Gross Migration, Housing and Urban Population Dynamics*

Morris A. Davis
University of Wisconsin
mdavis@bus.wisc.edu

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

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

August 13, 2014

Abstract

We quantify the influence of migration and housing on urban population dynamics using a dynamic general equilibrium model of cities which incorporates a new theory of migration motivated by patterns uncovered in a panel of US cities. Cities experience significant, near random walk productivity shocks, yet population is slow to adjust. While city populations are dominated by *net* migration, we show that measuring and modeling *gross* migration is essential to quantifying migration's impact on their dynamics. Housing is also a natural candidate for influencing population dynamics because it is difficult to move, costly to build quickly, and is very durable. We use our panel data along with additional micro- and macro-economic evidence to calibrate our model. The implied dynamic responses to productivity shocks of a city's population, gross migration, employment, wages, home construction and house prices closely resemble those we estimate with our panel. In the model costs of finding better cities to live and work drive slow population adjustments in the short run. Housing plays a very limited role. We also show that the model's slow response of population to past declines in productivity also account for persistent urban decline.

JEL Classification Numbers: E0, O4, R0
Keywords: Gross migration, housing, urban population dynamics, persistent urban decline, house prices

*An earlier draft of this paper circulated under the title "The Role of Housing in Labor Reallocation." We thank Marco Bassetto, Jeff Campbell, John Kennan, Erzo Luttmer, Nicolas Petrosky-Nadeau, Robert Shimer and seminar participants at various institutions and conferences for helpful comments. The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

1 Introduction

As we document in this paper, cities experience significant, random-walk-like productivity shocks, yet population is very slow to adjust to them. This evidence suggests there are substantial barriers to geographical labor reallocation within the US. What are these barriers? Urban population dynamics are the result of many individual decisions about whether to leave one's current location and if so where to move to. Hence the costs and benefits underlying gross migration decisions are a natural starting point for understanding slow population adjustment. Previous research suggests that housing may be important as well. The reasons are that it is difficult to move, requires time-to-build to accommodate an increasing population, and a large durable housing stock makes a city attractive to potential migrants.

We study how migration and housing frictions drive slow population adjustments by developing a quantitative dynamic general equilibrium model of cities and comparing it to panel data from 365 US cities over the period 1985-2007. Our model incorporates a new theory of migration between cities motivated by evidence from these data. We find that gross in-migration is linearly increasing in net migration and that in-migration rises and out-migration falls following positive productivity shocks. These findings suggest it is important to model the gross flows underlying net migration and that an empirically plausible model of gross migration should embody some form of directed migration which allows in-migration to vary systematically with net migration. Moreover, we show that the observed gross migration is extremely informative about the costs of adjusting population through net migration.

Building on Kennan and Walker (2011) (hereafter KW), large gross flows arise in our model from idiosyncratic shocks to workers' tastes for their current cities over others. When considering a move workers understand the distribution of city types but not the location of any particular type of city. They can pay a variable cost to find a particular city type, a form of directed migration, or pay a fixed cost to be randomly allocated. Costly directed migration captures the myriad ways in which workers find a given type of city, including through active recruitment by firms and cities and informal networking. Random matching accounts for intangible factors that influence location choices such as to be near family members. Since employing more of the existing population is a natural substitute for migration to accommodate higher labor demand, labor supply is endogenous in our model as well.

We integrate this theory of migration and labor supply into an otherwise familiar generalization of the neoclassical growth model. Each city produces a unique good using locally supplied workers and freely mobile equipment capital, subject to idiosyncratic productivity shocks. These locally produced goods are imperfectly substitutable inputs into the production of the freely tradeable final goods, consumption and equipment. Individuals enjoy consumption and housing services in the same city they work with housing services derived from locally produced, immobile and durable residential structures.

The model is calibrated using observations different from those we use for its validation. Four features of the calibration are new. First, we obtain the migration technology's parameters by matching the linear relationship between gross in-migration and net migration and the net benefits to migration estimated by KW. Second, we calibrate the model's productivity process using our panel, thereby pinning down its exogenous source of persistence and variability. Third, we obtain the substitutability of intermediate goods by matching the empirical cross-section distribution of population, verifying that our model is consistent with Zipf's law.[1] Finally, labor supply is calibrated to cities' wage elasticity of employment relative to population we estimate from our panel.

We validate the calibrated model by comparing how population, gross in- and out-migration, employment, wages, residential investment and house prices respond to productivity shocks with estimates of the same objects based on our panel. The model does surprisingly well along these dimensions and importantly it is consistent with the slow response of population even though this evidence is not targeted in our calibration. We also find that the model is broadly consistent with the unconditional volatility, persistence, and contemporaneous co-movement of the key variables, and that productivity shocks alone account for about 60% of the variation in local population growth.

Having established the empirical relevance of our model, we use it to examine how migration and housing influence population adjustments. We find that workers' costs of finding new cities to live and work essentially account for the slow population adjustments we observe on their own. Housing plays a very limited role. In the absence of migration frictions, costs of producing durable, immobile housing do lower the amplitude and increase the persistence of population's response to a productivity shock. However, if migration frictions are already present, adding those for housing does virtually nothing to influence population dynamics.

There is evidence of persistent urban decline in our panel with many cities experiencing declining populations throughout the sample period. Glaeser and Gyourko (2005) argue that housing's immobility and durability account for persistent urban decline and they dismiss

[1] Zipf's law is that in its upper tail city population is distributed exponentially with an exponent close to one. See for example Gabaix (1999) and Eeckhout (2004).

a role for migration (see pp. 368–369). The declining cities in our sample also experience declining (relative) productivity, suggesting that our model might account for persistent urban decline in addition to short run population dynamics. We find that it does, through the slow response of population to past declines in productivity. This suggests that migration frictions are a major factor in urban decline even though, as emphasized by Glaeser and Gyourko (2005), rates of in-migration remain high in declining cities.

Our model builds on an extensive empirical and theoretical microeconomic literature on migration, surveyed by Greenwood (1997) and Lucas (1997). KW is an important recent contribution. They analyze individual migration decisions in the face of wage shocks, idiosyncratic location-preference shocks, and moving costs, but without housing or equilibrium interactions. KW calculate the speed of adjustment of state populations to permanent wage changes using their estimated model. Despite using very different methodologies we find similarly slow population adjustments. In addition, we show how and why their microeconomic estimates of moving costs based on inter-*state* migration can be used to calibrate a general equilibrium model of inter-*city* migration.

Our model is a dynamic version of the classic Roback (1982) and Rosen (1979) static model of cities with costless mobility. Recent contributions using this approach include Albouy (2009) and Diamond (2012). Because it is static, the Roback-Rosen model does not address migration or population adjustments to shocks. Van Nieuwerburgh and Weil (2010) introduce dynamics to this framework. Their model has implications for net migration, but not gross flows. Coen-Pirani (2010) also constructs a dynamic Roback-Rosen model, studying gross population flows among US states in a framework similar to the one used by Davis, Faberman, and Haltiwanger (2011) and others to study gross worker flows among firms. Our empirical work demonstrates that gross population flows are quite dissimilar to gross worker flows, which motivates taking a different approach to modeling migration.

This paper also contributes to the literature by introducing a city's dynamic response to an identified productivity shock as a model validation tool and by estimating the underlying stochastic process for productivity.[2] Model validation in the literature emphasizes unconditional cross-sectional and time-series patterns which in some cases are also used to identify model parameters. In contrast we calibrate parameters using statistics different from those we use to validate our model. Furthermore, while much of the literature relies on idiosyn-

[2] Lloyd-Ellis, Head, and Sun (2014) estimate the responses of population, residential construction and house prices to personal income shocks identified using a panel VAR and a Choleski decomposition of the residual variance-covariance matrix. They abstract from migration decisions and equilibrium interactions among cities.

cratic productivity shocks to drive variation, it does not provide evidence on the nature of these shocks as we do.[3]

The recent housing boom and bust has prompted a growing literature that quantifies how housing frictions impede geographical labor reallocation and leads to persistently high aggregate unemployment. For example, Karahan and Rhee (2012), Lloyd-Ellis and Head (2012) and Nenov (2012) study how house price collapses limit labor reallocation through disincentives to migrate arising from home ownership and local search frictions. We abstract from these factors and emphasize instead costly migration between geographically distinct labor and housing markets. Nenov (2012) and Karahan and Rhee (2012) model gross migration, but they do not address the linear relationship between gross and net migration that we find is essential to quantifying the speed of urban population adjustment.[4]

The rest of the paper is organized as follows. Section 2 describes new empirical evidence from our panel of cities on migration and the responses of population and gross migration to productivity shocks. After this we use two simplified versions of our quantitative model to describe our approach to migration and the possible role for housing in slowing population adjustments. Section 5 introduces the quantitative model and Section 6 describes how we calibrate it. Section 7 validates the calibrated model by comparing its predictions to dynamics we estimate from our panel and quantifies the influence of migration costs and housing on short and long run urban population dynamics. The last section concludes.

