Journal of Economic Dynamics & Control 133 (2021) 104234

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Migration and urban economic dynamics

Morris A. Davis[a], Jonas D.M. Fisher[b], Marcelo Veracierto[b,1,*]

[a] Rutgers University, United States
[b] Federal Reserve Bank of Chicago, United States

ARTICLE INFO

Article history:
Received 16 August 2021
Accepted 30 August 2021
Available online 14 September 2021

JEL classification:
E0
O4
R0

Keywords:
Persistent urban decline
Urban economic dynamics
Gross migration
Net migration

ABSTRACT

What generates the large amount of heterogeneity and persistence in U.S. city growth rates? To answer this question we construct a dynamic general equilibrium model of urban migration that is consistent with the approximately linear relation between gross and net migration rates that we uncover in a panel of 381 metropolitan statistical areas. Consistency with this relation together with the model's structure of moving costs delivers a parsimonious reduced form in which competitive equilibrium allocations are given by the solution to a city's social planner problem subject to quadratic population adjustment costs. A calibrated version of the model indicates that empirically measured total factor productivity shocks account for most of the short-run and long-run population dynamics observed in U.S. data.

1. Introduction

Between 1985 and 2013, the aggregate population in the U.S. grew at a fairly constant rate of about 1% a year. However, over this span individual cities often grew at rates that were far from that trend. Indeed, the standard deviation of cities' annual growth rates is 1.24%, seven times larger than that of aggregate population. A quarter of the observations have cities growing at least 0.76% below the aggregate population's rate of growth, while another quarter have them growing at least 0.58% above it. Some of these deviations are extremely persistent. For example, the 20 cities with the lowest population growth rates between 1985 and 2013 contracted at an annual rate of 1.5% relative to the aggregate population. What drives this amount of heterogeneity and persistence in the growth rates of cities? In this paper we explore the role of city-level total factor productivity (TFP) shocks.[2] We consider two related questions: (1) How much of the heterogeneity in cities' growth rates can be accounted for by idiosyncratic shocks to TFP? (2) And to what extent can the most extreme cases of urban decline be accounted for by TFP?

* Corresponding author.
E-mail addresses: morris.a.davis@rutgers.edu (M.A. Davis), jfisher@frbchi.org (J.D.M. Fisher), mveracierto@frbchi.org (M. Veracierto).

[1] We thank Marco Bassetto, Jeff Campbell, Gilles Duranton, Edward Glaeser, John Kennan, Erzo Luttmer, Fernando Parro, Nicolas Petrosky-Nadeau, Robert Shimer, and seminar participants at various institutions and conferences for helpful comments. The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

[2] TFP shocks, while hard to interpret, have been widely used not only in the macro literature but in the urban literature. The reason why we focus on them and abstract from other shocks (such as shocks to amenity levels) is because we can measure them relatively easily but not other shocks. Observe that to the extent that other shocks are independent of TFP shocks, their presence would not affect the contribution of TFP shocks to urban dynamics.

https://doi.org/10.1016/j.jedc.2021.104234

In order to study urban population dynamics it is crucial to analyze internal migration decisions; births, deaths, and international migration flows generally play a minor role (e.g., Turek, 1985). For this reason we begin our investigation by characterizing gross and net urban migration flows in a panel of 381 U.S. cities over the period 1985–2013. As is well known, there is a large amount of churning behind the net population changes that we observe. Even in cities that experience no changes to their total population levels, on average about 5.5% of their previous year's population leave and are replaced with people arriving from other cities. Conditioning on a same net change in total population, we find a significant amount of heterogeneity in arrival and departure rates across cities. However, within cities, these arrival and departure rates vary systematically with net population growth rates. In particular, we uncover an approximately linear relation between these rates. This relation indicates not only that cities grow by both increasing arrivals and decreasing departures, but more importantly, its linear form is extremely informative about the type of adjustment costs to net population changes that individual cities face.

To show this, we start our theoretical analysis by considering a simple static model of gross migration. The model has a continuum of cities and a measure one of agents initially distributed across them. All cities produce the same consumption good using a production function that has labor as its only input. However, the cities differ in terms of their idiosyncratic productivity levels. An economy-wide social planner must decide how many people to remove from each city and how many people to bring in. Removing people from a city generates costs that are described by a convex cost function (costs are allowed to be negative over certain ranges). The costs of bringing people into a city are described by another convex cost function. We show that when both of these cost functions are quadratic, arrival rates (and departure rates) are linearly related to net population changes, making the model consistent with the empirical evidence. We also show that the economy-wide planning problem can be decomposed into a series of city-level planning problems, one for each city.[3] The solutions to these city-level planning problems correspond to those of the economy-wide planning problem when certain side conditions are satisfied. Most importantly, we show that when the cost functions of bringing in and removing people from cities are both quadratic, the city-level social planners effectively face quadratic adjustment costs to changing the total population level of their cities.

Because adjustment costs to net population changes play a crucial role in a dynamic setting and migration decisions are inherently forward-looking, we extend the simple model to an infinite horizon framework (essentially, we embed it into the neoclassical growth model). A key difference with the static economy is that cities are now subject to idiosyncratic productivity shocks that change over time. Another difference is that we introduce an endogenous labor supply margin, which allows cities to adjust to their productivity shocks not only by changing their sizes (extensive margin) but by changing their employment levels (intensive margin).

Instead of just specifying social planning problems that face reduced-form cost functions for bringing people into cities and removing them, we find it valuable to provide microeconomic foundations to those planning problems (even in a highly stylized way) in order to facilitate economic interpretation and provide a justification for the social planning problems considered. For this reason, we consider a competitive equilibrium in which fully insured agents make their own migration decisions. Each agent receives an idiosyncratic shock to the value of staying in their initial location versus leaving. The agents that decide to leave have two ways of reallocating to a different city (both ways assume that agents do not know the economic conditions of the individual cities, only their distribution). The first way is through undirected search: The agent can select a city at random and move there (after incurring a fixed disutility cost). The second way is by joining formal recruitment markets, where agents are contacted by recruiters offering the opportunity to move to their cities of origin. The recruitment activity is costly: The effort level of a recruiter determines how many agents they will be able to contact. In order to preserve the abstraction of a competitive equilibrium, we assume that when an agent meets a recruiter, the economic conditions of the destination city are perfectly known to the agent (or, alternatively, that the economic conditions are contractable), but not the identity of the destination city. Once an agent accepts the offer made by a recruiter, the identity of the recruiter's city of origin is revealed to the agent, allowing them to move to that particular city. At equilibrium, agents are indifferent between doing undirected search or joining any of the formal recruiting markets in the economy.

Since this is a convex economy with no distortions, the welfare theorems apply and the equilibrium allocation can be characterized as the solution to a recursive city-planner problem and some side conditions.[4] We provide sufficient conditions (for the distribution of idiosyncratic shocks to the value of staying in versus leaving a city and for the technology that recruiters operate) that give rise to quadratic cost functions for total arrivals to and total departures from a city. Similarly to the static economy, these functional forms are needed to make the model consistent with the linear relation between arrival rates (and departure rates) and net population changes that we uncovered in U.S. data (as long as undirected search is not too large, the linear relation is obtained).

While we present our dynamic model using ex-ante identical cities, we discuss how the model can be extended to accommodate for the vast amount of cross-sectional heterogeneity that we observe. However, these extensions turn out to be completely immaterial to the main questions addressed in the paper. The reason is that because the quadratic adjustment

[3] The structure of these sub-planning problems is the same for all cities. Only the initial population and idiosyncratic productivity levels that the different city planners take as given vary.

[4] Similarly to the static model, each city has a social planner solving this common recursive problem. Only the values of the state variables differ across cities.

costs on net population changes are exactly the same across the different types of cities, the cross-sectional heterogeneity does not affect how cities respond to their idiosyncratic shocks to TFP.

Having determined the functional forms needed to make the model consistent with the empirical evidence on gross-migration flows, we turn to calibrate parameter values. A key ingredient for our quantitative analysis is the stochastic process followed by idiosyncratic TFP. We measure TFP at the city level using wage and employment data from our constructed panel of U.S. cities, together with the first-order conditions of the competitive firms in our model. We find that TFP growth rates are well described by a first order autoregressive (AR(1)) process with a positive serial correlation. Since this positive serial correlation introduces a non-stationarity and our model requires a stationary TFP process in levels, we follow Gabaix (1999b) by introducing a negative drift and a reflecting barrier. Such structure is enough to generate a stationary TFP process and is consistent with obtaining a reasonable fit of AR(1) growth rates over small samples.[5] Having estimated our TFP process, two other important parameters for our quantitative exercise are the quadratic adjustment cost coefficient on net population changes (that the city-level social planners face) and the elasticity of labor supply. In order to calibrate them, we first estimate the impulse responses of population, employment, and wages to a one standard-deviation innovation to TFP. We then select the quadratic adjustment cost coefficient to reproduce the log population response on impact; we also select the labor supply elasticity to reproduce the impact response of the ratio of the log wage and the log employment-to-population ratio. While the impact responses of these variables are therefore pre-determined, their responses at all other horizons serve as a test for the model.

We find that the model does an excellent job at reproducing the impulse response function for population levels at all time horizons that our data allows us to estimate. The impulse response functions for employment, wages, arrival rates, and departure rates are less satisfactory, but are still surprisingly reasonable given the simple structure of our model. Importantly, the arrival rate rises and the departure rate falls after a positive change in TFP, just as the model predicts.

With respect to the first of the two questions that constitute the main focus of the paper, we find that idiosyncratic TFP shocks account for 77% of the variance of annual population growth rates. To answer the second question, we feed into the model the average TFP path for the 20 cities with the largest urban decline in our sample. We find that by the end of our sample period, our model predicts a contraction in the population level of these cities of 24%, which is quite close to the decline of 20% that we actually observed. These findings indicate that idiosyncratic shocks to TFP account for the bulk of urban dynamics.

The paper is organized as follows. In Section 2, we discuss the related literature. In Section 3, we present our empirical findings. In Section 4, we describe a simple static model of gross migration. We extend the static model to a dynamic setting in Section 5. Then, in Section 6 we parametrize the model. We present the results in Section 7. Finally, we present our conclusions in Section 8. All appendices to the paper are provided as online supplementary material.

2. Relationship to the literature

There is a large literature analyzing urban growth.[6] Our paper contributes to the strand of this literature that studies the role of exogenous shocks on city growth using dynamic general equilibrium models, including papers by Eeckhout (2004), Gabaix (1999a,b), and Rossi-Hansberg and Wright (2007). This literature does not directly measure the shocks that drive urban growth. Instead, it characterizes the theoretical properties that the shocks must have in order to reproduce certain features of the long run size distribution of cities (in particular, Zipf's Law). City-level TFP shocks drive urban growth in our framework and one of our contributions is to show how to measure them using readily available data on employment and wages.[7] Another difference with this literature is that we introduce migration costs, with implications both for gross and net migration dynamics. Finally, to the best of our knowledge, our paper is the first to quantify the role of TFP shocks in explaining the large amount of heterogeneity and persistence in U.S. city growth rates.

Our paper also contributes to the large literature on U.S. internal migration. Much of this literature focuses on migration between states. The seminal Blanchard and Katz (1992) paper uses state-level panel data to argue that persistent labor demand shocks are accommodated over the long run through migration. In our framework persistent shocks to TFP drive labor demand at the city level, and the long-run response occurs through migration as well. In another seminal paper, Kennan and Walker (2011) uses individual-level data and a partial equilibrium discrete choice framework to estimate inter-state migration propensities as a function of individual characteristics, relative wages, and migration costs. While we account for the equilibrium interactions, focus on cities (not states), and use a very different modeling framework, we reach remarkably similar conclusions regarding the speed of adjustment to shocks that persistently change wages in different locations.

