Migration and Urban Economic Dynamics*

Morris A. Davis
Rutgers University
morris.a.davis@rutgers.edu

Jonas D.M. Fisher
Federal Reserve Bank of Chicago
jfisher@frbchi.org

Marcelo Veracierto
Federal Reserve Bank of Chicago
mveracierto@frbchi.org

June 18, 2020

Abstract

We document that within cities gross migration is linearly related to net migration, there is vast heterogeneity in gross migration and productivity across cities, and gross migration is diminishing in city size. We build a parsimonious general equilibrium model of migration that accounts for these observations and use it to study the role of migration in urban economic dynamics. Because of the linearity in migration, competitive equilibrium urban dynamics can be found by solving the planning problem of a representative city with quadratic adjustment costs in population that depend on the migration slope. For plausible adjustment costs we find: (1) population, gross migration, employment, and wage dynamics after productivity shocks are close to those we estimate, and (2) slowly growing productivity and costly migration account for much of the persistent urban decline we observe.

JEL Classification Numbers: E0, O4, R0
Keywords: persistent urban decline, urban economic dynamics, gross migration, net migration

*Corresponding author: Jonas Fisher. 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 the those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

1. Introduction

Even with a secular trend toward lower rates of internal migration millions of American residents continue to migrate between cities every year. Net migration is the primary source of population changes in urban areas.[1] Therefore, any questions about the growth and decline of cities require an understanding of migration. How do cities respond to productivity changes or changes in taxes and amenities? What is the long term impact on cities of the secular decline of industries or firms' location decisions? We develop a parsimonious quantitative general equilibrium model of migration that can be used to address these and other questions in which the costs of migration are part of the answer. Then we use the model to investigate urban economic dynamics in response to productivity shocks.

We characterize urban gross and net migration in a panel of 381 US cities over the period 1985–2013. Our analysis uncovers a striking relationship between gross and net migration: within cities the rate of in-migration (and therefore also out-migration) is linear in the net migration rate with a common slope. This common linearity within cities exists alongside vast heterogeneity in population, gross migration, and productivity across cities. We also find that larger cities tend to have lower gross migration rates. A credible model of urban migration should be consistent with these basic facts.

Cities in our model are combinations of permanent attractiveness, for example due to geography and weather, productivity, and worker attachment. For parsimony we model worker attachment to a city as a reduced form, but our approach encapsulates forces that are intuitive. The idiosyncratic attachment of a person to a city should be related to its permanent attractiveness and productivity. For instance, people may be more attached to a city if it provides specific varieties of certain goods and services that are closer to their ideal preferences or provides a greater variety of job opportunities to match with their skills. Since larger cities provide a wider variety of goods, services, and jobs, and since the long-run average size of a city is endogenously related to its permanent attractiveness and productivity, we connect attachment to these permanent characteristics. This allows the model to reproduce the observed cross-sectional heterogeneity in population, gross migration,

[1] Turek (1985) is the only study we are aware of that quantifies the contribution of net internal migration relative to births, deaths and immigration. He finds net migration is the main source of regional variation in population growth in the US.

productivity, and the fact that larger cities tend to have lower rates of gross migration.

Our model's cities are comprised of firms and workers. Firms produce the economy's freely tradeable final good with capital and labor subject to city-wide idiosyncratic productivity shocks. Labor is locally supplied but can move at a cost between cities while capital is perfectly mobile. Labor supply within a city is endogenous because employing more of the existing population is a substitute for migration to accommodate higher labor demand. Blanchard and Katz (1992) identify this as one of the main channels through which a region's employment responds in the years immediately following a labor demand shock.

While our model captures the observed cross-sectional heterogeneity we show that for understanding urban economic dynamics it is sufficient to study a representative city. We are able to do this because the model is consistent with the common migration linearity. The remaining structure of the model and the linearity imply a reduced form in which equilibrium outcomes are found by solving a city planning problem in population subject to quadratic adjustment costs. The size of these costs depends on parameters that can be directly connected to the slope of the in-migration versus net migration relationship and the average costs of migration for those who move. The former we estimate with our data and the latter we obtain from the seminal micro study by Kennan and Walker (2011). In principle moving costs depend on the distribution of permanent types. We are able to reduce our analysis to that of a single representative city by showing that these moving costs in the model can be estimated from our migration data using a short list of unconditional moments.

We apply our model to address two sets of questions. First, How do the costs of migration influence a city's economic dynamics in response to idiosyncratic productivity shocks? We compare our model to estimates of the dynamic responses of a city's population, gross migration, employment and wages to a local productivity shock. Our model with empirically plausible adjustment costs comes close to replicating the empirical dynamics of the average city.

Second, we examine the model's implications for the evidence of highly persistent urban decline and cities growing faster than they shrink documented by Glaeser and Gyourko (2005). The model has considerable persistence in urban decline due to slowly declining productivity relative to other cities, although the model predicts declines 25-50 faster than in the data. Quadratic adjustment costs imply population growth is only skewed in the model if productivity growth is similarly skewed. We measure productivity to be close to symmetric and if anything negatively skewed.

This paper is related to several strands of the literature. Blanchard and Katz (1992), Glaeser, Scheinkman, and Shleifer (1995) and Howard (2019) are earlier empirical studies of migration at the city and regional levels. In particular, Blanchard and Katz (1992) establish that persistent regional labor demand shocks are accommodated over the long run through migration. We contribute to this literature by presenting new empirical evidence on the nature of gross and net migration within and across cities.

Using the migration slope in our measurement of migration costs is similar to the approach taken by Sahin, Song, Topa, and Violante (2014) in the context of firm-worker matching. They identify parameters in a hiring cost function using the slope of the linear relationship between the job filling and employment growth rates estimated at the firm-level by Davis, Faberman, and Haltiwanger (2013).

The magnitude of the migration slope we estimate implies that a city's population adjusts to shocks by movements in arrivals and departures in opposite directions. Such a negative correlation might seem surprising given Kennan and Walker (2011)'s finding that recent arrivals are more likely to leave a city than long-time residents and Coen-Pirani (2010)'s demonstration that this phenomenon can induce a positive correlation at the state-level. In our model a positive labor demand shock raises wages, making incumbents who are on the margin of leaving a city less likely to do so. At the same time the higher wages attract more migrants. Consistent with the operation of this mechanism, we estimate a positive productivity shock leads city's arrival rate to rise and departure rate to fall. This corroborates Howard (2019)'s finding that gross migration responds oppositely in response to the demand for a city's output.

Our modeling of migration shares some features with the literature. As in Kennan and Walker (2011), Gordon and Guerron-Quintana (2019) and others the decision to leave a city is driven by idiosyncratic shocks to workers' attachment to their current city. Location choices have a directed component as in Kennan and Walker (2011). Aside from realism, we include a directed component because we find that arrival rates are increasing in net migration. If migration was not at least in part directed then arrival rates would be independent of net migration. In our framework it is costly to direct migration. This captures the myriad ways workers find new cities to live and work, including via friends and family, professional networks, vacancy postings, head-hunters, cities' promotional advertising, and active recruiting by workers whose primary responsibility is some other productive activity.[2]

Coen-Pirani (2010) models gross worker flows among US states in a search and matching framework taken from studies of gross worker flows between firms. We observe stark differences with gross population flows between cities that leads us to a different approach to modeling migration. Rappaport (2004) studies a two-city equilibrium model with exogenously given quadratic costs of population adjustment. He does not quantify these costs. While the reduced form of our model similarly involves quadratic adjustment costs, we derive them from the underlying gross migration choices and are able to measure the magnitude of these costs by exploiting our linearity finding.

Other work that studies migration in a structural framework includes Gordon and Guerron-Quintana (2019), Karahan and Rhee (2019), Lloyd-Ellis and Head (2012) and Nenov (2015). These papers emphasize the interaction of migration and housing. We abstract from housing to isolate the role of migration in urban economic dynamics. Gordon and Guerron-Quintana (2019), Nenov (2015) and Karahan and Rhee (2019) model gross migration, but they do not address the linear relationship between gross and net migration.

2. Empirical evidence on urban migration

We characterize gross and net urban migration using an annual panel of 381 Metropolitan Statistical Areas (MSAs) for the period 1985 to 2013.[3] These data cover roughly 80% of the U.S. population. The Office of Management and Budget defines a MSA as "one or more adjacent counties or county equivalents that have at least one urban core area of at least 50,000 population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties." While they sometimes include

[2]Our model abstracts from the pervasive decline in gross migration studied by Karahan and Rhee (2014) and Kaplan and Schulhofer-Wohl (2015). The conclusion discusses how our model can be used to shed light on the determinants of the trend.

[3]All our data is described in Appendix A.

multiple legally incorporated municipal entities, for convenience we refer to our MSAs as cities. Cities are a natural unit of analysis to study internal migration as they represent geographically distinct labor markets.[4]

We calculate annual inter-city migration rates by aggregating county-level data published by the Internal Revenue Service to the city level.[5] 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 and the city's *net migration rate* is simply the difference between its arrival and departure rates, $n_{it} = a_{it} - l_{it}$.

To characterize gross and net inter-city migration we estimate the following relationship:

$$a_{it} = \phi_i + \beta n_{it} + T_t + \varepsilon_{it}, \tag{1}$$

where ϕ_i denotes a city fixed effect, T_t denotes a time effect, and ε_{it} is an error term. We include time effects since gross migration rates fluctuate over the business cycle and have been falling over our sample period.[6] We abstract from these dynamics. Including city effects allows us to identify the within-city relationship between gross and net migration and to quantify the extent of heterogeneity in gross migration across cities. As discussed by Kaplan and Schulhofer-Wohl (2015) there are idiosyncracies with the IRS migration data which suggest it is measured with error. This leads us to estimate (1) by 2SLS using log housing permits from the Census Bureau as an instrument for n_{it}. The F-statistic from the first stage regression is 246.5. We estimate $\beta = 0.782$ with clustered-by-city standard error equal to 0.046.[7]

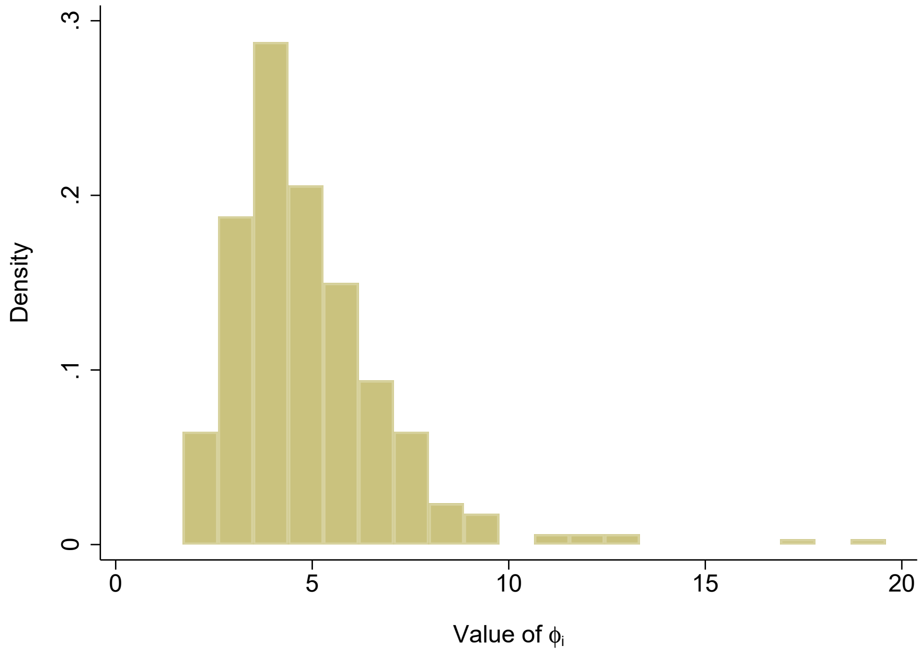
[4]MSAs are similar to The United States Department of Agriculture's market-oriented delineation of counties called Commuting Zones (CZs) which covers the universe of US counties while MSAs exclude rural counties. We chose to work with MSAs primarily because urban and rural areas might have fundamentally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

[5]The Current Population Survey and the American Community Survey are used widely to construct annual state-level migration rates, but small samples substantially limit their coverage of cities. Our key migration finding, in particular linearity, holds at the state level in all three data-sets.

[6]See Saks and Wozniak (2011) for evidence on the cyclicality of gross migration and Molloy, Smith, and Wozniak (2011) and Kaplan and Schulhofer-Wohl (2015) for evidence of its trend.

[7]The OLS estimate is 0.718. Since the measurement errors in the net migration and arrival rates are positively correlated the theoretical sign of the bias in the OLS estimate is ambiguous.

Figure 1: Gross migration varies greatly across cities

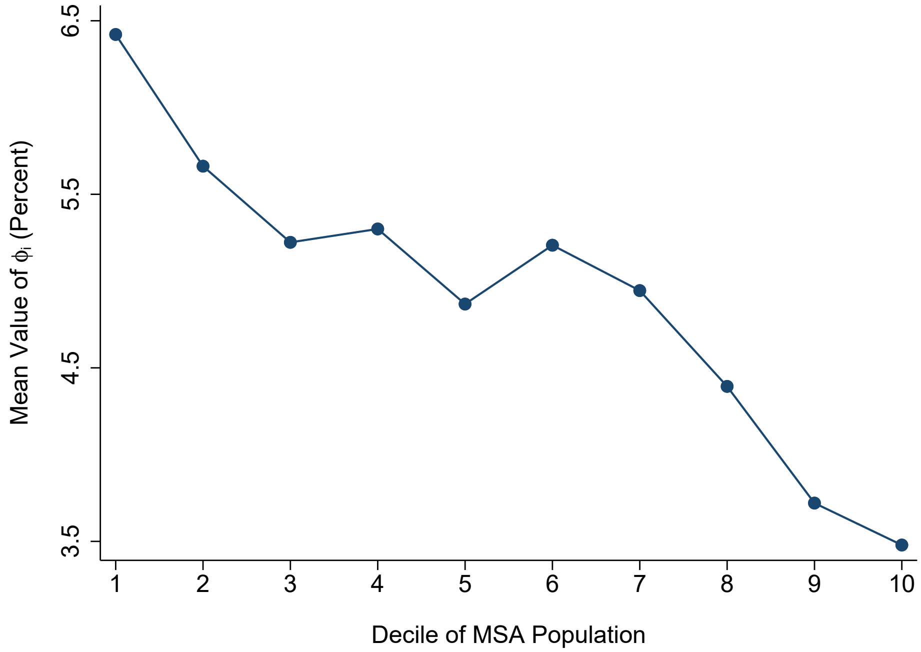


Note: Twenty bin histogram of city effects ϕ_i from equation (1).

Since the net migration rate is the difference between the arrival and departure rates the fact that our estimate of β lies between zero and one implies that within cities arrival and departure rates are negatively correlated. This stands in contrast to Coen-Pirani (2010) who argues that within-state arrival and departure rates are positively correlated. Coen-Pirani (2010) does not estimate a fixed effects regression and instead studies simple correlations after first conditioning on observable characteristics. We find a positive unconditional correlation in our raw migration data at both the city and state level. The mean gross migration rates are almost perfectly correlated with our estimated city effects and the city effects account for about 70% of the total variation in gross migration. This likely explains why our results differ from Coen-Pirani (2010)'s.

Figure 1 displays the histogram of the estimated city effects. Given the high correlation of the city effects with gross migration and their similar variances this figure illustrates the vast heterogeneity in gross migration among US cities. The fixed effects range from 2% to 20% with most of the mass in the range of 2-10%. Figure 2 shows that the heterogeneity in gross migration between cities is tightly connected to population. This figure shows the mean fixed effect within each decile of average city population. Generally gross migration

Figure 2: Gross migration is negatively related to population

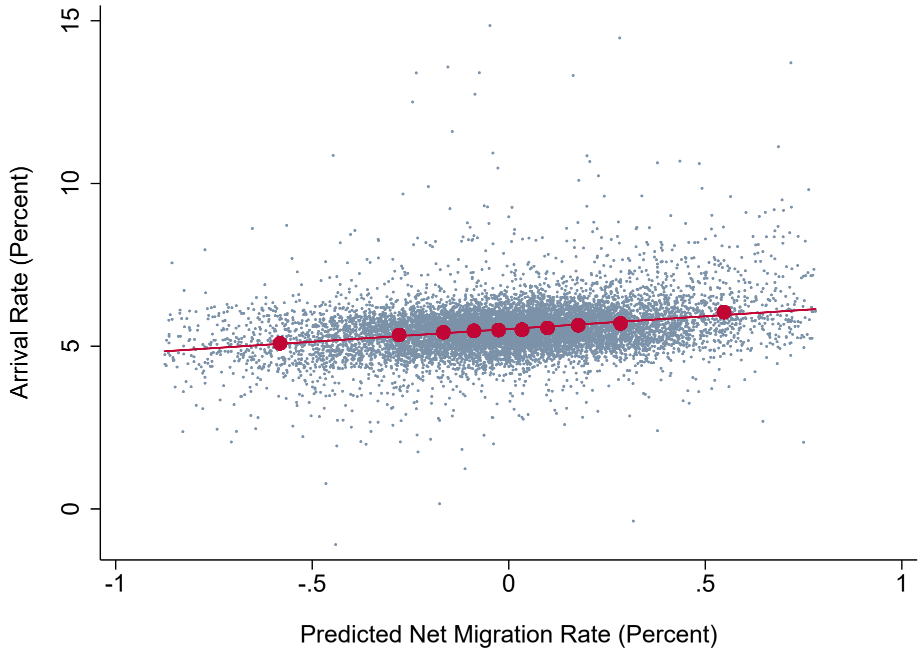


Note: Mean value of the city effects ϕ_i from equation (1) within each decile of cities' average population.

is smaller the larger the city. For example, the average fixed effect for the 10 smallest cities is 7.9% while the average fixed effect for the 10 largest cities is 2.8%. This close connection between city size and gross migration helps to explain the positive unconditional correlation between arrival and departure rates.