2 Empirical Evidence

This section describes empirical evidence to motivate our analysis and guide our modeling of migration. We work with an annual panel data set covering 1985 to 2007 that includes population, net and gross migration, employment, wages, residential construction, and house prices for 365 Metropolitan Statistical Areas (MSAs) comprising about 83% of the aggregate population.[5] We consider MSAs because they represent geographically distinct labor markets. They are defined as a region with a relatively high population density at its core and close economic ties throughout as measured by commuting patterns. Such regions are not

[3] An exception is Karahan and Rhee (2012) who estimate an auto-regressive process in the level of GDP per worker using a short panel of cities.

[4] There is also an empirical literature that investigates the effects of housing related financial frictions on mobility. See for example Ferreira, Gyourko, and Tracy (2011), Modestino and Dennett (2012) and Schulhofer-Wohl (2012). We abstract from financial frictions in this paper.

[5] See Davis, Fisher, and Veracierto (2011) for a detailed description of these data.

legally incorporated as a city or town would be, nor are they legal administrative divisions like counties or sovereign entities like states. A typical MSA is centered around a single large city that wields substantial influence over the region, *e.g.* Chicago, although some MSAs contain more than one large city with no single municipality holding a substantially dominant position, *e.g.* Dallas–Fort Worth. With these caveats, for convenience we refer to our MSAs as cities.

2.1 Gross Versus Net Migration

We use IRS data to calculate city-level net and gross migration rates because these data have wide coverage of US cities. Due to limited sample sizes gross migration rates can only be calculated for a small number of cities using other available data including the Current Population Survey and the American Community Survey. State-level migration rates can be calculated using these surveys. These data yield very similar results to those we obtain with city-level and state-level migration rates calculated using the IRS data.[6]

Let a_{it} and l_{it} denote the number of people flowing into and out of city i in year t and x_{it} the population of that city at the beginning of the same year. For an individual city the *arrival* (in-migration) rate is a_{it}/x_{it} and the *leaving* (out-migration) rate is l_{it}/x_{it}. The difference between the arrival and leaving rates is the net migration rate. Gross migration rates fluctuate over the business cycle and have been falling over our sample period.[7] To abstract from these dynamics we subtract from each city's gross rate in a year the corresponding cross section average in that year and define net migration as the difference between these measures.

Figure 1 contains plots of gross and net migration rates by population decile with the time effects removed. Net migration is essentially unrelated to city size. This reflects Gibrat's law for cities, that population growth is independent of city size. However, the arrival and leaving rates are clearly diminishing in city size. This is an interesting finding worthy of further study, but its presence conflates cross-city variation with the within-city dynamics we are interested in. Therefore, after removing time fixed effects, for every city we subtract from each year's arrival and leaving rate the time series average of the sum of the arrival

[6]As emphasized by Kaplan and Schulfofer-Wohl (2012) there are some drawbacks to using the IRS data: tax filings under-represent the poor and elderly; addresses on tax forms are not necessarily home addresses; and tax returns may be filed late.

[7]See Saks and Wozniak (2011) for evidence on the cyclical characteristics of migration and Karahan and Rhee (2013), Molloy, Smith, and Wozniak (2011) and Kaplan and Schulfofer-Wohl (2012) for studies of its trend.

and leaving rates for that city. This removes city fixed effects in gross migration without affecting net migration rates.

Figure 1: Gross and Net Migration Rates by Population Decile

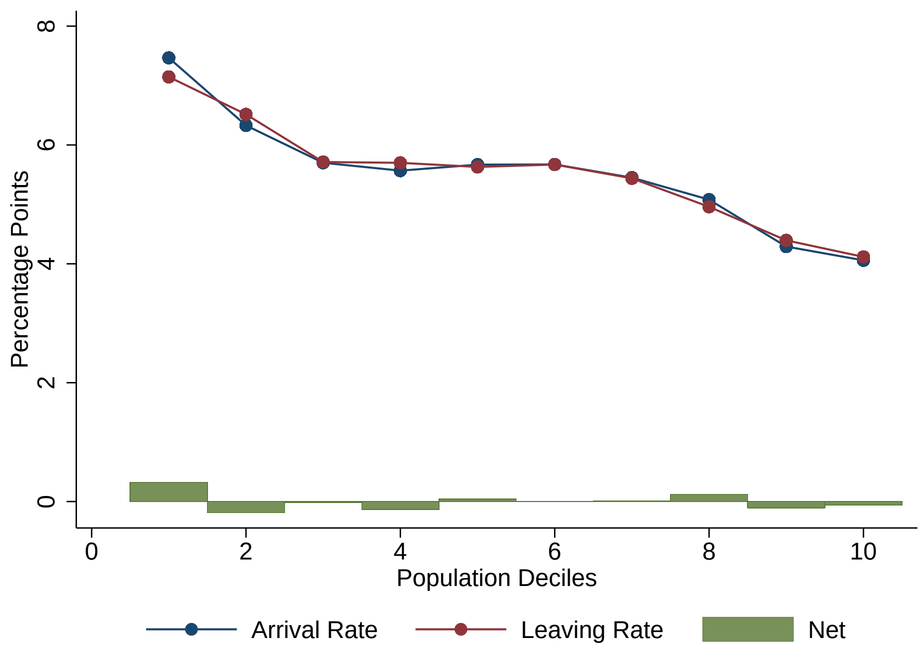


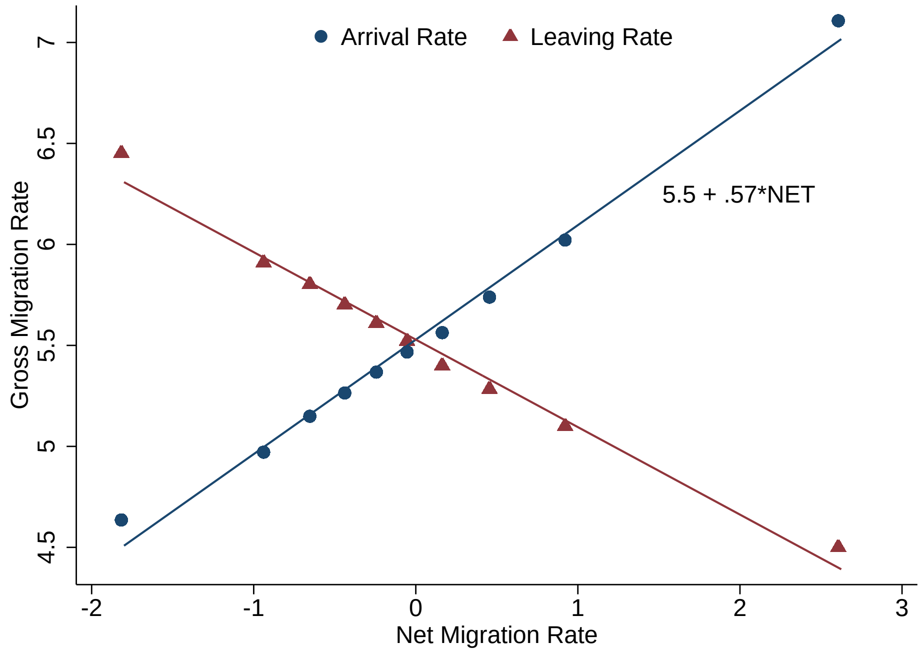
Figure 2 displays mean arrival and leaving rates against mean net migration for each net migration decile, after removing both time and city fixed effects and adding back the corresponding unconditional mean to the gross migration rates. Notice first that gross migration is large and far exceeds the amount necessary to account for net migration. For example, in a city with zero net migration an average of 11% of the population either arrives to or leaves from a city during a year.

Second, the arrival rate is monotonically increasing and the leaving rate is monotonically decreasing in net migration. This suggests it is important to model both gross migration margins in order to understand urban population dynamics. The rising arrival rate at a function of net migration is prima facie evidence that migration involves a directed component. Otherwise gross arrivals would be independent of net migration.

Third, and most striking, the gross migration rates fall almost exactly on the corresponding regression lines.[8] This finding sharply contrasts with what we observe for worker flows.

[8]We obtain virtually identical regression lines when we use all the data rather than first taking averages of deciles and when we estimate using data from the first 5 years of the sample and the last five years of the

Figure 2: Gross Migration Rates by Net Migration



In particular, Davis, Faberman, and Haltiwanger (2006) find that hires are flat and close to zero for negative net flows and linearly increasing in net flows otherwise; separations are essentially the mirror image. These differences suggests models designed to explain worker flows are not suitable for understanding population flows – we require a different theory. Our theory of migration and calibration of the underlying parameters are tied closely to the linearity finding.