Caliendo et al. (2019) is another important related paper that considers inter-state migration. It embeds discrete migration choices within a general equilibrium model with a rich production structure. They analyze the impact of the rise in trade with China on employment and welfare (finding that migration decisions play an important role). We could have used a version of their model to analyze the questions addressed in our paper. Instead we introduce a micro-founded model that is quite different. The main difference is with respect to the structure of moving costs. Caliendo et al. (2019) assume that

[5] Gabaix (1999b) used a random walk with negative drift and a reflecting barrier to generate a cross-sectional city size distribution consistent with Zipf's law.

[6] See Duranton and Puga (2014) and Gabaix and Ioannides (2004) for recent surveys.

[7] In related work, Cingano and Schivardi (2004) measures TFP at the city level in Italy by aggregating measures of firm-level TFP.

when moving decisions are made, agents compare the realized idiosyncratic value of living in their state of origin with the realized idiosyncratic value of living in each of the other states in the economy (all i.i.d. idiosyncratic shocks being perfectly known). In addition, they introduce moving costs that are common to all agents, but depend on the state of origin and the state of destination. For both of these reasons, their moving costs are specific to the different origin-destination pairings. On the contrary, we assume that idiosyncratic moving costs solely depend on the city of origin and that the common moving costs solely depend on the city of destination. While we do not argue which specification may be more realistic, ours is more parsimonious. The reason is that it allows us to reduce the economy-wide social planning problem into a city-level social planning problem, which has a well-defined cost function for removing people out of the city that is separate from a cost function for bringing people in. Moreover, these cost functions are common across all cities. This specification allows us to focus on the functional forms that these cost functions must take in order to generate the linear relation between gross and net migration rates that we find in the data. It is an open question whether the Type 1 Extreme Value distribution that Caliendo et al. (2019) specify for their idiosyncratic value shocks (which they need for tractability) is flexible enough to generate that linear relation.

Coen-Pirani (2010) is also closely related to our paper. Similar to Caliendo et al. (2019), it uses a Lucas-Prescott islands economy to analyze migration patterns across U.S. states. However, contrary to Caliendo et al. (2019), agents ignore the idiosyncratic values that they could get at other locations: Agents need to move there to learn. In addition, there is a fixed cost to move that is exactly the same across all agents and origin-destination combinations. Thus, similar to our paper, Coen-Pirani (2010) assumes moving costs that are origin-destination independent.[8] A key difference with our paper, however, is that Coen-Pirani (2010) assumes persistent idiosyncratic shocks. As a consequence, when states experience positive productivity shocks and attract more migrants, a fraction of the new arrivals receive low and persistent realizations of the idiosyncratic shock that make them leave immediately. This mechanism is crucial for reproducing a key feature that Coen-Pirani (2010) argues is present in the U.S. data—that inter-state arrivals and departures are positively correlated within states. We could generate this feature by grouping the cities in our model into states and introducing a migration structure and idiosyncratic shocks at the state level similar to Coen-Pirani (2010).[9] However, this state-level migration structure and the idiosyncratic shocks would greatly complicate the analysis without generating any further insights to those already found in Coen-Pirani (2010). For this reason, we abstract from inter-state migration patterns and focus on city-level dynamics instead. It is at the city level that our paper provides novel insights about population dynamics.

Several papers have focused on modelling the well-known secular decline in gross inter-state migration—for example Kaplan and Schulhofer-Wohl (2017) and Karahan and Rhee (2013). Not surprisingly, we find a similar trend in gross inter-city migration. We abstract from this trend, as well as the business cycle, so we can focus on within-city population adjustments. In the conclusion we discuss how our model could be modified to generate a secular decline in gross migration.

Inter-city migration has been an active area of inquiry as well. Howard (2020) studies within-city variation in gross migration from an empirical perspective. He finds that arrival rates rise and departure rates fall in response to an increase in labor demand due to a shock to the demand for locally produced goods. We corroborate this finding by showing that arrival rates rise and departure rates fall after a TFP shock that increases labor demand. Rappaport (2004) studies rare permanent large changes to a city's TFP in a neoclassical growth framework in which migration costs are assumed to be linear in net migration. In our framework the linear relationship between gross and net migration we uncover in U.S. data implies that migration costs are quadratic in net migration.

Other quantitative-theoretic work on inter-city migration involves its interaction with housing. Gordon and Guerron-Quintana (2019) studies how debt and taxation influence the distribution of the population across cities. Karahan and Rhee (2019) study the role of the housing bust on inter-city reallocation following the Great Financial Crisis. Nenov (2015) quantifies the impact of housing markets on worker reallocation across cities and finds that their effect is small. Van Nieuwerburgh and Weil (2010) investigate the secular trend toward higher house price dispersion across U.S. cities, relating it to a coincident increase in wage dispersion. We abstract from housing to isolate the role of migration costs on city dynamics.

Finally, our paper is also related to the empirical study of urban decline by Glaeser and Gyourko (2005). They argue that urban decline is persistent because as a city's population declines so does its housing costs (because housing investment is irreversible and the housing stock is very durable). These lower housing costs reduce the incentive for people to leave, slowing the decline of population. They consider a role for migration, but dismiss its importance because gross migration remains high in declining cities. By abstracting from housing and finding that our model comes close to matching the rate of urban decline in response to falling relative TFP, we show that migration costs may play an important role in explaining persistent urban decline after all.

[8] As a consequence, his economy-wide social planner problem could also be decomposed into a state-level planning problem and side conditions.

[9] Directly introducing their migration structure and idiosyncratic shocks at the city level would counter-factually imply that only the cities with the highest expected value receive migrants (this is a direct consequence of the directed search structure of a Lucas–Prescott islands economy). On the contrary, in U.S. data all cities receive migrants all the time. Moreover, the model would be inconsistent with the negative relation between arrival and departure rates that we find at the city level.

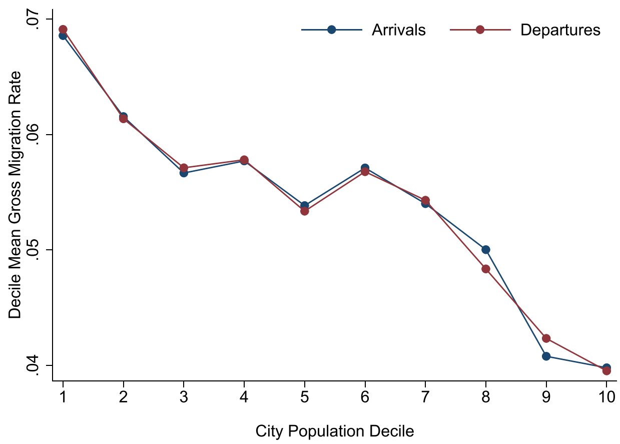

Fig. 1. Gross migration is negatively related to city size. Note: This figure displays mean values of gross migration within each decile of cities' average population.

3. Empirical evidence on inter-city migration

This section describes three features of inter-city migration: Gross migration is approximately linear in net migration; gross migration generally is declining in city size; and there is a large amount of cross-sectional variation in gross migration. We characterize gross and net inter-city migration using an annual panel of 381 metropolitan statistical areas (MSAs) for the period 1985 through 2013.[10] These data cover roughly 80% of the U.S. population. The Office of Management and Budget defines an MSA as a core of one or more adjacent counties (or county equivalents) that have at least one densely settled area, with a population of at least 50,000, plus any other adjacent counties that have a high degree of social and economic integration with the core as measured by commuting ties. While they sometimes include multiple legally incorporated municipal entities, for convenience we refer to our MSAs as cities.[11]

We calculate annual inter-city gross migration rates by aggregating to the city level county-to-county gross migration flows published by the Internal Revenue Service (IRS). The IRS data is unique for its universal coverage of cities and its availability over a relatively long sample period. However, as discussed by Kaplan and Schulhofer-Wohl (2017), migration flows are likely measured with error using IRS data.[12] We will discuss how measurement error might influence our findings later on. The *American Community Survey* can be used to measure migration between cities, but because of small sample sizes, gross migration rates can be calculated reliably for only a small number of cities. Furthermore, these data are only available for a short period. Another popular source of data for studying migration is the *Current Population Survey* (CPS). The CPS can be used to calculate inter-state migration, but not inter-city migration. While we focus on city-level migration, our three main qualitative findings for cities also hold for inter-state migration calculated using the CPS and the IRS data.

City i's *arrival rate* in year t, a_{it}, is defined as the total number of people who move to the city from any other city within the year divided by the city's beginning-of-year population, multiplied by 100. The corresponding *departure rate*, l_{it}, is defined similarly in terms of people leaving a city. Gross migration rates fluctuate over the business cycle and have been falling over our sample period.[13] To abstract from these dynamics we subtract from a city's annual gross rate the corresponding cross-section average, i.e. we remove time fixed effects. A city's net migration rate, n_{it}, is defined as the difference between these two gross rates so that it too is free of a time fixed effect.

Fig. 1 plots gross migration rates by population decile with only time effects removed and after adding back the unconditional average gross migration rate (i.e., the average gross migration rate across cities over the entire sample period). The deciles are ranked from lowest to highest populations. The difference between the arrival and departure rates for each decile is the mean net migration rate for that decile. Notice that net migration is essentially unrelated to city size. This fact reflects Gibrat's law for cities—that population growth is independent of city size. However, both the arrival and departure rates are

[10] All of our data is described in Appendix A.

[11] MSAs are similar to the United States Department of Agriculture's market-oriented delineation of counties called Commuting Zones (CZs), which covers the universe of US counties, whereas MSAs do not. We chose to work with MSAs because urban and rural areas might have fundamentally different migration dynamics and the employment and wage variables we require for our analysis are not available for CZs.

[12] They emphasize three sources of error: they cover only people with incomes high enough to file taxes and so undercount the poor and elderly; they track mailing addresses rather than home addresses; and they can be distorted by changes in household formation and in the time of year when people file their returns. Furthermore, to measure migration from one year to the next the IRS matches the tax returns of individual tax filers. If a tax filer does not file taxes in consecutive years they (and their dependents) will not be included in the migration counts.

[13] See Saks and Wozniak (2011) for evidence on the cyclicality of gross migration and Molloy et al. (2011) and Kaplan and Schulhofer-Wohl (2017) for evidence of its trend.

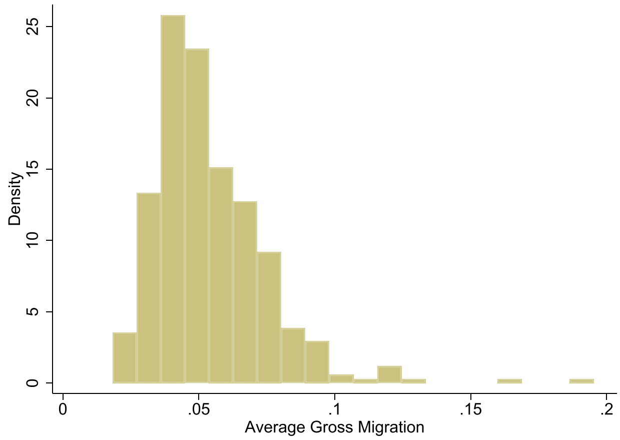

Fig. 2. Heterogeneity in gross migration across cities. Note: Twenty bin histogram of city fixed effects, where the fixed effect is measured as the average gross migration rate.

clearly not independent of city size: Smaller cities generally have higher migration rates than larger cities. For example, the average gross migration rate for the 10 smallest cities is 7.9% while for the 10 largest cities it is 2.8%.