Despite the vast heterogeneity in gross migration between cities the within-city patterns of gross and net migration are strikingly similar. This is evident in Figure 3. The small (blue) circles show the within-city variation of the arrival rates and the net migration rates predicted by the first stage regression after removing city and time effects and adding back the unconditional mean arrival rate. The (red) line is the fitted regression line from the 2SLS regression and the large (red) circles correspond to mean arrival rates and mean net migration rates for each decile of net migration. The tight bunching of the small circles around the regression line and the large circles lining up along the regression line strongly suggest that cities' gross migration rates are the same linear function of their net migration rates up to a constant. This is reflected in the large within-city R^2 from the estimation of equation (1), .63. Within-city variation of predicted net migration accounts for nearly 10%

Figure 3: Gross migration is linear in net migration



Note: Small (blue) circles are the scatter plot of arrival rates and net migration rates predicted by housing permits, after removing city and time effects and adding back the unconditional mean arrival rate. The large (red) circles are the corresponding mean arrival rates for each decile of net migration. The (red) line is the fitted regression using all the observations. To conserve on white space the scatter plot excludes the one percent largest and smallest net migration rates.

of the total variation in arrival rates, which is more than double that for the time effects.[8] Overall Figure 3 validates the relationship we postulated in equation (1): Within cities gross migration is linearly related to net migration.

To summarize, our analysis of inter-city gross and net migration has revealed that gross migration rates vary greatly between cities, they are negatively related to city size, and within cities gross migration is essentially a linear function of net migration. A theory of urban dynamics should be consistent with these patterns.

3. The model economy

This section describes our model of urban dynamics and how it can replicate the empirical evidence presented above. While the model allows for many heterogeneous cities and

[8]The time trend in the IRS data is less pronounced than in the Current Population Survey for states so this relative variance contribution may be overstated. Note that the remaining variation in gross migration is accounted for by the regression residual (10.3%) and covariance terms (4.3%).

can replicate the cross-sectional evidence, we show that focusing on a representative city is sufficient for studying urban dynamics.

3.1. Environment

Time is infinite and discrete. The economy is populated by a representative household constituted by a measure P of individuals. At the beginning of every period, they are distributed in some way across a measure M of geographically distinct cities, where they must decide whether to stay or leave. While migrating to a new city takes place within the same time period, it is a costly activity. An agent can migrate through two possible channels. The first alternative is to do it individually. In this case, after payment of a fixed utility cost τ, the agent randomly draws the name of another household member and moves to the city where that person was located at the beginning of the period. We interpret these moves as taking place for personal reasons (e.g. to live close to family and friends).[9] Observe that as a consequence of this mechanism, larger cities will receive a proportionately larger fraction of the total number of agents moving for personal reasons.

The second alternative is to join an economy-wide pool of migrants, which is targeted by recruiters located in the destination cities. 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 migrants that a recruiter can attract depends on their own effort according to the following technology:

$$a_t = H^{-\upsilon} d_t^{\upsilon}, \tag{2}$$

where a_t is the number of attracted migrants, d_t is the recruiting effort measured in terms of lost utility, $H > 0$, and $0 < \upsilon < 1$.

There is a finite number of permanent city types, with λ^m denoting the total measure of cities of type m. Cities differ in terms of their permanent characteristics and certain temporary idiosyncratic characteristics, all of which we describe next. First, we assume that

[9] Alternatively, these moves can be interpreted as taking place for economic reasons but without having a job lined-up beforehand. In the absence of a direct hire, a person can always choose to move to another city at random and try their luck there. We assume that this random selection of cities is size-weighted (i.e. New York City is more likely to be selected than Ithaca, NY). Despite this alternative interpretation, for simplicity we refer to this type of move throughout the paper as happening for "personal reasons".

each city of type m provides a flow utility $\varsigma(m)$ to every agent living in that city after migration is completed. We interpret $\varsigma(m)$ to be the permanent attractiveness of the city, such as its weather and geographic location.

Second, we assume that cities differ in terms of their productivity levels. Specifically, we assume that firms in a city of type m can produce the economy's consumption good using the following production function:

$$y_t = \kappa(m) s_t n_t^{\alpha} k_t^{\gamma},$$

where y_t is output, $\kappa(m)$ is the permanent productivity level associated with the city's type m, 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, $\alpha > 0$, $\gamma > 0$, and $\alpha + \gamma < 1$.

Third, 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 uniform distribution on the interval $[-\psi_1(m), -\psi_1(m) + 2\psi_2]$, where $\psi_1(m) > 0$ and $\psi_2 > 0$. The uniform distribution helps to rationalize the linear relationship between gross and net migration we found in the data.

Observe that the support of this distribution is constant in width but the bounds depend on a city's permanent type m. The reason for doing this is that the idiosyncratic attachment of a person to a city could be related to its permanent attractiveness and productivity. For instance, agents may be more attached to a city if it provides specific varieties of certain goods and services (such as restaurants, music, and people) that are closer to their ideal preferences or provides a greater variety of job opportunities to match with their skills. Since larger cities provide a wider variety of goods, services, and jobs, and since 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. Notice with a larger $\psi_1(m)$, more individuals experience utility gains from staying put, and therefore, other things equal, more will choose to remain in the city.

In addition to deciding whether to stay or move, agents must decide whether to work their unit endowment of labor or not.[10] These decisions, which are made after migration, depend on their idiosyncratic disutility of working ω_t. We assume that ω_t is i.i.d. across individuals and over time, and that it is drawn from a distribution function with density $f(x) = Ax^{\theta}$ on the interval $\left[0, \left(\frac{\theta+1}{A}\right)^{\frac{1}{\theta+1}}\right]$, where $A > 0$ and $\theta > -1$.

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

$$u_t = U(c_t) + \varsigma_t - d_t - \xi_t \chi_t^l - \tau \chi_t^p - \omega_t \chi_t^n, \tag{3}$$

where c_t is their consumption, U is an increasing and strictly concave utility function, ς_t is the permanent attractiveness of the city of destination, d_t is the recruiting effort 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^p indicates if the agent performed a personal move, 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 $\bar{\beta}$.

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

$$K_{t+1} = (1-\delta) K_t + I_t, \tag{4}$$

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

[10] Allowing for endogenous labor supply is important for studying urban dynamics, since labor in a city can adjust to shocks via the extensive margin (population adjustments) and the intensive margin (employment adjustments).

3.2. Competitive equilibrium

In order to simplify the description of a competitive equilibrium we assume that all the insurance in the economy is done within the representative household.[11] As a result, independently of idiosyncratic histories, every agent in the household consumes the same. Also, we simplify the structure of the labor markets to a bare minimum. In particular, we assume that each city has its own recruitment market in which potential migrants and recruiters from that city meet to trade employment opportunities. Once a migrant purchases an employment opportunity (i.e. is recruited), they gain access to the city's spot labor market as long as they remain in the city.[12] Each agent in the economy-wide pool of migrants is free to enter any recruitment market. It is important to observe that, since the cities are heterogeneous, the price of an employment opportunity will generally differ across cities.

Cities at date t are indexed by a triplet (x_0, s_0, m) and by a history s^t, where x_0 is the total population of the city at the beginning of date 0, s_0 is its idiosyncratic productivity shock at date 0, m is its permanent type, and $s^t = (s_0, \ldots, s_t)$ is its history of productivity shocks between dates 0 and t.[13] We denote the date-t city type by $z^t = (s^t, x_0, s_0, m)$. A measure μ_t describes the total number of cities across the different z^t. Starting from the initial μ_0 at date 0, $\{\mu_t\}_{t=1}^{\infty}$ satisfies that $\mu_t\left(z^t\right) = \mu_{t-1}\left(z^{t-1}\right) Q(s_t, s_{t-1})$ for every date t and history z^t.

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, m) 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

[11] Alternatively we could assume that individual agents trade lotteries. Such a scenario would be much more complicated to describe but would lead to exactly the same equilibrium allocation as here.

[12] 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. In each city there would be a continuum of competitive markets 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 Alvarez and Veracierto (2012) and Veracierto (2016). Nevertheless, the equilibrium allocations would be exactly the same as here.

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

enters linearly in the flow utility (3) and since the recruitment technology (2) has decreasing returns to scale, 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 in city m at the beginning of the period and given an attachment threshold $\bar{\xi}$, the total attachment losses incurred by the household can then be written as

$$\left(\int_{-\psi_1(m)}^{\bar{\xi}} \xi \frac{1}{2\psi_2} d\xi\right) x = -\left[\psi_1\left(m\right)\lambda_l - \psi_2\lambda_l^2\right] x, \tag{5}$$

where λ_l is the fraction of agents that leave.

Third, since after migration has finished 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 in city m and given a labor-supply threshold $\bar{\omega}$, the total disutility of working incurred by the household can then be written as

$$\left(\int_0^{\bar{\omega}} \omega A\omega^{\theta} d\omega\right) p = B\lambda_e^{\nu} p \tag{6}$$

where $\nu = \frac{\theta+2}{\theta+1} > 1$, $B = A^{1-\nu}\left(\theta+1\right)^{\pi}\left(\theta+2\right)^{-1} > 0$, and λ_e is the fraction of agents that work.[14]

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

$$\max \sum_{t=0}^{\infty} \bar{\beta}^t \left\{U\left(c_t\right)P - \tau\Lambda_t\right.$$

$$+\sum_{z^t}\left[\varsigma\left(m\left(z^t\right)\right)\ p_t\left(z^t\right) - B\left(\frac{e_t\left(z^t\right)}{p_t\left(z^t\right)}\right)^{\nu} p_t\left(z^t\right) - H\left(\frac{a_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^{\frac{1}{\upsilon}} p_{t-1}\left(z^{t-1}\right)\right.$$

[14] Appendix B provides a detailed derivation of equations (5) and (6).

$$+\left[\psi_1\left(m\left(z^t\right)\right)\frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}-\psi_2\left(\frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^2\right]p_{t-1}\left(z^{t-1}\right)\right]\times\mu_t\left(z^t\right)\Bigg\} \quad (7)$$

subject to

$$p_t\left(z^t\right)=p_{t-1}\left(z^{t-1}\right)+b_t\left(z^t\right)+\frac{\Lambda_t}{P}p_{t-1}\left(z^{t-1}\right)-l_t\left(z^t\right), \quad (8)$$

$$\sum_{z^t}b_t\left(z^t\right)\mu_t\left(z^t\right)+\Lambda_t=\sum_{z^t}l_t\left(z^t\right)\mu_t\left(z^t\right), \quad (9)$$

$$c_tP+\sum_{z^t}q_t\left(z^t\right)b_t\left(z^t\right)\mu_t\left(z^t\right)+I_t$$

$$\leq \sum_{z^t}q_t\left(z^t\right)a_t\left(z^t\right)\mu_t\left(z^t\right)+\sum_{z^t}w_t\left(z^t\right)e_t\left(z^t\right)\mu_t\left(z^t\right)+r_tK_t+\Pi_t, \quad (10)$$

and equation (4), where Λ_t is the total number of agents that move for personal reasons, $m\left(z^t\right)$ describes the m-component of z^t, 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 employment opportunities sold, b_t is the total number of employment opportunities bought, q_t is the price of an employment opportunity, 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.[15] The optimal static decisions of the representative household are already incorporated in this decision problem.

The constraints begin with equation (8) which states that the end-of-period population level at a city with history z^t is given by the beginning-of-period population, plus the total number of agents that purchase employment opportunities in the city, plus the total number of agents that are attracted to the city through personal moves, minus the total number of agents that leave. Observe that Λ_t/P is the expected number of household members that each agent attracts through personal moves. Equation (9) states that the total number of agents leaving their cities either perform personal moves or purchase employment opportunities at some other city. Equation (10) is the budget constraint of the household. It states that total consumption, plus the total value of all employment opportunities purchased, plus investment cannot exceed the household's income. This includes the total value of all employment opportunities sold, the total value of salaries earned, the rental of capital, and the profits of firms.

[15] To simplify notation, we assume that $p_{t-1}\left(s^{t-1},x_0,s_0,m\right)$ at $t=0$ simply denotes x_0.

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

$$\pi_t\left(z^t\right) = \max\left\{\kappa\left(m\left(z^t\right)\right) s_t n_t\left(z^t\right)^{\alpha} k_t\left(z^t\right)^{\gamma} - w_t\left(z^t\right) n_t\left(z^t\right) - r_t k_t\left(z^t\right)\right\}.$$

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

$$\begin{aligned} n_t\left(z^t\right) &= e_t\left(z^t\right), && (11)\\ b_t\left(z^t\right) &= a_t\left(z^t\right), && (12)\\ K_t &= \sum_{z^t} k_t\left(z^t\right)\mu_t\left(z^t\right) && (13)\\ c_t P + I_t &= \sum_{z^t} \kappa\left(m\left(z^t\right)\right) s_t n_t\left(z^t\right)^{\alpha} k_t\left(z^t\right)^{\gamma} \mu_t\left(z^t\right), && (14) \end{aligned}$$

where equation (11) is market clearing in the labor market of each city, equation (12) is market clearing for the employment opportunities of each city, and equations (13) and (14) are the economy-wide market clearing conditions for capital and the consumption good, respectively.

3.3. 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 equation (7), subject to equations (8), (9), and (11)-(14). While this gets rid of prices, it still seems a complicated problem to solve. Fortunately, the model structure is such that the economy-wide social planning problem can be decomposed into a series of sub-planning problems (one for each city), together with 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 convenience, we describe it in recursive form.[16]

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

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

$$V(x,s,m) = \max \left\{ \varphi\kappa(m)\, sn^{\alpha}k^{\gamma} + \varsigma(m)\, p - B\left(\frac{n}{p}\right)^{\nu} p - H\left(\frac{a}{x}\right)^{\frac{1}{\upsilon}} x - \varphi\eta\left(a + \frac{\Lambda}{P}x\right) \right.$$

$$\left. + \varphi\eta l + \left[\psi_1(m)\frac{l}{x} - \psi_2\left(\frac{l}{x}\right)^2\right] x - \varphi rk + \bar{\beta}\sum_{s'} V(p, s', m)\, Q(s', s) \right\} \tag{15}$$

subject to

$$p = x + a + \frac{\Lambda}{P}x - l, \tag{16}$$

where φ is the shadow price of the consumption good, η is the shadow consumption price of a reallocated 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.

For a given permanent type m, the optimal decision rules to the city planner's problem generate an invariant distribution μ^m across duples (x, s) that satisfies the following equation:

$$\mu^m(B, s') = \int_{\{(x,s):\ p(x,s,m)\in B\}} Q(s', s)\, d\mu^m, \tag{17}$$

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

$$\begin{aligned}
\varphi &= U'(c), && (18)\\
cP + \delta\sum_m \lambda^m \int k(x,s,m)\, d\mu^m &= \sum_m \lambda^m \int \kappa(m)\, sn(x,s,m)^{\alpha} k(x,s,m)^{\gamma}\, d\mu^m, && (19)\\
\sum_m \lambda^m \int a(x,s,m)\, d\mu^m + \Lambda &= \sum_m \lambda^m \int l(x,s,m)\, d\mu^m, && (20)\\
\tau &= \sum_m \lambda^m \int \frac{1}{\upsilon} H\left(\frac{a(x,s,m)}{x}\right)^{\frac{1}{\upsilon}-1} \frac{x}{P} d\mu^m. && (21)
\end{aligned}$$

Equation (18) says that the shadow price of the consumption good equals the marginal utility of consumption. Equation (19) is consumption good feasibility. Equation (20) says that the total number of agents arriving to cities equals the total number leaving them. Equation (21) 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/P, 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 $\frac{1}{\upsilon}H\left(a/x\right)^{\frac{1}{\upsilon}-1}$. At the optimum, marginal costs are equated to marginal benefits.[17]

3.4. Consistency with empirical migration patterns

Assuming an interior solution, the city planner's first order conditions for a and l imply that

$$2\psi_2\frac{a}{x}+H\frac{1}{\upsilon}\left(\frac{a}{x}\right)^{\frac{1}{\upsilon}-1}=\psi_1\left(m\right)+2\psi_2\left(\frac{p-x}{x}\right)-2\psi_2\frac{\Lambda}{P}. \tag{22}$$

Thus, in general the model implies a non-linear relation between the arrival rate to a city $a/x+\Lambda/P$ and its net migration rate $\left(p-x\right)/x$. However, Section 2 showed strong evidence that in U.S. data this relation is linear. To make the model consistent with this empirical finding we must set $\upsilon=0.5$. In this case, equation (22) implies that the arrival rate is given by

$$\frac{a}{x}+\frac{\Lambda}{P}=\frac{\psi_1\left(m\right)}{2\left(\psi_2+H\right)}+\frac{H}{\left(\psi_2+H\right)}\frac{\Lambda}{P}+\frac{\psi_2}{\left(\psi_2+H\right)}\left(\frac{p-x}{x}\right), \tag{23}$$

and, using equation (16), that the departure rate is

$$\frac{l}{x}=\frac{\psi_1\left(m\right)}{2\left(\psi_2+H\right)}+\frac{H}{\left(\psi_2+H\right)}\frac{\Lambda}{P}-\frac{H}{\left(\psi_2+H\right)}\left(\frac{p-x}{x}\right). \tag{24}$$

Hereafter, the assumption of $\upsilon=0.5$ will always be made and, therefore, the linear relations in equations (23) and (24) will hold. It follows that the model can rationalize equation (1). The coefficient $\frac{\psi_2}{\psi_2+H}$ in (23) corresponds to the migration slope β in (1). If in addition we associate each city in our data with a permanent type m, then the fixed effects in (1) are

[17] Observe that the city planner's problem with side conditions provide the basis for a straightforward algorithm for computing a steady state: 1) for fixed values of (Λ,φ,η,r) solve the city planner's problems (15), 2) compute the associated invariant distributions μ^m in (17), and 3) verify that the side conditions (18)-(21) are satisfied. If they are not, go back to 1) with new values for (Λ,φ,η,r). See Appendix D for how we solve the model.

given by $\frac{\psi_1(m)}{2(\psi_2+H)} + \frac{H}{(\psi_2+H)}\frac{\Lambda}{P}$.[18]

Both the costly recruiting technology and the structure of idiosyncratic attachment shocks play a crucial role in delivering the linear relations (23) and (24). If the recruiting technology was not available (i.e. if $H = \infty$) then the arrival rate would be the same in all cities (it would be equal to the fraction of the population moving for personal reasons Λ/P) and all the net population adjustments would take place through changes in the departure rate. This would be highly counterfactual. On the other extreme, if it was completely free to recruit people (i.e. if $H = 0$) then the city's planner would always set the departures rate at the point of maximum total idiosyncratic benefits of moving $0.5 \cdot \psi_1(m) / (\psi_2 + H)$ and make all net population adjustments through the arrivals margin. This would also be highly counterfactual. Eliminating the idiosyncratic attachment costs (i.e. setting $\psi_1(m) = \psi_2 = 0$) would make the city's planner choose $a/x = 0$ (to avoid incurring any recruiting effort), and rely exclusively on the departures margin to adjust the city's population. Again, this would be highly counterfactual.[19]

Also, for gross migration in the model to be consistent with the empirical evidence of Section 2, the net migration rates $(p - x)/x$ cannot be too negative. The reason is that if $(p - x)/x$ was negative enough, city planners would set a/x to zero and rely on the departure rate l/x to make up for the difference. This means, that for sufficiently negative net migration the arrival rate become constant (and equal to Λ/P). Since Section 2 showed that the relation between arrival rates and net migration rates is monotone, this would be counterfactual. In what follows, we assume that net migration rates are always sufficiently large to make a/x positive.