Finally, the negative relationship between the arrival and leaving rates appears to be inconsistent with Coen-Pirani (2010)'s finding of a positive correlation between the two gross migration rates at the state level. The difference is not because we study cities instead of states. It arises from our removal of city-specific fixed effects from the gross migration rates. As suggested by Figure 1, when we do not remove these effects the gross migration rates are strongly positively correlated.[9]

sample. We also find qualitatively similar results when we regress gross on net migration separately for each city in our sample.

[9]Coen-Pirani (2010) removes cross-sectional variation in the occupational characteristics of states prior to his analysis.

2.2 Responses of Population and Gross Migration to TFP Shocks

We estimate dynamic responses of city-level variables to local productivity shocks by exploiting the first order conditions of final good producers and intermediate good in the quantitative model described in Section 5. These conditions can be used to obtain a measure of total factor productivity (TFP) using data on wages and employment from which we estimate a stochastic process for its growth. We estimate the dynamic response of a variable to TFP shocks by regressing it on current and lagged values of the innovations derived from the estimated TFP growth process.

There are N cities that each produce a distinct intermediate good used as an input into the production of final goods. The production function for a representative firm producing intermediate goods in city i at date t is

$$y_{it} = s_{it} n_{y,it}^{\theta} k_{y,it}^{\gamma}, \tag{1}$$

where s_{it} is exogenous TFP for the city, $n_{y,it}$ is employment, $k_{y,it}$ is equipment, $\theta > 0$, $\gamma > 0$, and $\theta + \gamma \leq 1$.[10] The output of the final good at date t, Y_t is produced using inputs of city-specific intermediate goods according to

$$Y_t = \left[\sum_{i=1}^{N} y_{it}^{\chi} \right]^{\frac{1}{\chi}}, \tag{2}$$

where $\chi \leq 1$.

Our measurement of city-specific TFP relies on the following definition. For any variable v_{it}:

$$\hat{v}_{it} \equiv \ln v_{it} - \frac{1}{N} \sum_{j=1}^{N} \ln v_{jt}. \tag{3}$$

Subtracting the mean value of $\ln v_{jt}$ in each period eliminates variation due to aggregate shocks, allowing us to focus on within-city dynamics. Under the assumption of perfectly mobile equipment the rental rate of equipment is common to all cities. It then follows from the first order conditions of competitive final and intermediate good producers that

$$\Delta \hat{s}_{it} = \frac{1 - \gamma\chi}{\chi} \Delta \hat{w}_{it} + \frac{1 - \theta\chi - \gamma\chi}{\chi} \Delta \hat{n}_{y,it}, \tag{4}$$

[10] The additional subscripts on employment and equipment are used later to distinguish between employment and equipment used in the production of intermediate goods and residential construction.

where Δ is the first difference operator and w_i denotes the wage in city i. Applying the first difference operator addresses non-stationarity over the sample period. Our quantitative model appears non-stationary over samples of similar length, but nevertheless implies a stationary distribution in the levels of its variables.[11]

Given values for χ, θ and γ and data on wages and employment equation (4) can be used to measure $\Delta \hat{s}_{it}$, the growth rate of city-specific TFP.[12] Below we calibrate θ and γ using traditional methods and find a value for χ to match the model to Zipf's law. With calibrated values $\chi = 0.9$, $\theta = 0.66$ and $\gamma = .235$ we estimate a first order auto-regression in $\Delta \hat{s}_{it}$ with a statistically significant auto-correlation coefficient equal to 0.24 and standard deviation of the error term equal to 0.015. Wooldridge (2002)'s test of the null hypothesis of no first order serial correlation in the residuals yields a p-value of 0.28, which confirms that this specification is a good fit for the data.

A natural concern about measuring TFP with (4) is that it ignores agglomeration. Davis, Fisher, and Whited (2014) find statistically significant agglomeration effects in a model where agglomeration affects TFP endogenously through an externality in output per acre of land. It is straightforward to modify equation (4) to include agglomeration like this and it leads to the same measurement equation for the exogenous component of TFP except that the coefficients on wage and employment growth include the parameter governing the magnitude of the externality. When we re-estimate the TFP process using Davis et al. (2014)'s estimate of the externality parameter we find the serial correlation coefficient and the innovation standard deviation are a little different, falling to 0.20 and 0.013. While we do not include agglomeration in our model, we expect that doing so would reconcile the two sets of estimates but have little impact on our other results.[13]

We now show how to use the estimated TFP process (without agglomeration) to identify the dynamic responses of variables to exogenous local TFP shocks. Let e_{it} denote the residual from the estimated TFP growth auto-regression. Then, we estimate the dynamic response

[11]First differencing removes fixed effects if they are present in the data and so our empirical analysis is robust to them. However, as in Gabaix (1999), our quantitative model addresses the cross-section of city populations without appealing to fixed effects.

[12]We use total employment to measure $n_{y,it}$. In our quantitative model total employment includes workers in the residential construction and intermediate good sectors. Residential construction employment data are not available, but for some cities it is for the construction sector as a whole. When we redo our empirical analysis subtracting construction employment from the total we obtain almost identical results because construction is a small fraction of total employment. We measure TFP in the same way in the data and the model, with one exception pointed out below.

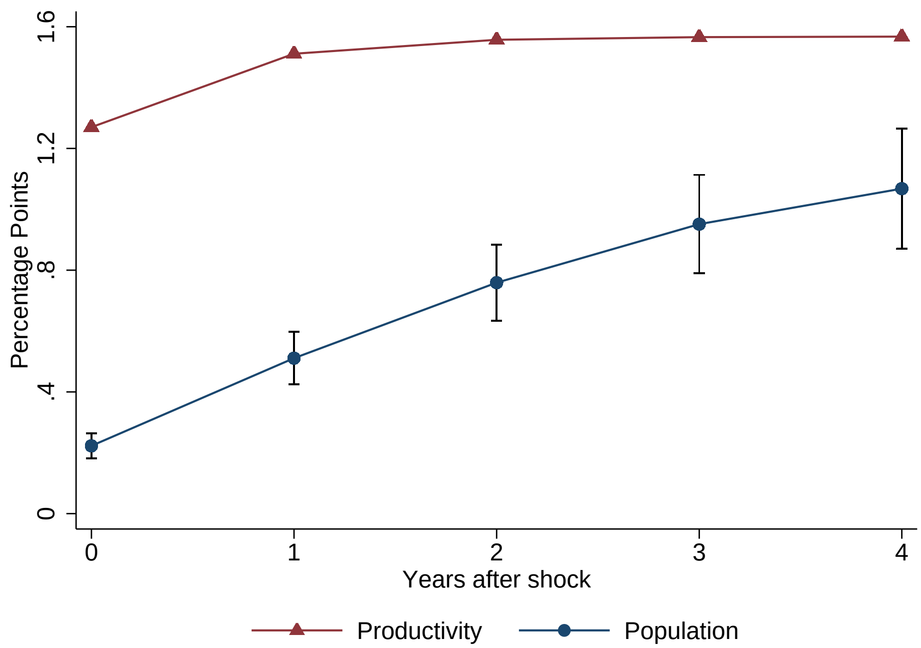
[13]Verifying this conjecture is beyond the scope of this paper. In Davis et al. (2014)'s model the externality amplifies the response of TFP to an exogenous TFP shock and makes it more persistent.

to a TFP shock of variable $\Delta\hat{x}_{it}$ as the coefficients b_0, b_1, $\ldots$, b_4 from the following panel regression:

$$\Delta\hat{x}_{it} = \sum_{l=0}^{4} b_l e_{it-l} + u_{it} \tag{5}$$

where u_{it} is an error term which is orthogonal to the other right-hand-side variables under the maintained hypothesis that the process for TFP growth is correctly specified. The dynamic response of $\hat{x}_{it}$ is obtained by summing the estimated coefficients appropriately. For the gross migration rates we replace $\Delta\hat{x}_{it}$ with the rates themselves (transformed as described above) in (5) and identify the dynamic responses with the estimated coefficients directly.