This wide variation in gross migration by city size suggests substantial cross-sectional heterogeneity in gross migration. Such cross-sectional variation could mask the within-city dynamics we are interested in. To assess the extent of the heterogeneity we estimate gross migration fixed effects by city. We identify a city's gross migration fixed effect with its average gross migration rate, i.e., the time average of the sum of its arrival and departure rates divided by two. Note that the correlation between the mean arrival and mean departure rates calculated separately is very high, 0.9. The fixed effects are intended to capture the wide variety of observed and unobserved factors that influence the average gross migration rate in a city. Later on in the paper we use our model to shed light on what these factors might be. Fig. 2 displays a 20 bin histogram of the estimated city fixed effects. The fixed effects range from 2% to 20% with most of the mass in the range of 4% to 10%. This figure confirms that there is a vast amount of heterogeneity in gross migration across cities. Indeed, this variation accounts for 75% and 92% of the variances of the unadjusted arrival and departure rates, respectively.

From hereon we consider within-city variation in gross migration obtained by subtracting both time and city fixed effects from the raw gross migration rates in each city. These adjustments do not affect net migration so that Gibrat's law continues to hold. The within-city variances account for 15% and 12% of the variances of the unadjusted arrival and departure rates. Time effects account for relatively smaller shares—3.9% and 2.9%— of the variances of these two unadjusted rates.[14]

In contrast to the vast heterogeneity in gross migration rates across cities, the within-city relationship between gross and net migration is essentially homogeneous. This is evident in Fig. 3, which shows the scatter plot of arrival and net migration rates for each city-year observation after removing time and city fixed effects (the small blue solid circles). The red line corresponds to the fitted values of the ordinary least squares (OLS) regression of the arrival rates on the net migration rates. This regression yields a slope equal to 0.59 (0.01) and an R^2 equal to 0.61.[15] The red circles correspond to mean arrival rates for 15 percentiles of net migration. The red circles lining up along the regression line and the high R^2 from the OLS regression suggest within-city migration dynamics are well approximated by a unique linear relationship between gross and net migration rates.

Because the net migration rate is the difference between the arrival and departure rates, the fact that the slope of the regression line lies between zero and one implies that departure rates tend to co-vary negatively with net migration. This suggests that when a city receives a shock that leads to an increase (decrease) in net migration, arrival rates tend to rise (fall) and departure rates tend to fall (rise). Consistent with this we find a negative within-city correlation between arrival and departure rates.[16]

[14] The shares of the unconditional variances of gross migration do not sum to one because of the covariance terms. The within-city shares are likely biased upward because of measurement error; the time effects shares might be downward biased, as Kaplan and Schulhofer-Wohl (2017) show that the IRS-based migration rates understate the downward trend in gross migration relative to that from the CPS.

[15] As mentioned earlier, gross migration measured using IRS data is subject to measurement error. To address this we ran a 2SLS regression using log housing permits from the U.S. Census Bureau as an instrument for n_{it}. The F-statistic from the first stage regression is 58.6. The 2SLS slope is equal to 0.58 (0.02). Since the measurement error in the net migration and arrival rates are most likely positively correlated, the theoretical sign of the bias in the OLS estimate is ambiguous.

[16] The within-city correlation is small, -0.05, but it is dramatically different from the correlation without the removal of time and city fixed effects, which is 0.86. In Appendix B we discuss how measurement error can obscure a within-city correlation much closer to -1. Coen-Pirani (2010) considers five-year gross inter-state migration from the 1970–2000 decennial censuses and finds that the unconditional within-state correlation is positive, about 0.6 (Table 2). In a regression of the departure rate on current and lagged arrival rates he finds a partial correlation coefficient of about -0.2 after removing time and

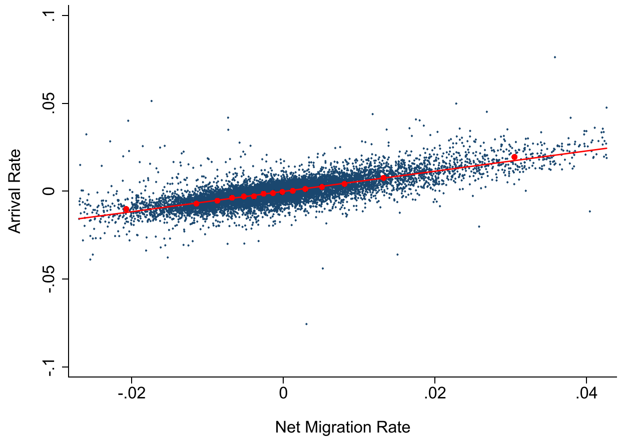

Fig. 3. Arrival rates are approximately linear in net migration rates. Notes: Small (blue) circles are the scatter plot of arrival rates and net migration rates, after removing city and time fixed effects. The larger (red) circles are the corresponding mean arrival rates for each of the 15 percentiles of net migration. The (red) line is the predicted values of the ordinary least squares (OLS) regression of the arrival rates on the net migration rates. To conserve on white space the scatter plot excludes the one percent largest and smallest net migration rates.

To summarize, we find that gross migration rates are negatively related to city size, that there is substantial heterogeneity in gross migration across cities, and that within cities gross inter-city migration is approximately linear in net migration. We now consider a theory of gross migration that is consistent with these patterns.

4. A simple static model of gross migrations

We start by considering a static model because it provides the most elemental structure for our theory of gross migration. The economy consists of a large number of geographically distinct cities with initial population x. Each city produces the consumption good using the following production function:

$$y = sF(n),$$

where s is the city's TFP level; n is labor; and F is a strictly increasing, strictly concave, and continuously differentiable function. There is a representative household that has a continuum of members initially distributed across the different city types (x, s) according to a measure μ. Each household member supplies one unit of labor inelastically and values consumption according to a strictly increasing, strictly concave, and continuously differentiable utility function U. The household must decide how many of its members to remove from each city and how many to bring in. The reallocation process is costly. In particular, if the household removes l agents from a city with initial population x, it creates a total disutility cost $G(l, x)$ to the household. Similarly, if a household members are brought into a city, it generates a total disutility cost $D(a, x)$ to the household. Both G and D are assumed convex and continuously differentiable with respect to l and a, respectively. Production takes place only after the reallocation process has been completed.

A social planner maximizes the utility of the representative household subject to feasibility constraints. Their problem is the following:

$$\max \left\{ U(C) - \int [D[a(x, s), x] + G[l(x, s), x]] d\mu \right\}, \tag{1}$$

subject to

$$p(x, s) \leq x + a(x, s) - l(x, s), \tag{2}$$

$$\int a(x, s) d\mu \leq \int l(x, s) d\mu, \tag{3}$$

$$C \leq \int sF[p(x, s)] d\mu, \tag{4}$$

and non-negativity constraints on all choice variables. Eq. (2) states that the population level of a city can be no greater than its initial population level, plus arrivals, minus departures. Eq. (3) states that the total number of arrivals across all cities

state fixed effects (Table 6). Aggregating our IRS data to the state level and removing time and state fixed effects, we find a within-state one-year gross migration correlation equal to -0.34. Using CPS state-level data, we find that this correlation is -0.16.

cannot exceed the total number of departures across all cities. Eq. (4) restricts consumption to be no greater than aggregate production, taking into account that each individual supplies one unit of labor inelastically, i.e., $n(x,s) = p(x,s)$.

The first order conditions to this problem are the following:

$$U'(C) - \varphi = 0, \tag{5}$$

$$\varphi s F'[p(x,s)] - \varphi\xi(x,s) = 0, \tag{6}$$

$$-\frac{\partial D[a(x,s),x]}{\partial a} + \varphi\xi(x,s) - \varphi\eta \leq 0, \ (= 0, \text{ if } a(x,s) > 0), \tag{7}$$

$$-\frac{\partial G[l(x,s),x]}{\partial l} - \varphi\xi(x,s) + \varphi\eta \leq 0, \ (= 0, \text{ if } l(x,s) > 0), \tag{8}$$

where $\varphi\xi(x,s)$, $\varphi\eta$, and φ are the Lagrange multipliers of Eqs. (2), (3), and (4), respectively.

Now, consider the following social planner's problem at a city of type (x,s):

$$\max\{\varphi s F(p) - \varphi\eta a + \varphi\eta l - D(a,x) - G(l,x)\}, \tag{9}$$

subject to

$$p \leq x + a - l, \tag{10}$$

where φ and η are taken as given. This city planner chooses how many agents to bring in and how many to remove, in order to maximize the contribution to social welfare made by the city. This contribution is the output produced by the city (evaluated at the economy-wide shadow value of consumption φ), minus both the net arrivals to the city (evaluated at the economy-wide shadow value of a reallocating agent $\varphi\eta$) and the disutility costs created by the gross arrivals and departures. Defining $\varphi\xi$ to be the Lagrange multiplier for Eq. (10), we see that the first-order conditions to the city social planner's problem are given by Eqs. (6)–(8). This indicates that the solution to the economy-wide social planner's problem—i.e., Eqs. (1)–(4)—can be obtained from the solutions to the city social planning problems described by Eqs. (9) and (10) (one for each city type (x,s)), plus the side conditions that φ and η satisfy Eqs. (3) and (5) (where C is given by Eq. (4)). This decomposition into city social planner problems will be extremely useful in determining the empirically relevant type of adjustment costs that cities face.

To see this, it will be important to first determine which functional forms for D and G are consistent with the empirical findings of Section 3. Recall that in that section we found strong evidence of a linear relation between arrival rates and net population growth rates. In particular, we found that

$$\frac{a}{x} = \phi + \vartheta\left(\frac{p-x}{x}\right), \tag{11}$$

with $0 < \vartheta < 1$, provides a good description of the data at the city level.[17] We also found that for every city, arrivals and departures were always positive throughout the whole panel. Assuming an interior solution for a and l, from Eqs. (7), (8), and (10), we get that in our model arrivals are related to net population changes according to

$$\frac{\partial D(a,x)}{\partial a} = -\frac{\partial G(a+x-p,x)}{\partial l}. \tag{12}$$

Observe that the only way that Eq. (12) can be made consistent with Eq. (11) is if D and G have the following quadratic forms:

$$\begin{aligned} D(a,x) &= \phi_2\tfrac{a^2}{x} + \phi_1 a + \phi_0, \\ G(l,x) &= \lambda_2\tfrac{l^2}{x} + \lambda_1 l + \lambda_0, \end{aligned}$$

where $\phi_2 > 0$, $\lambda_2 > 0$, and $\phi_1 + \lambda_1 < 0$.[18] Under these quadratic forms we get, from Eqs. (10) and (12), that arrival and departure rates are related to net population growth rates according to

$$\frac{a}{x} = \frac{\lambda_2}{\phi_2+\lambda_2}\left(\frac{p-x}{x}\right) - 2\left(\frac{\phi_1+\lambda_1}{\phi_2+\lambda_2}\right), \tag{13}$$

$$\frac{l}{x} = -\frac{\phi_2}{\phi_2+\lambda_2}\left(\frac{p-x}{x}\right) - 2\left(\frac{\phi_1+\lambda_1}{\phi_2+\lambda_2}\right). \tag{14}$$

[17] At this point we are abstracting from city-level fixed-effects in ϕ. These fixed effects will be introduced in Section 6.