Another empirical pattern uncovered in Section 2 is that larger cities tend to have lower arrival and departure rates (see Figure 2). Equation (23) and (24) imply that the variation of average gross migration rates across permanent city types m is directly determined by $\psi_1(m)$. This gives the model enough flexibility to make it fully consistent with Figure 2.

[18]The fact that we only have access to noisy measures of the migration variables justifies the error term in (1).

[19]In these arguments we implicitly assume that the net migration rates $(p - x)/x$ are sufficiently close to zero. If this was not the case, the gross migration rates described here would have to be modified somewhat to handle the constraint that a/x and l/x cannot be negative. However, the gross migration rates would still be highly counterfactual. See Appendix C.6 for the details.

To see this, associate each city in our data with a permanent type m and assume that the values of H and Λ/P are somehow given to us. Observe that the empirical estimate obtained in Section 2 for the slope in equation (1) corresponds to the slope $\psi_2/(\psi_2 + H)$ in equation (23). Given H, this estimate would then determine the value of ψ_2. The empirical estimate for the fixed effect of city m in equation (1) would then determine $\psi_1(m)$. Conditioning on the average TFP level $\kappa(m)$ of city m (to be measured in Section 4) and given other model parameters, the city planner's problem could then be used to solve for the permanent attractiveness $\varsigma(m)$ needed to generate the average size of city m observed in the data. Doing this for each city m would guarantee that Figure 3 is obtained. Thus our model can simultaneously account for the three main findings of Section 2. It also suggests a natural source of the negative relationship between population and gross migration: larger cities have lower migration rates because current residents have fewer reasons to leave.

3.5. Using the model to study urban economic dynamics

Substituting the linear gross migration rules (23) and (24) into equation (15), and maximizing with respect to n and k, reduces the city planner's problem to the following:

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

$$+ \Phi\left(m, \frac{\Lambda}{P}\right)x + \Gamma\left(m, \frac{\Lambda}{P}\right)(p - x) - \frac{H\psi_2}{H + \psi_2}\left(\frac{p - x}{x}\right)^2 x + \bar{\beta}\sum_{s'} V(p, s', m)\,Q(s', s)\Bigg\} \tag{25}$$

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

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

Writing the city planner's problem as in equation (25) is extremely informative and highlights three fundamental issues raised in the paper.[20] First, the linearity of the gross migration rules, which is strongly indicated by the empirical evidence, implies quadratic

[20]See Appendix C.7 and Appendix C.8 for the derivation of equation (25).

adjustment costs to population changes $(p-x)/x$. This emphasizes the crucial importance of explicitly modeling the gross migration process in order to understand urban dynamics, since it has direct implications for it.

Second, the linear gross migration rule estimated from the data not only informs on the qualitative nature of the population adjustment costs at the city level, but on their magnitude. The coefficient $H\psi_2/(H+\psi_2)$ on the quadratic adjustment term in equation (25) is equal to H times the slope of the linear arrival rate rule, equation (23). We obtained an estimate of this slope from Section 2.

Third, given H and the estimated slope of arrival rate rule, all the cross-sectional heterogeneity due to differences in permanent types m is irrelevant for understanding city dynamics: The coefficient $H\psi_2/(H+\psi_2)$ on the quadratic adjustment term in (25) is common to all permanent types m. This suggests the possibility of simplifying the analysis of urban dynamics in a crucial way. If we could somehow calibrate H without having to solve the full-blown model with heterogeneous permanent types m, there would be no need to ever consider this version of the model. We could safely proceed with our analysis of urban dynamics by assuming that all cities have exactly the same permanent type m. The following section shows how we can do this within a comprehensive calibration strategy.

4. Calibration

We now calibrate the steady state competitive equilibrium to U.S. data. In addition to specifying the permanent characteristics of cities, the stochastic process for idiosyncratic productivity and the utility function we need to calibrate the following 9 parameters: H, ψ_2, τ, γ, α, δ, $\bar{\beta}$, B, and ν. These include the migration parameters, the factor shares in production, the depreciation rate, the discount factor, and the parameters governing labor supply. We first discuss the migration parameters and then the parameters governing production and preferences.

4.1. Migration Parameters

We calibrate the migration parameters H and ψ_2 using the migration slope and by matching a micro estimate of the ratio of average net cost of migration of those who move to average wages from Kennan and Walker (2011).[21] They identify moving costs using a detailed structural model of individual choice and data on moves. We cannot replicate their estimation strategy in our model, but we can calculate the average moving costs of those who move. We match this to the Kennan-Walker statistic.[22]

In our model economy the average moving cost of agents who move is given by

$$J_c = \frac{c\sum_m \lambda_m \int \left(-\psi_1(m)\frac{l(x,s,m)}{x} + \psi_2\left(\frac{l(x,s,m)}{x}\right)^2\right) x d\mu^m}{\sum_m \lambda_m \int l(x,s,m) d\mu^m}$$

$$+\frac{\sum_m \lambda_m \int q(x,s,m)a(x,s,m)d\mu^m + c\tau\Lambda}{\sum_m \lambda_m \int a(x,s,m)d\mu^m + \Lambda}.$$

The first term is the average idiosyncratic attachment costs incurred at the cities of origin, while the second term is the average recruitment costs and the disutility of personal moves incurred at the cities of destination (all expressed in consumption units).

We show in Appendix F that since the equilibrium price of a recruitment opportunity is $q(x,s,m) = 2cHa(x,s,m)/x$ and equation (21) holds, $J_c/(c\psi_2)$ can be written in terms of the fraction of the total population doing personal moves Λ/P, the slope $\psi_2/(\psi_2 + H)$ in equation (23), the average size of cities P/M, and five unconditional moments that we can estimate with our data. In particular,

$$\begin{aligned}\frac{J_c}{c\psi_2} \quad = \quad & \frac{2\frac{H}{\psi_2}E\frac{a(x,s,m)^2}{x}}{Ea(x,s,m) + \frac{\Lambda}{P}\frac{P}{M}} + \frac{2\frac{\Lambda}{P}\frac{H}{\psi_2}Ea(x,s,m)}{Ea(x,s,m) + \frac{\Lambda}{P}\frac{P}{M}} \\ & - \frac{\frac{2}{\left(\frac{\psi_2}{\psi_2+H}\right)}E\left[\phi(m)\, l(x,s,m)\right]}{El(x,s,m)} + 2\frac{\frac{\Lambda}{P}}{\left(\frac{\psi_2}{\psi_2+H}\right)} - 2\frac{\Lambda}{P} + \frac{E\frac{l(x,s,m)^2}{x}}{El(x,s,m)}\end{aligned}$$

where $\phi(m)$ denotes the city-type-specific intercept term in equation (23) and $M = \sum \lambda^m$. As we previously noted we can use the estimated slope in equation (1) to obtain a value

[21] Their estimates imply a ratio of -1.9. The numerator is the entry in the row and columns titled 'Total' in Table V and the denominator is the wage income of the median AFQT scorer aged 30 in 1989 reported in Table III. The negative value of the estimate indicates that individuals receive benefits to induce them to move.

[22] Kennan and Walker (2011)'s estimates are based on the frequency of inter-*state* moves, while our model describes inter-*city* moves. It is possible that they would have estimated different moving costs with inter-city data as these moves are more frequent. In Appendix E we study a calibrated variant of Kennan and Walker (2011)'s model and find that any bias is likely to be small.

for $\psi_2/(\psi_2+H)$. The estimated fixed effects in (1) provide values for the intercepts $\phi(m)$ assuming we associate each city in our data with a permanent type. We can also use our data to obtain P/M, l, and a. P and l can be measured directly. We can obtain a from the total arrivals to a city $a+x\Lambda/P$ and its previous population level x, by conditioning on some value for Λ/P. Thus, conditional on Λ/P, $J_c/(c\psi_2)$ can be estimated using our data.

Average wages J_w satisfy

$$\frac{J_w}{c}=\frac{\alpha}{\frac{N}{P}\cdot\frac{cP}{Y}}.$$

J_w/c can be obtained from the values for the labor share α, the employment-population ratio N/P and the consumption to output ratio cP/Y that we discuss below.

The ratio of average moving costs to average wages $J=J_c/J_w$ satisfies

$$J=\psi_2\cdot\frac{J_c}{c\psi_2}\cdot\frac{1}{\frac{J_w}{c}}. \tag{26}$$

Given our empirical measures for $J_c/(c\psi_2)$ and J_w/c and associating J with the Kennan-Walker statistic, equation (26) gives a calibrated value for ψ_2. Our estimated value for the slope $\psi_2/(\psi_2+H)$ then implies a calibrated value for H. The associated value of τ can then be obtained from equation (21).

While our calibration of moving costs is conditional on Λ/P, it turns out that within a reasonable range its actual value has negligible effects on the calibration. For this reason, in what follows we only consider results for $\Lambda/P=0.003$, which is the lowest arrival rate in our sample (i.e. it is the largest value of Λ/P consistent with having a positive value for a in every city). The associated values for H, ψ_2 and τ are 13.5, 48.5 and 0.0026, respectively.[23]

Since we are able to calibrate H without having to solve the full-blown model with heterogenous permanent types m, we can follow the suggestion for simplifying the analysis of urban dynamics described at the end of Section 3.5. Therefore we work with a version of the model in which there is only one permanent city type m. We normalize the permanent productivity $\kappa(m)$ to one, the permanent attractiveness $\varsigma(m)$ to zero, and set $\psi_1(m)=6.615$ to generate an average arrival rate (and departure rate) of 5.4%, which is the average gross

[23]The "reasonable range" of values for Λ/P that we have referred to has 0.003 as its largest value. Its lowest value is 0. The values for H, ψ_2 and τ associated with $\Lambda/P=0$ are virtually identical to their benchmark values.

migration rate in our data.

4.2. Production and preference parameters

In selecting the remaining parameter values we 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 model 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.[24] Calibrating to an annual interest rate of 4 percent requires a time discount factor $\bar{\beta}$ equal to 0.96. In steady state capital's income share in the model satisfies

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

We use this equation to calibrate $\gamma = 0.175$ using our calibrated values for $\bar{\beta}$ and δ and an average capital-output ratio of 0.755. The parameter α is chosen to reproduce a labor share of 0.64. Note that with δ and K/Y in hand we can obtain the consumption-output ratio that we used to calibrate the migration costs above.

We measure city-level TFPs using data on wages and employment, and 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, while we are calibrating a deterministic steady 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

[24] All statistics mentioned in this section are based on annual data between 1985 and 2013 (the same time period used in Section 2).

abstract from trend and cycle. In particular, for any variable x_{it} in city i at date t we define

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

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 and 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 (such as permanent productivity levels $\kappa(m)$).[25] 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:

$$\Delta \hat{s}_{i,t} = \hat{\rho}\Delta \hat{s}_{i,t-1} + e_{i,t}. \tag{27}$$

Our OLS estimate of $\hat{\rho}$ is a tightly estimated 0.272, while our estimate of the standard deviation of $e_{i,t}$ is $\sigma_e = 0.0129$.

While it is reassuring that the stochastic process for the growth rates $\Delta \hat{s}_{it}$ is tightly estimated, it implies a non-stationary stochastic process for the idiosyncratic productivity levels. This is problemetic for our theoretical model, since 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:

$$\ln s_{t+1} = \max\left\{g + (1+\rho)\ln s_t - \rho \ln s_{t-1} + \varepsilon_{t+1}, \ln s_{\min}\right\}, \tag{28}$$

where $\varepsilon_{t+1} \sim N(0, \sigma_\varepsilon)$, $g < 0$ and $\rho > 0$.[26] Observe that with this stochastic process, TFP growth rates are approximately AR(1), while TFP levels are stationary due to the negative

[25] Measuring fixed effects as the time series means of $\hat{s}_{it}$ we find considerable heterogeneity in productivity much like with gross migration. This is not surprising given that our measure of a city's TFP is directly related to its wages and employment.

[26] This process is similar to the one used by Gabaix (1999), except that he assumes $\rho = 0$.

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 equation (28), normalize $\ln s_{\min}$ to zero, and select values for ρ and σ_{ε} so that when we run the regression given by equation (27) on the values of s_t obtained from simulating the stochastic process (28), we reproduce the empirical estimates of $\hat{\rho}$ and $\hat{\sigma_e}$. Observe that selecting a very negative value for g would make cities reflect the barrier $\ln s_{\min}$ often and would create very little heterogeneity in long-run growth rates, while selecting a value of g close to zero would do the opposite. Therefore, we choose the negative drift g to reproduce a standard deviation of ten-year population growth rates equal to 0.10 from our empirical panel of cities. The resulting value is $g = -0.0017$.

We assume that the utility function $U(c)$ is logarithmic. The first-order condition for employment e_t in the representative household problem (7) is then given by

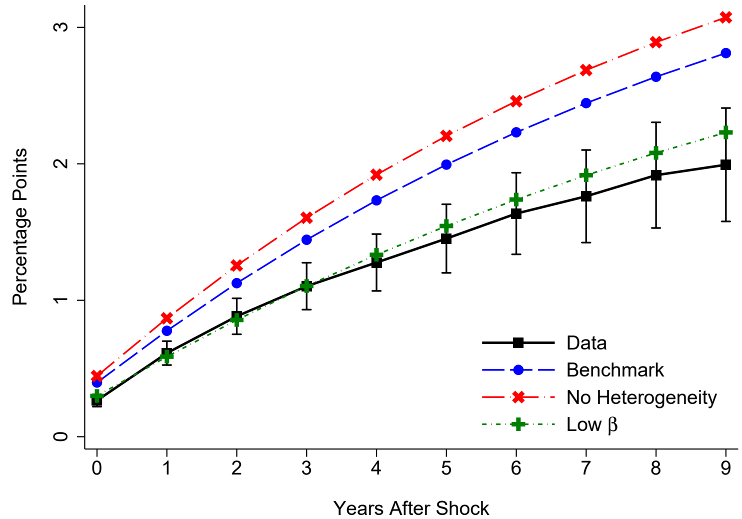
$$w_t = cB\pi \left(\frac{e_t}{p_t}\right)^{\nu-1}. \tag{29}$$

Thus, $1/(\nu - 1)$ is the Frisch labor supply elasticity of the representative household. Estimating the impulse responses for the log of w_t and the log of e_t/p_t to a one-standard deviation innovation in city-level TFP, we select $\nu = 3.7$ to reproduce the ratio of those responses on impact. (Our estimation of impulse responses is described in detail in Section 5.) In turn, the labor supply disutility B is selected to reproduce an aggregate employment to population ratio equal to 0.62.

5. Results

We now apply our model to study the role of migration in urban economic dynamics. First, we compare our model to estimates of the dynamic responses of population, gross migration, employment and wages to TFP shocks. Second, we examine the extent to which migration and the evolution of a city's productivity can account for two facts documented by Glaeser and Gyourko (2005): persistent urban decline and cities grow faster than they shrink.

Figure 4: Dynamic response of population to a TFP shock



Note: The 'Data' line with whiskers shows the point estimates with 2 standard error bands. The 'Benchmark' line corresponds to the calibration discussed in the previous section. The 'No heterogeneity' line comes from calibrating the model under the assumption that all cities are of the same permanent type. The 'Low β' line is based on calibrating using the lower bound of the confidence interval for the migration slope β rather than the point estimate.

5.1. Dynamic responses to TFP Shocks

We use local projections to estimate the dynamic response of a variable to a TFP shock. Specifically, the responses of the non-stationary variables – population, employment and wages – j years after a TFP shock are identified with the coefficients from linear projections of log differences of the variables between date $t + j$ and date $t - 1$ on the date t values of the residuals from equation (27) determined by our estimated value of $\hat{\rho}$. For the arrival and departure rates we use the coefficients from projections of the levels of the rates at $t + j$ on the date t residuals. Standard errors are clustered by city.