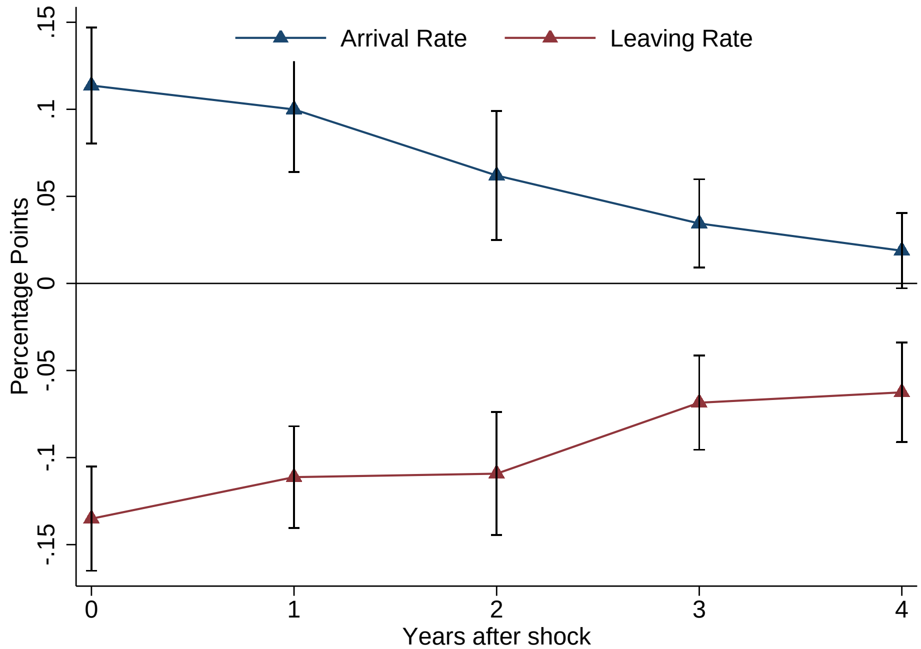
Figure 3: Responses of TFP and Population



Note: Point estimates along with 2 standard error bands.

Figure 3 displays the percentage point deviation responses of TFP and population to a 1 standard deviation impulse to TFP. This plot establishes the claim made in the introduction that TFP responds much like a random walk, rising quickly to its new long run level, and that population responds far more slowly. Figure 4 shows that the adjustment of population occurs along both the arrival and leaving margins, as suggested by our earlier discussion of Figure 2. On impact the arrival rate jumps up and the leaving rate jumps down and then both slowly return toward their long run levels. The indicated sampling uncertainty suggests that the arrival and leaving margins are about equally important for population adjustments

Figure 4: Responses of Arrival and Leaving Rates



Note: Point estimates along with 2 standard error bands.

to TFP shocks. Evidently improving local prospects influence population dynamics both by encouraging workers not to move and by attracting new workers to the city.

3 Modeling Migration

We now introduce our theory of migration by studying a simple, static model which abstracts from housing, equipment, and labor supply. This simplified model is used to develop intuition about migration choices; to describe how and why we can reproduce the relationships depicted in Figure 2; to establish that modeling gross migration is essential to understanding urban population dynamics; and to describe a simple decentralization that is helpful for calibrating our dynamic quantitative model. The main results extend to our quantitative model.

3.1 A Static Model

The economy consists of a large number of geographically distinct cities with initial population x. In each city there are firms which produce identical, freely tradeable consumption goods with the technology sn^{θ}, where s is a city-wide TFP shock, n is labor and $0 < \theta < 1$. There is a representative household with a unit continuum of members that are distributed across city types $z = (s, x)$ according to the measure μ. Each household member enjoys consumption, C, and supplies a unit of labor inelastically. After the TFP shocks have been realized, but before production takes place, the household decides how many of its members leave each city and how many of those chosen to leave move to each city. Once these migration decisions have been made, production and consumption take place.

The leaving decision is based on each household member receiving a *location-taste* shock ψ, with measure μ_l, that subtracts from their utility of staying in the city in which they are initially located. This kind of shock is used by KW in their measurement of migration costs as well as by Karahan and Rhee (2012) and Nenov (2012). To match the empirical relationship between gross and net migration we make a parametric assumption on the distribution of individual location-taste shocks in a city of type z:

$$\int_{-\infty}^{\bar{\psi}(l(z)/x)} \psi d\mu_l = -\psi_1 \frac{l(z)}{x} + \frac{\psi_2}{2} \left(\frac{l(z)}{x} \right)^2$$

where the parameters ψ_1 and ψ_2 are both non-negative and $\bar{\psi}\left(l(z)/x\right)$ is defined by

$$\frac{l(z)}{x} = \int_{-\infty}^{\bar{\psi}(l(z)/x)} d\mu_l.$$

This parameterization is U-shaped starting at the origin meaning that the first workers to leave a city are those for whom leaving raises their utility. As more people leave the remaining inhabitants are those who have a strong preference for staying. These features are consistent KW's evidence that individuals who move receive substantial non-pecuniary benefits and that non-movers would find it extremely costly if they were forced to move. For example, many individuals move to be near family members or find it very costly to move because they are already near family members. Subject to these shocks, the household determines how many of its members from each city must find new cities. Household members chosen to find new cities are called *leavers*.

When deciding where to send its leavers the household knows the distribution of city types μ but not the location of any specific type z. However, it can find a particular type by obtaining a *guided trip*, a form of directed migration. We choose a functional form for producing guided trips to match the evidence on gross and net migration. By giving up u units of utility each household member can produce $\sqrt{2}A^{-1/2}u^{1/2}$, $A \geqslant 0$, guided trips to her initial location. Therefore, to attract $a(z)$ workers to a city of the indicated type the household must incur a total utility cost of $(A/2)\left(a(z)/x\right)^2 x$.

The production of guided trips encompasses the many ways in which workers find new cities to live and work, including via informal contacts between friends and family, professional networks, specialized firms like head-hunters, advertising that promotes cities as desirable places to live and work, firms' human resource departments, and via recruiting by workers whose primary responsibility is some other productive activity. Clearly some of these activities are part of recruiting workers within a local labor market and as such would be included in any measurement of the vacancy costs typically assumed in models of labor market search and matching. Our approach can be thought of as including the portion of these activities devoted to attracting workers to a local labor market from other locations.

If a household member does not obtain a guided trip it can migrate to another city using *undirected migration*. Specifically, by incurring a utility cost τ a leaver is randomly allocated to another city in proportion to its initial population. Including undirected migration captures the idea that choosing to move to a particular city is often the outcome of idiosyncratic factors other than wages or housing costs that are difficult to model explicitly, such as attractiveness of amenities and proximity to family members.[14] Furthermore, it is natural to let people move to a location without forcing them to find someone to guide them.

We characterize allocations in this economy by solving the following planning problem:

$$\max_{\substack{\{C,\Lambda,a(z),\\ l(z),p(z)\}}} \left\{ \ln C - \int \left[\frac{A}{2}\left(\frac{a\left(z\right)}{x}\right)^2 x + \left(-\psi_1 \frac{l\left(z\right)}{x} + \frac{\psi_2}{2}\left(\frac{l\left(z\right)}{x}\right)^2\right) x \right] d\mu - \tau\Lambda \right\} \tag{6}$$

subject to

$$p\left(z\right) \leq x + a\left(z\right) + \Lambda x - l\left(z\right), \forall\, z \tag{7}$$

$$\int \left[a\left(z\right) + \Lambda x\right] d\mu \leq \int l\left(z\right) d\mu \tag{8}$$

[14]KW include both undirected and directed migration. It is undirected because to learn a location's permanent component of wages workers have to migrate there. It is directed because workers retain information about locations to which they have previously migrated and include this information in their current migration decision along with expectations about locations they have not visited already.

$$C \leq \int sp(z)^{\theta} \, d\mu \tag{9}$$

and non-negativity constraints on the choice variables. The variable Λ is the fraction of the household that engages in undirected migration. Since these workers are allocated to cities in proportion to their initial populations, Λ also corresponds to the share of a city's initial population that migrates to that city within the period. Constraint (7) states that population in a city is no greater than the initial population plus arrivals through guided trips and undirected migration minus the number of workers who migrate out of the city. Constraint (8) says that total arrivals can be no greater than the total number of workers who migrate out of cities and (9) restricts consumption to be no greater than total production, taking into account that each individual supplies a unit of labor inelastically, $n(z) = p(z)$, $\forall z$.

3.2 The Importance of Both Gross Migration Margins

It is necessary to include frictions on both gross migration margins to match the evidence depicted in Figure 2. Suppose $A = 0$ so that guided trips can be produced at no cost, but that household members continue to be subject to location-taste shocks, $\psi_1 > 0$ and $\psi_2 > 0$. Then it is straightforward to show

$$\begin{aligned} \frac{a(z)}{x} &= \max\left\{\frac{p(z) - x}{x} + \frac{\psi_1}{\psi_2}, 0\right\}; \\ \frac{l(z)}{x} &= \max\left\{\frac{\psi_1}{\psi_2}, -\left(\frac{p(z) - x}{x}\right)\right\}. \end{aligned}$$

Observe that as long as the net population growth rate, $(p(z) - x)/x$, is not too negative, the planner sets the leaving rate, $l(z)/x$ at the point of maximum benefits, ψ_1/ψ_2, and adjusts population using the arrival rate, $a(z)/x$, only. In this situation the leaving rate is independent of net population adjustments, contradicting Figure 2.