[18] Under these quadratic forms, Eq. (12) becomes $\frac{\phi_2+\lambda_2}{2}\frac{a}{x} + \phi_1 = \frac{\lambda_2}{2}\left(\frac{p-x}{x}\right) - \lambda_1$. The convexity of D and G (which requires $\phi_2 > 0$ and $\lambda_2 > 0$) then implies equation (11) with $0 < \vartheta < 1$. Eqs. (7) and (8) with $a > 0$ and $l > 0$ imply that $\frac{\phi_2}{2}\frac{a}{x} + \phi_1 = -\frac{\lambda_2}{2}\frac{l}{x} - \lambda_1$, which requires that $\phi_1 + \lambda_1 < 0$.

Substituting Eqs. (13) and (14) into Eq. (9) and using (10), we get that the social planner's problem at a city of type (x, s) can be written as follows:

$$\max_p \left\{ \varphi s F(p) - \varphi \eta (p - x) + \Phi x + \Gamma (p - x) - \frac{\phi_2 \lambda_2}{(\phi_2 + \lambda_2)} \left(\frac{p - x}{x} \right)^2 \right\},$$

where Φ and Γ are constants.[19] We have thus obtained an important result: The empirical linear relation between arrival rates and net population growth rates that we uncovered in Section 3 implies quadratic adjustment costs to net population changes at the city level. Thus, the uncovered relation provides crucial information for the analysis of urban dynamics.

5. The dynamic model

The quadratic adjustment costs to net population changes implied by the linear relation between arrival rates and net population growth rates play a particularly important role in a dynamic setting. Given this and the fact that migration decisions are inherently forward-looking, we extend the simple model of the previous section to an infinite horizon framework. A key difference with the static economy is that cities are subject to idiosyncratic productivity shocks that change over time. Another difference is that we introduce an endogenous labor supply margin, which allows cities to adjust to their productivity shocks not only by changing their sizes (extensive margin), but by changing their employment levels (intensive margin). The last difference is that we include capital as an input into production. Instead of limiting ourselves to specifying social planning problems that face reduced-form cost functions for bringing in and removing people from cities, we find it valuable to provide microeconomic foundations to those planning problems (even in a highly stylized way) in order to facilitate economic interpretation. The reader comfortable with staying at the level of social planning problems may skip to Section 5.3 without loss of continuity.

5.1. Environment

Time is infinite and discrete. Similar to the static model previously described, the dynamic economy is populated by a representative household constituted by a continuum of members. At the beginning of every period, these agents are distributed in some way across a continuum of cities, where they must decide whether to stay or leave. The agents that decide to leave have two ways of reallocating to a different city within the same time period (both ways assume that agents do not know the economic conditions of the individual cities, only their distribution). The first way is through undirected search: The agent can randomly select a city and move there after incurring a disutility cost τ. We assume that the random selection of cities is size-weighted (e.g., New York City is more likely to be selected than Ithaca, NY).[20] As a consequence, larger cities end up receiving a proportionately larger fraction of the total number of random searchers in the economy.

The second way of reallocating to a different city is by joining one of the many recruitment markets in the economy. Each type of city (to be described in Section 5.2) has its own recruitment market where agents are contacted by recruiters offering slots to move to a city of that type (i.e., city types are contractable).[21] Any person located in a given city at the beginning of the period can act as a recruiter for that city. The total number of agents that a recruiter is able to contact in the recruitment market for their type of city is given by the following technology:

$$a_t = R(d_t), \tag{15}$$

where a_t is the number of contacted agents, d_t is the recruiting effort of the recruiter (measured in terms of lost utility), and R is a strictly increasing, strictly concave, and continuously differentiable function. Once an agent accepts the offer made by a recruiter, the identity of the recruiter's city of origin is revealed to the agent, allowing them to move to that particular city. In the competitive equilibrium to be described in the next section, the prices of the different types of recruitment slots adjust in such a way that agents become indifferent between doing undirected search or joining any of the recruitment markets in the economy.

Before moving decisions are made, we assume that each agent located in a city at the beginning of the period receives an idiosyncratic shock ξ_t that determines how attached they are to that city. In particular, ξ_t represents the one-time idiosyncratic utility loss that the agent would experience if they were to move out of the city. We assume that this idiosyncratic shock (which can be positive or negative) is i.i.d. across individuals and over time, and is drawn from a distribution with continuous density function $\psi > 0$ defined over an interval $[\xi_{\min}, \xi_{\max}]$.

In addition to deciding whether to stay or move, agents must decide whether to work or not. These decisions, which are made after the migration process has been completed, depend on the realization of their idiosyncratic disutility of working

[19] See Appendix C for a complete derivation.

[20] In particular, we assume that agents doing undirected search randomly draw the name of another household member and moves to the city where that person was located at the beginning of the period. For this reason, one can think that part of the random moves in the economy are motivated for family reasons. In the U.S. about 27% of people who move do so at least in part for family reasons (this is based on CPS data).

[21] Recruiters are meant to be a highly stylized representation of the recruitment activities of firms. These activities can be either performed by their own human resources departments and employees, or by outside recruiters (e.g., recruitment/staffing agencies) that are contracted to facilitate their hiring processes. In either case, the recruitment activities that we focus on are those directed to finding suitable candidates from other cities.

$\omega_t > 0$. We assume that ω_t is i.i.d. across individuals and over time, and that it is drawn from a distribution with continuous density function φ defined over an interval $[\omega_{\min}, \omega_{\max}]$.

The realized flow utility that an agent obtains at the end of the period is then given by

$$u_t = U(C_t) - d_t - \xi_t \chi_t^l - \tau \chi_t^s - \omega_t \chi_t^n, \tag{16}$$

where C_t is their consumption level; U is a strictly increasing, strictly concave, and continuously differentiable utility function; d_t is the recruiting effort that the agent makes in the city of origin; ξ_t is the realized idiosyncratic attachment to the city of origin; χ_t^l indicates if the agent moved; χ_t^s indicates if the agent performed undirected search; ω_t is the realized idiosyncratic disutility of work in the city of destination; and χ_t^n indicates if the agent worked in the city of destination. Observe that if the agent does not move within the period, the city of origin and the city of destination are the same. The agents' preferences are described by the expected discounted value of these flow utilities, using a discount factor β.

Each city has a representative firm that produces the consumption good using the following production technology:

$$y_t = s_t F(n_t, k_t),$$

where y_t is output; s_t is a city-level idiosyncratic productivity shock that follows a finite Markov process with transition matrix Q; n_t is labor; k_t is capital; and F is a strictly increasing, strictly concave, and continuously differentiable function.

Capital is assumed to be freely movable across the cities. The aggregate stock of capital follows a standard law of motion

$$K_{t+1} = (1-\delta)K_t + I_t, \tag{17}$$

where I_t is gross investment and δ is the depreciation rate.

5.2. Competitive equilibrium

The type of a city at date t is given by a real number x_0 and a history s^t, where x_0 is the total population of the city at the beginning of date 0 and where $s^t = (s_0, \ldots, s_t)$ is its history of productivity shocks between dates 0 and t.[22] We denote the date-t city-type by $z^t = (s^t, x_0)$. A measure μ_t describes the total number of cities across the different z^t. Starting from an initial μ_0 at date 0, $\{\mu_t\}_{t=1}^{\infty}$ satisfies that $\mu_t(z^t) = \mu_{t-1}(z^{t-1})Q(s_t, s_{t-1})$ for every date t and history z^t.

In order to simplify the description of a competitive equilibrium, we assume that agents obtain full insurance within their households.[23] As a result, independent of their idiosyncratic histories, every agent in the representative household consumes the same. Also, we simplify the structure of the labor markets to a bare minimum by assuming that each city operates spot labor markets. When an agent purchases a recruitment slot from a recruiter, they gain access to the spot labor markets of the recruiter's city of origin.[24]

At the beginning of period 0, the total number of agents that the representative household has at each city of type (x_0, s_0) is equal to x_0. Thereafter, the representative household must decide the recruitment, labor-supply, and migration actions of each of these agents, while perfectly insuring their consumption. Before describing the household's dynamic problem, it is convenient to characterize some of its static decisions. First, since the recruiting effort d_t enters linearly in the flow utility (16) and since the recruitment technology (15) is strictly concave, the representative household will always direct all of its members in a given city to put in exactly the same amount of recruiting effort.

Second, since at the beginning of the period all of its agents in a given city are identical except for their idiosyncratic attachments to the city, the representative household will always move out the agents with the lowest idiosyncratic attachments first, following a simple threshold rule. Given a total number of agents x_t in the city at the beginning of the period and given an attachment threshold $\bar{\xi}_t$, the total attachment losses incurred by the household are given by

$$\left(\int_{\xi_{\min}}^{\bar{\xi}_t} \xi \ \psi(\xi) d\xi\right) x_t. \tag{18}$$

Defining $l_t = x_t \int_{\xi_{\min}}^{\bar{\xi}_t} \psi(\xi) d\xi$ as the total number of agents that leave, we can invert this expression to define the threshold $\bar{\xi}_t$ in terms of the fraction of agents that move l_t/x_t. Thus, Eq. (18) can be written as a function $G(l_t, x_t)$. We assume that ψ is such that G is convex and continuously differentiable.

[22] To simplify the notation, we assume that in addition to there being a finite number of idiosyncratic productivity levels there is a finite number of values for x_0.

[23] Assuming a large household is a simple way of introducing complete markets. Alternatively, we could assume that individual agents trade lotteries that allow them to obtain full insurance through the marketplace (lotteries would be necessary because of indivisibilities in the commodity space at the individual level). Such a scenario would be much more complicated to describe, but it would lead to exactly the same equilibrium allocation as the one here.

[24] Instead of assuming spot labor markets in each city, we could alternatively specify an equilibrium with long-term labor contracts. These long-term contracts could be modeled as stopping-times specifying contingencies under which a worker would stop working for an employer. There would be a continuum of competitive markets, one for each possible type of stopping time, and the recruiters would intermediate the stopping times traded between workers and firms. Not surprisingly, describing such an equilibrium is much more complicated (see, e.g., Alvarez and Veracierto, 2012 and Veracierto, 2016). Nevertheless, the equilibrium allocations would be exactly the same as those here.

Third, since after the migration process has been completed all agents in a given city are identical except for their disutility of working, the representative household will always put to work the agents with the lowest disutility of working first, again following a simple threshold rule. Given a total (post-migration) number of agents p_t in the city and given a labor-supply threshold $\bar{\omega}_t$, the total disutility of working incurred by the household can then be written as

$$\left(\int_{\omega_{\min}}^{\bar{\omega}_t} \omega\varphi(\omega)d\omega\right)p_t. \tag{19}$$

Defining $e_t = p_t \int_{\omega_{\min}}^{\bar{\omega}_t} \varphi(\omega)d\omega$ as the total number of agents that work, we can again invert this expression to define the threshold $\bar{\omega}_t$ in terms of the employment-to-population ratio, e_t/p_t, and write Eq. (19) as a function $J(e_t, p_t)$. We assume that φ is such that J is convex and continuously differentiable.