Figure 4 shows our estimated response of population and responses in the model under different calibrations. Equation (27) implies that after a shock TFP rises rather quickly to its new long run level; after 4 years TFP is already within 0.5% of it.[27]

If there were no costs to migration population would track TFP and therefore reach its long run level almost immediately. Yet the empirical response of population is very slow

[27]The discussion loosely identifies the long run level of TFP with its peak response, while equation (28) indicates that the long run TFP level must be $\ln s_m in$. The reason for taking this liberty is that the drift in equation (28) is very small, so TFP stays close to its peak response for a long time.

and is still rising 9 years after the shock. The response of population from the benchmark calibration is also slow, with a half life of over 10 years. However it is still very much faster than the data and lies far outside the confidence bounds of our point estimates.

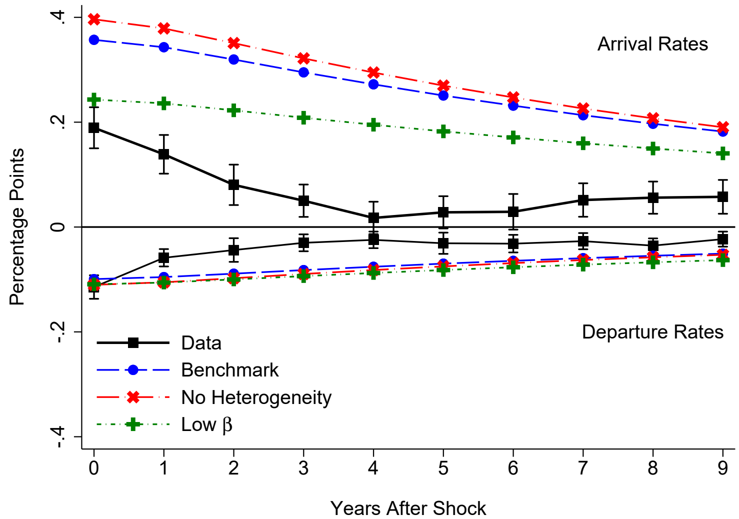
We have explored some alternative calibrations of the migration parameters to address the robustness of this finding. When considering a different value for a given migration parameter we re-calibrate the other migration parameters using the benchmark calibration procedure. The speed of population adjustment in the model is increasing in the targeted value of the migration slope β from equation (1) and this slope is estimated with uncertainty.[28] Therefore we considered the 'Low β' case in which we calibrate to the value of the slope at the lower end of its 95% confidence interval. As shown in Figure 4 the model's population response is attenuated substantially and now lies well within the confidence bands for the estimated responses. From this perspective the model does well at accounting for the data.

We also calibrated the migration parameters holding β at its benchmark value but dropping the Kennan-Walker statistic from the calibration and increasing the recruiting parameter H until we matched the impact period population response in the model to the data. When we do this the model's population response lies virtually on top of the estimated response (not shown). However, the model-implied Kennan-Walker statistic is fifty percent larger.

Our calibration strategy takes into account cross-sectional heterogeneity in gross migration without having to solve the full model with heterogeneity. How sensitive are our findings to accounting for this heterogeneity? The final experiment we ran was to calibrate the migration parameters under the assumption that there is only one type of city. In this 'No Heterogeneity' case we ignore the empirical moments underlying the benchmark calibration and simply adjust the recruiting parameter H to match the model-implied Kennan-Walker

[28] The speed of adjustment is determined by the quadratic adjustment cost coefficient $H \cdot \beta = \frac{H}{1+H/\psi_2}$ in the city planner's reduced form problem. When we change β, our calibration of H changes to match the Kennan-Walker statistic. In fact, when we increase β, our calibrated value of H decreases so much that $H \cdot \beta$ decreases and population adjustments become less costly. To get intuition for this, consider what happens when $\beta \to 1$. This requires that $H/\psi_2 \to 0$. If H converges to some positive value, then ψ_2 would have to go to infinity. However, the expression for J_c/c indicates that this would go to infinity as well. Since J_w/c is pinned down by data, this would mean that the Kennan-Walker statistic $J = J_c/J_w$ would go to infinity and we would be missing our calibration target. Thus, $H \to 0$ when $\beta \to 1$ and so $H \cdot \beta \to 0$. Therefore, as $\beta \to 1$ and we follow our calibration strategy adjustment costs go to 0.

Figure 5: Dynamic response of gross migration to a TFP shock



Note: Note: The 'Data' line with whiskers shows the point estimates with 2 standard error bands. The 'Benchmark' line corresponds to the calibration discussed in the previous section. The 'No heterogeneity' line comes from calibrating the model under the assumption that all cities are of the same permanent type. The 'Low β' line is based on calibrating using the lower bound of the confidence interval for the migration slope β rather than the point estimate.

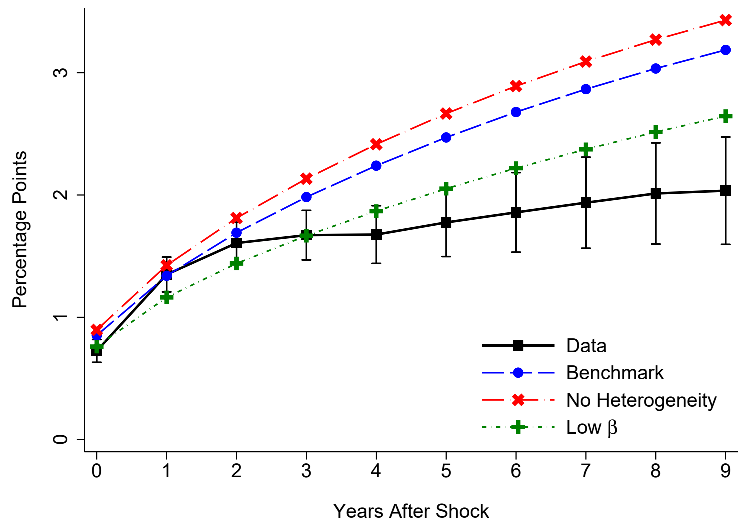
statistic to its estimated counterpart. Ex ante it was not clear how accounting for heterogeneity would influence our findings. As shown in Figure 4 the impact was to move the model closer to the data.

The population responses arise from variation in gross migration.[29] Figure 5 shows the estimated responses of gross migration along with the corresponding responses in the model under the same calibrations underlying Figure 4. The figure shows that regardless of the calibration the model replicates the key qualitative features of the empirical responses: Arrivals rise while departures decline, arrivals rise by more than departures fall, and both responses are very persistent. However the model predicts more pronounced responses than in the data, particularly for arrivals. Nevertheless given its simplicity the model does surprisingly well.

Population adjusts as workers move or stay put to take advantage of the higher wages that result from more efficient production opportunities. However employment can rise even if population does not through more intensive use of the existing workforce. Figure 6 shows

[29]Empirically international migration and changes in births and deaths play a role as well. Our analysis abstracts from these sources of population adjustments because they are likely to have second order effects.

Figure 6: Dynamic response of employment to a TFP shock

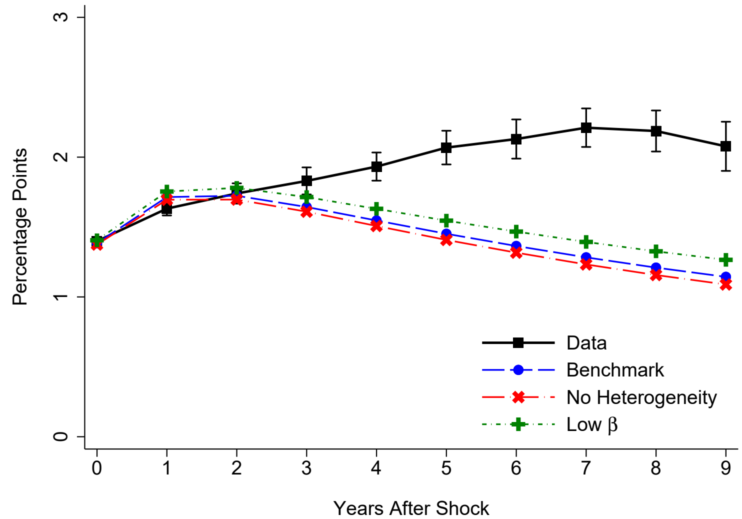


Note: Note: The 'Data' line with whiskers shows the point estimates with 2 standard error bands. The 'Benchmark' line corresponds to the calibration discussed in the previous section. The 'No heterogeneity' line comes from calibrating the model under the assumption that all cities are of the same permanent type. The 'Low β' line is based on calibrating using the lower bound of the confidence interval for the migration slope β rather than the point estimate.

the empirical response of employment is initially much faster than for population but they end up increasing at the same rate 9 years out. So initially employment responds in part via a higher employment rate and later on entirely through the addition of workers to the city. The model's endogenous labor supply allows it to capture these features of the data. The 'Low β' response is within the confidence band initially but does not tail off like the data and so the employment rate remains elevated after 9 years.

The response of wages to a positive idiosyncratic TFP shock is shown in Figure 7. If workers were perfectly mobile like capital then wages would be equalized across cities and wages would not respond to a productivity shock. Thus their response is informative about labor scarcity due to migration and other frictions. Empirically the wage response is statistically significant and persistently positive throughout, rising by more than population before eventually tailing of in year 8. The labor scarcity is persistent. The labor supply elasticity parameter ν was calibrated to match the empirical wage response in the initial period. Thereafter it rises like the data initially but tails off much earlier. Equation (29) shows the wage response in the model is governed by the employment to population ratio.

Figure 7: Dynamic response of the wages to a TFP shock



Note: Note: The 'Data' line with whiskers shows the point estimates with 2 standard error bands. The 'Benchmark' line corresponds to the calibration discussed in the previous section. The 'No heterogeneity' line comes from calibrating the model under the assumption that all cities are of the same permanent type. The 'Low β' line is based on calibrating using the lower bound of the confidence interval for the migration slope β rather than the point estimate.

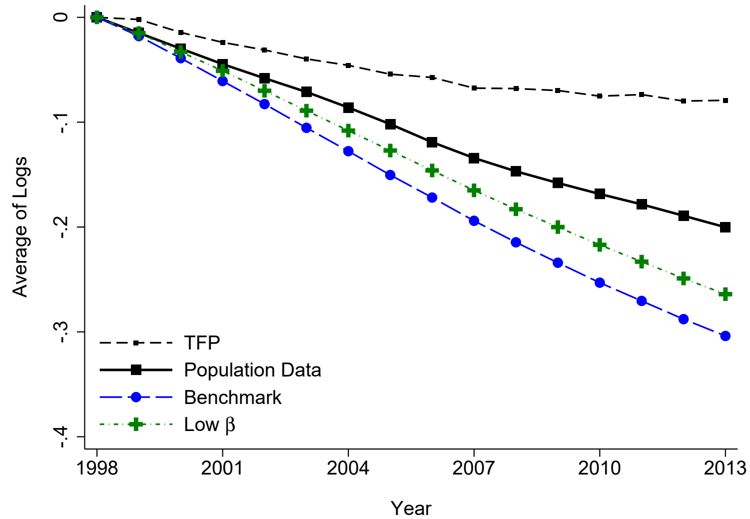
The responses of this variable in the model and the data can be read off of Figure's 4 and 6. In the data the employment to population ratio rises and follows a hump-shape. The model gets the level and trajectory of this response about right but it is more persistent than in the data (not shown).

Overall these economic dynamics illustrate the operation of the key mechanisms in the model. Having multiple margins to respond to the increase in TFP, agents exploit them all. Labor supply of the existing population can increase and the workforce can increase through migration. For the latter, given the better wages there is less incentive for people to move out of the city and the higher wages make it easier to attract workers to the city from elsewhere. The similarity of the responses in the model compared to the data suggests these margins are prominent empirically as well.

5.2. Migration, Urban Decline, and Asymmetric Growth

Consistent with Glaeser and Gyourko (2005) in our data some cities experience highly persistent decline and growing cities tend to grow faster than shrinking cities shrink. We now

Figure 8: Persistent urban decline



Note: The figure displays averages of log variables of 20 cities with the largest population declines over 1985-2013. The 'TFP' and 'Population' lines are averages from the data. The 'Benchmark' line is averages of population implied by the model with the TFP paths from the data under the benchmark calibration. The 'Low β' line is the analogous population path based on calibrating the migration slope using the lower bound of its confidence interval rather than its point estimate.

examine the extent to which our model is consistent with this evidence.

To address the urban decline question we focus on the 20 cities that experience the largest population declines in our sample. Their average TFP path is fed into the model starting from an initial TFP level assumed to have held for a long time.[30] We then feed into the model the complete TFP path starting in 1985, but focus on the resulting average population dynamics starting in 1998. The reason for doing this is that with adjustment costs, population adjustments depend on past TFP shocks.[31] For each city we calculate the predicted path for log population starting in 1998, average over these paths, and compare the result to the same object constructed using the data.

Figure 8 shows the average log paths for TFP and population for the data and the model. TFP falls by 0.08 log units from 1998-2013 and population falls by about 0.2 log points, along a path that starts to flatten after 10 years. The benchmark path for population flattens only late and falls by 0.3 overall, 50 percent faster than the data. With 'Low β' the difference

[30]The initial TFP level is selected to be the smallest value consistent with the TFP path not hitting the reflecting barrier.

[31]We can construct TFP going back to 1970. Adding the additional data does not change the estimated idiosyncratic TFP process but we have yet to explore it in this experiment.

is 25 percent and only 15 percent after 10 years. So the model does predict very persistent population declines but these are (somewhat) faster than in the data.

There are two key factors driving the model dynamics: persistent declines in TFP taken from the data and the slow response of population to declines in TFP prior to 1998. The impact of past TFP declines on current population is demonstrated by the faster rate of population decline relative to TFP – in the short run population's response to a TFP shock is much smaller than TFP's, but over longer horizons it responds by much more.

Glaeser and Gyourko (2005) document that population growth is positively skewed so that growing cities tend to grow faster than shrinking cities shrink. In our model of quadratic adjustment costs the only way this can happen is if productivity is similarly skewed. Productivity in the model, other than the small drift term, follows a symmetric process, but our measurement of productivity does not depend on this assumption. Therefore we can test our model by examining the empirical distribution of productivity. We measure productivity growth to be essentially symmetric in our 1985-2013 sample. Using data back to 1970 the distribution is moderately skewed to the right.

6. Conclusion

Migration is integral to urban economic dynamics. We have presented a parsimonious dynamic general equilibrium model of migration consistent with the cross sectional patterns of migration, population and productivity and applied it to study urban economic dynamics driven by idiosyncratic productivity changes. The model's parsimony makes it amenable to study other questions about urban economic dynamics.

Our model has faster urban decline than we see empirically and symmetric population growth. Glaeser and Gyourko (2005) argue that the growing supply of housing relative to population as a city shrinks lowers housing costs relative to growing cities and discourages out-migration. This limits net migration relative to growing cities and so is a possible source of both persistent decline and skewness. It is straightforward to introduce housing into our framework and quantify the role of housing in urban dynamics.[32]

[32] Davis, Fisher, and Veracierto (2013) studied a quantitative model in which housing does not slow population adjustments very much once migration costs are taken into account. That result likely depends on the way housing was modeled.

Our model abstracts from the pervasive secular trend in gross migration. However, it is flexible enough to shed light on the source of the trend. Steady state migration is determined by $\psi_1(m)$. The trend can be analysed by considering a transition down toward a stable long run value of $\psi_1(m)$. In this case the gross migration decision depends on the rate of convergence to and distance from the steady state. Calculating these dynamics is left to future work but we already know now what the model implies: The secular decline in migration is due to a rising attachment to location. This suggests future work to understand the trend might benefit from focusing on the determinants of attachment, such as the examples given in the introduction.

References

Alvarez, F. and M. Veracierto (2012). Fixed-term employment contracts in an equilibrium search model. *Journal of Economic Theory 147*(5), 1725–1753.

Blanchard, O. J. and L. F. Katz (1992). Regional evolutions. *Brookings Papers on Economic Activity 1992*(1), 1–75.

Coen-Pirani, D. (2010). Understanding gross worker flows across US states. *Journal of Monetary Economics 57*, 769–784.

Davis, M., J. Fisher, and M. Veracierto (2013). Gross migration, housing and urban population dynamics. Chicago Fed working paper no. 2013-19.

Davis, S., J. Faberman, and J. Haltiwanger (2013). The establishment-level behavior of vacancies and hiring. *Quarterly Journal of Economics 128*(2), 581–622.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *Quarterly Journal of Economics 114*(3), 739–767.

Glaeser, E. L. and J. Gyourko (2005). Urban decline and durable housing. *Journal of Political Economy 113*(2), 345–375.

Glaeser, E. L., J. A. Scheinkman, and A. Shleifer (1995). Economic growth in a cross-section of cities. *Journal of Monetary Economics 36*(1), 117–143.

Gordon, G. and P. Guerron-Quintana (2019). On regional borrowing, default, and migration. Richmond Fed Working Paer No. 19-04.

Howard, G. (2019). The migration accelerator: Labor mobility, housing, and aggregate demand. *American Economic Journal: Macroeconomics*.

Kaplan, G. and S. Schulhofer-Wohl (2015). Understanding the long-run decline in interstate migration. Minneapolis Fed Working Paper 697.

Karahan, F. and S. Rhee (2014). Population aging, migration spillovers, and the decline in interstate migration. Federal Reserve Bank of New York working paper.

Karahan, F. and S. Rhee (2019). Geographical reallocation and unemployment during the great recession: The role of the housing bust. *Journal of Economic Dynamics and Control 100*, 47–69.

Kennan, J. (2008). Average switching costs in dynamic logit models. University of Wisconsin-Madison manuscript.

Kennan, J. and J. R. Walker (2011). The effect of expected income on individual migration decisions. *Econometrica 79*(1), 211–251.

