frac{(1-\alpha)\cdot L_g(t)\cdot\theta}{T_g}\right)^{\theta}\cdot W_g(t) \tag{14}$$

Combining them with the demand for consumption goods and housing, I derive the flow utility a worker with productivity level z receives when living in region g:

$$U_g(z,t) \;=\; \underbrace{D_g(t)}_{Amenities} \;\cdot\; \underbrace{z\cdot\frac{\Gamma_g}{\Gamma_U}\cdot\Big(\pi_{gg}(t)\Big)^{\frac{\alpha}{1-\sigma}}\cdot\Big(M_g(t)\Big)^{\frac{\alpha}{\varepsilon-1}}\cdot\Big(L_g(t)\Big)^{-\theta(1-\alpha)}}_{Real\ Wage} \tag{15}$$

where $\Gamma_U = (\frac{\varepsilon}{\varepsilon-1})^{\alpha}((1-\alpha)\theta)^{\theta(1-\alpha)}$ and $\Gamma_g = A_g^{\alpha}T_g^{\theta(1-\alpha)}$. Note the real wage is a function of the mass of firms and total efficiency units in the regional market, as well as those in other localities through the own trade share $\pi_{gg}(t)$.

4.3 Dynamic Equilibrium and Balanced Growth Path

Now that we defined a static equilibrium for every t we can study the dynamics of the model, which involves spatial labor reallocation through migration as well as the evolution of the mass of firms within each region.

The following Hamilton-Jacobi-Bellman (HJB) equation summarizes the household's

recursive problem (I derive this equation in Appendix A.1.1):

$$(\delta+\rho)V_g(z,t) = U_g(z,t)+\dot{V}_g+\lambda\textstyle\sum_i m_{gi}(z,t)\Big[\delta_{gi}(z,t)V_i(z,t)-V_g(z,t)\Big] \tag{16}$$

where $\delta_{gi}(z,t)=\frac{1}{G}\cdot\kappa_{gi}\cdot m_{gi}(z,t)^{-(\frac{1}{\nu}+1)}$ and the migration shares are given by:

$$m_{gd}(z,t)=\frac{[\kappa_{gd}\cdot V_d(z,t)]^\nu}{\sum_i[\kappa_{gi}\cdot V_i(z,t)]^\nu} \tag{17}$$

This is the probability that a worker with productivity z moves from region g to d, conditional on having a migration opportunity mediated by λ.

Recall that the distribution of individual productivities $\psi_g(z,t)$ will determine regional labor and population. Thus the evolution of these variables will depend on the Kolmogorov forward equation below (see Appendix A.1.2 for a derivation):

$$\begin{aligned}\dot{\psi}_g(z,t) &= b_g(t)\cdot\frac{N_g(t)}{N(t)}\cdot f(z)-\delta\psi_g(z,t)\\ &- \lambda\Big[(1-m_{gg}(z,t))\psi_g(z,t)-\textstyle\sum_{i\neq g}m_{ig}(z,t)\psi_i(z,t)\Big]\\ &- \psi_g(z,t)\frac{\dot{N}(t)}{N(t)}\end{aligned} \tag{18}$$

Finally, equation (11) establishes a relation between the fixed cost, labor demand and the mass of firms. Namely, regions with a lower $F_g(t)$ will have a smaller individual labor requirement for each producer which tends to reduce total demand $L_g^y(t)$ in that market. Although this cost is fixed at any point t, equation (9) states its evolution will depend on that of the mass of firms $M_g(t)$. By combining both I obtain:

$$\dot{M}_g(t)=\left(\frac{\alpha+\theta(1-\alpha)}{\varepsilon}\right)L_g(t)M_g(t)^\eta \tag{19}$$

This differential equation determines how the regional mass of firms changes over time. I can now define a dynamic equilibrium.

Definition 2. *A **dynamic equilibrium** is i) a value function $V_g(z,t)$, ii) a distribution $\psi_g(z,t)$ and iii) migration shares $m_{gd}(z,t)$ for $g,d\in G$, $z\in\Omega_z$ and $t\in\mathbb{R}_+$, as well as iv) functions for labor, population and mass of firms $\{L,N,M\}$ and v) wages $\{W\}$ such that:*

1. *M evolves according to equation (19);*
2. *V and m solve the HJB equation in (16), taking $\{L,N,M\}$ and $\{W\}$ as given;*

3. *the densities ψ evolve according to the Kolmogorov equation (18) $\forall r, g \in G$, taking migration decisions as given;*

4. *given ψ, populations and total skills satisfy the following equations:*

$$N_g(t) = N(t) \cdot \int_{\Omega_z} \psi_g(z,t) dz \qquad L_g(t) = N(t) \cdot \int_{\Omega_z} z \cdot \psi_g(z,t) dz \qquad N(t) = \sum_{g \in G} N_g(t)$$

5. *the resulting prices and allocations constitute a static equilibrium $\forall t \in \mathbb{R}_+$.*

Balanced Growth Path A balanced growth path (BGP) is a special type of dynamic equilibrium where all variables grow at a constant rate. To define such a solution, I detrend each variable $x(t)$ by rewriting it relative to its long-run growth rate along the BGP $x(t) = \tilde{x}(t) \cdot e^{\gamma_x t}$. For all derivations see Appendix A.1.3. The result is a set of detrended equilibrium conditions.

The first is a differential equation describing the evolution of regional varieties: the detrended version of equation (19) in terms of growth rates. That is

$$\gamma_{\tilde{M}_g}(t) + \gamma_M = \left(\frac{\alpha + \theta(1-\alpha)}{\varepsilon} \right) \tilde{L}_g(t) \tilde{M}_g(t)^{\eta - 1}. \tag{20}$$

From this equation I also find $\gamma_L = \gamma_M(1-\eta)$, thereby connecting the growth rate of $M_g(t)$ and that of $L_g(t)$ along the BGP. I also detrend the HJB in equation (16) to obtain:

$$\begin{aligned} (\delta + \rho - \gamma_v)\tilde{V}_g(z,t) &= \tilde{U}_g(z,t) + \partial_t \tilde{V}_g(z,t) \\ &+ \lambda \textstyle\sum_i m_{gi}(z,t) \Big[\delta_{gi}(z,t) \tilde{V}_i(z,t) - \tilde{V}_g(z,t) \Big], \end{aligned} \tag{21}$$

where the effective discount rate is now $(\delta + \rho - \gamma_v)$ with $\gamma_v = (\frac{\alpha}{\varepsilon - 1} - (1-\eta)(\phi + \theta(1-\alpha)))\gamma_M$.

The detrended Kolmogorov forward equation (18) is now:

$$\begin{aligned} \partial_t \tilde{\psi}_g(z,t) &= b_g(t) \frac{\tilde{N}_g(t)}{\tilde{N}(t)} f(z) - \delta \tilde{\psi}_g(z,t) \\ &- \lambda \Big[(1 - m_{gg}(z,t)) \tilde{\psi}_g(z,t) - \textstyle\sum_{i \neq g} m_{ig}(z,t) \tilde{\psi}_i(z,t) \Big] \\ &- \Big(\tilde{\gamma}_N(t) + \gamma_N \Big) \tilde{\psi}_g(z,t). \end{aligned} \tag{22}$$

By detrending this equation I find $\gamma_\psi = 0$, which implies from the definition of $L_g(t)$, that labor grows at the same rate as population along the BGP so that $\gamma_L = \gamma_N$. The detrended

versions of these variables are are defined by

$$\tilde{N}_g(t) = \tilde{N}(t) \int_{z \in \Omega_z} \tilde{\psi}_g(z,t) dz \qquad \& \qquad \tilde{L}_g(t) = \tilde{N}(t) \int_{z \in \Omega_z} z \tilde{\psi}_g(z,t) dz.$$

Using population dynamics, I compute the growth rate of $\tilde{N}_g(t)$ as a function of regional birth rates and δ:

$$\frac{\dot{\tilde{N}}(t)}{\tilde{N}(t)} = \frac{\sum_g b_g(t) \cdot \tilde{N}_g(t)}{\tilde{N}(t)} - \delta - \gamma_N \tag{23}$$

Note that once the overall economy reaches the BGP, all detrended variables $\tilde{x}(t)$ are constant so that $\tilde{x}(t) = \bar{x}\ \forall t$. I now formally define this type of equilbrium below.

Definition 3. *A **balanced growth path** is a dynamic equilibrium in which population $N_g(t)$ grows at a constant rate γ_N across locations, with constant growth rates for all equilibrium variables and detrended functions $\{\tilde{V}, m, \tilde{\psi}\}$ such that:*

$$\begin{aligned} V_g(z,t) &= e^{\gamma_v t} \bar{V}_g(z) \\ m_{ig}(z,t) &= \bar{m}_{ig}(z) \\ \psi_g(z,t) &= \bar{\psi}_g(z) \end{aligned}$$

where $\{V, m, \psi\}$ solve the dynamic equilibrium, $\{\bar{V}, \bar{m}, \bar{\psi}\}$ are the detrended value function, migration shares and skill distributions at the BGP, $\gamma_v = (\frac{\alpha}{\varepsilon - 1} - (1-\eta)(\phi + \theta(1-\alpha)))\gamma_M$ and $\gamma_N = \gamma_M(1-\eta)$.