The representative household seeks to maximize the total welfare of its members by solving the following problem:

$$\max \sum_{t=0}^{\infty} \beta^t \left\{U(C_t) - \tau\Lambda_t + \sum_{z^t}\left[-J\left(e_t\left(z^t\right), p_t\left(z^t\right)\right) - p_{t-1}\left(z^{t-1}\right)R^{-1}\left(\frac{a_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right) - G\left(l_t\left(z^t\right), p_{t-1}\left(z^{t-1}\right)\right)\right]\mu_t\left(z^t\right)\right\} \tag{20}$$

subject to

$$p_t\left(z^t\right) = p_{t-1}\left(z^{t-1}\right) + b_t\left(z^t\right) + \Lambda_t p_{t-1}\left(z^{t-1}\right) - l_t\left(z^t\right), \tag{21}$$

$$\sum_{z^t} b_t\left(z^t\right)\mu_t\left(z^t\right) + \Lambda_t = \sum_{z^t} l_t\left(z^t\right)\mu_t\left(z^t\right), \tag{22}$$

$$C_t + \sum_{z^t} q_t\left(z^t\right)b_t\left(z^t\right)\mu_t\left(z^t\right) + I_t \le \sum_{z^t} q_t\left(z^t\right)a_t\left(z^t\right)\mu_t\left(z^t\right) + \sum_{z^t} w_t\left(z^t\right)e_t\left(z^t\right)\mu_t\left(z^t\right) + r_t K_t + \Pi_t, \tag{23}$$

and Eq. (17), where Λ_t is the total number of agents that do undirected search, p_t is the end-of-period population level, e_t is employment, l_t is the total number of agents that leave, a_t is the total number of recruitment slots sold, b_t is the total number of recruitment slots bought, q_t is the price of a recruitment slot, w_t is the wage rate, r_t is the economy-wide rental rate of capital, and Π_t is the economy-wide profits of firms.[25] The optimal static decisions of the representative household are already incorporated in this decision problem.

The constraints begin with Eq. (21), which states that the end-of-period population level at a city of type z^t is given by the beginning-of-period population, plus the total number of agents that purchase recruitment slots to the city, plus the total number of agents that arrive to the city through undirected search, minus the total number of agents that leave. Eq. (22) states that the total number of agents leaving their cities either do undirected search or purchase recruitment slots to other cities. Eq. (23) is the budget constraint of the household. It states that total consumption, plus the total value of all recruitment slots purchased, plus investment cannot exceed the household's income. This includes the total value of all recruitment slots sold, the total value of salaries earned, the rental of capital, and the profits of firms.

In each city with history z^t, there is a representative firm that solves the following static problem:

$$\pi_t\left(z^t\right) = \max\left\{s_t F\left[n_t\left(z^t\right), k_t\left(z^t\right)\right] - w_t\left(z^t\right)n_t\left(z^t\right) - r_t k_t\left(z^t\right)\right\}.$$

Firms purchase labor services in their local spot labor market and rent capital in the economy-wide capital market. Note that $\Pi_t = \sum_{z^t} \pi_t(z^t)\mu_t\left(z^t\right)$.

The market clearing conditions are

$$n_t\left(z^t\right) = e_t\left(z^t\right), \tag{24}$$

$$b_t\left(z^t\right) = a_t\left(z^t\right), \tag{25}$$

$$K_t = \sum_{z^t} k_t\left(z^t\right)\mu_t\left(z^t\right), \tag{26}$$

$$C_t + I_t = \sum_{z^t} s_t F\left[n_t\left(z^t\right), k_t\left(z^t\right)\right]\mu_t\left(z^t\right), \tag{27}$$

where Eq. (24) is market clearing in the labor market of each type of city; Eq. (25) is market clearing in the recruitment market of each type of city; and Eqs. (26) and (27) are the economy-wide market clearing conditions for capital and the consumption good, respectively.

[25] To simplify the notation, we assume that $p_{t-1}\left(s^{t-1}, x_0\right)$ at $t = 0$ simply denotes x_0.

5.3. The city planner's problem and steady state

Since this is a convex economy with no distortions, the welfare theorems hold. As a consequence, equilibrium allocations can be found by solving the economy-wide planning problem, which is to maximize the objective function in Eq. (20), subject to Eqs. (21), (22), and (24)–(27). While this gets rid of prices, it still seems to be a complicated problem to solve. Fortunately, similar to the static model described in Section 4, the economy-wide social planning problem can be decomposed into a city-level planning problem and certain side conditions. Since we focus on steady-state equilibria in the rest of the paper, we describe this decomposition only at the steady state. For the sake of convenience, we describe it in recursive form.[26]

The state of a city is given by a pair (x, s), where x is the total population at the beginning of the period, and s is its current productivity shock. The city planner's problem is the following:

$$V(x,s) = \max\left\{\varphi sF[n,k] - J(n,p) - xR^{-1}\left(\frac{a}{x}\right) - G(l,x) - \varphi\eta(a+\Lambda x) + \varphi\eta l - \varphi rk + \beta\sum_{s'} V\left(p,s'\right)Q\left(s',s\right)\right\} \tag{28}$$

subject to

$$p = x + a + \Lambda x - l, \tag{29}$$

where φ is the shadow price of the consumption good, η is the shadow consumption price of a reallocating agent, and r is the shadow consumption price of capital. The economy-wide values $(\Lambda, \varphi, \eta, r)$ are taken as given by the city planner.[27]

The optimal decision rules to the city planner's problem generate an invariant distribution μ across duples (x, s) that satisfies the following equation:

$$\mu\left(B,s'\right) = \int_{\{(x,s):\ p(x,s)\in B\}} Q\left(s',s\right)d\mu, \tag{30}$$

for every Borel set B and next-period productivity s'.

The side conditions that guarantee that the solution to the city planner's problem corresponds to the solution of the economy-wide planner's problem are

$$\varphi = U'(C), \tag{31}$$

$$C + \delta\int k(x,s)d\mu = \int sF[n(x,s),k(x,s)]d\mu, \tag{32}$$

$$\int a(x,s)d\mu + \Lambda = \int l(x,s)d\mu, \tag{33}$$

$$\tau = \int \left(R^{-1}\right)'\left(\frac{a(x,s)}{x}\right)xd\mu. \tag{34}$$

Eq. (31) says that the shadow price of the consumption good equals the marginal utility of consumption. Eq. (32) is consumption good feasibility. Eq. (33) says that the total number of agents arriving to cities equals the total number leaving them. Eq. (34) is the first-order condition for Λ in the economy-wide social planner's problem. To understand it, observe that the marginal cost of increasing Λ is simply the disutility cost τ. The marginal benefit of increasing Λ is that each city sees its arrivals for personal reasons increase by x, which allows it to cut its recruiting activity by exactly the same amount (and leave its population unchanged). The associated marginal saving in recruitment effort is given by the derivative of R^{-1} evaluated at $\frac{a(x,s)}{x}$. At the optimum, marginal costs are equated to marginal benefits.[28]

6. Parameterization

This section discusses how we determine the different functional forms in the model, how the model can be extended to make it consistent with the heterogeneity described in Section 3, and how we calibrate parameter values.

[26] See Appendix D for a detailed discussion of the claims made in this section.

[27] For the reader that has skipped Sections 5.1 and 5.2, output is now produced using labor n and capital k according to F. Capital is freely movable across cities and depreciates at a constant rate δ. The function J describes the total disutility of employing n agents out of the p end-of-period inhabitants. The term xR^{-1} describes the total disutility of bringing a arrivals to the city (similar to the function D in the static economy). G is the total disutility generated by the total departures l from the city. Q describes transition probabilities for the idiosyncratic productivity shocks. A difference with the static economy is that, in addition to the a arrivals decided by the city planner, the city attracts Λx agents out of the economy-wide total number of agents Λ doing undirected search. Eqs. (28) and (29) are analogous to Eqs. (9) and (10).

[28] Observe that the city planner's problem with side conditions provides the basis for a straightforward algorithm for computing a steady state: (1) For fixed values of $(\Lambda, \varphi, \eta, r)$, solve the city planner's problem (28) and (29), (2) compute the associated invariant distribution μ in (30), and (3) verify that the side conditions (31)–(34) are satisfied. If they are not, go back to 1) with new values for $(\Lambda, \varphi, \eta, r)$.

6.1. Functional forms

A quick inspection of Eqs. (28)–(34) reveals that there are five functional forms to be determined—those for U, F, J, R^{-1}, and G. For determining U and F, we follow the macroeconomic literature in choosing them to be consistent with "balanced growth path observations". One of these observations is that over long periods of time, the share of labor in national income has remained fairly stable. A convenient functional form for F that is consistent with this observation is Cobb–Douglas. In particular, we assume that

$$F(n,k) = n^{\alpha} k^{\gamma}. \tag{35}$$

If we introduced an aggregate TFP variable (common to all cities), which multiplies the production function F and grows at a constant rate, we would obtain essentially the same structure as the neoclassical growth model. Therefore, we determine the functional form for U using another of the balanced growth observation commonly used in the calibration of that model—that over long periods of time there has been no trend in labor supply while wages have been growing at a fairly constant rate. To be consistent with this observation, preferences must be such that income effects exactly cancel substitution effects. Under separable preferences, this requires U to be logarithmic.

No balanced growth observation sheds light on the functional form for J. For the sake of parsimony, we thus assume that

$$J(n,p) = B\left(\frac{n}{p}\right)^{\nu} p. \tag{36}$$

Since the first order condition for n in the city planner's problem (28) and (29) is $\varphi s \frac{\partial F}{\partial n} = \frac{\partial J}{\partial n}$, and the spot wage rate of the city is $s\frac{\partial F}{\partial n}$, Eq. (36) implies a constant elasticity of labor supply. Recall that under the microeconomic structure of Section 5, the total disutility of labor supply $J(n,p)$ satisfied Eq. (19). Eq. (36) can thus be obtained by assuming a density function $\varphi(\omega) = A\omega^{\theta}$ defined on the interval $\left[0, \left(\frac{\theta+1}{A}\right)^{\frac{1}{\theta+1}}\right]$, with $A > 0$ and $\theta > -1$. The implied values for ν and B are $\frac{\theta+2}{\theta+1} > 1$ and $A^{1-\nu}(\theta+1)^{\pi}(\theta+2)^{-1} > 0$, respectively.[29]

The functional forms for R^{-1} and G are selected to be consistent with the empirical evidence of Section 3. Defining $D(a,x) = xR^{-1}\left(\frac{a}{x}\right)$ and letting $\varphi\xi(x,s)$ be the Lagrange multiplier for Eq. (29), the first order conditions for $a(x,s)$ and $l(x,s)$ in the city planner's problem (28) and (29) are identical to Eqs. (7) and (8). Using Eq. (29) and assuming interior solutions for $a(x,s)$ and $l(x,s)$, it is then straightforward to verify (similar to what we did in Section 4) that for the model to be consistent with the linear relation between arrival rates and net population growth rates uncovered in our empirical analysis, D and G must have quadratic forms. For this reason, we assume that R^{-1} and G are given by

$$R^{-1}\left(\frac{a}{x}\right) = H\left(\frac{a}{x}\right)^2, \tag{37}$$

$$G(l,x) = \left[-\psi_1 \frac{l}{x} + \psi_2\left(\frac{l}{x}\right)^2\right]x, \tag{38}$$

where $H, \psi_1, \psi_2 > 0$. Under these functional forms, the total arrival rate is given by

$$\frac{a}{x} + \Lambda = \frac{\psi_2}{(\psi_2+H)}\left(\frac{p-x}{x}\right) + \frac{\psi_1}{2(\psi_2+H)} + \frac{H}{(\psi_2+H)}\Lambda, \tag{39}$$

and the total departure rate is given by

$$\frac{l}{x} = -\frac{H}{(\psi_2+H)}\left(\frac{p-x}{x}\right) + \frac{\psi_1}{2(\psi_2+H)} + \frac{H}{(\psi_2+H)}\Lambda. \tag{40}$$

Both Eqs. (39) and (40) are linear functions of the net population growth rate $(p-x)/x$. In terms of the microeconomic structure of Section 5, recall that $G(l,x)$ satisfied Eq. (18). Thus, Eq. (38) can be obtained by assuming that ψ is a uniform density function over the interval $[-\psi_1, -\psi_1 + 2\psi_2]$.[30]

6.2. Heterogeneity

Section 3 described a large amount of heterogeneity in our empirical panel of cities. The cities vary widely in terms of their average sizes and their gross migration rates. Moreover, we showed that (on average) larger cities tend to have lower arrival and departure rates (Fig. 1). Cities also turn out to vary quite significantly in terms of their average TFP levels (TFP measurement will be described in Section 6.3). In this section we discuss how the model described so far, which has ex-ante identical cities, can be extended to be consistent with all of this heterogeneity.