Now suppose that there are no location-taste shocks, $\psi_1 = \psi_2 = 0$, but it is costly to create guided trips, $A > 0$. In this case we find

$$\begin{aligned} \frac{l(z)}{x} &= \max\left\{-\left(\frac{p(z) - x}{x} - \Lambda\right), 0\right\}; \\ \frac{a(z)}{x} &= \max\left\{\frac{p(z) - x}{x} - \Lambda, 0\right\}. \end{aligned}$$

Without taste shocks the planner always goes to a corner: when net population growth is positive the leaving rate is set to zero, and when net population growth is negative the arrival rate is set to zero. Clearly the relationship between gross and net migration in this situation also contradicts Figure 2. We conclude that to be consistent with the relationship between gross and net migration, it is necessary to include frictions on both gross migration margins.

3.3 Migration Trade-offs

Now we use the planner's first order conditions to illustrate the trade-offs involved in allocating workers across cities. To begin we note that for the model to be consistent with Figure 2, the number of workers leaving a city and the number arriving to the same city using guided trips must both be strictly positive, $l(z) > 0$ and $a(z) > 0$. The reason we require $l(z) > 0$ is that gross out-migration is always positive in Figure 2. The reason we require $a(z) > 0$ is that otherwise there would be intervals of net migration in which arrival rates are constant, equal to Λ, which is also inconsistent with Figure 2. Therefore, unless otherwise noted, from now on we assume that $a(z) > 0$ and $l(z) > 0$.

Combining the first order conditions for Λ and $a(z)$ we obtain

$$\tau = \int Aa(z)\, d\mu.$$

This equation describes the trade-off between using guided trips and undirected migration. The marginal cost of raising the fraction of household members engaged in undirected migration is equated to the average marginal cost of allocating those household members using guided trips. The averaging reflects the fact that undirected migration allocates workers in proportion to each city's initial population. The first order conditions for $a(z)$ and $l(z)$ imply that

$$A\frac{a(z)}{x} = \psi_1 - \psi_2 \frac{l(z)}{x}.$$

Intuitively, migration out of a city increases to the point where the marginal benefits of doing so (recall that the location-taste shocks initially imply benefits to leaving a city) are equated with the marginal cost of attracting workers into the city.

Finally, it is helpful to consider the first order condition for population, $p(z)$:

$$\xi(z) = s\theta p(z)^{\theta-1}. \tag{10}$$

The term $\xi(z)$ is the Lagrange multiplier corresponding to (7) and therefore measures the value to the planner of an additional worker in a type-z city. Equation (10) says the shadow value of bringing an extra worker to a city equals its marginal product of labor. Combining (10) with the first order conditions for $a(z)$ and $l(z)$ we find that absent migration frictions, $A = \psi_1 = \psi_2 = 0$, efficiency involves equating cities' marginal products of labor so that the value of an additional worker is equated also. This contrasts with the Roback (1982) and Rosen (1979) model with free mobility in which equilibrium allocations are obtained by equating the level of utility across cities. The difference arises from our assumption of perfect consumption insurance. The same three first order conditions imply that migration frictions drive a wedge between marginal products of labor because heterogeneous initial populations imply differential costs of moving workers around. Nevertheless, taking into account the net costs of migration, workers are indifferent to where they move (the quantitative model developed below shares this property.)

3.4 Connecting Figure 2 to Population Adjustments

The planner's first order conditions reveal how gross migration relates to net migration. From the first order conditions for $a(z)$ and $l(z)$ and the population constraint, (7), it is straightforward to show that

$$\frac{a\left(z\right)}{x} + \Lambda = \frac{\psi_1}{A+\psi_2} + \frac{A}{A+\psi_2}\Lambda + \frac{\psi_2}{A+\psi_2}\left(\frac{p\left(z\right)-x}{x}\right). \tag{11}$$

The arrival rate is an affine function of the net migration rate $\left(p(z)-x\right)/x$ with the linear coefficient satisfying $0 < \psi_2/(\psi_2 + A) < 1$. Similarly the leaving rate is given by:

$$\frac{l\left(z\right)}{x} = \frac{\psi_1}{A+\psi_2} + \frac{A}{A+\psi_2}\Lambda - \frac{A}{A+\psi_2}\left(\frac{p\left(z\right)-x}{x}\right). \tag{12}$$

The leaving rate is also is an affine function of the net migration rate with the linear coefficient satisfying $-1 < -A/(\psi_2+A) < 0$. Equations (11) and (12) establish that gross migration in the model can be made consistent with Figure 2. This result is the underlying reason for our specifications of the location-taste shocks and the production of guided trips. Clearly, the relationship between gross and net migration depicted in Figure 2 places strong restrictions on the nature of migration frictions and will be useful in quantifying those frictions.

Modeling both gross migration margins is important for replicating Figure 2, but it also plays a crucial role in determining the speed of population adjustments. This can be seen by substituting for $a(z)$ and $l(z)$ in the original planning problem using (11) and (12), which simplifies it to

$$\max_{\{p(z),\Lambda\}}\left\{\ln\left(\int sp(z)^{\theta}\,d\mu\right)-\int\left[\Phi(\Lambda)+\frac{1}{2}\frac{A\psi_2}{A+\psi_2}\left(\frac{p(z)-x}{x}\right)^2 x\right]d\mu-\tau\Lambda\right\}$$

subject to:

$$\int p(z)\,d\mu \;=\; \int x d\mu \tag{13}$$

with non-negativity constraints on the choice variables and where $\Phi(\Lambda)$ is a quadratic function in Λ involving the underlying structural parameters ψ_1, ψ_2 and A. In deriving this simplified planning problem we have used the fact that (7) and (8) reduce to (13) and that this constraint holds with equality at the optimum. Similarly we have used (9) to substitute for consumption in the planner's objective function.

When the planning problem is written in this way we see that population adjustments do not depend directly on $a(z)$ and $l(z)$. Nevertheless modeling these decisions is crucial for understanding population dynamics because the coefficient that determines the speed of population's adjustment to shocks, $A\psi_2/(A+\psi_2)$, involves parameters governing them. Also notice that the reduced form costs of adjusting population are quadratic. This is a direct consequence of specifying the location-taste shocks and guided trip production function to reproduce Figure 2. In other words Figure 2 implies quadratic adjustment costs in net population adjustments.

Finally, notice that as long as $a(z) > 0$ and $l(z) > 0$, assumed in the statement of the simplified planning problem, population adjustments are independent of the undirected migration decision. Undirected migration is determined by the solution to

$$\tau = \frac{d\Phi(\Lambda)}{d\Lambda}.$$

Therefore when arrivals are always strictly positive undirected migration plays no role in net population adjustments. In the more general quantitative model arrivals are set to zero in especially undesirable cities. Still, for most cases arrivals are strictly positive so that undirected migration is essentially irrelevant for our results. This is a useful property given that there is little evidence on the magnitude of undirected migration in the data. Nevertheless we include it because, as noted previously, otherwise workers could only move by obtaining a guided trip and we think this is implausible.

3.5 One Possible Decentralization

Decentralizing the planning problem is useful for calibrating the quantitative model. The key challenge involved is how to treat guided trips. One valid approach is to have guided trips allocated entirely within the household through home production without any market interactions. We view guided trips in the model as an amalgam of both market and non-market activities and so we think a more natural approach involves both activities. We now consider such a decentralization.

Markets are competitive. Firms in a city of type z hire labor at wage $w(z)$ and produce consumption goods to maximize profits. Household members initially located in a type-z city produce $a_m(z)$ guided trips to that city which they sell to prospective migrants at price $q(z)$. The household also home produces guided trips, denoted by $a_h(z)$. Let $m(z)$ denote the total number of guided trips to z-type cities purchased by household members in the market.

The representative household solves the following optimization problem

$$\max_{\substack{\{C,\Lambda,m(z),\\ a_m(z),a_h(z),\\ l(z),p(z)\}}} \left\{ \ln C - \int \left[\frac{A}{2} \left(\frac{a_m(z) + a_h(z)}{x} \right)^2 x + \left(-\psi_1 \frac{l(z)}{x} + \frac{\psi_2}{2} \left(\frac{l(z)}{x} \right)^2 \right) x \right] d\mu - \tau\Lambda \right\} \tag{14}$$

subject to:

$$\begin{aligned} C + \int q(z)\, m(z) d\mu &= \int q(z)\, a_m(z) d\mu + \int w(z)\, p(z) d\mu + \int \Pi(z) d\mu && (15) \\ p(z) &= x + m(z) + a_h(z) + \Lambda x - l(z), \; \forall z && (16) \\ \int [m(z) + a_h(z) + \Lambda x]\, d\mu &= \int l(z)\, d\mu && (17) \end{aligned}$$

along with non-negativity constraints on the choice variables. Equation (15) is the household's budget constraint where Π denotes profits from owning the firms. Equation (16) states that the population of a city after migration equals the initial population plus migrants from guided trips and undirected migration less the initial population that migrates out of the city. Finally, equation (17) states that the household members that migrate to cities must equal the number of household members that migrate out of cities.