Along the BGP, the detrended HJB (21) becomes:

$$(\delta + \rho - \gamma_v)\bar{V}_g(z) = \bar{U}_g(z) + \lambda \sum_i \bar{m}_{gi}(z) \left[\bar{\delta}_{gi}(z) \bar{V}_i(z) - \bar{V}_g(z) \right] \tag{24}$$

whereas the Kolmogorov equation (22) is now

$$0 = \bar{b}_g \frac{\bar{N}_g}{\bar{N}} f(z) - (\delta + \gamma_N)\bar{\psi}_g(z) - \lambda(1 - \bar{m}_{gg}(z))\bar{\psi}_g(z) + \lambda \sum_{i \neq g} \bar{m}_{ig}(z)\bar{\psi}_i(z) \tag{25}$$

Finally, to find the population growth rate γ_N, I use the fact that $\dot{\tilde{N}} = 0$ along the BGP in equation 23. This implies:

$$\gamma_N = \frac{\sum_g \bar{b}_g}{\bar{N}} - \delta \tag{26}$$

Note that I could also derive this by integrating the BGP Kolmogorov equation (25) over all z and summing over g, as well as using the fact that along the BGP there is no net migration.

4.4 Model Solution

To compute the solution of this model, I adapt the algorithms in Achdou et al. (2022). This was done in a spatial setting in Crews (2023) to compute the steady state for a centralized economy. However, since an estimation of the structural parameters through impulse response matching requires a simulation of dynamic outcomes, I implement a fixed-point procedure that solves the transition path of the economy as well as a BGP algorithm. To simplify notation I group all general parameters into $\mathbf{\Theta}$ and regional fundamentals into $\mathbf{\Theta}_G$. The full algorithms are detailed in Appendix B.

5 Structural Estimation

With the framework in the previous section I can study the effects of local migration and relocation shocks on aggregate outcomes such as output and productivity. To carry this out, I first estimate the structural parameters so that the model is consistent with the impulse responses in Section 3.

5.1 Estimation Strategy

The model parametrization includes a tuple of regional fundamentals $\{T_g, A_g, D_g\}$ as well as 14 structural parameters: those I set exogenously $\mathbf{\Theta}_1 = \{\alpha, \rho, \sigma, \zeta_\tau, \gamma_L, \bar{N}\}$ and those I estimate within the simulation

$$\mathbf{\Theta}_2 = \{\varepsilon, \eta, \nu, \phi, \zeta_m, \theta, \lambda, \delta\}.$$

Given a set of parameters I compute the initial and final steady states, as well as the transition path that results from an exogenous series of regional birth rates $\{b_g(t)\}$ that take the economy away from the first BGP. This produces a simulated dataset I then use to replicate the shift-share IV and estimate the SVAR-IV described in Section 2.

Clustering Commuting Zones Solving the theoretical model is time-consuming and the cost increases dramatically with the number of regions. Therefore I need to reduce this set from the 660 CZs used in the empirical part to estimate the structural parameters of the theory. To carry this out, I applied a K-means clustering procedure based on geographical coordinates and the variables in the VAR.

In particular, for a given number of clusters, I use the dataset with all CZs in 2010 and perform a K-means clustering algorithm where the "features" are the variables in the empirical VAR as well as the CZ coordinates. This produces a set of clusters containing CZs that are similar in terms of migration, output and other economic variables, but that are also geographically close to each other. This rules out, for example, a cluster with some

CZs in California as well as Florida. I then aggregate the VAR variables at the cluster level and re-estimate the SVAR-IV to compare the IRF coefficients with the ones in Section 3. I repeat these steps for a number of clusters ranging from 15 to 100, and choose the grouping that produces the coefficients closest to the ones in the baseline estimation. The result is set of 35 CZ-clusters.

Exogenous Parameters I set $\rho = 0.1$ as in Crews (2023) and $\sigma = 2$ following the results in Boehm, Levchenko and Pandalai-Nayar (2023). I target the average share of personal consumption expenditures on housing over the time period and set $\alpha = 0.89$. For the initial distribution of productivities $f(z)$, I use a truncated $Lognormal(\mu_f, \sigma_f)$ distribution using the estimates of the mean and coefficient of variation from Huggett, Ventura and Yaron (2006) to set μ_f and σ_f. To identify available land for housing T_g, I use raster data from the USGS National Land Cover as well as Data Elevation Models. Specifically, for each county I find the area (in square miles) that does not contain open water, wetlands or perennial snow as well as with an inclination lower than 15% following architectural guidelines and Saiz (2010).

I follow Peters (2022) parameterize bilateral trade and migration costs as functions of distance by setting $\tau_{gd} = (d_{gd}/d_{min})^{\zeta_\tau}$ and $\kappa_{gd} = (d_{gd}/d_{min})^{\zeta_m}$, with $d_{gg} = d_{min}$. Since CZ group one or more counties, I can compute the average distance between them. Therefore I use the 5% quintile of these within-CZ average distances to set $d_{min} = 32$mi. Although I estimate ζ_m through impulse response matching, I estimate ζ_τ by regressing bilateral prices on distance, similar to Castro-Vincenzi et al. (2024).

In particular, I use the Commodity Flow Survey (CFS) datasets from 2012 and 2017, which provide information on domestic freight shipping within the US. For each shipment i, I observe its weight and value, as well as the great circle distance $d_{i,t}$ between shipment origin and destination. Thus, since in the model $p_{gd}(t) \propto \tau_{gd}$, I estimate the following regression:

$$\log p_{i,t} = \beta_0^p + \zeta_\tau \log(d_{i,t}) + \boldsymbol{X}_{i,t} + \epsilon_{i,t} \tag{27}$$

where $p_{i,m,n,t}$ is the value per pound of the shipment and $\boldsymbol{X}_{i,t}$ are a set of controls such as destination, origin and quarter fixed effects. The estimates are reported in Table 5 in the Appendix and I use the estimate in the column (4).

Finally, I assume the economy is at a BGP in 2002 and set γ_L to 1.25% by targeting the average growth rate of the civilian labor force in the 10 years preceding my sample. For total population I use the sum of wage employment in that year from the sample in Section 2 so that $\bar{N} \approx 134$ mil. workers.

Internal Estimation For a given set $\boldsymbol{\Theta}_2$, I estimate the remaining regional fundamentals $\{A_g, D_g\}$ by targeting output per worker and population shares in 2002 while solving the

initial steady state. I use a modified version of Algorithm 1 where I solve for the fixed point of parameters in addition to endogenous variables. This also allows me to find λ and δ. I impute the latter by using equation 26 and the birth rates $\bar{b}_g^1$ observed 25 years before the first year in my sample. These are the initial elements in the series $\{b_g(t)\}$ I will use to simulate the transition. On the other hand, to infer the probability of migration opportunities I target the sum of migration inflows into all regions:

$$\underbrace{\frac{\sum_g \bar{m}_g}{\bar{N}}}_{Observed} = \lambda \cdot \underbrace{\sum_g \int \sum_{i \neq g} \bar{m}_{ig}(z) \cdot \bar{\psi}_i(z) dz}_{Model\ Variable} \tag{28}$$

Once this is complete, I compute the second endpoint for the transition: the final BGP. In terms of parameters, the only difference with the initial steady state is that I use the lagged birth rates corresponding to the last year of my sample $\bar{b}_g^2$.

With both endpoints in hand, I can now use Algorithm 2 to compute the transition dynamics of the economy. In particular, I leave all parameters constant except birth rates $\{b_g(t)\}$, which are the same as the ones I use to build the shift-share instrument. This results in a panel series of 17 years (including the first and last that I use for the BGP computation). Note it is unlikely the economy reaches the second steady state after this period. Therefore, I assume $b_g(t) = \bar{b}_g^2$ for all t thereafter and set the total periods to a higher number ($T = 200$).

To estimate $\boldsymbol{\Theta}_2$, I follow an impulse response matching procedure. Specifically, I repeat the internal estimation steps described above for a sequence of parameters in a Sobol grid and seek to minimize the following objective function:

$$J = \min_{\boldsymbol{\Theta}_2} (\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}(\boldsymbol{\Theta}_2))' W^{-1} (\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}(\boldsymbol{\Theta}_2)) \tag{29}$$

where $\hat{\boldsymbol{\Gamma}}$ is a vector with the IRFs estimated using the 35 clusters, $\boldsymbol{\Gamma}(\boldsymbol{\Theta}_2)$ are the model generated IRFs. The weighting matrix W^{-1} is diagonal and contains the reciprocals of coefficient variance, similar to Christiano, Eichenbaum and Evans (2005).

5.2 Parameter Estimates

The results from the impulse response matching are displayed in Table 4 and Figure 7 shows the model fit.

Table 4 Structural Parameters

Parameter	Interpretation	Estimate
ε	*Elasticity of substitution*	6.57
η	*Inter-temporal elasticity*	0.83
ν	*Migration elasticity*	3.41
ϕ	*Elasticity of amenities*	2.90
θ	*Share of land in housing*	0.93
λ	*Migration Probability*	0.01

6 Population and Relocation Shocks

How would aggregate output per worker respond over time if some regions lost a share of their workers? Although the IRFs in Section 3 show what the effect of migration on local conditions is, they are not enough to infer the relevance of population flows on aggregate variables. Therefore, to answer this question, I use the calibrated model. Specifically, I consider two policy exercises separately. The first is a "deportation" shock that decreases the total number of workers in the economy by reducing the population in certain areas while keeping the size of the remaining regions constant. The second one is a "relocation" shock: instead of maintaining the population of other locations, I increase it by a uniform share so as to keep the total number of households constant.