[29] Appendix E.1 provides a detailed derivation of Eq. (36) from microeconomic primitives.

[30] Appendix E.2 provides a detailed derivation of Eq. (38) from microeconomic primitives.

A simple way of doing this is by introducing a finite number of city categories that differ in terms of the following three permanent characteristics.[31] First, each city of category m provides a flow utility $\varsigma(m)$ to every agent living in that city. $\varsigma(m)$ can be interpreted as the permanent attractiveness of the city, such as its weather and geographic location. Second, each city of category m has a total factor productivity level $\kappa(m)$ that enters multiplicatively in the production function (35). Third, the parameter ψ_1 that enters the cost function $G(l, x)$ in Eq. (38) is category-specific. To provide an economic interpretation to this assumption, we resort to the microeconomic structure of $G(l, x)$ provided in Section 5.1. In that section we assumed that before moving decisions are made, that each agent located in a city at the beginning of the period receives an idiosyncratic i.i.d. shock ξ_t that determines the one-time idiosyncratic utility loss that the agent would experience if they were to move out of the city. The previous section determined that, in order to be consistent with our empirical findings of Section 3, the ξ_t shocks must be drawn from a uniform distribution over the interval $[-\psi_1(m), -\psi_1(m)+2\psi_2]$. Thus, the assumption that $\psi_1(m)$ is category-specific amounts to assuming that the support of the distribution of ξ_t is category-specific. The rationale for doing this is then that the idiosyncratic attachment of a person to a city could be related (in a reduced form) to its permanent attractiveness and productivity. For instance, agents may be more attached to a city if it provides a wider variety of goods, services, and jobs. Because empirically larger cities do provide a wider variety of goods, services, and jobs and because the long-run average size of a city will be endogenously related to its permanent attractiveness and productivity levels, we allow the support of the idiosyncratic attachment shocks to be related to these permanent characteristics.

To see why this version of the model can be consistent with the vast amount of heterogeneity observed in the data, associate each city with a permanent category m. Now, observe that while the slope of Eq. (39) is common to all cities, its intercept is city-specific (since $\psi_1(m)$ is). Thus, conditional on values for H, ψ_2, and Λ, the fixed effects obtained in Section 3 could be used to determine the different values of $\psi_1(m)$ (the values for H and ψ_2 should be chosen to reproduce the common slope estimated in that regression). In turn, direct measures of TFP for each city in our sample could be used to determine each city's value of $\kappa(m)$. Conditional on all other parameter values in the model, the flow utilities $\varsigma(m)$ could then be chosen to reproduce the average size of each city in our sample. By construction, the model would then be consistent with all the heterogeneity in gross migration rates, TFP levels, and population sizes observed in the data, including the negative relation between average city sizes and gross migration reported in Fig. 1.

However, replicating all the cross-sectional heterogeneity observed in the data would be completely irrelevant for the main focus of this paper, which is analyzing the contribution of TFP shocks to urban population dynamics. The reason is that with all the heterogeneity introduced, the city planner' problem (28) and (29) would take the following form:[32]

$$V(x,s,m) = \max_p \Big\{ \varphi\hat{\kappa}(m)\hat{s}p^{\varrho} + \varsigma(m)p - \varphi\eta(p-x)$$

$$+ \Phi(m,\Lambda)x + \Gamma(m,\Lambda)(p-x) - \frac{H\psi_2}{\psi_2+H}\left(\frac{p-x}{x}\right)^2 x + \beta\sum_{s'} V\left(p,s',m\right)Q\left(s',s\right)\Big\}, \tag{41}$$

where Φ and Γ are terms that depend only on m and Λ, $\varrho = \frac{\sigma}{\sigma-\nu}(1-\nu)$ with $\sigma = \frac{\alpha}{1-\gamma}$, $\hat{\kappa}(m) = \kappa(m)^{-\frac{\pi}{(\sigma-\nu)(1-\gamma)}}$, and

$$\hat{s} = \varphi^{-\frac{\nu}{\sigma-\nu}-1}\left[r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\right]^{-\frac{\nu}{\sigma-\nu}} B^{\frac{\sigma}{\sigma-\nu}}\left[\left(\frac{\nu}{\sigma}\right)^{\frac{\sigma}{\sigma-\nu}} - \left(\frac{\nu}{\sigma}\right)^{\frac{\nu}{\sigma-\nu}}\right] s^{-\frac{\nu}{(\sigma-\nu)(1-\gamma)}}.$$

Since the coefficient $H\psi_2/(\psi_2+H)$ in the quadratic adjustment costs term in Eq. (41) is common to all permanent categories m, the population dynamics following a TFP shock would be independent of any heterogeneity. For this reason—and because it is far simpler to solve—for the rest of the paper we continue considering the model with ex-ante identical cities.

6.3. Calibration

In this section we calibrate the steady-state competitive equilibrium of the model to U.S. data. In addition to specifying the stochastic process for idiosyncratic productivity levels, we need to calibrate the following 10 parameters: H, ψ_1, ψ_2, τ, γ, α, δ, β, B, and ν. These comprise the migration parameters, the factor shares in production, the depreciation rate, the discount factor, and the parameters governing labor supply.

In order to calibrate the model we first need to specify the empirical counterpart to capital in the model. Since we assume that capital is freely movable across cities, it seems natural to abstract from land and structures. The empirical counterpart for the model's capital we work with is therefore identified with equipment and intellectual property products. The empirical counterpart for consumption is identified with personal consumption expenditures in nondurable goods and services. Output Y is then defined as the sum of this consumption measure and private fixed investment in equipment and intellectual property products.

Since at steady state $\delta = I/K$, we calibrate the depreciation rate using the average investment-capital ratio, which by our measurement is equal to 0.19.[33] Calibrating to an annual interest rate of 4% requires a time discount factor β equal to 0.96.

[31] Appendix D describes the heterogeneous version of the model in detail.

[32] See Appendix D.8 for a detailed derivation.

[33] All statistics mentioned in this section are based on annual data between 1985 and 2013 (the same time period used in Section 3).

In steady state the capital's income share satisfies

$$\gamma = \left(\frac{1}{\beta} - 1 + \delta\right)\frac{K}{Y}.$$

We use this equation to calibrate $\gamma = 0.175$ using our calibrated values for β and δ and an average capital-output ratio of 0.755. The parameter α is chosen to reproduce a labor share of 0.64.

We measure city-level TFPs using the first order conditions of firms and data on wages and employment. We then use these measured TFPs to estimate the stochastic process for the idiosyncratic productivity shock s_t. A slight complication is that wages and employment contain both trend and business cycle components, though we are calibrating a deterministic steady state with no growth. Fortunately, the Cobb–Douglas specification for the production function generates a log-linear relation between those variables that allows for a simple transformation of the data so that we can abstract from the trend and cycle. In particular, for any variable x_{it} in city i at date t, we define

$$\hat{x}_{it} = \ln x_{it} - \frac{1}{M}\sum_{j=1}^{M} \ln x_{jt}, \tag{42}$$

where M is the number of cities in our sample. Subtracting the mean value of $\ln x_{jt}$ in each period eliminates variation due to trend and business cycle dynamics. Using firms' first-order conditions and exploiting the fact that capital is perfectly mobile (so the rental rate of capital is the same in each city), we have that

$$\Delta\hat{s}_{it} = (1-\gamma)\Delta\hat{w}_{it} + (1-\alpha-\gamma)\Delta\hat{n}_{it},$$

where Δ is the first difference operator, and w_{it} and n_{it} denote wages and employment, respectively. Applying the first difference operator removes any fixed effects.[34] Given the values of α and γ already determined, this equation allows us to measure $\Delta\hat{s}_{it}$ using data on $\Delta\hat{w}_{it}$ and $\Delta\hat{n}_{it}$. Having done this, we run the following regression:

$$\Delta\hat{s}_{i,t} = \hat{\rho}\Delta\hat{s}_{i,t-1} + e_{i,t}. \tag{43}$$

Our OLS estimate of $\hat{\rho}$ is 0.27 (0.01), while our estimate of the standard deviation of $e_{i,t}$ is $\sigma_e = 0.013$.[35]

The stationary stochastic process for the growth rates $\Delta\hat{s}_{it}$ implies a non-stationary stochastic process for the idiosyncratic productivity levels. This is problematic for our theoretical model because it requires stationarity in levels. To overcome this difficulty we assume that there is a reflecting barrier for the idiosyncratic productivity levels. In particular, we assume the following stochastic process for s_t:

$$\ln s_{t+1} = \max\{g + (1+\rho)\ln s_t - \rho \ln s_{t-1} + \varepsilon_{t+1}, \ln s_{\min}\}, \tag{44}$$

where $\varepsilon_{t+1} \sim N(0, \sigma_\varepsilon)$, $g < 0$, and $\rho > 0$.[36] Observe that with this stochastic process, TFP growth rates are approximately AR(1), while TFP levels are stationary because of the negative drift and the reflection at the barrier $\ln s_{\min}$. Despite this, the model's endogenous variables will appear to be non-stationary over samples of similar length to our data. In our calibration we work with a finite approximation to the stochastic process in Eq. (44), normalize $\ln s_{\min}$ to zero, and select values for ρ and σ_ε so that when we run the regression given by Eq. (43) on the values of s_t obtained from simulating the stochastic process (44), we reproduce the empirical estimates of $\hat{\rho}$ and $\hat{\sigma_e}$. In turn, we choose g to reproduce a Zipf's law coefficient over TFP levels of −4.0, which we measure in our data. The required value of g is −0.00075. Since this value is so close to zero, the distortions to Eq. (43) introduced in Eq. (44) turn out to be negligible.

Under the preferences described by Eq. (36), the first-order condition for employment n_t in the city's planning problem (28) is given by

$$w_t = CB\pi\left(\frac{n_t}{p_t}\right)^{\nu-1}. \tag{45}$$

Thus, $1/(\nu-1)$ is the Frisch elasticity of labor supply. Estimating the impulse responses for the log of w_t and the log of n_t/p_t to a one-standard deviation innovation in city-level TFP, we select $\nu = 4.05$ to reproduce the ratio of those responses on impact (our estimation of the impulse responses is described in detail in Section 7.) In turn, the disutility of labor supply B is set at 1.29 to reproduce a ratio of aggregate employment to population equal to 0.62.