The unique competitive equilibrium is defined in the usual way. Using the market clearing condition $m(z) = a_m(z)$ and the first order conditions of the household's problem we verify that a competitive equilibrium only determines the total number of guided trips into a city, $a_m(z) + a_h(z)$. The composition of these guided trips between market and non-market activities is left undetermined.[15]

This particular decentralization makes it possible to calculate the total value of guided trips. In particular, as long as there are some guided trips purchased in the marketplace their total value is given by $q(z)a(z)$, with $a(z) = a_m(z) + a_h(z)$ and $q(z) = CAa(z)/x$. We use the total value of guided trips to calibrate our model to the estimate of average moving costs in KW.

4 Urban Population Dynamics with Housing

We expect housing to influence urban population dynamics for the reasons discussed in the introduction: it is costly to build quickly, durable and immobile. This section studies a simple model to explain why these factors may be important. The model borrows the geography and production structure from the previous section. There are three differences with that model: individuals have a preference for housing services; to emphasize the role of housing, the model excludes migration frictions; and to study dynamics the model introduces infinitely lived households.

To analyze this simple model it is convenient to exploit the fact, discussed further below in the context of the quantitative model, that the unique stationary competitive equilibrium can be obtained as the solution to a representative city planner's problem that maximizes local net surplus taking economy-wide variables as given, where the economy-wide variables are constrained to satisfy particular side conditions. Since here we are only interested in the qualitative implications of housing we ignore the side conditions and study the city planner's problem assuming the economy-wide variables are exogenous. When we analyze our quantitative model below we take into account the relevant side conditions so that the economy-wide variables are determined endogenously.

[15] For details see the online technical appendix.

Gross surplus in the representative city is given by

$$E_0 \sum_{t=0}^{\infty} \beta^t \left\{ s_t p_t^{\theta} + H \ln(\frac{h_t}{p_t}) p_t \right\}$$

where E_0 denotes the date $t = 0$ conditional expectations operator and $0 < \beta < 1$ is the household's time discount factor. Housing services are perfectly divisible so that each individual in the city enjoys h_t/p_t units of housing services where h_t denotes the local housing stock. The total utility individuals in the city derive from housing is given by $H \ln(h_t/p_t)p_t$, $H > 0$. We assume logarithmic preferences for housing here and below because they imply housing's share in household expenditures is constant across cities, which is consistent with evidence reported by Davis and Ortalo-Magné (2011). The planner must give up $\eta > 0$ units of surplus for each individual it brings to the city and employ in the production of consumption goods.

Within this framework we consider the speed of adjustment of population to a one time permanent change to TFP. To be concrete, suppose the city is in a steady state at $t = 0$ with $s = s_0$ and then at date $t = 1$ it faces a one-time unanticipated permanent change in TFP to $s = s^*$. We consider the adjustment of population to this unanticipated change in TFP under three scenarios for housing.

In the first scenario the planner can rent housing services from other cities at the exogenous price r_h. This assumption is equivalent to assuming that housing is perfectly mobile across cities. Equilibrium in this scenario is characterized by the first order conditions for population and housing:

$$H \ln(\frac{h_t}{p_t}) - H + s_t \theta p_t^{\theta-1} = \eta; \tag{18}$$

$$H \frac{p_t}{h_t} = r_h. \tag{19}$$

Condition (18) states that population is chosen to equate the marginal benefit of an additional individual working in the city to the cost of bringing that individual to the city, where the former is the sum of the marginal product of the individual plus the housing services she enjoys. The second condition equates the marginal utility of an extra unit of housing services with its cost. Replacing housing per individual in (18) using (19) yields

$$H \ln(\frac{H}{r_h}) - H + s_t \theta p_t^{\theta-1} = \eta. \tag{20}$$

The key feature of (20) is that it does not include the housing stock, h_t. This means that after the unanticipated change in TFP population jumps immediately to its new permanent level $p = p^*$ found as the solution to (20) with $s_t = s^*$. So, when housing is perfectly mobile it is irrelevant for population dynamics.

Now assume the city is endowed with h_0 units of housing at $t = 0$ and that housing is immobile, meaning that it cannot be rented from or to any other city. In addition, suppose the change in TFP at $t = 1$ coincides with the onset of a potentially different exogenous path of the local housing stock satisfying

$$\ln h_t - \ln h^* = \rho_h^{t-1} (\ln h_0 - \ln h^*) \tag{21}$$

for $t \geq 1$, $0 \leq \rho_h < 1$, and h^* is the new long run level for h. Equilibrium population is determined by (18) conditional on (21). We consider two cases for this scenerio.

First suppose that the local, immobile housing stock does not change with TFP, that is $h^* = h_0$. From (18) after the change in TFP p jumps immediately to its new level given by the value p^* that solves this equation. The new long run level of population depends on h_0 but the speed of adjustment to p^* is the same as when housing is perfectly mobile. In other words the presence of local, immobile housing is not sufficient for housing to affect population dynamics.

The second case is the new long run level of housing changes with TFP, $h^* \neq h_0$. We approximate the transition of population to its new steady state in this case by log linearizing (18) around $\ln p^*$ and $\ln h^*$. This yields

$$\ln p_t - \ln p^* = \frac{H}{H + s^* \theta (1-\theta) p^{*\theta - 1}} \rho_h^{t-1} (\ln h_0 - \ln h^*) .$$

In this case the speed of convergence of population to its new steady state p^* is directly related to the speed of convergence of housing to its new steady state through ρ_h. If the adjustment of housing is immediate, $\rho_h = 0$, then population's adjustment is instantaneous as in the case when $h^* = h_0$. If $0 < \rho_h < 1$ then population adjusts in proportion to the adjustment of housing.

We conclude that housing must be immobile and adjust slowly to changes in local productivity for it to affect population dynamics. It follows that a plausible quantitative analysis of urban population dynamics in response to TFP shocks must include endogenous immobile housing and include the possibility of its slow adjustment. Natural candidates for influencing the speed of adjustment of housing are construction depending on local resources and durability.

5 The Quantitative Model

This section describes the model we use to quantify gross migration and housing's influence on urban population dynamics, integrating and extending the models of the previous two sections. The model introduces dynamics to the gross migration decision; endogenizes housing; and includes a labor supply decision. The latter is introduced because changes in labor supply are a natural alternative to migration for a city to adjust to labor demand shocks. After describing the model environment we characterize the its unique stationary competitive equilibrium as the solution to a representative city planning problem with side conditions.

5.1 The Environment

As before the economy consists of a continuum of geographically distinct locations called cities that are subject to idiosyncratic TFP shocks. Cities are distinguished by their stock of housing, h, initial population, x, and the current and lagged TFP, s and s_{-1}. The measure over these state variables is given by μ.[16]

Within cities there are three production sectors corresponding to intermediate goods, housing services and construction. The representative firm of each sector maximizes profits taking prices as given. Intermediate goods are distinct to a city and imperfectly substitutable in the production of the freely tradeable final goods non-durable consumption and durable equipment. The technologies for producing intermediate and final goods are identical to those underlying our estimates of TFP, described in equations (1) and (2).[17] Housing services are produced by combining residential structures with land, b_r, according to $h^{1-\zeta}b_r^{\zeta}$, $0 < \zeta < 1$. Following the convention that the prime symbol denotes next period's value of a variable, residential structures evolve as

$$h' = (1 - \delta_h)\, h + n_h^{\alpha} k_h^{\vartheta} b_h^{1-\alpha-\vartheta}, \tag{22}$$

[16]Current and lagged TFP both appear in this list to accommodate the estimated TFP process described in Section 2.2. This is discussed further below.

[17]Equations (1) and (2) are written in terms of the location of a city, indexed by i, but here it is convenient to index them by the type of the city as represented by its state vector (h, x, s, s_{-1}).

where the factor shares are restricted to $\alpha > 0$, $\vartheta > 0$ and $\alpha + \vartheta < 1$, and $0 < \delta_h < 1$ denotes housing's depreciation rate. The last term in (22) represents housing construction. Local TFP s does not impact residential construction, reflecting our view that residential construction productivity is not a major source of cross-city variation in TFP. Equation (22) embodies our assumptions that residential structures are immobile, durable and costly to build quickly. The latter follows because residential construction requires local labor and land which have alternative uses in intermediate goods production and housing services. We assume that equipment used in production and construction is homogenous.

There is an infinitely lived representative household that allocates its unit continuum of members across the cities. The household faces the same migration choices described in Section 3, but being infinitely lived it takes into account the affects of current migration decisions on its members' allocation across cities in future periods. In particular, it is now bound by the constraint

$$x' = p \tag{23}$$

in each city where p continues to denote the post-migration population of a city. The household's members have logarithmic preferences for consumption and housing services in the city in which they are located. They also face a non-trivial labor supply decision. We assume that each period, after the migration decisions have been made, but before production and construction take place, individual household members receive a labor disutility shock φ with measure μ_n. Similar to our treatment of migration costs we make a parametric assumption for the average disutility of working. Specifically, if the household decides n of its members in a city will work for a year these costs are specified as

$$\int_{-\infty}^{\bar{\varphi}(n/p)} \varphi d\mu_n = \phi \left(\frac{n}{p}\right)^{\pi},$$

where $\phi > 0$, $\pi \geq 1$ and $\bar{\varphi}\,(n/p)$ is defined by

$$\frac{n}{p} = \int_{0}^{\bar{\varphi}(n/p)} d\mu_n.$$

The parameter π governs the elasticity of a city's labor supply with respect to the local wage.