Lloyd-Ellis, H. and A. Head (2012). Housing liquidity, mobility, and the labour market. *Review of Economic Studies 79*(4), 1559–1589.

Lyche, T. and K. Morken (2011). Spline methods. Department of Informatics, Centre of Mathematics for Applications, University of Oslo, manuscript.

Molloy, R., C. L. Smith, and A. Wozniak (2011). Internal migration in the united states. *Journal of Economic Perspectives 25*(2), 1–42.

Nenov, P. (2015). Regional reallocation and housing markets in an equilibrium model of frictional migration. *Review of Economic Dynamics 18*(4), 863–880.

Rappaport, J. (2004). Why are population flows so persistent? *Journal of Urban Economics 56*, 554–580.

Sahin, A., J. Song, G. Topa, and G. Violante (2014). Measuring mismatch in the US labor market. *American Economic Review 104*(11), 3529–3564.

Saks, R. E. and A. Wozniak (2011). Labor reallocation over the business cycle: New evidence from internal migration. *Journal of Labor Economics 29*(4), 697–739.

Shoenberg, I. (1967). On spline functions. In O. Shisha (Ed.), *Proceedings of a symposium held at Wright-Patterson Air Force Base*, pp. 255–291. New York: Academic Press.

Turek, J. H. (1985). The northeast in a national context: background trends in population, income, and empoloyment. In H. W. Richardson and J. H. Turek (Eds.), *Economic prospects for the northeast*. Philadelphia: Temple University Press.

Veracierto, M. (2016). Establishment dynamics, vacancies, and unemployment: A neoclassical approach. *International Economic Review 57*, 1201–1236.

Not for Publication Appendix

(For online publication)

A. Data

This section describes our data sources for our panel data on Metropolitan Statistical Areas (MSAs). The mappings of counties to MSAs we use are consistent with the definitions given by the U.S. Office of Management and Budget (OMB) as of December, 2009. As of that time, OMB defined 366 MSAs in the United States. Over our sample period, these MSAs account for about 80 percent of the aggregate population. We exclude all data from the New Orleans MSA from our study because of the disruption caused by Hurricane Katrina in 2005.

A.1. Calculating Gross Migration Flows with IRS Data

We construct data on gross MSA-level population inflows and outflows using county-to-county migration data based on tax records that is constructed by the Internal Revenue Service (IRS). These data are available annually from 1990 onwards on the IRS web site and are available from 1983 through through 1992 at the Inter-University Consortium for Political and Social Research (ICPSR) web site. The data cover the "filing year" period, not calendar year. For example, the data for 2007 approximately refer to migration over the period April, 2007 to April, 2008. The beginning of the sample period for all our data is 1985, dictated by the availability of the IRS data. We do not use IRS data that is available prior to 1985 because counties are not identified by FIPS codes in these years.

For each year, the IRS reports the migration data using two files, one for outflows and one for inflows. These files cover the experience of each county in the United States. Both the inflow and the outflow files report migrants in units of "returns" and in units of "personal exemptions." According to information from the IRS web site, the returns data approximates the number of households and the personal exemptions data approximates the population.[33] We use the exemptions data.

We define gross inflows into an MSA as the sum of all migrants into any county in that MSA, as long as the inflows did not originate from a county within the MSA. Analogously, we define gross outflows from an MSA as the sum of all migrants leaving any county in that MSA, as long as the migrants did not ultimately move to another county in the MSA. We exclude people migrating into- and out of the United States. But otherwise, for gross inflows the originating counties are not restricted to be part of one of the 365 MSAs, and for gross outflows the counties receiving the migrants are not restricted to be included in one of the 365 MSAs. Over our sample period, counties inside MSAs slightly increased in population, on-net, relative to counties outside of MSAs.

Define A_{it} as the number of new entrants to MSA i during year t, L_{it} as the number of people exiting, and $\bar{P}_t$ as all the people that did not move in or out. We compute the

[33] See `http://www.irs.gov/taxstats/article/0,,id=212683,00.html` for details.

beginning of year population X_{it} and end of year population P_{it} as

$$\begin{aligned} X_{it} &= \bar{P}_{it} + L_{it} \\ P_{it} &= \bar{P}_{it} + A_{it} \end{aligned}$$

Net migration in MSA i during year t is therefore $P_{it} - X_{it} = A_{it} - L_{it}$. Note that due to births and deaths and foreign migration, P_{it} is typically less than the beginning of year population in the subsequent year, X_{it+1}. The corresponding gross arrival *rate*, gross departure *rate* and net migration *rate* are defined as

$$a_{it} = \frac{A_{it}}{X_{it}} \tag{A.1}$$

$$l_{it} \quad \frac{L_{it}}{X_{it}}$$

$$n_{it} = a_{it} - l_{it}. \tag{A.2}$$

A.2. Other data

Our data on annual building permits for all units, by county, are from the SOCDS database as published by the Department of Housing and Urban Development. We compute the MSA total as the sum of the permits for all the counties in that MSA Employment and wages are derived from Table CA4 of the "Regional Economic Accounts: Local Area Personal Income and Employment" as produced by the Bureau of Economic Analysis (BEA). The population data (line 20) are mid-year estimates from the Census Bureau. For employment, we use "Wage and Salary Employment," line 7020. These are counts of full- and part- time jobs of salaried employees. We construct nominal average wage per job, "wages", as the sum of total wage and salary disbursements, line 50, and supplements to wages and salaries, line 60, divided by wage and salary employment.[34] Since our detrending removes time effects, we do not adjust wages for overall inflation.

The aggregate data used to calibrate the model is obtained from Haver Analytics.

B. Derivation of equations (5) and (6)

B.1. Derivation of equation (5)

Given a total number of household members x and a thresshold $\bar{\xi}$ for their idiosyncratic attachments, the total attachment losses incurred are

$$\left(\int_{-\psi_1}^{\bar{\xi}} \xi \frac{1}{2\psi_2} d\xi \right) x = \frac{1}{4\psi_2} \left(\bar{\xi}^2 - \psi_1^2 \right) x,$$

[34]The BEA also reports a broader measure of employment called "total employment" (line 7010) but this is the sum of Wage and Salary Employment and Proprietors Employment (line 7040). In CA4, Proprietors Income (line 70) is associated with Proprietors Employment. Since Proprietors Income includes some capital income, we use the narrower measures of employment and compensation in our study.

and the fraction of household members that move is

$$\lambda = \frac{\bar{\xi} + \psi_1}{2\psi_2}$$

Observe that

$$\begin{aligned}
-\psi_1\lambda + \psi_2\lambda^2 &= -\frac{\psi_1}{2\psi_2}\left(\bar{\xi} + \psi_1\right) + \psi_2 \frac{1}{4\psi_2^2}\left(\bar{\xi} + \psi_1\right)^2 \\
&= -\frac{\psi_1^2}{2\psi_2} - \frac{\psi_1}{2\psi_2}\bar{\xi} + \frac{1}{4\psi_2}\left(\psi_1^2 + 2\psi_1\bar{\xi} + \bar{\xi}^2\right) \\
&= -\frac{\psi_1^2}{2\psi_2} - \frac{\psi_1}{2\psi_2}\bar{\xi} + \frac{\psi_1^2}{4\psi_2} + \frac{\psi_1}{2\psi_2}\bar{\xi} + \frac{1}{4\psi_2}\bar{\xi}^2 \\
&= -\frac{\psi_1^2}{2\psi_2} + \frac{\psi_1^2}{4\psi_2} + \frac{1}{4\psi_2}\bar{\xi}^2 \\
&= \frac{1}{4\psi_2}\bar{\xi}^2 - \frac{\psi_1^2}{4\psi_2} \\
&= \frac{1}{4\psi_2}\left(\bar{\xi}^2 - \psi_1^2\right) \\
&= \int_{-\psi_1}^{\bar{\xi}} \xi \frac{1}{2\psi_2} d\xi
\end{aligned}$$

Hence,

$$\left(\int_{-\psi_1}^{\bar{\xi}} \xi \frac{1}{2\psi_2} d\xi\right) x = -\left[\psi_1\lambda - \psi_2\lambda^2\right] x$$

B.2. Derivation of equation (6)

Given a (post-reallocation) total number of household members p in city m and given a labor-supply threshold $\bar{\omega}$, the total disutility of working in the city is

$$\left(\int_0^{\bar{\omega}} \omega A \omega^\theta d\omega\right) p = \left(A \int_0^{\bar{\omega}} \omega^{\theta+1} d\omega\right) p = \left(A \left[\frac{1}{\theta+2}\omega^{\theta+2}\right]_0^{\bar{\omega}}\right) p = A\frac{1}{\theta+2}\bar{\omega}^{\theta+2} p$$

and the fraction of households members that work is

$$\eta = \int_0^{\bar{\omega}} A\omega^\theta d\omega = A\left[\frac{1}{\theta+1}\omega^{\theta+1}\right]_0^{\bar{\omega}} = A\frac{1}{\theta+1}\bar{\omega}^{\theta+1} \tag{B.1}$$

From equation (B.1) we then have that

$$\left(\frac{\theta+1}{A}\eta\right)^{\frac{1}{\theta+1}} = \bar{\omega} \tag{B.2}$$

Define $\pi = \frac{\theta+2}{\theta+1}$. Substituting equation (B.2) in equation (B.1) we then have:

$$\left(\int_0^{\bar{\omega}} \omega A \omega^{\theta} d\omega\right) p = A \frac{1}{\theta+2}\left(\frac{\theta+1}{A}\eta\right)^{\frac{\theta+2}{\theta+1}} p = \frac{A^{1-\pi}(\theta+1)^{\pi}}{\theta+2}\eta^{\pi} p.$$

C. Characterization of equilibrium and city-planner's problem

To somewhat simplify the analysis, we initially assume that there is no capital and that labor is inelastically supplied. Once we conclude the analysis of this scenario, we describe how it can be trivially extended to the case with capital and endogenous labor supply. Note that notation for some parameters is different from the main text in this appendix but is internally consistent and the connection to the main text should be transparent.

C.1. Competitive equilibrium

C.1.1. Representative household problem

$$MAX \sum_{t=0}^{\infty} \beta^t \left\{ U(c_t) P - \tau \Lambda_t + \sum_{z^t} \left[\varsigma(m)\ p_t\left(z^t\right) - H\left(\frac{a_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^{\frac{1}{\upsilon}} p_{t-1}\left(z^{t-1}\right) \right.\right.$$
$$\left.\left. + \left(\psi_1(m)\frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)} - \psi_2\left(\frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^2\right) p_{t-1}\left(z^{t-1}\right) \right] \times \mu_t\left(z^t\right) \right\}$$

subject to

$$p_t\left(z^t\right) = p_{t-1}\left(z^{t-1}\right) + b_t\left(z^t\right) + \frac{\Lambda_t}{P} p_{t-1}\left(z^{t-1}\right) - l_t\left(z^t\right) \tag{C.1}$$

$$a_t\left(z^t\right) \geq 0 \tag{C.2}$$

$$b_t\left(z^t\right) \geq 0 \tag{C.3}$$

$$l_t\left(z^t\right) \geq 0 \tag{C.4}$$

$$\sum \left[b_t\left(z^t\right) + \frac{\Lambda_t}{P} p_{t-1}\left(z^{t-1}\right) \right] \mu_t\left(z^t\right) = \sum l_t\left(z^t\right) \mu_t\left(z^t\right) \tag{C.5}$$

$$\begin{aligned} & c_t P + \sum q_t\left(z^t\right) b_t\left(z^t\right) \mu_t\left(z^t\right) \\ \leq\ & \sum q_t\left(z^t\right) a_t\left(z^t\right) \mu_t\left(z^t\right) \\ & + \sum w_t\left(z^t\right) p_t\left(z^t\right) \mu_t\left(z^t\right) + \Pi_t \end{aligned} \tag{C.6}$$

C.1.2. Firm's problem in city with history z^t

$$\max \left\{ \kappa(m)\, s_t n_t\left(z^t\right)^{\theta} - w_t\left(z^t\right) n_t\left(z^t\right) \right\}$$

C.1.3. Market clearing conditions

$$n_t\left(z^t\right) = p_t\left(z^t\right) \tag{C.7}$$

$$b_t\left(z^t\right) = a_t\left(z^t\right) \tag{C.8}$$

$$c_t P = \sum \kappa\left(m\right) s_t n_t\left(z^t\right)^{\theta} \mu_t\left(z^t\right) \tag{C.9}$$

$$\Pi_t = \sum \left[\kappa\left(m\right) s_t n_t\left(z^t\right)^{\theta} - w_t\left(z^t\right) n_t\left(z^t\right)\right] \mu_t\left(z^t\right) \tag{C.10}$$

C.1.4. First order conditions

$$\varphi_t = U'\left(c_t\right) \tag{C.11}$$

$$\sum\left[\varphi_t \xi_t(z^t) - \varphi_t \eta_t\right] \frac{p_{t-1}\left(z^{t-1}\right)}{P} \mu_t\left(z^t\right) \leq \tau, \;\; (= 0 \text{ if } \Lambda_t > 0) \tag{C.12}$$

$$\varphi_t q_t(z^t) \leq \frac{1}{\upsilon} H\left(\frac{a_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^{\frac{1}{\upsilon}-1}, \left(= 0 \text{ if } a_t\left(z^t\right) > 0\right) \tag{C.13}$$

$$-\varphi_t q_t(z^t) + \varphi_t \xi_t(z^t) - \varphi_t \eta_t \leq 0 \left(\text{with } = \text{ if } b_t\left(z^t\right) > 0\right) \tag{C.14}$$

$$\psi_1\left(m\right) - 2\psi_2\left(\frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right) \leq \varphi_t \xi_t(z^t) - \varphi_t \eta_t, \;\; \left(= 0 \text{ if } l_t\left(z^t\right) > 0\right) \tag{C.15}$$

$$\begin{aligned} \varphi_t \xi_t(z^t) &= \varphi_t w_t\left(z^t\right) + \varsigma\left(m\right) \\ &-\beta \varphi_{t+1} \eta_{t+1} \frac{\Lambda_{t+1}}{P} - \beta \sum_{s_{t+1}} H\left(\frac{\upsilon-1}{\upsilon}\right)\left(\frac{a_{t+1}\left(s^{t+1}, x_0, s_0, m\right)}{p_t\left(z^t\right)}\right)^{\frac{1}{\upsilon}} Q(s_{t+1}, s_t) \\ &+\beta \sum_{s_{t+1}} \psi_2 \left(\frac{l_{t+1}\left(s^{t+1}, x_0, s_0, m\right)}{p_t\left(z^t\right)}\right)^2 Q(s_{t+1}, s_t) + \beta \sum_{s_{t+1}} \varphi_{t+1} \xi_{t+1}(s^{t+1}, x_0, s_0, m)\left(1 + \frac{\Lambda_{t+1}}{P}\right) Q(s_{t+1}, s_t) \end{aligned} \tag{C.16}$$

$$w_t\left(z^t\right) = \kappa\left(m\right) s_t \theta n_t\left(z^t\right)^{\theta-1} \tag{C.17}$$

Remark: Since this is a convex economy with no distortions the Welfare Theorems hold.

C.2. Economy-wide Planning Problem in Sequential Form

The economy-wide social planner problem is the following:

$$\begin{aligned} MAX \sum_{t=0}^{\infty} \beta^t \Bigg\{ & U\left(c_t\right) P - \tau \Lambda_t + \sum_{z^t} \Bigg[\varsigma\left(m\right) \; p_t\left(z^t\right) - H\left(\frac{a_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^{\frac{1}{\upsilon}} p_{t-1}\left(z^{t-1}\right) \\ & + \left(\psi_1\left(m\right) \frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)} - \psi_2\left(\frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^2\right) p_{t-1}\left(z^{t-1}\right) \Bigg] \times \mu_t\left(z^t\right) \Bigg\} \end{aligned}$$

subject to:

$$p_t\left(z^t\right) = p_{t-1}\left(z^{t-1}\right) + a_t\left(z^t\right) + \frac{\Lambda_t}{P} p_{t-1}\left(z^{t-1}\right) - l_t\left(z^t\right) \tag{C.18}$$

$$a_t\left(z^t\right) \geq 0 \tag{C.19}$$

$$l_t\left(z^t\right) \geq 0 \tag{C.20}$$

$$\sum\left[a_t\left(z^t\right)+\frac{\Lambda_t}{P}p_{t-1}\left(z^{t-1}\right)\right]\mu_t\left(z^t\right)=\sum l_t\left(z^t\right)\mu_t\left(z^t\right) \tag{C.21}$$

$$c_t P \leq \sum \kappa\left(m\right)s_t p_t\left(z^t\right)^{\theta}\mu_t\left(z^t\right) \tag{C.22}$$

with μ_0 given and where $p_{t-1}\left(z^{t-1}\right)=x_0$ for $t=0$.