Observe that the coefficient in the quadratic adjustment costs term in the city's planning problem (41) is equal to H times the slope $\psi_2/(\psi_2 + H)$ in Eq. (39). Since, conditional on our estimate of 0.58 for the value of this slope (from Section 3), H determines the adjustment costs to population changes, we choose it to reproduce the impact response of population to a one-standard deviation innovation in city-level TFP. The value for ψ_2 is then obtained from H and our estimate for the slope $\psi_2/(\psi_2 + H)$. In principle, the value for the disutility of undirected search τ could be determined by reproducing an

[34] Measuring fixed effects as the time series means of $\hat{s}_{it}$, we find considerable heterogeneity in productivity, much like with gross migration.

[35] The reader may be concerned that there could be externalities due to agglomeration effects that our measurement of TFP may be confounding. Measuring TFP shocks in this case requires using an agglomeration elasticity in addition to factor shares. It turns out that when we use the agglomeration elasticity estimated in Davis et al. (2014) we find a very similar TFP process to the one reported here.

[36] The strategy of introducing a reflecting barrier to generate a stationary process in levels has been previously used in the seminal work of Gabaix (1999b). The only difference is that instead of having $\rho > 0$, Gabaix assumed a random walk ($\rho = 0$).

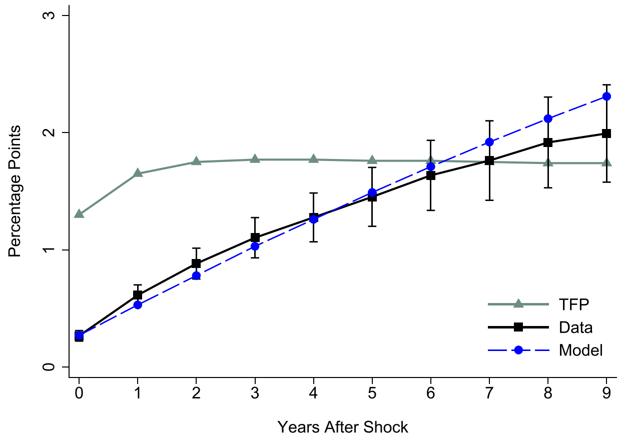

Fig. 4. Dynamic response of population to a TFP shock. Note: The "Data" line with whiskers shows the point estimates of the dynamic response of population to a one standard deviation positive TFP shock, accompanied by two-standard-error bands. The "TFP" line is the dynamic response of TFP in the model. The "Model" line is the dynamic population response predicted by the model.

empirically reasonable value for the fraction Λ of agents that perform that type of search. We don't have direct measures of Λ, but we can bound it between 0 and 0.003. The reason for the upper bound is that 0.003 is the lowest arrival rate in our sample (i.e., it is the largest value of Λ consistent with having a positive value for a in every city).[37] It turns out that our results are virtually identical for any value of Λ in that range. Taking $\Lambda = 0.003$ as our benchmark case, the associated values for H, ψ_2, and τ are 59.0, 81.5, and 6.0, respectively. In turn, the parameter ψ_1 is set at 14.8 to reproduce an average arrival rate of 0.054.

7. Results

In this section we use the calibrated version of the model to answer the two main questions that motivated our paper: (1) How much of the heterogeneity in cities' growth rates can be accounted for by idiosyncratic TFP shocks? (2) And to what extent can the most extreme cases of persistent urban decline be accounted for by reductions in TFP? However, in order to gain confidence in the quantitative predictions of the model, we first compare different impulse responses in the model economy with their empirical counterparts. These comparisons will serve as a validation test for the model.

We use local projection methods described in Jordà (2005) to estimate the impulse responses of different variables to a TFP shock. For each of population, employment, and wages we first apply the transformation in Eq. (42) and then we identify the response j years after a TFP shock with the coefficient obtained from linearly projecting the difference of the transformed variable between dates $t+j$ and $t-1$ on the date t value of the TFP residuals obtained from Eq. (43) (under our estimated value of $\hat{\rho}$). For the arrival and departure rates we first remove time and city fixed effects as in Section 3 and then we identify the response j years after a TFP shock with the coefficient obtained from projecting the level of the transformed variable at date $t+j$ on the date t value of the TFP residuals. Standard errors are clustered by city.

In order to have a clear picture of the exogenous variations that a city experiences after a positive TFP shock, the line with triangles in Fig. 4 displays the impulse response of TFP to a positive TFP shock equal to one standard deviation. We see that the TFP level jumps on impact (by the magnitude of the initial shock), continues to grow at a decreasing rate (given the positive serial correlation $\hat{\rho} > 0$), and reaches its peak level three years after the TFP shock before starting to slowly decline over time (as the effects of the negative drift g take hold of the dynamics). If there were no migration costs, population dynamics would exactly mirror the TFP dynamics (up to a factor of proportionality). However, in the data, the population dynamics are far from doing this. This is shown by the line with squares in Fig. 4, which displays the empirical impulse response of population to a TFP shock. We see that instead of following the (approximately) random-walk type of behavior displayed by TFP, population adjusts very little on impact, and grows monotonically over time. In fact, the population continues to grow nine years after the initial TFP shock. A crucial test for the model economy is if it is able to reproduce this type of sluggish population dynamics. The dotted line with circles in Fig. 4 displays the impulse response of population to a TFP shock estimated using synthetic data generated by the model. We see that the model population impulse response comfortably falls within the 95% confidence band of the empirical population impulse response. While we have used the empirical impact population response as a calibration target, it is important to note that

[37] Larger values for Λ would make a equal to zero over a range of contracting cities, introducing a flat portion to the relation between arrival rates and net population changes. As Fig. 3 has shown, this would be counterfactual.

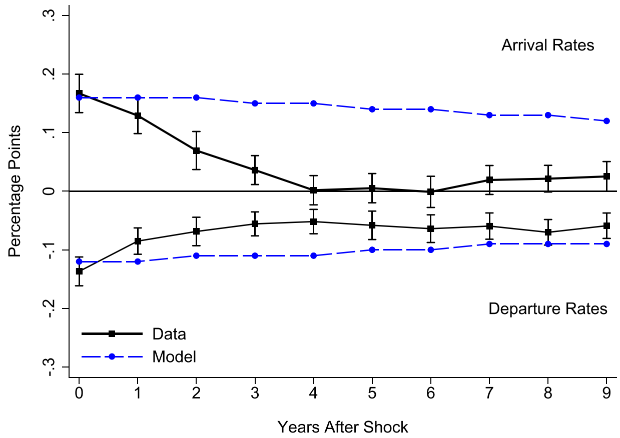

Fig. 5. Dynamic response of gross migration to a TFP shock. Note: The "Data" line with whiskers shows the point estimates of the dynamic response to a one standard deviation positive TFP shock accompanied by two-standard-error bands. The "Model" line is the dynamic response predicted by the model.

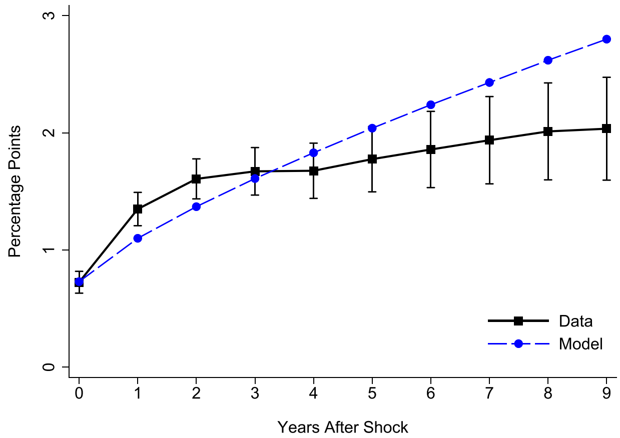

Fig. 6. Dynamic response of employment to a TFP shock. Note: The "Data" line with whiskers shows the point estimates of the dynamic response to a one standard deviation positive TFP shock accompanied by two-standard-error bands. The "Model" line is the dynamic response predicted by the model.

we have not targeted the population responses after the initial period. Thus Fig. 4 represents a successful test of the model economy.

Fig. 5 reports impulse responses for arrival and departure rates, both estimated from the data and the model economy. We see that in the data, following a TFP shock, arrivals rise while departures decline, arrivals rise more than departures fall, and the responses display significant persistence (especially for departures). In the model, the impact responses of arrivals and departures roughly match the impact responses of their empirical counterparts. However, the subsequent responses are more persistent in the model than in the data (especially for arrivals). Nevertheless, given its extreme simplicity the model seems to perform surprisingly well.

Fig. 6 shows the impulse responses for employment, both in the data and the model. We see that in the data, employment jumps on impact, but similar to population, its response is rather sluggish—after the initial year of the shock, employment continues to grow quite significantly. In the model, we see that the impact response of employment to the TFP shock coincides with the one observed in the data. However, the growth rate of employment after the initial response is somewhat more persistent than in the data. In fact, by the sixth year after the shock, employment in the model starts to slightly exceed the empirical confidence bands. Notwithstanding, in broad terms, employment in the model economy replicates reasonably well the sluggish response that we observe in the data.

The last impulse responses that we consider are those of wages. These are shown in Fig. 7, both for the data and the model economy. We see that in the data, wages jump on impact and continue to increase over time until they reach a peak level seven years after the initial TFP shock.[38] In the model, wages respond on impact as much as in the data and continue

[38] If workers were freely movable, wages would be equalized across all cities. Moreover, an idiosyncratic TFP shock would have no effect whatsoever on wages (only on population and employment). Thus, the persistent increase in wages after a TFP shock that we see in the data is indicative of substantial migration costs.

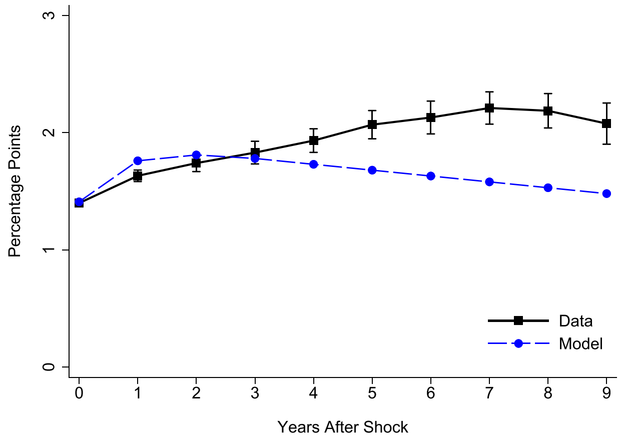

Fig. 7. Dynamic response of the wages to a TFP shock. Note: The “Data” line with whiskers shows the point estimates of the dynamic response to a one standard deviation positive TFP shock accompanied by two-standard-error bands. The “Model” line is the dynamic response predicted by the model.

to increase thereafter, but only for two years (after which, wages start to slowly decrease). That is, we see that the response of wages to the TFP shock is less persistent in the model than in the data. However, in levels, the response of wages in the model is close to what is observed.

Taken together, all these impulse response comparisons indicate that the empirical performance of the model is quite satisfactory, especially given the simplicity of the model (which, in reduced form, is summarized by the Bellman Eq. (41). Having established the empirical relevance of the model, we can now turn to answer the two questions that motivated the paper.