5.2 Stationary Competitive Equilibrium

This model has a unique stationary competitive equilibrium. Since it is a convex economy with no distortions, the welfare theorems apply and so the equilibrium allocation can be obtained by solving the problem of a social planner that maximizes the expected utility of the representative household subject to the resource feasibility constraints. However, it is more useful to characterize the equilibrium allocation as the solution to a representative city social planner's problem with side conditions. This approach to studying equilibrium allocation follows Alvarez and Shimer (2011) and Alvarez and Veracierto (2012).

The city planner enters a period with the state vector $z = (h, x, s, s_{-1})$. Taking as given aggregate output of final goods, Y, the marginal utility of consumption, λ, the shadow value of adding one individual to the city's population exclusive of the arrival and leaving costs, $\lambda\eta$, the shadow value of equipment, λr_k, the arrival rate of workers through undirected migration Λ, and the transition function for TFP, $Q(s'; s, s_{-1})$, the representative city planner solves

$$V(z) = \max_{\substack{\{n_y, n_h, k_y, k_h, \\ h', b_r, b_h, p, a, l\}}} \left\{ \lambda \frac{1}{\chi} Y^{1-\chi} \left[s n_y^{\theta} k_y^{\gamma} \right]^{\chi} + H \ln \left(\frac{h^{1-\zeta} b_r^{\zeta}}{p} \right) p - \phi (n_y + n_h)^{\pi} p^{1-\pi} \right.$$

$$-\lambda r_k (k_y + k_h) - \lambda \eta (a + \Lambda x - l)$$

$$\left. -\frac{A}{2} \left(\frac{a}{x} \right)^2 x - \left[-\psi_1 \frac{l}{x} + \frac{\psi_2}{2} \left(\frac{l}{x} \right)^2 \right] x + \beta \int V(z') \, dQ(s'; s, s_{-1}) \right\}$$

subject to

$$\begin{aligned} p &= x + a + \Lambda x - l \qquad (24) \\ n_y + n_h &\leq p \\ b_r + b_h &= 1 \end{aligned}$$

plus (22), (23), and non-negativity constraints on the choice variables.

The planner's objective is to maximize the expected present discounted value of net local surplus. To see this note that the first two terms are the value of intermediate good production and the housing services consumed in the city. The next five terms comprise the contemporaneous costs to the planner of obtaining this surplus: the disutility of sending the indicated number of people to work; the shadow cost of equipment used in the city; and the disutility of net migration inclusive of guided trip production and location-taste shocks.

The last term is the discounted continuation value given the updated state vector. Constraining the achievement of the city planner's objective are the local resource constraints, the housing and population transition equations and the non-negativity constraints on the choice variables. Note that in the statement of the land constraint we have normalized the local endowment of residential land to unity and used the fact that land used for current housing services cannot be built on in the same period.[18]

Let $\lambda\xi(z)$ denote the Lagrange multiplier corresponding to constraint (24) in the city planner's problem. This function represents the shadow value of bringing an additional individual to a type-z city. From the first order conditions of the city planner's problem it is easy to show that

$$\lambda\xi\left(z\right)=\begin{cases} A\left[\frac{a(z)}{x}\right]+\lambda\eta, \text{ if } a\left(z\right)>0, \\ \left[\psi_1-\psi_2\left(\frac{l(z)}{x}\right)\right]+\lambda\eta, \text{ if } l\left(z\right)>0. \end{cases} \tag{25}$$

which takes into account the fact that $a(z)=l(z)=0$ will never occur in equilibrium. Comparing equation (25) to the first order conditions for $a(z)$ and $l(z)$ in the static model of Section 3 we see that if gross migration rates are positive then the shadow value of a migrant is related to migration costs in the same way.

The unique stationary allocation is the solution to the city planner's problem that satisfies particular side conditions we now describe. To begin, let $\{n_y, n_h, k_y, k_h, h', b_r, b_h, p, a, l\}$ denote the optimal decision rules (which are functions of the state z) for the city planner's problem that takes $\{Y, \lambda, \eta, r_k, \Lambda\}$ as given and μ be the invariant distribution generated by the optimal decision rules $\{h', p\}$ and the transition function Q. In addition define the aggregate stock of equipment and per capita consumption:

$$\begin{aligned} K &= \int\left(k_y+k_h\right)d\mu; \\ C &= Y-\delta_k K, \end{aligned}$$

where $0<\delta_k<1$ denotes equipment's depreciation rate. Now suppose the following equa-

[18] In our calibration $0<\theta+\gamma<1$ which implies the presence of a fixed factor in the production of the city's intermediate good. As written the production function assumes that the supply of this fixed factor is constant (equal to one) across cities. One interpretation of this fixed factor is that it represents commercial land. Under this interpretation commercial land cannot be converted into residential land and vice versa.

tions are satisfied

$$Y = \left\{ \int \left[s n_y(z)^{\theta} k_y(z)^{\gamma} \right]^{\chi} d\mu \right\}^{\frac{1}{\chi}} \tag{26}$$

$$\lambda = \frac{1}{C} \tag{27}$$

$$\int a(z)\, d\mu + \Lambda = \int l(z)\, d\mu \tag{28}$$

$$r_k = \frac{1}{\beta} - 1 + \delta_k \tag{29}$$

$$\lambda \int [\xi(z) - \eta]\, x d\mu - \tau \leq 0, (= 0 \text{ if } \Lambda > 0) \tag{30}$$

Then $\{C, K, n_y, n_h, k_y, k_h, h', b_r, b_h, p, \Lambda, a, l\}$ is a steady state allocation.[19]

In the steady state the variables taken as given in the city planner's problem solve the side conditions given by (26)–(30). Equation (26) expresses aggregate output in terms of intermediate good production in each city. This equation is the theoretical counterpart to equation (2) used to estimate city-specific TFP. The marginal utility of consumption is given by equation (27). Equation (28) states that total in-migration equals total out-migration. Equation (29) defines the rental rate for equipment. The last side condition (30) determines steady state undirected migration. It turns out to be identical to the first order condition for Λ in the static model.

The function $\xi(z)$ in (30) represents the value to the city planner of bringing an additional individual to the city. It is central to the determination of migration in the model and can be shown to satisfy

$$\begin{aligned} \xi(z) = {} & C\phi \left[n_y(z) + n_h(z)\right]^{\pi} (\pi - 1)\, p(z)^{-\pi} + CH \ln\left(\frac{h(z)^{\varsigma}\, b_r(z)^{1-\varsigma}}{p(z)} \right) - CH \\ & + \beta \int \left(CA \left[\frac{a(z')}{p(z)}\right]^2 + C\psi_2 \left[\frac{l(z')}{p(z)}\right]^2 + \Lambda\left[\xi(z') - \eta\right] + \xi(z' \right) dQ(s'; s, s_{-1}) \end{aligned} \tag{31}$$

The value of bringing an additional worker to a city comprises four terms: the benefits of

[19] We prove this result in the online technical appendix. We take a traditional dynamic programming approach to solving the city planner's problem. This is complicated substantially by the fact that there are four state variables in the city planner's problem, two of them endogenous. Furthermore the TFP process has a large domain. We overcome the computational challenges of a large dimensional and high variance state space in two main ways. First we exploit a parsimonious spline method to approximate the planner's value function and within-period return function. Second we take advantage of the large number of processors contained in graphics cards. For details see the appendix below.

obtaining a better selection of worker disutilities given the same amount of total employment $n_y + n_h$; the benefits of the local housing services that the additional person will enjoy; the costs of reducing housing services for everybody else; and the expected discounted value of starting the following period with an additional person. This last term includes the benefits of having an additional person producing guided trips to the city, the benefits of obtaining a better selection of location-taste shocks (given the same number of individuals leaving the city), and the benefits of attracting additional people to the city through the undirected migration technology.

When there are no migration frictions, $A = \psi_1 = \psi_2 = 0$, equation (25) implies that the marginal value of bringing an additional individual to a city is equated across cities just as in the static version of the model, $\xi(z) = \eta$, $\forall z$ (see Section 3.3). However, unlike the static case this does not imply that wages are equated across cities. Instead, equation (31) says that the marginal savings in worker disutility plus the marginal impact on the utility of housing services is equated. When in addition to $A = \psi_1 = \psi_2 = 0$ housing structures are made perfectly mobile across cities, the same condition is obtained because land remains immobile. Finally, when land is also made mobile, then the marginal savings in work disutility and the marginal utility of housing services are each equated across cities.