FOC's:

$$\varphi_t = U'\left(c_t\right) \tag{C.23}$$

$$\sum\left[\varphi_t\xi_t(z^t)-\varphi_t\eta_t\right]\frac{p_{t-1}\left(z^{t-1}\right)}{P}\mu_t\left(z^t\right)\leq\tau,\ \ (=0 \text{ if } \Lambda_t>0) \tag{C.24}$$

$$\varphi_t\xi_t(z^t)-\varphi_t\eta_t \leq \frac{1}{\upsilon}H\left(\frac{a_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^{\frac{1}{\upsilon}-1},\ \ \left(=0 \text{ if } a_t\left(z^t\right)>0\right) \tag{C.25}$$

$$\psi_1\left(m\right)-2\psi_2\left(\frac{l_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)\leq\varphi_t\xi_t(z^t)-\varphi_t\eta_t,\ \ \left(=0 \text{ if } l_t\left(z^t\right)>0\right) \tag{C.26}$$

$$\begin{aligned}
\varphi_t\xi_t(z^t) &= \varphi_t\kappa\left(m\right)s_t\theta p_t\left(z^t\right)^{\theta-1}+\varsigma\left(m\right) \\
&-\beta\varphi_{t+1}\eta_{t+1}\frac{\Lambda_{t+1}}{P}-\beta\sum_{s_{t+1}}H\left(\frac{\upsilon-1}{\upsilon}\right)\left(\frac{a_{t+1}\left(s^{t+1},x_0,s_0,m\right)}{p_t\left(z^t\right)}\right)^{\frac{1}{\upsilon}}Q(s_{t+1},s_t) \\
&+\beta\sum_{s_{t+1}}\psi_2\left(\frac{l_{t+1}\left(s^{t+1},x_0,s_0,m\right)}{p_t\left(z^t\right)}\right)^2 Q(s_{t+1},s_t)+\beta\sum_{s_{t+1}}\varphi_{t+1}\xi_{t+1}(s^{t+1},x_0,s_0,m)\left(1+\frac{\Lambda_{t+1}}{P}\right)Q(s_{t+1},s_t)
\end{aligned} \tag{C.27}$$

C.3. City Planner's problem in sequential form

The city planner takes as given the sequence $\{\Lambda_t,\varphi_t,\eta_t\}_{t=0}^{\infty}$.

Define

$$\hat{\mu}\left(s^t\right)=Q\left(s_t;s_{t-1},s_{t-2}\right)...Q(s_2;s_1,s_0)Q(s_1;s_0,s_{-1}).$$

Then, the city planner's problem is the following:

$$\begin{aligned}
MAX\sum_{t=0}^{\infty}\beta^t\sum_{s^t}\Big\{&\varphi_t\kappa\left(m\right)s_t p_t\left(z^t\right)^{\theta}+\varsigma\left(m\right)p_t\left(z^t\right) \\
&-\varphi_t\eta_t\left[a_t\left(z^t\right)+\frac{\Lambda_t}{P}p_{t-1}\left(z^{t-1}\right)\right]+\varphi_t\eta_t l_t\left(z^t\right) \\
&-H\left(\frac{a_t\left(z^t\right)}{p_{t-1}\left(z^{t-1}\right)}\right)^{\frac{1}{\upsilon}}p_{t-1}\left(z^{t-1}\right)
\end{aligned} \tag{C.28}$$

$$+\left[\psi_1(m)\frac{l_t(z^t)}{p_{t-1}(z^{t-1})}-\psi_2\left(\frac{l_t(z^t)}{p_{t-1}(z^{t-1})}\right)^2\right]p_{t-1}(z^{t-1})\Bigg\}\hat{\mu}(s^t)$$

subject to:

$$p_t(z^t)=p_{t-1}(z^{t-1})+a_t(z^t)+\frac{\Lambda_t}{P}p_{t-1}(z^{t-1})-l_t(z^t) \tag{C.29}$$

$$a_t(z^t)\geq 0 \tag{C.30}$$

$$l_t(z^t)\geq 0 \tag{C.31}$$

FOC's:

$$\varphi_t\xi_t(z^t)-\varphi_t\eta_t\leq\frac{1}{\upsilon}H\left(\frac{a_t(z^t)}{p_{t-1}(z^{t-1})}\right)^{\frac{1}{\upsilon}-1},\quad (=0 \text{ if } a_t(z^t)>0) \tag{C.32}$$

$$\psi_1(m)-2\psi_2\left(\frac{l_t(z^t)}{p_{t-1}(z^{t-1})}\right)\leq\varphi_t\xi_t(z^t)-\varphi_t\eta_t,\quad (=0 \text{ if } l_t(z^t)>0) \tag{C.33}$$

$$\varphi_t\xi_t(z^t)=\varphi_t\kappa(m)s_t\theta p_t(z^t)^{\theta-1}+\varsigma(m)$$

$$-\beta\varphi_{t+1}\eta_{t+1}\frac{\Lambda_{t+1}}{P}-\beta\sum_{s_{t+1}}H\left(\frac{\upsilon-1}{\upsilon}\right)\left(\frac{a_{t+1}(s^{t+1},x_0,s_0,m)}{p_t(z^t)}\right)^{\frac{1}{\upsilon}}Q(s_{t+1},s_t)$$

$$+\beta\sum_{s_{t+1}}\psi_2\left(\frac{l_{t+1}(s^{t+1},x_0,s_0,m)}{p_t(z^t)}\right)^2Q(s_{t+1},s_t)+\beta\sum_{s_{t+1}}\varphi_{t+1}\xi_{t+1}(s^{t+1},x_0,s_0,m)\left(1+\frac{\Lambda_{t+1}}{P}\right)Q(s_{t+1},s_t) \tag{C.34}$$

with (x_0,s_0,m) given and where $p_{t-1}(z^{t-1})=x_0$ for $t=0$.

C.4. Equivalence between economy-wide social planner's problem and city planner's problem

Proposition 1: Let $\{C_t,\Lambda_t,p_t,a_t,l_t,\varphi_t,\xi_t,\eta_t\}_{t=0}^{\infty}$ be the unique solution to the economy-wide social planner problem with initial state μ_0.

Then, for each initial state (x_0,s_0,m), $\{p_t,a_t,l_t,\xi_t\}_{t=0}^{\infty}$ is the unique solution to the city planner's problem that takes $\{\Lambda_t,\varphi_t,\eta_t\}_{t=0}^{\infty}$ as given.

Proof: It follows from the fact that equations (C.18)-(C.20),(C.25)-(C.27) imply equations (C.29)-(C.34).

Proposition 2: For each initial state (x_0,s_0,m) let $\{p_t,a_t,l_t,\xi\}_{t=0}^{\infty}$ be the unique solution to the city planner's problem that takes $\{\Lambda_t,\varphi_t,\eta_t\}_{t=0}^{\infty}$ as given.

Define

$$c_tP=\sum\kappa(m)s_tp_t(z^t)^{\theta}\mu_t(z^t) \tag{C.35}$$

Suppose that

$$\varphi_t=U'(c_t) \tag{C.36}$$

$$\sum\left[\varphi_t\xi_t(z^t)-\varphi_t\eta_t\right]\frac{p_{t-1}(z^{t-1})}{P}\mu_t(z^t)\leq\tau,\quad (=0 \text{ if } \Lambda_t>0) \tag{C.37}$$

$$\sum \left[a_t\left(z^t\right) + \frac{\Lambda_t}{P} p_{t-1}\left(z^{t-1}\right)\right] \mu_t\left(z^t\right) = \sum l_t\left(z^t\right) \mu_t\left(z^t\right) \tag{C.38}$$

Then, $\{C_t, \Lambda_t, p_t, a_t, l_t, \varphi_t, \xi_t, \eta_t\}_{t=0}^{\infty}$ is the unique solution to the economy-wide social planner problem with initial state μ_0.

Proof: It follows from the fact that equations (C.35)-(C.38),(C.29)-(C.34) imply equations (C.18)-(C.27).

C.5. Steady state allocation

In order to define a steady sate allocation it will be convenient to work with a recursive formulation to the city planner's problem given by (C.28).

At a steady sate, the recursive city planner's problem is the following:

$$V\left(x, s, m\right) = \max \left\{ \varphi\kappa\left(m\right) sp^{\theta} + \varsigma\left(m\right) p - \varphi\eta\left(a + \frac{\Lambda}{P} x\right) + \varphi\eta l \right. \tag{C.39}$$

$$\left. -H\left(\frac{a}{x}\right)^2 x + \left[\psi_1\left(m\right)\frac{l}{x} - \psi_2\left(\frac{l}{x}\right)^2\right] x + \beta \sum_{s'} V\left(x', s', m\right) Q\left(s', s\right) \right\}$$

subject to

$$p = x + a + \frac{\Lambda}{P} x - l \tag{C.40}$$

$$a \geq 0 \tag{C.41}$$

$$l \geq 0 \tag{C.42}$$

$$x' = p \tag{C.43}$$

where the constant values (Λ, φ, η) are taken as given.

The first order conditions are the following:

$$\varphi\xi(x, s, m) - \varphi\eta \leq 2H\frac{a\left(x, s, m\right)}{x}, \ (= 0 \text{ if } a(x, s, m) > 0) \tag{C.44}$$

$$\psi_1\left(m\right) - 2\psi_2\frac{l\left(x, s, m\right)}{x} \leq \varphi\xi(x, s, m) - \varphi\eta, \ (= 0 \text{ if } l\left(x, s, m\right) > 0) \tag{C.45}$$

$$\varphi\xi(x, s, m) = \varphi\kappa\left(m\right) s\theta p^{\theta - 1} + \varsigma\left(m\right) + \beta \sum_{s'} V'\left(x', s', m\right) Q\left(s', s\right) \tag{C.46}$$

and the envelope condition is given by

$$V'\left(x, s, m\right) = H\left[\frac{a\left(x, s, m\right)}{x}\right]^2 + \psi_2\left[\frac{l\left(x, s, m\right)}{x}\right]^2 + 2H\frac{a\left(x, s, m\right)}{x}\frac{\Lambda}{P} + \varphi\xi(x, s, m) > 0 \tag{C.47}$$

Observe that the concavity of the return function implies that

$$V''\left(x, s, m\right) < 0. \tag{C.48}$$

The optimal decision rules to this problem generate an invariant distribution μ^m that satisfies the following recursion:

$$\mu^m(X, s') = \int_{\{(x,s):\ p(x,s,m)\in X\}} Q(s', s)\, d\mu^m. \tag{C.49}$$

Recall that λ is distribution of cities across permanent types m. Then, the side conditions in Proposition 2 guaranteeing a steady state solution to the economy-wide social planner's problem are the following:

$$cP = \sum_m \lambda(m) \int \kappa(m)\, sp(x,s,m)^\theta\, d\mu^m, \tag{C.50}$$

$$\varphi = U'(c), \tag{C.51}$$

$$\sum_m \lambda(m) \int [\varphi\xi(x,s,m) - \varphi\eta] \frac{x}{P} d\mu^m \leq \tau, \ \ (= 0 \text{ if } \Lambda > 0), \tag{C.52}$$

$$\sum_m \lambda(m) \int \left[a(x,s,m) + \frac{\Lambda}{P}x\right] d\mu^m = \sum_m \lambda(m) \int l(x,s,m)\, d\mu^m. \tag{C.53}$$

C.6. Linear relation between gross and net flows

The empirical evidence indicates that assuming that $l_t > 0$ and $a_t > 0$ in all cities is the relevant case. The reason for $l_t > 0$ is that departure rates are always positive in our dataset. The reason for $a_t > 0$ is that otherwise total arrival rates would be the same (equal to $\frac{\Lambda_t}{P}$) for some range of net population changes $(p_t - p_{t-1})/p_{t-1}$, and this would be highly counterfactual.

Hereon, we will then assume that $a_t > 0$ and that $l_t > 0$. Moreover, we will assume that $\upsilon = 0.5$.

Then, from equations (C.32) and (C.33) we have that

$$\frac{l_t}{p_{t-1}} = \frac{1}{2}\frac{\psi_1(m)}{\psi_2} - \frac{H}{\psi_2}\frac{a_t}{p_{t-1}}$$

Also, from equation (C.29) we have that

$$\frac{p_t - p_{t-1}}{p_{t-1}} = \frac{a_t}{p_{t-1}} + \frac{\Lambda_t}{P} - \frac{l_t}{p_{t-1}}$$

It follows that the total arrival rate is given by:

$$\frac{a_t}{p_{t-1}} + \frac{\Lambda_t}{P} = \frac{\psi_2}{\psi_2 + H}\left(\frac{p_t - p_{t-1}}{p_{t-1}}\right) + \frac{\psi_2}{\psi_2 + H}\left(\frac{1}{2}\frac{\psi_1(m)}{\psi_2} - \frac{\Lambda_t}{P}\right) + \frac{\Lambda_t}{P}, \tag{C.54}$$

which is a linear function of the net population change $(p_t - p_{t-1})/p_{t-1}$. Moreover, the linear coefficient is:

$$0 < \frac{\psi_2}{\psi_2 + H} < 1.$$

Also, the departure rate is given by:

$$\frac{l_t}{p_{t-1}} = \left(-\frac{H}{\psi_2 + H}\right)\left(\frac{p_t - p_{t-1}}{p_{t-1}}\right) + \frac{\psi_2}{\psi_2 + H}\left(\frac{1}{2}\frac{\psi_1(m)}{\psi_2} - \frac{\Lambda_t}{P}\right) + \frac{\Lambda_t}{P} \tag{C.55}$$

which also is a linear function of the net population change $(p_t - p_{t-1})/p_{t-1}$, with the linear coefficient given by

$$-1 < -\frac{H}{\psi_2 + H} < 0.$$

It turns out that the empirical evidence indicates a strong linear relation between arrival rates a_t/p_{t-1} and net population changes $(p_t - p_{t-1})/p_{t-1}$ in the cross section of cities, with slope between 0 and 1. Thus, the model is consistent with the empirical evidence.

The above analysis has shown that the model is consistent with the empirical evidence about the relation between gross and net population flows. We now explain why both types of moving costs (idiosyncratic attachment shocks and recruitment activities) are necessary to deliver that result. We do this by describing the consequences of removing each of these type of moving costs.

Free recruiting activities ($H = 0$) When $H = 0$, the FOC's (C.32) and (C.33) become:

$$\varphi_t \xi_t - \varphi_t \eta_t \leq 0, \quad (= 0 \text{ if } a_t > 0) \tag{C.56}$$

$$\psi_1(m) - 2\psi_2\left(\frac{l_t}{p_{t-1}}\right) \leq \varphi_t \xi_t - \varphi_t \eta_t, \quad (= 0 \text{ if } l_t > 0) \tag{C.57}$$

These conditions reveal the following lower bound on departure rates:

$$0 < \frac{\psi_1(m)}{2\psi_2} \leq \frac{l_t}{p_{t-1}}$$

Suppose that $\frac{\psi_1(m)}{2\psi_2} < \frac{l_t}{p_{t-1}}$.
Then, from equation (C.57) we have that

$$\psi_1(m) - 2\psi_2\left(\frac{l_t}{p_{t-1}}\right) = \varphi_t \xi_t - \varphi_t \eta_t < 0$$

From equation (C.56) we then have that

$$a_t = 0.$$

From equation (C.29) it then follows that

$$\frac{l_t}{p_{t-1}} = \frac{\Lambda_t}{P} - \frac{p_t - p_{t-1}}{p_{t-1}}$$

Suppose now that $\frac{\psi_1(m)}{2\psi_2} = \frac{l_t}{p_{t-1}}$.

From equation (C.57) we then have that

$$\psi_1(m) - 2\psi_2 \left(\frac{l_t}{p_{t-1}}\right) = \varphi_t \xi_t - \varphi_t \eta_t = 0.$$

From equation (C.56) it follows that $a_t \geq 0$. In particular, from equation (C.29) we have that

$$\frac{a_t}{p_{t-1}} = \frac{p_t - p_{t-1}}{p_{t-1}} + \frac{\psi_1(m)}{2\psi_2} - \frac{\Lambda_t}{P} \geq 0.$$

From this discussion it follows that:

$$\frac{a_t}{p_{t-1}} = \max\left\{\frac{p_t - p_{t-1}}{p_{t-1}} - \frac{\Lambda_t}{P} + \frac{\psi_1(m)}{2\psi_2}, 0\right\},$$

$$\frac{l_t}{p_{t-1}} = \max\left\{\frac{\psi_1(m)}{2\psi_2}, -\left(\frac{p_t - p_{t-1}}{p_{t-1}}\right) + \frac{\Lambda_t}{P}\right\}.$$

Observe that in this case, as long as the net population growth rate is not negative enough, that the city planner sets the departure rate at the point of maximum static benefits $\frac{\psi_1(m)}{2\psi_2}$ and lets the arrival rate make all the adjustment. A constant departure rate and an arrival rate that moves one-to-one with net population growth is highly counterfactural.

No idiosyncratic attachment shocks ($\psi_1(m) = \psi_2 = 0$) In this case the FOC's (C.32) and (C.33) become:

$$\varphi_t \xi_t - \varphi_t \eta_t \leq 2H \left(\frac{a_t}{p_{t-1}}\right), (= 0 \text{ if } a_t > 0) \tag{C.58}$$

$$0 \leq \varphi_t \xi_t - \varphi_t \eta_t, (= 0 \text{ if } l_t > 0) \tag{C.59}$$

Suppose that $(p_t - p_{t-1})/p_{t-1} - \Lambda_t/P > 0$.

From equation (C.29) we have that

$$\frac{p_t - p_{t-1}}{p_{t-1}} - \frac{\Lambda_t}{P} = \frac{a_t}{p_{t-1}} - \frac{l_t}{p_{t-1}} > 0. \tag{C.60}$$

Hence, $a_t > 0$ and from equation (C.58),

$$\varphi_t \xi_t - \varphi_t \eta_t = 2H \left(\frac{a_t}{p_{t-1}}\right) > 0$$

From equation (C.59) it follows that

$$l_t = 0$$

and from equation (C.60) that

$$\frac{a_t}{p_{t-1}} = \frac{p_t - p_{t-1}}{p_{t-1}} - \frac{\Lambda_t}{P} > 0.$$

Suppose now that $(p_t - p_{t-1})/p_{t-1} - \Lambda_t/P \leq 0$.
From equation (C.29) we have that

$$\frac{p_t - p_{t-1}}{p_{t-1}} - \frac{\Lambda_t}{P} = \frac{a_t}{p_{t-1}} - \frac{l_t}{p_{t-1}} < 0. \tag{C.61}$$

Hence, $l_t > 0$ and from equation (C.59),

$$0 = \varphi_t \xi_t - \varphi_t \eta_t.$$

From equation (C.58) it follows that

$$a_t = 0$$

and from equation (C.61) that

$$\frac{l_t}{p_{t-1}} = -\left(\frac{p_t - p_{t-1}}{p_{t-1}} - \frac{\Lambda_t}{P}\right).$$

From this discussion it follows that

$$\frac{a_t}{p_{t=1}} = \max\left\{\frac{p_t - p_{t-1}}{p_t} - \frac{\Lambda_t}{P}, 0\right\} \tag{C.62}$$

$$\frac{l_t}{p_{t-1}} = \max\left\{-\left(\frac{p_t - p_{t-1}}{p_t} - \frac{\Lambda_t}{P}\right), 0\right\} \tag{C.63}$$

Observe that in this case, if arrivals are positive then departures are zero and if departures are positive arrivals are zero. This is a direct consequence of arrivals being costly but not departures and the goal of economizing costs. However, this is highly counterfactual: In the data arrivals and departures are always positive.