An important element of these counterfactuals is that the local population share that is removed or relocated differs across treated regions so that these local shocks are a geographically heterogeneous. To calibrate these shares I use the estimates from Passel and Krogstad (2023) to identify the states with the highest number of undocumented immigrants. I then match these with one of the 35 clusters in my model and perform the counterfactuals. Consequently, we can think of the first shock as a general deportation of approximately 1% of the labor force. By region however, this percentage varies from 3% to 5%. On the other hand, in the second exercise these workers are redistributed to places with a lower incidence of undocumented immigration, in a similar manner to the transportation of migrants to sanctuary cities by some states in 2023 (see for example Goodman et al. (2024)).

6.1 Local Deportations

Figure 11 displays the effects of the initial shock. It reveals that remaining residents in regions directly impacted by the deportation experience less congested amenities and an average increase in real incomes of 1.8%, after adjusting for prices of consumption goods and housing services. As these regions become more attractive, more workers choose to move there, resulting in an immediate 20% increase in gross inflows and a 4% decrease in outflows, while the opposite trends are observed in non-treated areas. Over time, as households choose to leave these non-treated regions, their amenities also become less congested.

Over time, the effects of the initial shock begin to dissipate, and most variables stabilize at new levels relative to the original steady state. With fewer workers in the economy, amenities in the long run increase by 2.5%, and real incomes rise by 1%. This aligns with the long-term effect on housing prices, shown in Figure 9, which exhibit a 1% decrease. Notably, if the shock were to target high-skilled individuals instead of being uniformly distributed across the regional productivity distribution $\psi_g(z,t)$, housing prices would experience a smaller decline, while real incomes would be higher. Intuitively, such a policy would lower the average skill level in affected regions, thereby reducing total demand for housing services and exerting downward pressure on prices.

Local populations also converge to a new steady state that is 1% lower than the level observed before the policy was implemented. A smaller workforce implies reduced migration flows, as shown in Figure 11. Additionally, a lower labor supply means fewer firms can cover the fixed costs of operation, leading to a decrease in the number of available varieties. This effect is further amplified by the dynamic influence of $M_g(t)$ on fixed costs, as described in equation (9). Together, these mechanisms contribute to a steady decline in the mass of firms, illustrated in Figure 9. A similar process unfolds in unaffected areas, where population loss due to migration outflows also reduces labor and variety levels, ultimately decreasing output in those regions. Meanwhile, in the regions affected by the shock, migration inflows gradually increase the variety of goods over time. Despite this, both shocked and untreated markets converge to a steady-state level of output approximately 2% lower than the original steady state.

In terms of aggregate effects, the deportation shock results in a long-run decline in output per worker of 1%, as shown in Figure 10. This new steady state arises because workers exit local economies without being replaced elsewhere, unlike the "relocation" shock I explore later. Consequently, the overall economy is left with a smaller labor supply and, therefore, fewer varieties, which reduces productivity. This outcome is similar to an inverse version of the refugee allocation scenario in Peters (2022), where output per worker also rises in the long run. However, in this case, the effect is driven by agglomeration externalities across all markets, rather than being limited to the manufacturing sector. Additionally, it is worth noting the different impacts depending on the type of workers removed. If high-skilled workers leave, aggregate output per capita immediately decreases to the new level. By contrast, if the shock removes low-skilled labor, output per capita initially rises, as the average skill level in the economy increases. Over time, however, the same dynamics that reduce varieties come into play, though at a slower rate.

6.2 Relocating Workers

How would aggregate productivity react if instead of removing workers they were re-allocated to other sectors? Effectively, this policy also takes the economy out of its initial

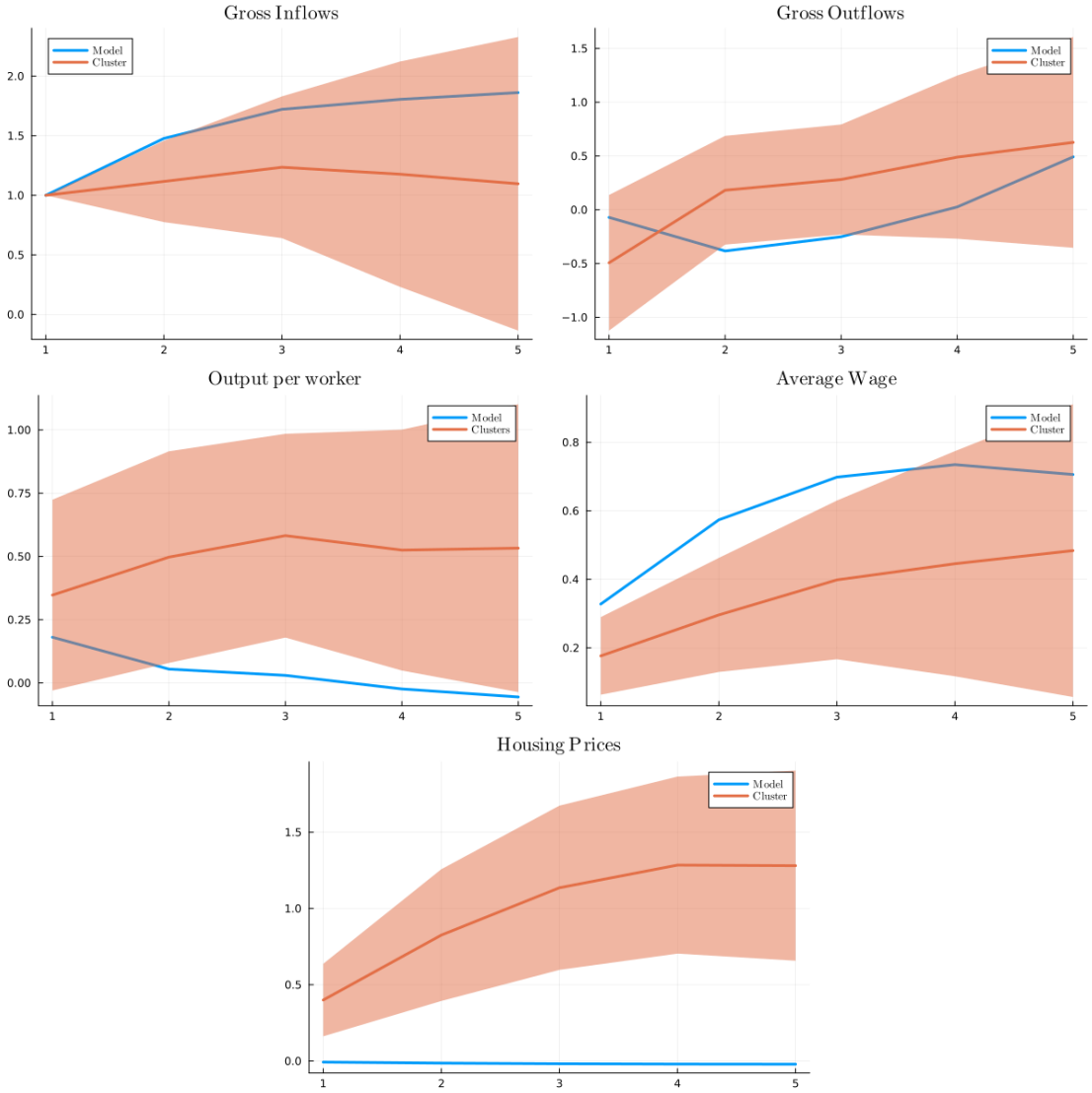
steady state but does not change the BGP towards which the economy transitions to in the long-run. Figure 11 shows that, similar to before, regions that lose their population see an increase of net inflows whereas the opposite is true of areas that receive the initial influx. Right after the shock, congested amenities and higher housing prices drive migrants back to areas whose population was relocated. This dynamic continues for 50 years until the number of workers in each places converges back.

The mass of firms in each region also transition back to the initial steady state, as well as output. However, unlike population, these variables take longer to converge. The number of producers is a slow moving variable that follows the differential equation 20. It therefore takes time to adjust, which is why the initial drop in shocked areas also occurs in a longer time frame relative to labor.