The first question—which is similar to the one asked in the early Real Business Cycle literature in the context of aggregate fluctuations—is this: How much of the variability in city-level population growth rates can be accounted by TFP shocks? Using the city-level population growth rates in our empirical panel of cities (which have been adjusted by removing time fixed effects), we compute a standard deviation of one-year growth rates equal to 1.23%. In contrast, using a large synthetic panel of cities drawn from the invariant distribution of our model economy, we compute a standard deviation of one-year growth rates equal to 1.09%. That is, TFP shocks account for 77% of the variance in one-year population growth rates. Over a time horizon of 10 years, the empirical standard deviation of population growth rates is 10.2%. TFP shocks account for essentially *all* of this variability.

The second question that motivated the paper is, to what extent can the most extreme cases of persistent urban decline be accounted by reductions in TFP? In order to answer this question we perform the following experiment. We rank all cities in our empirical panel in terms of their total population changes over the entire 1985–2013 period. We then focus on the 20 cities with the largest population declines. For this group of cities we average their TFP paths over the entire sample period and feed this path to the model economy. In particular, we endow a large synthetic panel of cities with a population level equal to one (the average population level in the model economy) and a TFP level equal to the midpoint in our computational grid, and feed these cities realizations of idiosyncratic TFP paths such that (in probability) they experience the same TFP path as the 20 cities with the largest population declines.[39]

In principle, we could address our second question of interest by comparing population paths over the entire 1985-2013 sample period. However, we believe that such a comparison would be flawed. The reason, is that since population responses to TFP shocks are sluggish (see Fig. 4), early on in the sample period the cities are affected by TFP shocks that took place well before the sample period. In order to avoid this source of noise, we split the sample period in two and compare population paths only during the second half. The results are reported in Fig. 8. The dotted line displays the average TFP path during 1998–2013 for the 20 cities that experienced the largest population declines over 1985–2013. We see that over this time period the average TFP level of these cities declined monotonically, resulting in a total accumulated decline of about 8%. The line with squares represents the average population path for these cities. We see that their average population level decreased at a fairly constant rate, resulting in a total decline of 20%. In contrast, the line with circles shows the average population path for our synthetic panel of cities. We see that the average population of these cities closely follows the empirical path until 2007, but starts to differ after that year. In particular, by 2013 the accumulated population decline reaches 24%. This slight over-prediction by the end of the sample is consistent with previous findings in Fig. 4—in particular, that the response of population to TFP shocks tends to be larger in the model than in the data after a few years.

This last feature of the model indicates that the quadratic adjustment costs to total population changes implied by the gross migration costs uncovered in this paper are only part of the story. The analysis suggests that other factors may also

[39] Since in computations we work with a grid of 25 values for the TFP levels, the realized path of TFP in the model is quite choppy. In order to reproduce the smooth empirical path of TFP levels for the 20 cities with the largest population declines, we can only do it in probability.

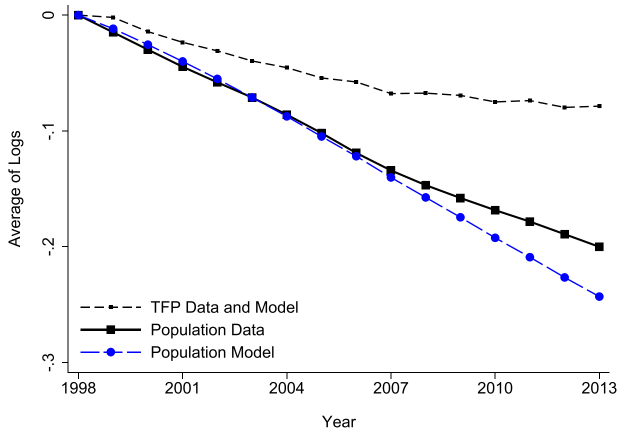

Fig. 8. Persistent urban decline. Note: The “TFP Data and Model” line corresponds to the path of average TFP of the 20 cities with the largest population declines in our data that we feed into the model to generate “Population Model.” The “Population Data” line corresponds to the average log population of the same 20 cities.

play a role in slowing down the adjustment of population to TFP changes, especially at long time horizons. Housing irreversibilities of the type emphasized by Glaeser and Gyourko (2005) seem a promising candidate.[40] Nevertheless, since our model misses the total population drop of the 20 cities with the largest population declines by such a small amount and since TFP shocks account for a large fraction of the total variability in population growth rates, our analysis suggests that TFP shocks represent a dominant source of urban dynamics.

8. Concluding remarks

In this paper we constructed a panel of 381 metropolitan statistical areas to investigate urban migration patterns, and found that within cities, gross migration rates are approximately linearly related to net migration rates. This linear relation has key implications for urban dynamics: In light of the model of gross migrations developed in this paper, consistency with that relation implies that cities face quadratic costs for adjusting their population levels. Our model is consistent not only with that linear relation, but with the the vast amount of cross-sectional heterogeneity in population, gross migration, and total factor productivity (TFP) found in the data. Measuring TFP shocks using our panel of cities, we found that the model reproduces quite well the empirical impulse responses of population, gross migration rates, employment, and wages to an idiosyncratic TFP shock. Given this important validation of the model, we then used it to measure the contribution of TFP shocks to urban population dynamics. We found that they account for the bulk of these dynamics, both in the short run and the long run.

In this paper we exclusively focused on TFP shocks. However, it would be extremely interesting to analyze other shocks, such as shocks to amenity levels, property taxes, local government expenditures, or local pension entitlements. If these shocks could be accurately measured their contribution to urban dynamics could be determined, similarly to what we did with TFP shocks. It could happen that some of these shocks together with TFP end up accounting for more than 100% of population volatility, in which case the presence of adjustment costs in addition to those considered in this paper would be strongly indicated. However, even if these alternative shocks end up being small and/or infrequent enough that they account for a small amount of regular urban dynamic, they could still generate dynamics that are quite relevant for understanding certain episodes observed. In either case, since our model of gross migrations is already able to reproduce the empirical impulse responses to TFP shocks, it should provide a useful building block for performing such type of analysis.

Having taken time fixed effects out of the gross migration rates obtained in our empirical panel, we have abstracted from the downward trend in gross migration rates that the previous literature has emphasized. In principle, our model could mechanically generate those observations by incorporating exogenous shifts to the left in the support of the distribution of idiosyncratic attachment shocks (i.e., a downward drift in ψ_1 towards some long-run value). However, this would leave open the question of how plausible those exogenous shifts are. To address this question, one would have to develop microeconomic foundations for the determinants of the idiosyncratic attachments to cities and evaluate how these may have changed over time (complementing other channels that the literature has already emphasized). For instance, one would have to explore how idiosyncratic attachments depend on agglomeration effects generated by productivity and amenity levels. We leave this for future research.

[40] Davis et al. (2013) studied a version of our model with housing. They find that housing does little to limit population adjustments, but this result likely depends on the implausible assumption that houses are infinitely divisible.

Finally, we have considered an economy with full insurance. A priori, this doesn't necessarily represent a serious limitation to our analysis: In many contexts, being able to save in a safe asset allows agents to self-insure to such an extent that an equilibrium with incomplete markets looks very close to an equilibrium with complete markets. Having said this, it would be extremely interesting to extend our model by introducing some realistic form of market incompleteness and see how our results may change. Since this would greatly complicate the analysis, we also leave it for future research.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.jedc.2021.104234.

References

Alvarez, F., Veracierto, M., 2012. Fixed-term employment contracts in an equilibrium search model. J. Econ. Theory 147 (5), 1725–1753.
Blanchard, O.J., Katz, L.F., 1992. Regional evolutions. Brook. Pap. Econ. Act. 1992 (1), 1–75.
Caliendo, L., Dvorkin, M., Parro, F., 2019. Trade and labor market dynamics: general equilibrium analysis of the China trade shock. Econometrica 87 (3), 741–835.
Cingano, F., Schivardi, F., 2004. Identifying the sources of productivity growth. J. Eur. Econ. Assoc. 2 (4), 720–742.
Coen-Pirani, D., 2010. Understanding gross worker flows across US states. J. Monet. Econ. 57, 769–784.
Davis, M., Fisher, J., Veracierto, M., 2013. Gross migration, housing and urban population dynamics. Chicago Fed working paper no. 2013-19.
Davis, M., Fisher, J., Whited, T., 2014. Macroeconomic implications of agglomeration. Econometrica 82 (2), 731–764.
Duranton, G., Puga, D., 2014. The growth of cities. In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth. In: Handbooks in Economics, Vol. 2. Elsevier Science, North-Holland, North-Holland, pp. 781–853.
Eeckhout, J., 2004. Gibrat's law for (all) cities. Am. Econ. Rev. 94 (5), 1429–1451.
Gabaix, X., 1999a. Zipf's Law and the growth of cities. Am. Econ. Rev. 89 (2), 129–132.
Gabaix, X., 1999b. Zipf's law for cities: an explanation. Q. J. Econ. 114 (3), 739–767.
Gabaix, X., Ioannides, Y.M., 2004. The evolution of city size distributions. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. In: Handbooks in Economics, Vol. 2. Elsevier Science, North-Holland, North-Holland, pp. 2341–2378.
Glaeser, E.L., Gyourko, J., 2005. Urban decline and durable housing. J. Polit. Econ. 113 (2), 345–375.
Gordon, G., Guerron-Quintana, P., 2019. On regional borrowing, default, and migration. Richmond Fed Working Paer No. 19-04.
Howard, G., 2020. The migration accelerator: labor mobility, housing, and aggregate demand. Am. Econ. J. Macroecon. 12 (4), 147–179.
Jordà, O., 2005. Estimation and inference of impulse responses by local projections. Am. Econ. Rev. 95, 161–182.
Kaplan, G., Schulhofer-Wohl, S., 2017. Understanding the long-run decline in interstate migration. Int. Econ. Rev. 58 (1), 57–94.
Karahan, F., Rhee, S., 2013. Population aging, migration spillovers, and the decline in interstate migration. FRB of New York Staff Report No. 605.
Karahan, F., Rhee, S., 2019. Geographical reallocation and unemployment during the great recession: the role of the housing bust. J. Econ. Dyn. Control 100, 47–69.
Kennan, J., Walker, J.R., 2011. The effect of expected income on individual migration decisions. Econometrica 79 (1), 211–251.
Molloy, R., Smith, C.L., Wozniak, A., 2011. Internal migration in the united states. J. Econ. Perspect. 25 (2), 1–42.
Nenov, P., 2015. Regional reallocation and housing markets in an equilibrium model of frictional migration. Rev. Econ. Dyn. 18 (4), 863–880.
Rappaport, J., 2004. Why are population flows so persistent? J. Urban Econ. 56, 554–580.
Rossi-Hansberg, E., Wright, M.L.J., 2007. Urban structure and growth. Rev. Econ. Stud. 74 (2), 597–624.
Saks, R.E., Wozniak, A., 2011. Labor reallocation over the business cycle: new evidence from internal migration. J. Labor Econ. 29 (4), 697–739.
Turek, J.H., 1985. The northeast in a national context: background trends in population, income, and employment. In: Richardson, H.W., Turek, J.H. (Eds.), Economic Prospects for the Northeast. Temple University Press, Philadelphia.
Van Nieuwerburgh, S., Weil, P.-O., 2010. Why has house price dispersion gone up? Rev. Econ. Stud. 77, 1567–1606.
Veracierto, M., 2016. Establishment dynamics, vacancies, and unemployment: a neoclassical approach. Int. Econ. Rev. 57, 1201–1236.