C.7. Moving costs imply quadratic adjustment costs on population changes

Suppose that $a(x,s,m) > 0$ and $l(x,s,m) > 0$. Then, from equations (C.54) and (C.55) we have that

$$\frac{a(x,s,m)}{x} + \frac{\Lambda}{P} = \frac{\psi_2}{\psi_2 + H}\left(\frac{p(x,s,m) - x}{x}\right) + \frac{\psi_2}{\psi_2 + H}\left(\frac{1}{2}\frac{\psi_1(m)}{\psi_2} - \frac{\Lambda}{P}\right) + \frac{\Lambda}{P} \tag{C.64}$$

$$\frac{l(x,s,m)}{x} = \left(-\frac{H}{\psi_2 + H}\right)\left(\frac{p(x,s,m) - x}{x}\right) + \frac{\psi_2}{\psi_2 + H}\left(\frac{1}{2}\frac{\psi_1(m)}{\psi_2} - \frac{\Lambda}{P}\right) + \frac{\Lambda}{P} \tag{C.65}$$

Substituting these expressions into the moving costs described by the second line of equation (C.39) gives

$$-H\left(\frac{a}{x}\right)^2 x + \left[\psi_1(m)\frac{l}{x} - \psi_2\left(\frac{l}{x}\right)^2\right] x$$

$$
\begin{aligned}
&= -H\left[\frac{\psi_1(m)}{2(H+\psi_2)} - \frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P} + \frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2 x \\
&\quad +\psi_1(m)\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P} - \frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right] x \\
&\quad -\psi_2\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P} - \frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2 x
\end{aligned}
$$

$$
\begin{aligned}
&= -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)} - \frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2 \\
&\quad -Hx2\left[\frac{\psi_1(m)}{2(H+\psi_2)} - \frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]\left[\frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)\right] \\
&\quad -Hx\left[\frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2 \\
&\quad +\psi_1(m)\,x\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P}\right] \\
&\quad -\psi_1(m)\,x\frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right) \\
&\quad -\psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2 \\
&\quad +\psi_2 x2\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]\left[\frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right] \\
&\quad -\psi_2 x\left[\frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&= -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)} - \frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2 + \psi_1(m)\,x\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P}\right] \\
&\quad -\psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2 \\
&\quad +\left\{-\psi_1(m)\,x\frac{H}{H+\psi_2} + \psi_2 x2\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]\frac{H}{H+\psi_2}\right\}\left(\frac{p-x}{x}\right) \\
&\quad -Hx2\left[\frac{\psi_1(m)}{2(H+\psi_2)} - \frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]\frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right) \\
&\quad -\psi_2 x\left[\frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2 - Hx\left[\frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2
\end{aligned}
$$

$$
= -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)} - \frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2 + \psi_1(m)\,x\left[\frac{\psi_1(m)}{2(H+\psi_2)} + \frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]
$$

$$
\begin{aligned}
&-\psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2\\
&+\left\{-\psi_1(m)\,x\frac{H}{H+\psi_2}+\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}-\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]\frac{2H\psi_2 x}{H+\psi_2}\right\}\left(\frac{p-x}{x}\right)\\
&-\psi_2 x\left[\frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2-Hx\left[\frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
= \; & -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)}-\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2+\psi_1(m)\,x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]-\\
& \psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2\\
& +\left\{-\psi_1(m)\,x\frac{H}{H+\psi_2}+\frac{2H\psi_2 x}{H+\psi_2}\frac{\Lambda}{P}\right\}\left(\frac{p-x}{x}\right)\\
& -\psi_2 x\left[\frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2-Hx\left[\frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
= \; & -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)}-\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2+\psi_1(m)\,x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]\\
& -\psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2\\
& +\left\{-\psi_1(m)+2\psi_2\frac{\Lambda}{P}\right\}\frac{H}{H+\psi_2}(p-x)\\
& -\psi_2 x\left[\frac{H}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2-Hx\left[\frac{\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
= \; & -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)}-\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2+\psi_1(m)\,x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]-\\
& \psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2\\
& +\left\{-\psi_1(m)+2\psi_2\frac{\Lambda}{P}\right\}\frac{H}{H+\psi_2}(p-x)\\
& -\left\{\psi_2\left(\frac{H}{H+\psi_2}\right)^2+H\left(\frac{\psi_2}{H+\psi_2}\right)^2\right\}x\left(\frac{p-x}{x}\right)^2
\end{aligned}
$$

$$
= \; -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)}-\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2+\psi_1(m)\,x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]-
$$

$$\psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2$$
$$+\left\{-\psi_1(m)+2\psi_2\frac{\Lambda}{P}\right\}\frac{H}{H+\psi_2}(p-x)$$
$$-\frac{\psi_2 H^2+H\psi_2^2}{(H+\psi_2)^2}x\left(\frac{p-x}{x}\right)^2$$

$$\begin{aligned}
= \quad & -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)}-\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2+\psi_1(m)x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]- \\
& \psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2 \\
& +\left\{-\psi_1(m)+2\psi_2\frac{\Lambda}{P}\right\}\frac{H}{H+\psi_2}(p-x) \\
& -\frac{H\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)^2 x
\end{aligned}$$

Then, the city social planner's problem in equation (C.39) can be written as:

$$V(x,s,m)=\max\Big\{\varphi\kappa(m)sp^{\theta}-\varphi\eta(p-x)+\varsigma(m)p \tag{C.66}$$
$$-Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)}-\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2+\psi_1(m)x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]-\psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2$$
$$+\left\{-\psi_1(m)+2\psi_2\frac{\Lambda}{P}\right\}\frac{H}{H+\psi_2}(p-x)-\frac{H\psi_2}{H+\psi_2}\left(\frac{p-x}{x}\right)^2 x+\beta\sum_{s'}V(x',s',m)Q(s',s)\Big\}$$

subject to:

$$a\geq 0 \tag{C.67}$$
$$l\geq 0 \tag{C.68}$$
$$x'=p \tag{C.69}$$

Define

$$\begin{aligned}
\Phi\left(\frac{\Lambda}{P},m\right) \quad = \quad & -Hx\left[\frac{\psi_1(m)}{2(H+\psi_2)}-\frac{\psi_2}{H+\psi_2}\frac{\Lambda}{P}\right]^2+\psi_1(m)x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right] \\
& -\psi_2 x\left[\frac{\psi_1(m)}{2(H+\psi_2)}+\frac{H}{H+\psi_2}\frac{\Lambda}{P}\right]^2
\end{aligned}$$

$$\Gamma\left(\frac{\Lambda}{P},m\right)=\left\{-\psi_1(m)+2\psi_2\frac{\Lambda}{P}\right\}\frac{H}{H+\psi_2}$$

Then the city social planner's problem can be written as:

$$V(x,s,m) = \max\left\{\varphi\kappa(m)\, sp^{\theta} - \varphi\eta(p-x) + \varsigma(m)\, p\right. \tag{C.70}$$

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

subject to:

$$a \geq 0 \tag{C.71}$$

$$l \geq 0 \tag{C.72}$$

$$x' = p \tag{C.73}$$

Thus, the structure of moving costs imply quadratic adjustment costs on population changes. Observe that the slope of equation (C.64) is extremely informative about the magnitude of these quadratic costs.

C.8. Equivalence with endogenous labor supply and mobile capital

For simplicity we derive the equivalence under the assumption of a single permanent city type. The extension to multiple types is trivial.

The city planner's problem for the model with endogenous labor supply and mobile capital is the following:

$$V(x,z) = \max\left\{\phi z n^{\alpha} k^{\gamma} - B n^{\pi} p^{1-\pi} - \phi r k - \phi\eta\left(a + \frac{\Lambda}{P}x\right) + \phi\eta l\right. \tag{C.74}$$

$$\left. -H\left(\frac{a}{x}\right)^2 x + \left[\psi_1 \frac{l}{x} - \psi_2\left(\frac{l}{x}\right)^2\right] x + \beta\sum_{z'} V(x',z')\, Q(z',z)\right\}$$

subject to

$$p = x + a + \frac{\Lambda}{P}x - l \tag{C.75}$$

$$a \geq 0 \tag{C.76}$$

$$l \geq 0 \tag{C.77}$$

$$x' = p \tag{C.78}$$

where the constant values (Λ, ϕ, η, r) are taken as given.

The first order condition with respect to k is

$$k = \left(\frac{z n^{\alpha}\gamma}{r}\right)^{\frac{1}{1-\gamma}}$$

Therefore,

$$\phi z n^{\alpha} k^{\gamma} - \phi r k \quad = \quad \phi\left[z n^{\alpha} k^{\gamma} - r k\right]$$

$$
\begin{aligned}
&= \phi\left[zn^{\alpha}\left(\frac{zn^{\alpha}\gamma}{r}\right)^{\frac{\gamma}{1-\gamma}} - r\left(\frac{zn^{\alpha}\gamma}{r}\right)^{\frac{1}{1-\gamma}}\right] \\
&= \phi z^{\frac{1}{1-\gamma}} n^{\frac{\alpha}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)
\end{aligned}
$$

Define

$$\sigma = \frac{\alpha}{1-\gamma}.$$

Then, the return function can be written as follows:

$$\phi zn^{\alpha}k^{\gamma} - Bn^{\pi}p^{1-\pi} - \phi rk = \phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right) n^{\sigma} - Bn^{\pi}p^{1-\pi}$$

Taking first order conditions with respect to n we get

$$\phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right) \sigma n^{\sigma-1} = B\pi n^{\pi-1}p^{1-\pi}$$

Hence,

$$n = \left[\frac{B\pi p^{1-\pi}}{\phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\sigma}\right]^{\frac{1}{\sigma-\pi}}$$

Therefore the return function can be written as

$$
\begin{aligned}
&\phi zn^{\alpha}k^{\gamma} - Bn^{\pi}p^{1-\pi} - \phi rk \\
= \;& \phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\left[\frac{B\pi p^{1-\pi}}{\phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\sigma}\right]^{\frac{\sigma}{\sigma-\pi}} \\
&-B\left[\frac{B\pi p^{1-\pi}}{\phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\sigma}\right]^{\frac{\pi}{\sigma-\pi}} p^{1-\pi} \\
= \;& \left[\phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\right]^{-\frac{\pi}{\sigma-\pi}}\left(\frac{B\pi}{\sigma}\right)^{\frac{\sigma}{\sigma-\pi}} p^{\frac{\sigma}{\sigma-\pi}(1-\pi)} \\
&-\left[\phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\right]^{-\frac{\pi}{\sigma-\pi}} B\left(\frac{B\pi}{\sigma}\right)^{\frac{\pi}{\sigma-\pi}} p^{(1-\pi)\frac{\sigma}{\sigma-\pi}}
\end{aligned}
$$

Thus, the return function is

$$
\begin{aligned}
&\phi zn^{\alpha}k^{\gamma} - Bn^{\pi}p^{1-\pi} - \phi rk \\
= \;& \left[\phi z^{\frac{1}{1-\gamma}} r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\right]^{-\frac{\pi}{\sigma-\pi}} B^{\frac{\sigma}{\sigma-\pi}}\left[\left(\frac{\pi}{\sigma}\right)^{\frac{\sigma}{\sigma-\pi}} - \left(\frac{\pi}{\sigma}\right)^{\frac{\pi}{\sigma-\pi}}\right] p^{\frac{\sigma}{\sigma-\pi}(1-\pi)} \\
= \;& \phi^{-\frac{\pi}{\sigma-\pi}}\left[r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}\right)\right]^{-\frac{\pi}{\sigma-\pi}} B^{\frac{\sigma}{\sigma-\pi}}\left[\left(\frac{\pi}{\sigma}\right)^{\frac{\sigma}{\sigma-\pi}} - \left(\frac{\pi}{\sigma}\right)^{\frac{\pi}{\sigma-\pi}}\right] z^{-\frac{\pi}{(\sigma-\pi)(1-\gamma)}} p^{\frac{\sigma}{\sigma-\pi}(1-\pi)}
\end{aligned}
$$

$$= \phi\phi^{-\frac{\pi}{\sigma-\pi}-1}\left[r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}}-\gamma^{\frac{1}{1-\gamma}}\right)\right]^{-\frac{\pi}{\sigma-\pi}}B^{\frac{\sigma}{\sigma-\pi}}\left[\left(\frac{\pi}{\sigma}\right)^{\frac{\sigma}{\sigma-\pi}}-\left(\frac{\pi}{\sigma}\right)^{\frac{\pi}{\sigma-\pi}}\right]z^{-\frac{\pi}{(\sigma-\pi)(1-\gamma)}}p^{\frac{\sigma}{\sigma-\pi}(1-\pi)}$$

Define

$$\begin{aligned}
\theta &= \frac{\sigma}{\sigma-\pi}(1-\pi) \\
s &= \phi^{-\frac{\pi}{\sigma-\pi}-1}\left[r^{-\frac{\gamma}{1-\gamma}}\left(\gamma^{\frac{\gamma}{1-\gamma}}-\gamma^{\frac{1}{1-\gamma}}\right)\right]^{-\frac{\pi}{\sigma-\pi}}B^{\frac{\sigma}{\sigma-\pi}}\left[\left(\frac{\pi}{\sigma}\right)^{\frac{\sigma}{\sigma-\pi}}-\left(\frac{\pi}{\sigma}\right)^{\frac{\pi}{\sigma-\pi}}\right]z^{-\frac{\pi}{(\sigma-\pi)(1-\gamma)}} \\
\varphi &= \phi.
\end{aligned}$$

Then the city-planner's problem for the model with endogenous labor supply and mobile capital maps exactly to the reduced model analyzed in this appendix.

D. Solving the Model

While representing the solution of the economy-wide social planner's as the solution to a city planner's problem plus side conditions is a huge simplification, computing the solution to the city planner's problem remains a nontrivial task.

The value function of the city planner's problem has one endogenous variables and two exogenous state variables. Each exogenous state variable takes values in a finite grid but this grid cannot be too coarse if the resulting discrete process is to represent the original AR(2) in a satisfactory way (in computations, we use 25 values for each exogenous state). To simplify the task of computing the value function we use spline approximation on the endogenous state variable.

Cubic spline interpolation is usually used in these cases. A difficulty with these methods is that they do not necessarily preserve the shape of the original function, or if they do (as with Schumacher shape-preserving interpolation) it is somewhat difficult to compute. For these reasons, we use a local method that does not interpolate the original function but that approximates it while preserving shape (monotonicity and concavity). An additional benefit is that it is extremely simple to compute (there is no need to solve a system of equations). The method is known as the Shoenberg's variation diminishing spline approximation. It was first introduced by Shoenberg (1967) and is described in a variety of sources (e.g. Lyche and Morken (2011)).

For a given continuous function f on an interval $[a, b]$, let p be a given positive integer, and let $\tau = (\tau_1, ..., \tau_{n+p+1})$ be a knot vector with $n \geq p+1$, $a \leq \tau_i \leq b$, $\tau_i \leq \tau_{i+1}$, $\tau_{p+1} = a$ and $\tau_{n+1} = b$. The *variation diminishing spline approximation of degree p to f* is then defined as

$$S_p(x) = \sum_{j=1}^{n} f\left(\tau_j^*\right) B_{jp}(x)$$

where $\tau_j^* = (\tau_{j+1} + ... + \tau_{j+p})/p$ and $B_{jp}(x)$ is the jth B-spline of degree p evaluated at x. The B-splines are defined recursively as follows

$$B_{jp}(x) = \frac{x-\tau_j}{\tau_{j+p}-\tau_j}B_{j,p-1}(x) + \frac{\tau_{j+1+p}-x}{\tau_{j+1+p}-\tau_{j+1}}B_{j+1,p-1}(x)$$

with

$$B_{j0}(x) = \begin{cases} 1, \text{ if } \tau_j \leq x < \tau_{j+1} \\ 0, \text{ otherwise} \end{cases}$$

We solve the city planner's problem by iterating on its first-order and envelope conditions, and use the variation diminishing splines to approximate the derivatives of the value function. As a consequence, the critical property of the spline approximations for our purposes is that they preserve the strict monotonicity of the derivatives of the value function (since the value function is strictly concave). However, the fact that the variation diminishing splines also preserve the concavity of the original function f (e.g. Lyche and Morken (2011), Section 5.2) and that their definition can be easily generalized to functions of more than one variable using tensor products (e.g. Lyche and Morken (2011), Section 7.2.1), should make this class of spline approximations extremely useful in a variety of other settings.

In actual computations we worked with variation diminishing spline approximations of degree $p = 3$.

Statistics under the invariant distribution were computed using Monte Carlo simulations. This part of the computations was offloaded to graphic cards to exploit their massively parallel capabilities. To avoid costly computations similar to those encountered in the evaluation of the return function, cubic spline approximations were used for the decision rules.

Speeding up the solution to the city planner's problem and Monte Carlo simulations was convenient since finding solutions (C, Λ, η) to the side conditions requires solving the city planner's problem and simulating its solution several times.