7 Conclusion

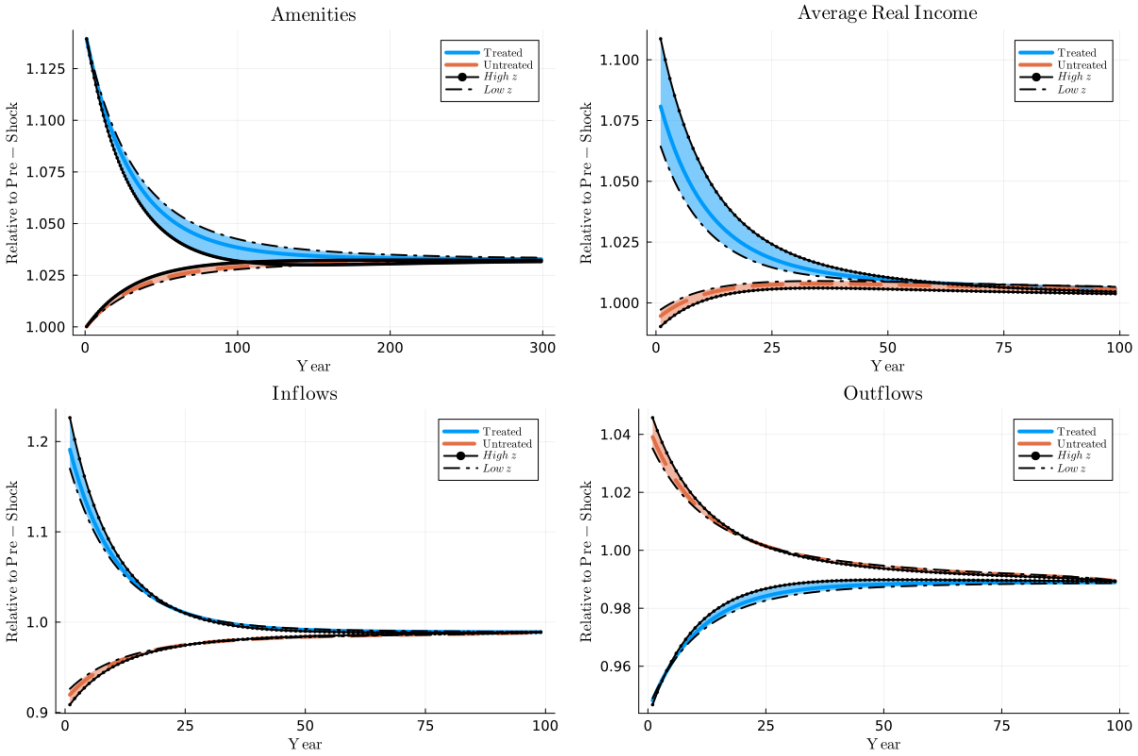
This paper studies the dynamic effect of increasing population inflows on local markets. Most of the previous research on labor mobility and its impact on economic outcomes focused on either short or long-term responses, but did not consider transitional dynamics. Although some papers address this element, their identification of structural migration shocks relies on timing assumptions, whereas the analysis above adopts an SVAR-IV approach using instruments based on weather variations. The results reveal labor productivity, wages and employment opportunities in urban economies respond positively to a 1% increase in population inflows.

FIGURE 7. MODEL FIT



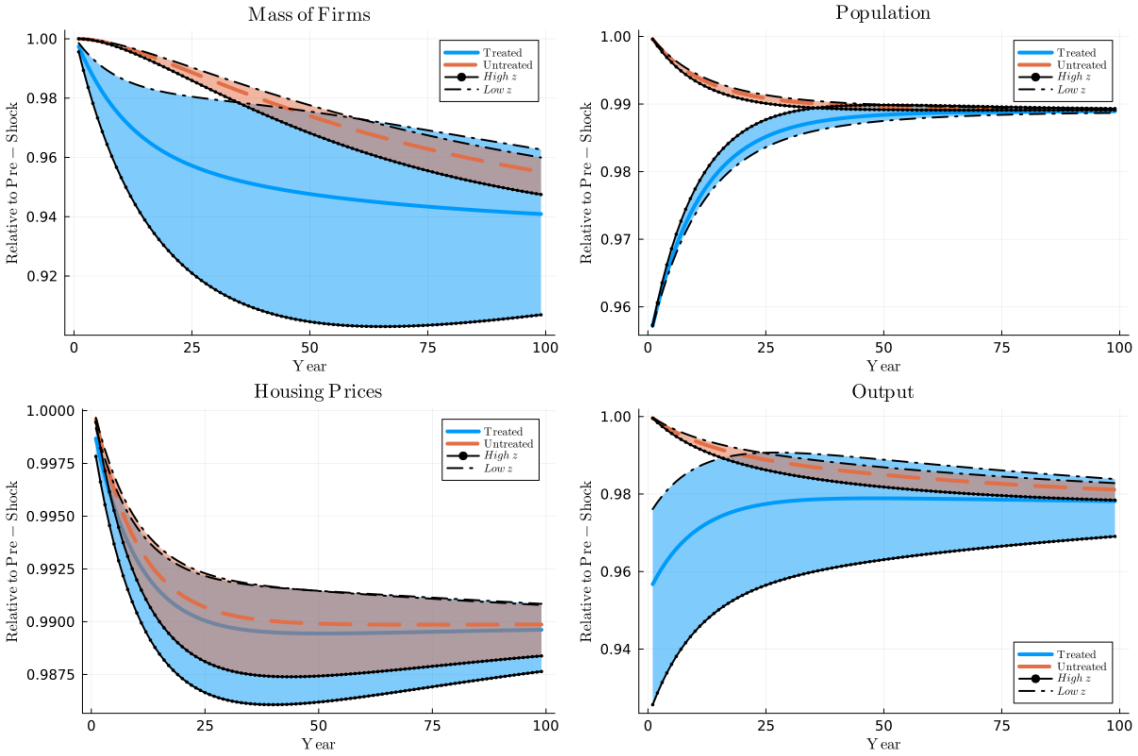
Note: The red line represents the cumulative impulse response (CIRF) estimated from the 35 K-means clusters. The red shade indicates a 90% confidence level. Standard errors are generated by Monte-Carlo with 200 repetitions. The blue line is the CIRF estimated by simulating the model with the baseline parametrization obtained from indirect inference.

FIGURE 8. DEPORTATION SHOCK: UTILITY AND POPULATION FLOWS



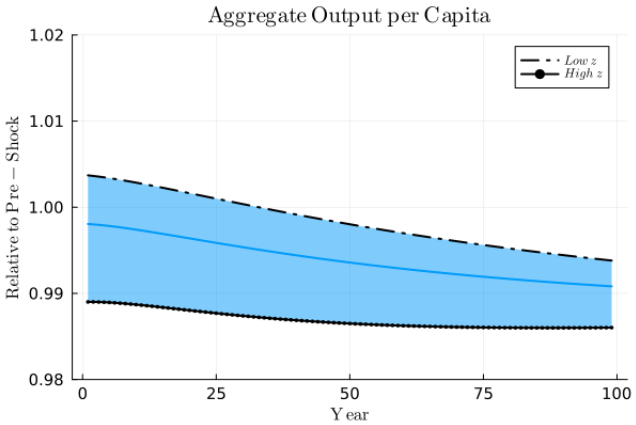
Note: Responses of amenities, real income, and gross migration flows to the "deportation" shock where heterogeneous regional population shares are removed from the economy. The blue line represents the response in regions that experience a population reduction ("treated") whereas the red line represents the effect in the remaining ones ("untreated"). In both settings, the policy removes a uniform share across the regional productivity distribution, whereas the bands represent the cases where only the top 10% ("High z") or bottom 10% ("Low z") are removed. Variables x_t are plotted relative to their pre-shock levels $\hat{x}_t = x_t / x_{pre-shock}$.

FIGURE 9. DEPORTATION SHOCK: POPULATION, VARIETIES AND OUTPUT



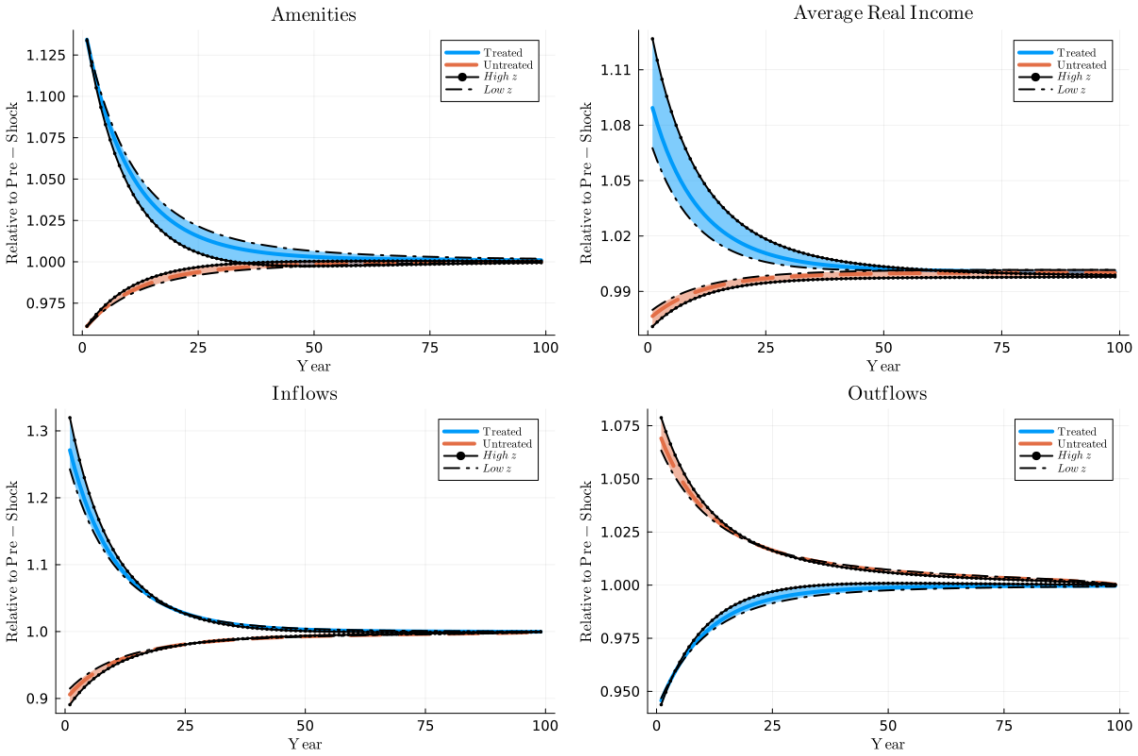
Note: Responses of varieties, population, housing prices and output to the "deportation" shock where heterogeneous regional population shares are removed from the economy. The blue line represents the response in regions that experience a population reduction ("treated") whereas the red line represents the effect in the remaining ones ("untreated"). In both settings, the policy removes a uniform share across the regional productivity distribution, whereas the bands represent the cases where only the top 10% ("High z") or bottom 10% ("Low z") are removed. Variables x_t are plotted relative to their pre-shock levels $\hat{x}_t = x_t / x_{pre-shock}$.