The source code, which is written in CUDA Fortran, is available upon request. Compiling it requires the PGI Fortran compiler. Running it requires at least one NVIDIA graphic card with compute capability higher than 2.0. In our case we used a system with two Tesla V100 graphic cards, totalling over 10,000 NVIDIA CUDA cores.

E. Inter-state and Inter-city Migration Costs in the Kennan and Walker (2011) Model

We now justify our conclusion that it is valid to apply Kennan and Walker (2011)'s estimate of migration costs in our environment. The argument is based on a calibrated model that incorporates the essence of the individual discrete choice problem studied by Kennan and Walker (2011) within an equilibrium setting.

There are N locations called cities. Each city i is associated with a wage that is fixed over time, w_i.[35] A person living in city i receives the wage and then receives a vector of i.i.d. preference shocks, one for each city including the person's current city, $e = (e_1, e_2, \ldots, e_N)$. After receiving the preference shocks, the person decides whether to move. The expected value of living in city i before the shocks are realized but after the wage is paid is

$$V_i \quad = \quad E\left[\max_{j \in \{1,\ldots,N\}} \left\{\frac{w_i}{\alpha} - \frac{c(i,j)}{\alpha} + e_j + \beta V_j\right\}\right]$$

[35]For simplicity we abstract from idiosyncratic wage shocks included by Kennan and Walker (2011). Kennan and Walker (2011) assume that individuals are finitely lived and only know the permanent component of wages of their current city and any city they have lived in previously. In an infinite horizon context individuals eventually live in every city and therefore have knowledge of the complete wage distribution.

Let s denote the state (a unique grouping of cities) containing city i and s' the state containing j. The moving cost function $c(i,j)$ is

$$c(i,j) = \begin{cases} 0, & \text{if } i = j \\ c_1, & \text{if } i \neq j \text{ and } s = s' \\ c_2, & \text{if } i \neq j \text{ and } s \neq s' \end{cases}$$

People pay no moving costs if they do not move, and in-state moving costs c_1 may be different than out-of-state moving costs c_2. Allowing c_1 to be different from c_2 is in the spirit of Kennan and Walker (2011)'s finding that moving costs increase with distance moved.[36]

Following Kennan and Walker (2011) we assume that the preference shocks are drawn from the Type 1 Extreme Value Distribution. Given a wage for each location w_i and the parameters of the model, α, β, c_1 and c_2, we compute the value functions using backwards recursion. We start with a guess of the expected value functions for every $j = 1, \ldots, N$. Call the current guess of the expected value function at location j as $\widehat{V}_j$. We then update the guess at each $i = 1, \ldots, N$

$$\widetilde{V}_i = \log\left\{\sum_{j=1}^{N}\left[\exp\left(\frac{w_i}{\alpha} - \frac{c(i,j)}{\alpha} + \beta\widehat{V}_j\right)\right]\right\} + \zeta$$

where ζ is Euler's constant and $\widetilde{V}_i$ is the updated guess. We repeat this entire process until the expected value functions have converged, that is until $\widehat{V}_i$ is equal to $\widetilde{V}_i$ at each of the $i = 1, \ldots, N$ cities.

We set $N = 365$. For each city, we set w_i equal to the average wage in the corresponding MSA in 1990 (the year Kennan and Walker (2011) use to calculate average state wages) in thousands of dollars. For states that span multiple MSAs, we set the state s as the state where most of the population of the MSA lives in 1990.[37]

We assume that $\beta = 0.96$, leaving three parameters to be estimated: α, which scales the shocks into dollar equivalents, and the two moving costs c_1 and c_2. We estimate these parameters by targeting three moments: the average rate of individual migration across all MSAs, 4.47 percent, the average rate of across-state migration, 3.0 percent, and the average flow benefit scaled by average wage experienced by migrants, 1.9. For a worker moving from city i to city j, the flow benefit (scaled to dollars) is $\alpha(e_j - e_i) - c(i,j)$. Our target value of the average flow benefits of across-state movers scaled by average wage is taken from estimates produced by Kennan and Walker (2011) (see footnote 21 in the main text.) We use data from the IRS for 1990 to compute across-MSA and across-state migration rates. Our estimate of the across-state migration rate is almost identical to the estimate reported in Table VIII, page 239 by Kennan and Walker (2011) of 2.9 percent.

We compute all three moments analytically. The probability agents migrate to location

[36] See the estimate of γ_1 in Table II on page 230 of their paper.

[37] Some MSAs span multiple states and this may introduce some error because within-MSA across-state moves that are truly within MSA will be misclassified as moves to a new labor market.

j given their current location of i, $\gamma(j,i)$ has the straightforward expression

$$\gamma(j,i) \quad = \quad \frac{\exp\left(\frac{w_i}{\alpha} - \frac{c(i,j)}{\alpha} + \beta V_j\right)}{\sum_{k=1}^{N}\left[\exp\left(\frac{w_i}{\alpha} - \frac{c(i,k)}{\alpha} + \beta V_k\right)\right]}.$$

We construct the $N \times N$ matrix Γ, with individual elements $\gamma(j,i)$, and determine the steady state distribution of population across metro areas, the N-dimensional vector ρ, such that $\rho = \Gamma\rho$. Given ρ, we compute the probability of any move at the steady-state population distribution as

$$\sum_{i=1}^{N} \rho(i) \left[\sum_{j \neq i} \gamma(j,i)\right]$$

and the probability of an across-state move as

$$\sum_{i=1}^{N} \rho(i) \left[\sum_{j \neq i, s' \neq s} \gamma(j,i)\right] .$$

For the third moment, it can be shown that the expected increase in continuation value from a worker choosing the optimal location as compared to an arbitrary location is a function of the probability the worker chooses the arbitrary location. For example, for a worker that optimally moves to location j, the expected increase in value, inclusive of flow utility and discounted future expected value, over remaining in the current location i is $-\log\gamma(i,i)/(1-\gamma(i,i))$, see Kennan (2008). The expected increase in current flow payoff for all moves from j to i is therefore

$$\frac{-\log\gamma(i,i)}{1-\gamma(i,i)} - \beta(V_j - V_i).$$

The average of this second term across all moves, from i to all $j \neq i$, is

$$\frac{\sum_{j \neq i} \gamma(j,i)\beta(V_j - V_i)}{\sum_{k \neq i} \gamma(k,i)} \quad = \quad \frac{\sum_{j \neq i} \gamma(j,i)\beta(V_j - V_i)}{1-\gamma(i,i)}.$$

The denominator is the probability a move occurs. Thus, conditional on moving, the average benefit of all moves that occur relative to staying put is

$$\sum_i \left(\frac{\rho(i)}{1-\gamma(i,i)}\right)\left(-\log\gamma(i,i) - \sum_{j \neq i} \gamma(j,i)\beta(V_j - V_i)\right).$$

We divide this expression by average wage (appropriately scaled), evaluated at the steady state: $\sum_i \rho(i) w_i/\alpha$.

We use the Nelder-Meade algorithm to search for parameters and we match our 3 target moments exactly. Our parameter estimates are $c_2 = 76.7$, $c_2 = 116.6$, and $\alpha = 17.6$. For reference, the mean wage at the steady state population distribution is 39.051 (\$39,051). Our estimates of c_1 and c_2 imply that in-state and out-of-state moving costs are twice and three times average wages, respectively. These large costs generate low mobility rates in the face of large permanent wage differentials across metro areas. However, the estimated value of α implies that the mean and variance of the preference shocks are large. This large variance generates shocks large enough to induce people to move given the high costs of moving.

To determine the size of the bias in the Kennan and Walker (2011) estimates from using across-state moves, rather than across-MSA moves, we run 100 simulations of the model, simulating 600,000 people per MSA in each run. This generates approximately 27,000 moves to any MSA and 20,000 out of state moves for each MSA in the simulation. In each simulation run, we compute the economy-wide average flow benefits to across-state movers scaled by average wages. Averaged across the 100 simulations, the average benefit to across-state movers scaled by average wage is exactly 2.0, 0.1 higher than the average simulated benefits accruing to all movers. The bias is therefore 5%.[38] We find that a bias of this size does not affect our conclusions.

F. Calibration of the recruitment efficiency parameter H

In this appendix it will be useful to allow for the measure of cities in the economy to be different than one. In particular, we assume that

$$\sum_m \lambda(m) = M.$$

We continue to assume that

$$\sum_m \lambda(m) \int x_0 \mu_0(s_0, x_0, s_0, m) = P.$$

In the model the average moving cost of agents that actually move is given by

$$J_c = \frac{c \sum_m \lambda(m) \int \left(-\psi_1(m) \frac{l(x,s,m)}{x} + \psi_2 \left(\frac{l(x,s,m)}{x} \right)^2 \right) x d\mu^m}{\sum_m \lambda(m) \int l(x,s,m) d\mu^m} + \frac{\sum_m \lambda(m) \int q(x,s,m) a(x,s,m) d\mu^m + c\tau\Lambda}{\sum_m \lambda(m) \int a(x,s,m) d\mu^m + \Lambda}.$$

while average wages are given by

$$J_w = \alpha \frac{Y}{N}.$$

The Kennan-Walker statistic J is then given by

$$J = \frac{J_c}{J_w}$$

[38] Measured across the 100 runs, the standard deviation of the percent of the bias is 0.2%, the minimum bias is 4.6% and the maximum bias is 5.7%.

Observe that

$$q(x,s,m) = 2cH\frac{a(x,s,m)}{x}$$

and, from equation (21), that

$$\tau = \frac{1}{P}\sum_m \lambda(m)\int 2Ha(x,s,m)\,d\mu^m$$

Also define the intercept of the adjustment rule

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

as

$$\begin{aligned}\phi(m) &= \frac{\psi_1(m)}{2(\psi_2+H)} + \frac{H}{(\psi_2+H)}\frac{\Lambda}{P} \\ &= \frac{\psi_2}{\psi_2+H}\left(\frac{1}{2}\frac{\psi_1(m)}{\psi_2} - \frac{\Lambda}{P}\right) + \frac{\Lambda}{P}.\end{aligned}$$

Observe that $\phi(m)$ is the fixed effect term in our empirical analysis.

Then,

$$\begin{aligned}J_c &= \frac{2cH\sum_m \lambda(m)\int \frac{a(x,s,m)^2}{x}d\mu^m + c\tau\Lambda}{\sum_m \lambda(m)\int a(x,s,m)d\mu^m + \Lambda} + \frac{c\sum_m \lambda(m)\int\left(-\psi_1(m)\frac{l(x,s,m)}{x} + \psi_2\left(\frac{l(x,s,m)}{x}\right)^2\right)xd\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m} \\ &= \frac{2cH\sum_m \lambda(m)\int \frac{a(x,s,m)^2}{x}d\mu^m + c\tau\Lambda}{\sum_m \lambda(m)\int a(x,s,m)d\mu^m + \Lambda} - \frac{c\sum_m \lambda(m)\int \psi_1(m)\frac{l(x,s,m)}{x}xd\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m} \\ &\quad + \frac{c\sum_m \lambda(m)\int \psi_2\left(\frac{l(x,s,m)}{x}\right)^2 xd\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m}\end{aligned}$$

It follows that

$$\begin{aligned}\frac{J_c}{\psi_2} &= \frac{2c\frac{H}{\psi_2}\sum_m \lambda(m)\int \frac{a(x,s,m)^2}{x}d\mu^m + c\frac{\tau}{\psi_2}\Lambda}{\sum_m \lambda(m)\int a(x,s,m)d\mu^m + \Lambda} - \frac{c\sum_m \lambda(m)\int \frac{\psi_1(m)}{\psi_2}\frac{l(x,s,m)}{x}xd\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m} \\ &\quad + \frac{c\sum_m \lambda(m)\int \left(\frac{l(x,s,m)}{x}\right)^2 xd\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m} \\ &= \frac{2c\frac{H}{\psi_2}\sum_m \lambda(m)\int \frac{a(x,s,m)^2}{x}d\mu^m + c\frac{\tau}{\psi_2}\Lambda}{\sum_m \lambda(m)\int a(x,s,m)d\mu^m + \Lambda} \\ &\quad - \frac{c\sum_m \lambda(m)\int 2\left[\frac{\phi(m)-\frac{\Lambda}{P}}{\left(\frac{\psi_2}{\psi_2+H}\right)} + \frac{\Lambda}{P}\right]\frac{l(x,s,m)}{x}xd\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m} + \frac{c\sum_m \lambda(m)\int \left(\frac{l(x,s,m)}{x}\right)^2 xd\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m}\end{aligned}$$

$$
\begin{aligned}
= & \frac{2c\frac{H}{\psi_2}\sum_m \lambda(m)\int \frac{a(x,s,m)^2}{x}d\mu^m}{\sum_m \lambda(m)\int a(x,s,m)d\mu^m+\Lambda}+\frac{c\frac{\Lambda}{P}\frac{H}{\psi_2}\sum_m \lambda(m)\int 2a(x,s,m)\,d\mu^m}{\sum_m \lambda(m)\int a(x,s,m)d\mu^m+\Lambda} \\
& -\frac{c\sum_m \lambda(m)\int 2\left[\frac{\phi(m)-\frac{\Lambda}{P}}{\left(\frac{\psi_2}{\psi_2+H}\right)}+\frac{\Lambda}{P}\right]\frac{l(x,s,m)}{x}x d\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m}+\frac{c\sum_m \lambda(m)\int \left(\frac{l(x,s,m)}{x}\right)^2 x d\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{J_c}{c\psi_2} = & \frac{2\frac{H}{\psi_2}\frac{1}{M}\sum_m \lambda(m)\int \frac{a(x,s,m)^2}{x}d\mu^m}{\frac{1}{M}\sum_m \lambda(m)\int a(x,s,m)d\mu^m+\frac{\Lambda}{P}\frac{P}{M}}+\frac{\frac{\Lambda}{P}\frac{H}{\psi_2}\frac{1}{M}\sum_m \lambda(m)\int 2a(x,s,m)\,d\mu^m}{\frac{1}{M}\sum_m \lambda(m)\int a(x,s,m)d\mu^m+\frac{\Lambda}{P}\frac{P}{M}} \\
& -\frac{\frac{2}{\left(\frac{\psi_2}{\psi_2+H}\right)}\frac{1}{M}\sum_m \lambda(m)\int \phi(m)\,l(x,s,m)\,d\mu^m}{\frac{1}{M}\sum_m \lambda(m)\int l(x,s,m)d\mu^m} \\
& -\frac{\sum_m \lambda(m)\int 2\left[\frac{\Lambda}{P}-\frac{\frac{\Lambda}{P}}{\left(\frac{\psi_2}{\psi_2+H}\right)}\right]\frac{l(x,s,m)}{x}x d\mu^m}{\sum_m \lambda(m)\int l(x,s,m)d\mu^m} \\
& +\frac{\frac{1}{M}\sum_m \lambda(m)\int \frac{l(x,s,m)^2}{x}d\mu^m}{\frac{1}{M}\sum_m \lambda(m)\int l(x,s,m)d\mu^m} \\
= & \frac{2\frac{H}{\psi_2}E\frac{a(x,s,m)^2}{x}}{Ea(x,s,m)+\frac{\Lambda}{P}\frac{P}{M}}+\frac{2\frac{\Lambda}{P}\frac{H}{\psi_2}Ea(x,s,m)}{Ea(x,s,m)+\frac{\Lambda}{P}\frac{P}{M}} \\
& -\frac{\frac{2}{\left(\frac{\psi_2}{\psi_2+H}\right)}E\left[\phi(m)\,l(x,s,m)\right]}{El(x,s,m)}+2\frac{\frac{\Lambda}{P}}{\left(\frac{\psi_2}{\psi_2+H}\right)}-2\frac{\Lambda}{P}+\frac{E\frac{l(x,s,m)^2}{x}}{El(x,s,m)}
\end{aligned}
$$

Also, observe that the slope of the arrival rate rule is

$$\frac{\psi_2}{\psi_2+H}=\frac{1}{1+\frac{H}{\psi_2}}$$

Thus, the slope determines $\frac{H}{\psi_2}$. In particular

$$\frac{H}{\psi_2}=\frac{1}{slope}-1$$

In turn, we can write

$$\frac{J_w}{c}=\alpha\frac{Y}{N}=\frac{\alpha}{\left(\frac{N}{P}\right)\left(\frac{cP}{Y}\right)}.$$

It follows that

$$\frac{1}{\psi_2}J=\frac{1}{\psi_2}\frac{J_c}{J_w}=\left(\frac{J_c}{c\psi_2}\right)\frac{1}{\left(\frac{J_w}{c}\right)} \tag{F.1}$$

Conditional on some value for $\frac{\Lambda}{P}$, we can obtain $\frac{\nu_n}{c\psi_2}$ from the the slope of the arrival rate rule $\frac{\psi_2}{\psi_2+H}$, the average population size of cities $\frac{P}{M}$, and the following cross-sectional moments:

$$E\left[\frac{a^2}{x}\right], E\left[a\right], E\left[\phi l\right], E\left[l\right], \text{ and } E\left[\frac{l^2}{x}\right].$$

Observe that given some empirical measure of total arrivals in a city $a + \frac{\Lambda}{P}x$, a can be obtained from the $\frac{\Lambda}{P}$ value being conditioned on and the size of the city in the previous period x. Thus, empirical counterparts for these moments are readily obtainable.

In turn, $\frac{\nu_d}{c}$ is obtained from the labor share α, the fraction of total population that works $\frac{N}{P}$ and the aggregate consumption/output ratio $\frac{cP}{Y}$.

Given a value for Kennan and Walker's estimate, which we assocate with ν, equation (F.1) then implies a value for ψ_2. Given this value for ψ_2, the slope of the arrival rate rule $\frac{\psi_2}{\psi_2+H}$ then implies a value for H.

Thus, H can be fully calibrated (conditional on some value for $\frac{\Lambda}{P}$) using measurable empirical moments.