FIGURE 10. DEPORTATION SHOCK: AGGREGATE OUTPUT



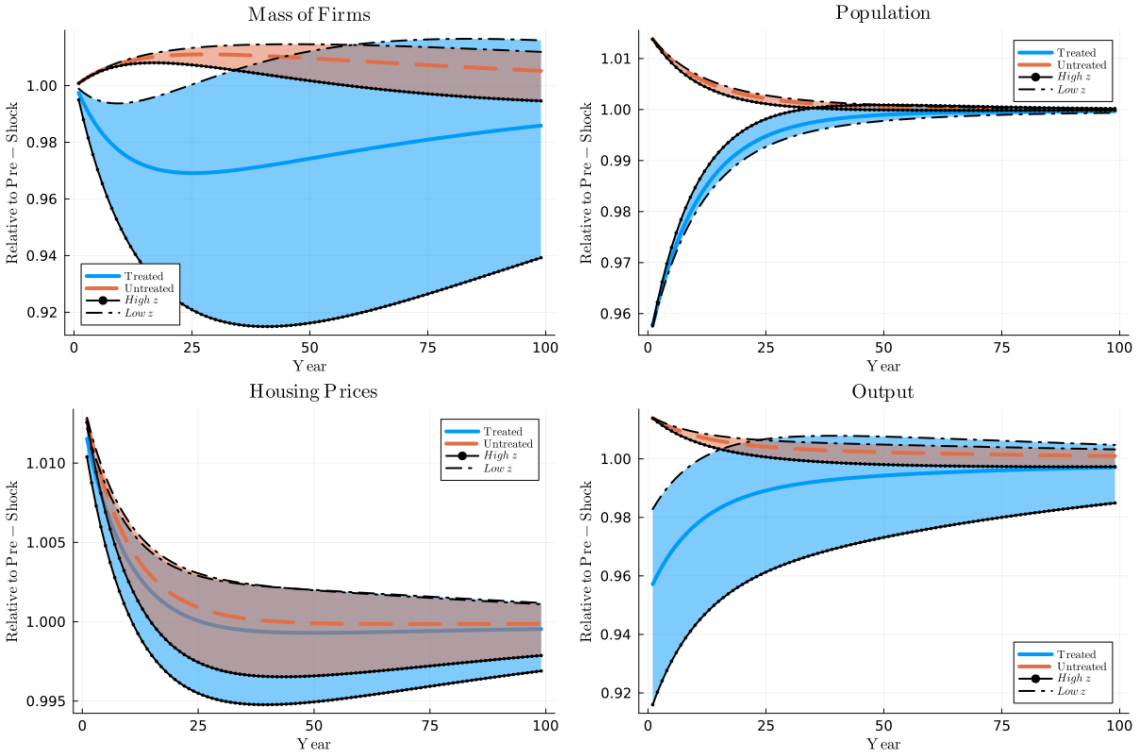
Note: Response of aggregate output per worker to the "deportation" shock where heterogeneous regional population shares are removed from the economy. The blue line represents the response when the shock removes a uniform share across the regional productivity distribution, whereas the bands represent the cases where only the top 10% ("High z") or bottom 10% ("Low z") are removed. Output per worker is plotted relative to its pre-shock levels $\hat{x}_t = x_t / x_{pre-shock}$.

FIGURE 11. RELOCATION SHOCK: UTILITY AND POPULATION FLOWS



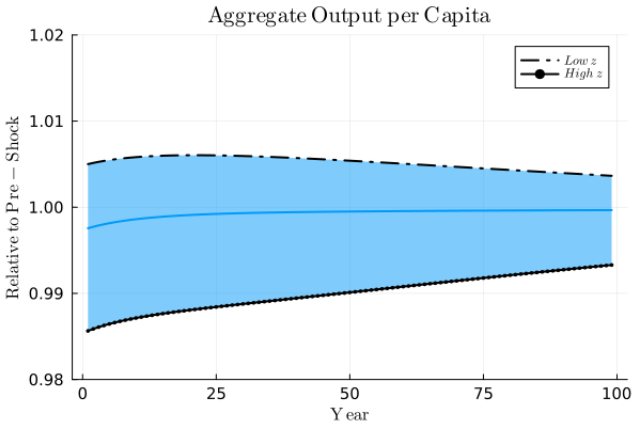
Note: Responses of amenities, real income, and gross migration inflows to the "deportation" shock where heterogeneous regional population shares are removed from the economy. The blue line represents the response in regions that experience a population reduction ("treated") whereas the red line represents the effect in the remaining ones ("untreated"). In both settings, the policy removes a uniform share across the regional productivity distribution, whereas the bands represent the cases where only the top 10% ("High z") or bottom 10% ("Low z") are removed. Variables x_t are plotted relative to their pre-shock levels $\hat{x}_t = x_t / x_{pre-shock}$.

FIGURE 12. RELOCATION SHOCK: POPULATION, VARIETIES AND OUTPUT



Note: Responses of varieties, population, housing prices and output to the "deportation" shock where heterogeneous regional population shares are removed from the economy. The blue line represents the response in regions that experience a population reduction ("treated") whereas the red line represents the effect in the remaining ones ("untreated"). In both settings, the policy removes a uniform share across the regional productivity distribution, whereas the bands represent the cases where only the top 10% ("High z") or bottom 10% ("Low z") are removed. Variables x_t are plotted relative to their pre-shock levels $\hat{x}_t = x_t / x_{pre-shock}$.

FIGURE 13. RELOCATION SHOCK: AGGREGATE OUTPUT



Note: Response of aggregate output per worker to the "deportation" shock where heterogeneous regional population shares are removed from the economy. The blue line represents the response when the shock removes a uniform share across the regional productivity distribution, whereas the bands represent the cases where only the top 10% ("High z") or bottom 10% ("Low z") are removed. Output per worker is plotted relative to its pre-shock levels $\hat{x}_t = x_t / x_{pre-shock}$.

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2022. "Income and wealth distribution in macroeconomics: A continuous-time approach." *The review of economic studies*, 89(1): 45–86.

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-share designs: Theory and inference." *The Quarterly Journal of Economics*, 134(4): 1949–2010.

Altonji, Joseph G., and David Card. 1991. "The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives." *Immigration, Trade, and the Labor Market*, 201–234. University of Chicago Press.

Autor, David H, and David Dorn. 2013. "The growth of low-skill service jobs and the polarization of the US labor market." *American economic review*, 103(5): 1553–1597.

Barcellos, Silvia Helena. 2010. "The dynamics of immigration and wages." *Rand Work. Pap. WR-755, Santa Monica, CA*.

Boehm, Christoph E, Andrei A Levchenko, and Nitya Pandalai-Nayar. 2023. "The long and short (run) of trade elasticities." *American Economic Review*, 113(4): 861–905.

Borjas, George J. 2003. "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.

Borjas, George J. 2013. "The analytics of the wage effect of immigration." *IZA Journal of Migration*, 2(1): 1–25.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *The Review of Economic Studies*, 89(1): 181–213.

Boubtane, Ekrame, Dramane Coulibaly, and Christophe Rault. 2013. "Immigration, growth, and unemployment: Panel VAR evidence from OECD countries." *Labour*, 27(4): 399–420.

Burchardi, Konrad B, and Tarek A Hassan. 2013. "The economic impact of social ties: Evidence from German reunification." *The Quarterly Journal of Economics*, 128(3): 1219–1271.

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

Castro-Vincenzi, Juanma, Gaurav Khanna, Nicolas Morales, and Nitya Pandalai-Nayar. 2024. "Weathering the storm: Supply chains and climate risk." National Bureau of Economic Research.

Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans. 2005. "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of political Economy*, 113(1): 1–45.

Crews, Levi. 2023. "A Dynamic Spatial Knowledge Economy." Working Paper.

Davis, Morris A, Jonas DM Fisher, and Toni M Whited. 2014. "Macroeconomic implications of agglomeration." *Econometrica*, 82(2): 731–764.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of urban agglomeration economies." In *Handbook of regional and urban economics*. Vol. 4, 2063–2117. Elsevier.

Edo, Anthony. 2019. "The impact of immigration on the labor market." *Journal of Economic Surveys*, 33(3): 922–948.

Gertler, Mark, and Peter Karadi. 2015. "Monetary policy surprises, credit costs, and economic activity." *American Economic Journal: Macroeconomics*, 7(1): 44–76.

Goodman, J. David, Keith Collins, Edgar Sandoval, and Jeremy White. 2024. "Bus by Bus, Texas' Governor Changed Migration Across the U.S." *New York Times*.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings." *Journal of Political Economy*, 118(3): 536–598.

Howard, Greg. 2020. "The migration accelerator: Labor mobility, housing, and demand." *American Economic Journal: Macroeconomics*, 12(4): 147–179.

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2006. "Human capital and earnings distribution dynamics." *Journal of Monetary Economics*, 53(2): 265–290.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin. 2019. "Demographic origins of the startup deficit." National Bureau of Economic Research.

Kline, Patrick, and Enrico Moretti. 2014. "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority." *The Quarterly journal of economics*, 129(1): 275–331.

Lucas, Robert E Jr. 1988. "On the mechanics of economic development." *Journal of monetary economics*, 22(1): 3–42.

Mertens, Karel, and Morten O Ravn. 2013. "The dynamic effects of personal and corporate income tax changes in the United States." *American economic review*, 103(4): 1212–47.

Moretti, Enrico. 2004. "Human capital externalities in cities." In *Handbook of regional and urban economics*. Vol. 4, 2243–2291. Elsevier.

Moretti, Enrico. 2010. "Local multipliers." *American Economic Review*, 100(2): 373–77.

Passel, Jeffrey S, and Jens M Krogstad. 2023. "What we know about unauthorized immigrants living in the US." *Pew Research Center*, 6.

Peters, Michael. 2022. "Market Size and Spatial Growth—Evidence From Germany's Post-War Population Expulsions." *Econometrica*, 90(5): 2357–2396.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of political Economy*, 98(5, Part 2): S71–S102.

Saiz, Albert. 2010. "The geographic determinants of housing supply." *The Quarterly Journal of Economics*, 125(3): 1253–1296.

Shimer, Robert. 2001. "The impact of young workers on the aggregate labor market." *The Quarterly Journal of Economics*, 116(3): 969–1007.

Stock, James H, and Mark W Watson. 2018. "Identification and estimation of dynamic causal effects in macroeconomics using external instruments." *The Economic Journal*, 128(610): 917–948.

Uzawa, Hirofumi. 1965. "Optimum technical change in an aggregative model of economic growth." *International economic review*, 6(1): 18–31.

A Theory Appendix

A.1 Dynamic Equilibrium Solution

A.1.1 Derivation of the HJB

Following similar steps to those in Crews (2023) we find the following Bellman in discrete time:

$$V_g(z,t) = U_g(z,t) + [(1-\rho)(1-\delta)]\mathbb{E}_t\left[\tilde{\xi}(t+1)\tilde{\kappa}(t+1)V_{g_{t+1}}(z,t+1)\right]$$

The expectation is with respect to the location in the next period since the households does not know whether or not it will have the opportunity to move (which occurs with probability λ), and what the preference shock will be. Thus we can rewrite the equation above as:

$$V_g(z,t) \;=\; U_g(z,t) + [(1-\rho)(1-\delta)]\left\{(1-\lambda)V_g(z,t+1) + \lambda\mathbb{E}_{\xi_i}\left[\max_i \xi_i\kappa_{g,i}V_i(z,t+1)\right]\right\}$$

To simplify notation define $V_{gi,t+1} = \kappa_{g,i}V_i(z,t+1)$. Since ξ_i follows a Frechet distribution with shape parameter ν and scale parameter[9] $s = \Gamma_\nu^{-1}$, we can find an expression for:

$$\begin{aligned}
\tilde{F}(y) &= P\left(\max_i\{\xi_i V_{gi,t+1}\} \le y\right) \\
&= P\left(\xi_1 \cdot V_{g1,t+1} \le y \cap \ldots \cap \xi_G \cdot V_{gG,t+1} \le y\right) \\
&= P\left(\xi_1 \le \frac{y}{V_{g1,t+1}}\right)\cdots P\left(\xi_G \le \frac{y}{V_{gG,t+1}}\right) \\
&= \exp\left(-(\frac{y}{s\cdot V_{g1,t+1}})^{-\nu}\right)\cdots\exp\left(-(\frac{y}{s\cdot V_{gG,t+1}})^{-\nu}\right) \\
&= \exp\left(-\left(\frac{y}{s}\right)^{-\nu}\cdot\textstyle\sum_{i=1}^{G}(V_{gi,t+1})^{\nu}\right)
\end{aligned}$$

Notice we can rewrite this as:

$$\tilde{F}(y) = \exp\left(-\left[\frac{y}{s\cdot\Phi}\right]^{-\nu}\right)$$

where $\Phi = [\sum_{i=1}^{G}(V_{gi,t+1})^{\nu}]^{\frac{1}{\nu}}$. Thus $y \sim Frechet(s\cdot\Phi,\nu)$ and therefore:

$$\begin{aligned}
E_{\xi_i}[\max_i \xi_i V_{gi,t+1}] &= s\cdot\Phi\cdot\Gamma_\nu \\
&= \left[\textstyle\sum_{i=1}^{G}(V_{gi,t+1})^{\nu}\right]^{\frac{1}{\nu}}
\end{aligned}$$

[9] Where $\Gamma_\nu = \Gamma(1-\frac{1}{\nu})$.

where $\Gamma_\nu = \Gamma\left(1 - \frac{1}{\nu}\right)$. Also note that, for any destination $d \in G$ (and dropping t for simplicity), migration flows are given by:

$$m_{gd}(z, t+1) = \frac{[V_{gd,t+1}]^\nu}{\sum_i [V_{gi,t+1}]^\nu}$$

So the expectation above becomes:

$$E_{\xi_i}[\max_i \xi_i V_{gi,t+1}] = \left[\frac{[V_{gd,t+1}]^\nu}{m_{gd}(z, t+1)}\right]^{\frac{1}{\nu}} = m_{gd}(z, t+1)^{-\frac{1}{\nu}} \cdot V_{gd,t+1}$$

Note that, since this applies $\forall d \in G$, we can express the expectation as:

$$E_{\xi_i}[\max_i \xi_i V_{gi,t+1}] = \frac{1}{G}\sum_i m_{gi}(z, t+1)^{-\frac{1}{\nu}} \cdot V_{gi,t+1}$$

Hence the Bellman in discrete time is given by:

$$V_g(z,t) = U_g(z,t) + [(1-\rho)(1-\delta)]\Big\{ \begin{array}{l} (1-\lambda)V_g(z, t+1) \\ +\lambda\frac{1}{G}\sum_i m_{gi}(z, t+1)^{-\frac{1}{\nu}} \cdot \kappa_{gi} V_i(z, t+1) \end{array} \Big\}$$

To convert to continuous time, consider a period of length Δ to the Bellman becomes:

$$V_g(z,t) = U_g(z,t) + [(1-\Delta\rho)(1-\Delta\delta)]\Big\{ \begin{array}{l} (1-\Delta\lambda)V_g(z, t+\Delta) \\ +\Delta\lambda\frac{1}{G}\sum_i m_{gi}(z, t+\Delta)^{-\frac{1}{\nu}} \cdot \kappa_{gi} V_i(z, t+\Delta) \end{array} \Big\}$$

After multiplying everything, substracting $V_g(z,t)$, dividing by Δ and taking the limit $\Delta \to 0$ we obtain the HJB in equation (16):

$$(\delta + \rho + \lambda)V_g(z,t) \quad = \quad U_g(z,t) + \dot{V}_g + \lambda\left[\frac{1}{G}\sum_i m_{gi}(z,t)^{-\frac{1}{\nu}} \cdot \kappa_{gi} V_i(z,t)\right]$$

A.1.2 Derivation of the Kolmogorov forward equation

By definition the distribution of efficiency units z at any point in time t and region g is given by:

$$\Psi_g(z,t) = \mathbb{P}\{Z \leq z, g(t) = g\}$$

Therefore, the number of people in region g with efficiency units less than or equal to z will be given by $N_g(z,t) = \Psi_g(z,t)N(t)$, so that:

$$\dot{N}_g(z,t) = \dot{\Psi}_g(z,t) \cdot N(t) + \Psi_g(z,t) \cdot \dot{N}(t)$$

On the other hand, in discrete time, after each interval of length Δ, $E_g(t)$ workers are born each period, which draw their skills from the distribution $F(z)$. Furthermore, a share of $(1-\delta)$ households remain alive in $t+\Delta$. Of those, λ will not be able to move whereas $(1-\lambda)$ from other regions can migrate to g from $i \in G$, and do so with probability $m_{ig}(z, t+\Delta)$. Consequently, the total share of workers with skill z moving from i to g is given by $\Delta\lambda m_{ig}(z, t+\Delta)$.

Combining all these elements, we obtain the following evolution of $N_g(z,t)$:

$$N_g(z,t+\Delta) = \Delta E_g(t)\cdot F(z) + (1-\Delta\delta)\Big[(1-\Delta\lambda)N_g(z,t) + \Delta\lambda\textstyle\sum_i m_{ig}(z,t+\Delta)N_i(z,t)\Big]$$

Subtracting $N_g(z,t)$, dividing by Δ and taking the limit $\Delta \to \infty$ we have:

$$\dot{N}_g(z,t) = E_g(t)\cdot F(z) - \delta N_g(z,t) - \lambda\Big[(1-m_{gg}(z,t))N_g(z,t) - \textstyle\sum_{i\neq g} m_{ig}(z,t)N_i(z,t)\Big]$$

where $\dot{\Psi} = \partial_t \Psi$. Finally, combine both expressions for $\dot{N}_g(z,t)$ and take derivatives with respect to z to obtain the Kolmogorov forward equation in equation (18):

$$\begin{aligned}\dot{\psi}_g(z,t) &= \frac{E_g(t)}{N(t)}\cdot f(z) - \delta\psi_g(z,t)\\ &- \lambda\Big[(1-m_{gg}(z,t))\psi_g(z,t) - \textstyle\sum_{i\neq g} m_{ig}(z,t)\psi_i(z,t)\Big]\\ &- \psi_g(z,t)\frac{\dot{N}(t)}{N(t)}\end{aligned}$$

where $\bar{z}$ is the maximum productivity level.

A.1.3 Detrended Variables and Balanced Growth Path

For any variable $y(t)$, denote the detrended version as $\tilde{y}(t) = e^{-\gamma_y t}y(t)$, where γ_y is the growth rate of y, and in the case of population, set $N(0) = N_0$ as its initial condition. Also, let $\bar{y}$ be the same variable along the balanced growth path. In a BGP, the productivity distribution is stationary and the mass of intermediate firms grows at a common constant rate $\gamma_M = \dot{M}_g(t)/M_g(t)$. In this case, the evolution of $M_g(t)$ implies $\gamma_M = F_g(t)\cdot M_g(t)^\eta$ and therefore the growth rate γ_L is determined by combining this with the production labor demand equation:

$$\begin{aligned}\gamma_M &= F_g(t)\cdot M_g(t)^\eta\\ \gamma_M &= \left(\frac{L_g^y(t)}{\varepsilon M_g(t)}\right)M_g(t)^\eta\\ \gamma_M &= \left(\frac{1}{\varepsilon}\right)\tilde{L}_g^y(t)\cdot(\tilde{M}_g(t))^{\eta-1}e^{[\gamma_L+\gamma_M(\eta-1)]t}\end{aligned}$$

which implies $\gamma_L = \gamma_M(1-\eta)$. Recall from the labor market clearing condition:

$$L_g(t) = L_g^y(t) + L_g^h(t)$$

so that:

$$L_g^y(t) = 1 - (1-\theta)(1-\alpha)L_g(t) = [\alpha + \theta(1-\alpha)]L_g(t)$$

So that:

$$\gamma_M = \left(\frac{\alpha+\theta(1-\alpha)}{\varepsilon}\right) L_g(t) \cdot (\tilde{M}_g(t))^{\eta-1}$$

From the definition of $L_g(t)$ we can find γ_L:

$$\begin{aligned} L_g(t) &= \left(N(t) \cdot \int_{\Omega_z} z^\beta \psi_g(z,t) dz\right)^{\frac{1}{\beta}} \\ \tilde{L}_g(t) \cdot e^{\gamma_L t} &= e^{\frac{1}{\beta}(\gamma_N + \gamma_\psi)t} \left(\tilde{N}(t) \cdot \int_{\Omega_z} z^\beta \psi_g(z,t) dz\right)^{\frac{1}{\beta}} \end{aligned}$$

which implies $\gamma_L = \dfrac{1}{\beta}\gamma_N$ (we will see later that $\gamma_\psi = 0$). We can use the solution for $U_g(\cdot)$ from the static equilibrium to find detrended flow utility:

$$\begin{aligned} U_g(z,t) &= D_g(t) \cdot z^\beta \cdot \frac{\Gamma_g}{\Gamma_U} \cdot \left(\pi_{gg}(t)\right)^{\frac{\alpha}{1-\sigma}} \left(M_g(t)\right)^{\frac{\alpha}{\varepsilon-1}} \left(L_g(t)\right)^{(1-\beta)-\theta(1-\alpha)} \\ \tilde{U}_g(z,t) e^{\gamma_U t} &= e^{[\gamma_M \frac{\alpha}{\varepsilon-1} + \gamma_L((1-\beta)-\theta(1-\alpha)) - \gamma_N \phi]t} \\ &\cdot \quad D_g \tilde{N}_g(t)^{-\phi} z^\beta \cdot \frac{\Gamma_g}{\Gamma_U} \cdot \left(\pi_{gg}(t)\right)^{\frac{\alpha}{1-\sigma}} \left(\tilde{M}_g(t)\right)^{\frac{\alpha}{\varepsilon-1}} \left(\tilde{L}_g(t)\right)^{(1-\beta)-\theta(1-\alpha)} \end{aligned}$$

which, along with the expressions $\gamma_N = \beta\gamma_L$ and $\gamma_L = \gamma_M(1-\eta)$, implies:

$$\gamma_U = [\tfrac{\alpha}{\varepsilon-1} - (1-\eta)(\beta(1+\phi) + \theta(1-\alpha) - 1)]\gamma_M$$

Detrended Equations For the value function, recall that by definition in a BGP, $V_g(z,t) = e^{\gamma_v t}\tilde{V}_g(z,t)$. To find the detrended HJB, we can then rewrite each element in the HJB:

$$\begin{aligned} (\delta+\rho)V_g(z,t) &= (\delta+\rho)e^{\gamma_v t}\tilde{V}_g(z,t) \\ U_g(z,t) &= e^{[\frac{\alpha}{\varepsilon-1} - (1-\eta)(\beta(1+\phi)+\theta(1-\alpha)-1)]\gamma_M t}\tilde{U}_g(z,t) \\ \partial_t V_g(z,t) &= \gamma_v e^{\gamma_v t}\tilde{V}_g(z,t) + e^{\gamma_v t}\partial_t \tilde{V}_g(z,t) \end{aligned}$$

Combining all this into the HJB we obtain:

$$\begin{aligned} (\delta+\rho)e^{\gamma_v t}\tilde{V}_g(z,t) &= e^{[\frac{\alpha}{\varepsilon-1} - (1-\eta)(\beta(1+\phi)+\theta(1-\alpha)-1)]\gamma_M t}\tilde{U}_g(z,t) \\ &+ \gamma_v e^{\gamma_v t}\tilde{V}_g(z,t) + e^{\gamma_v t}\partial_t \tilde{V}_g(z,t) \\ &+ e^{\gamma_v t}\lambda \textstyle\sum_i m_{gi}(z,t)\Big[\delta_{gi}(z,t)\tilde{V}_i(z,t) - \tilde{V}_g(z,t)\Big] \end{aligned}$$

This implies:

$$\gamma_v = \left(\frac{\alpha}{\varepsilon-1} - (1-\eta)(\beta(1+\phi)+\theta(1-\alpha)-1)\right)\gamma_M$$

which results in the detrended HJB:

$$(\delta+\rho-\gamma_v)\tilde{V}_g(z,t) = \tilde{U}_g(z,t) + \partial_t\tilde{V}_g(z,t) + \lambda\sum_i m_{gi}(z,t)\left[\delta_{gi}(z,t)\tilde{V}_i(z,t) - \tilde{V}_g(z,t)\right]$$

For the detrended Kolmogorov equation consider each element of the equation:

$$\begin{aligned}
\partial_t\psi_g &= \gamma_\psi e^{\gamma_\psi t}\tilde{\psi}_g(z,t) + e^{\gamma_\psi t}\partial_t\tilde{\psi}_g(z,t) \\
\frac{E_g(t)}{N(t)}f(z) - \delta\psi_g(z,t) &= \frac{\tilde{E}_g(t)}{\tilde{N}(t)}f(z) - \delta\tilde{\psi}_g(z,t)e^{\gamma_\psi t} \\
\lambda(1-m_{gg}(z,t))\psi_g(z,t) &= \lambda(1-m_{gg}(z,t))\tilde{\psi}_g(z,t)e^{\gamma_\psi t} \\
\lambda\sum_{i\neq g} m_{ig}(z,t)\psi_i(z,t) &= \lambda\sum_{i\neq g} m_{ig}(z,t)\tilde{\psi}_i(z,t)e^{\gamma_\psi t} \\
\frac{\dot{N}(t)}{N(t)}\psi_g(z,t) &= \left(\tilde{\gamma}_N(t)+\gamma_N\right)\tilde{\psi}_g(z,t)e^{\gamma_\psi t}
\end{aligned}$$

where $\tilde{\gamma}_N(t) = \dot{\tilde{N}}(t)/\tilde{N}(t)$ is the growth rate of the detrended total population. We can obtain this result since $N(t) = \tilde{N}(t)e^{\gamma_N t}$, which implies:

$$\dot{N}(t) = \frac{dN(t)}{dt} = \frac{d\tilde{N}(t)}{dt}\cdot e^{\gamma_N t} + e^{\gamma_N t}\gamma_N\tilde{N}(t)$$

so that:

$$\frac{\dot{N}(t)}{N(t)} = \frac{\dot{\tilde{N}}(t)}{\tilde{N}(t)} + \gamma_N$$

Combining them all we obtain:

$$\begin{aligned}
\gamma_\psi e^{\gamma_\psi t}\tilde{\psi}_g(z,t) + e^{\gamma_\psi t}\partial_t\tilde{\psi}_g(z,t) &= \frac{\tilde{E}_g(t)}{\tilde{N}(t)}f(z) - \delta\tilde{\psi}_g(z,t)e^{\gamma_\psi t} \\
&- \lambda[1-m_{gg}(z,t)]\tilde{\psi}_g(z,t)e^{\gamma_\psi t} \\
&+ \lambda\sum_{i\neq g} m_{ig}(z,t)\tilde{\psi}_i(z,t)e^{\gamma_\psi t} - \left(\tilde{\gamma}_N(t)+\gamma_N\right)\tilde{\psi}_g(z,t)e^{\gamma_\psi t}
\end{aligned}$$

which implies that $\gamma_\psi = 0$. Thus, using this, we obtain the detrended Kolmogorov equation:

$$\begin{aligned}\partial_t \tilde{\psi}_g(z,t) &= \frac{\tilde{E}_g(t)}{\tilde{N}(t)} f(z) - \delta \tilde{\psi}_g(z,t) \\ &- \lambda \Big[(1 - m_{gg}(z,t)) \tilde{\psi}_g(z,t) - \textstyle\sum_{i \neq g} m_{ig}(z,t) \tilde{\psi}_i(z,t) \Big] - \Big(\tilde{\gamma}_N(t) + \gamma_N \Big) \tilde{\psi}_g(z,t)\end{aligned}$$

Detrended growth Note that, using population dynamics, we can derive the detrended growth rate. By definition:

$$\dot{N}(t) = \sum_g (E_g(t) - \delta \cdot N_g(t))$$

Therefore:

$$\frac{\dot{N}(t)}{N(t)} = \frac{\sum_g (E_g(t) - \delta \cdot N_g(t))}{N(t)} = \frac{\sum_g \tilde{E}_g(t)}{\tilde{N}(t)} - \delta$$

where the second equality follows from $E_g(t) = e^{\gamma_N t} \tilde{E}_g(t)$ and $N(t) = \sum_g N_g(t)$. Therefore, combining this with equation above we have:

$$\frac{\dot{\tilde{N}}(t)}{\tilde{N}(t)} = \frac{\sum_g \tilde{E}_g(t)}{\tilde{N}(t)} - \delta - \gamma_N$$

BGP Equations In the BGP we have $\tilde{y}(t) = \bar{y}$. Thus the HJB and Kolmogorov forward equations become:

$$(\delta + \rho - \gamma_v) \bar{V}_g(z) = \bar{U}_g(z) + \lambda \textstyle\sum_i \bar{m}_{gi}(z) \Big[\bar{\delta}_{gi}(z) \bar{V}_i(z) - \tilde{V}_g(z) \Big]$$

$$0 = \frac{\bar{E}_g}{\bar{N}} f(z) - (\delta + \gamma_N) \bar{\psi}_g(z) - \lambda (1 - \bar{m}_{gg}(z)) \bar{\psi}_g(z) + \lambda \textstyle\sum_{i \neq g} \bar{m}_{ig}(z) \bar{\psi}_i(z)$$

Note that along the BGP there is no net migration so that:

$$\sum_g \int \lambda [1 - \bar{m}_{gg}(z)] \bar{\psi}_g(z) dz = \sum_g \int \lambda \sum_{i \neq g} \bar{m}_{ig}(z) \bar{\psi}_i(z) dz$$

Also recall that $N(t) \Psi_g(\bar{z}) = N(t) \int \psi_g(z,t) = N_g(t)$ so that $\bar{\Psi}_g(\bar{z}) = \bar{N}_g / \bar{N}$. Thus, to derive the equilibrium population growth rate γ_N, integrate the BGP KF equation over z

and sum across regions g to obtain:

$$\begin{aligned} 0 &= \sum_g \frac{\bar{E}_g}{\bar{N}} - \sum_g \delta \int \bar{\psi}_g(z)dz - \gamma_N \\ \gamma_N &= \sum_g \frac{\bar{E}_g}{N_0} - \sum_g \delta \frac{\bar{N}_g}{N_0} \\ \gamma_N &= \sum_g \frac{\bar{E}_g}{N_0} - \delta \end{aligned}$$

where we also use the fact that $\int f(z)dz = 1$ and imposing $\bar{N} = N_0$.

B Computation Appendix

Algorithm 1: Balanced Growth Path

Data: parameters $\{\boldsymbol{\Theta}, \boldsymbol{\Theta}_G\}$; productivity distribution $f(z)$
Result: solution functions $\{\bar{V}, \bar{\psi}, \bar{m}\}$
Initialize $\{\boldsymbol{W}^0, \boldsymbol{L}^0, \boldsymbol{N}^0\}$;
while *not converged* **do**
  Use $\{\boldsymbol{W}^n, \boldsymbol{L}^n, \boldsymbol{N}^n\}$ to compute shares π^n_{gd} and flow utility $\bar{U}^n_g(z)$;
  Solve the HJB in equation (24). Recover migration shares $\bar{m}^n_{gd}(z)$;
  Use $\bar{m}^n_{gd}(z)$ to solve the Kolmogorov forward equation (25). Recover the distribution $\bar{\psi}^n_g(z)$;
  Use $\bar{\psi}^n_g(z)$ to compute labor supply and population:

$$\hat{L}^n_g = N \cdot \int_{\Omega_z} z\bar{\psi}^n_g(z)dz \qquad \hat{N}^n_g = N \cdot \int_{\Omega_z} \bar{\psi}^n_g(z)dz$$

  then use $\{\hat{\boldsymbol{L}}^n\}$ to compute the wages that satisfy:

$$\hat{W}^n_g \hat{L}^n_g = \sum_d \pi^n_{gd} \hat{W}^n_d \hat{L}^n_d$$

  this is the trade balance in equation (12) ;
  if $\{\boldsymbol{W}^n, \boldsymbol{L}^n, \boldsymbol{N}^n\}$ *close to* $\{\hat{\boldsymbol{W}}^n, \hat{\boldsymbol{L}}^n, \hat{\boldsymbol{N}}^n\}$ **then**
    converged;
  else
    compute $\{\boldsymbol{W}^{n+1}, \boldsymbol{L}^{n+1}, \boldsymbol{N}^{n+1}\}$ as linear combination of old and new values;
  end
end

Algorithm 2: Transition Dynamics

Data: endpoint solutions $\{\bar{V}^1, \bar{\psi}^1, \bar{m}^1\}$ and $\{\bar{V}^2, \bar{\psi}^2, \bar{m}^2\}$; parameter series $\{\boldsymbol{\Theta}(t), \boldsymbol{\Theta}_G(t)\}$; productivity distribution $f(z)$

Result: solution functions $\{\tilde{V}(t), \tilde{\psi}(t), m(t)\}$

Initialize $\{\boldsymbol{W}^0(t), \boldsymbol{L}^0(t), \boldsymbol{N}^0(t)\}$;

while *not converged* **do**

Use $\{\boldsymbol{N}^j(t)\}$ to compute total population $\hat{N}^j(t)$;

Use $\{\boldsymbol{L}^j(t)\}$ to compute a transition path $\{\hat{\boldsymbol{M}}^j(t)\}$ according to equation (20) using the initial condition from the first endpoint;

Use $\{\boldsymbol{W}^j(t), \boldsymbol{L}^j(t), \boldsymbol{N}^j(t)\}$ and $\{\hat{\boldsymbol{M}}^j(t)\}$ to compute shares $\pi^j_{gd}(t)$ and flow utility $\tilde{U}^j_g(t)$;

Solve the HJB equation (21) by iterating backwards from $\bar{V}^2$. Recover migration shares $m^j_{gd}(z,t)$;

Use $m^j_{gd}(z,t)$ to solve the Kolmogorov forward equation (22) by iterating forward from $\bar{\psi}^1$. Recover $\psi^j_g(z,t)$;

Use $\tilde{\psi}^j_g(z,t)$ to compute labor supply and population for each t:

$$\hat{L}^j_g(t) = \hat{N}^j(t) \cdot \int_{\Omega_z} z \cdot \tilde{\psi}^j_g(z,t) dz \qquad \hat{N}^n_g = \hat{N}^j(t) \cdot \int_{\Omega_z} \tilde{\psi}^n_g(z,t) dz$$

then use $\{\hat{\boldsymbol{L}}^j(t)\}$ to compute wages that satisfy:

$$\hat{W}^j_g(t)\hat{L}^n_g(t) = \sum_d \pi^j_{gd} \hat{W}^j_d(t)\hat{L}^j_d(t)$$

this is the detrended trade balance equation.;

if $\{\boldsymbol{W}^j(t), \boldsymbol{L}^j(t), \boldsymbol{N}^j(t)\}$ *close to* $\{\hat{\boldsymbol{W}}^j(t), \hat{\boldsymbol{L}}^j(t), \hat{\boldsymbol{N}}^j(t)\}$ **then**

converged;

else

compute $\{\boldsymbol{W}^{j+1}(t), \boldsymbol{L}^{j+1}(t), \boldsymbol{N}^{j+1}(t)\}$ as linear combination of old and new values;

end

end

C Estimation Appendix

Table 5 Regression of Prices on Distance

	$\log p_{i,m,n,j,t}$			
	(1)	(2)	(3)	(4)
ζ_τ	0.283***	0.169***	0.283***	0.168***
	(0.001)	(0.001)	(0.001)	(0.001)
Same State	-0.467***	-0.210***	-0.467***	-0.211***
	(0.002)	(0.002)	(0.002)	(0.002)
Orig. State	✓			
Dest. State	✓			
Orig. State * Prod.		✓		
Dest. State * Prod.		✓		
Orig. State * Quarter			✓	
Dest. State * Quarter			✓	
Orig. State * Prod. * Quarter				✓
Dest. State * Prod. * Quarter				✓
N	10,526,115	10,526,114	10,526,115	10,525,995
R^2	0.114	0.396	0.115	0.400