6 Calibration

We now calibrate the steady state competitive equilibrium to U.S. data.[20] Our calibration has two important characteristics. First, the city-specific TFP process is chosen to match our estimates presented in Section 2.2 thereby pinning down the model's exogenous source of persistence and volatility. Second, the calibration targets for the remaining parameters involve features of the data that are not primary to our study. So, for instance, we do not choose parameters to fit our estimated response of population to a TFP shock. The model's response of population to a TFP shock is the consequence of the estimated TFP process and the remaining parameters that are chosen to fit other features of the data.

In addition to specifying the stochastic process for TFP we need to find values for 16 parameters:

$$\theta, \gamma, \alpha, \vartheta, \delta_k, \delta_h, \beta, H, \zeta, \pi, \phi, \psi_1, \psi_2, A, \tau, \chi.$$

These include the factor shares in production and construction, depreciation rates for equip-

[20]Except where noted the aggregate data used to calibrate our model is obtained from Haver Analytics.

ment and structures, the discount factor, the housing coefficient in preferences, land's share in housing services, and the parameters governing labor supply, migration, and intermediate goods' substitutability in final goods production.

We calibrate these parameters conditional on a given quantity of undirected migration Λ determined by τ. For larger values of τ undirected migration is relatively small so that $a(z) > 0, \ \forall\, z$. In these cases the behavior of the model is invariant to the specific value of τ. For smaller values of τ undirected migration is large and $a(z) = 0$ for some z. In these cases the behavior of the model is affected. It turns out that even for seemingly large steady state Λ corner solutions for $a(z)$ are either non-existent or extremely rare. We set our baseline so that the undirected arrival rate is 3.8%, roughly 70% of all moves.[21]

The baseline calibration for the assumed value of τ is summarized in Table 1. There we indicate for each parameter the proximate calibration target, the actual value for the target we obtain in the baseline calibration, and the resulting parameter value. In the remainder of this section we discuss the calculations underlying Table 1. We begin with the novel aspects of our calibration which involve the parameters governing migration, the city-level TFP process, the elasticity of substitution of city-specific intermediate goods in final good production, and labor supply.

6.1 Migration Parameters

Section 3.4 establishes that the migration parameters A, ψ_1 and ψ_2 are central to determining the speed of population's adjustment to TFP shocks in the model. Fortunately there is evidence at hand that makes assigning values to these parameters straightforward. First, conditional on a value for A reproducing reproducing Figure 2 pins down ψ_1 and ψ_2. To reproduce Figure 2 we set the constant and slope coefficients in equation (11) to their empirical counterparts displayed in Figure 2.[22] In particular

$$\begin{aligned}\frac{\psi_1}{A+\psi_2} + \frac{A}{A+\psi_2}\Lambda &= 5.5;\\ \frac{\psi_2}{A+\psi_2} &= 0.57.\end{aligned}$$

[21] The specific value is $\tau = 1$. For this value the baseline calibration has 0.3% of city-year observations involving zero arrivals.

[22] The constant term for arrivals in Figure 2 is the sample average gross migration rate, but gross migration is declining over our sample. Our measure of migration costs depends on this choice and so in principle our findings do as well. We examined the implications of calibrating to the average gross migration rate at the start and end of our sample and found that our results are substantively the same.

Table 1: Baseline Calibration

Parameter	Parameter Description	Calibration Target	Target Value	Actual Value	Parameter Value
θ	Labor's share in intermediate goods	$\int w\left[n_y+n_h\right]d\mu/GDP$	0.64	0.64	0.66
α	Labor's share in construction	$\int n_h d\mu/\int\left[n_y+n_h\right]d\mu$	0.042	0.042	0.41
β	Discount factor	Real interest rate	0.04	0.04	0.9615
γ	Intermediate goods' equipment share	K_y/GDP	1.53	1.53	0.235
δ_k	Depreciation rate of equipment	$\delta_k K/GDP$	0.16	0.16	0.104
ϑ	Equipment's share in construction	K_h/GDP	0.022	0.022	0.05
δ_h	Depreciation rate of structures	I/GDP	0.064	0.064	0.045
ζ	Land's share in housing services	$\int q^b b_r d\mu/\left[\int q^h h d\mu+\int q^b b_r d\mu\right]$	0.37	0.36	0.215
H	Housing coefficient in preferences	$\int q^h h d\mu/GDP$	1.55	1.50	0.205
ϕ	Labor disutility	$\int\left[n_y+n_h\right]d\mu/\int p d\mu$	0.63	0.63	1.61
π	Labor supply elasticity	$\partial\ln\left[n_y+n_h/p\right]/\partial\ln w$	0.24	0.25	5.0
ψ_1	Taste shock slope	Mean arrivals	5.5	5.5	6.12
ψ_2	Taste shock curvature	Slope of arrivals versus net	0.57	0.57	44.4
A	Guided trip cost	Average moving costs/average wages	-1.9	-1.9	33.5
χ	Intermediate goods' complimentarity	Zipf's law for population	-1.0	-1.3	0.9
g	Drift in technology	Zipf's law for TFP	-3.5	-3.4	-0.0017
ρ	TFP lag coefficient	Serial corr. of TFP growth	0.24	0.22	0.35
σ	TFP innovation std. err.	TFP growth innovation std. err.	0.015	0.015	0.019

Note: The calibration is based on $\tau=1$. See the text for details. The actual values for the serial correlation and innovation of TFP growth are based on simulations. The underlying parameter values for the TFP process are somewhat different due because of the discrete approximation we use.

To identify A we take advantage of KW's estimate of the average net cost of migration for those who move. Specifically, we match the statistic defined as the average net cost of migration of those who move divided by average wages where we take the latter from KW as well.[23] It is straightforward to replicate their concept of moving costs in our model. In KW, net moving costs sum two components of the utility flow of an individual in the period of a move. One component called "deterministic moving costs" is a function of the distance of the move, whether the move is to a location previously visited or not, the age of the mover, and the size of the destination location. The second component is the difference between idiosyncratic benefits in the current and destination location. We interpret guided trips in our model as representing the first component and the location-taste shocks the second one. Consequently we measure average moving costs of individuals who move as

$$\frac{\int q(z)a(z)\ d\mu + C\tau\Lambda}{\int a(z)\ d\mu + \Lambda} + \frac{C\int\left(-\psi_1\frac{l(z)}{x} + \psi_2\left(\frac{l(z)}{x}\right)^2\right)xd\mu}{\int l(z)d\mu}.$$

Average wages are simply

$$\frac{\int w(z)\left[n_y(z) + n_h(z)\right]d\mu}{\int\left[n_y(z) + n_h(z)\right]d\mu},$$

where wages in a type-z city, $w(z)$, equal the marginal product of labor.

There are two potential drawbacks to using KW's estimate of moving costs. First, KW identify moving costs using individual-level data. Since individual behavior is not observable in our model we cannot replicate their estimation strategy. Still, our model implies a value for KW's moving cost statistic so it is natural to take advantage of their estimates. Of more concern is the fact that KW estimate moving costs using data on the frequency of inter-*state* moves, while our quantitative model describes inter-*city* moves. Inter-city moves are more frequent than moves between states. Consequently it is possible that KW would have estimated a different value for moving costs had they had data on all inter-city moves, in which case we would be calibrating our model to the wrong value. This suggests it is important to quantify any bias in KW's estimate arising from their focus on inter-state moves only. We do this using a calibrated variant of their model which suggests any bias is likely to be small. Details of how we arrive at this conclusion are in the appendix below.

[23] Using KW's estimates, average net moving costs of those who move divided by average wages equal -1.9. This value equals the ratio -\$80,768/\$42,850. 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.

We conclude that using KW's estimate is valid in our context.

6.2 TFP and Substitutability of City-specific Goods

The calibration of the substitution parameter χ and the stochastic process for TFP are interconnected because χ is used to measure TFP. When we measure TFP using the procedure described in Section 2.2 its growth rate is well-represented as a stationary AR(1) process which is non-stationary in levels and therefore inconsistent with a steady state. To overcome this we assume a reflecting barrier process for TFP:

$$\ln s_{t+1} = \max\left\{g + (1+\rho)\ln s_t - \rho \ln s_{t-1} + \varepsilon_{t+1}, \ln s_{\min}\right\}. \tag{32}$$

where $\varepsilon_{t+1} \sim N(0, \sigma^2)$, $g < 0$ and $\rho > 0$. With this process TFP growth is approximately AR(1), while its level is stationary due to having a negative drift and being reflected at the barrier $\ln s_{\min}$ (which we normalize to zero).[24]

The case $\rho = 0$ was used by Gabaix (1999) to explain the cross section distribution of cities by population. In this case the invariant distribution has an exponential upper tail given by

$$\Pr\left[s_t > b\right] = \frac{d}{b^{\omega}}$$

for scalars d and b. A striking characteristic of cities is that when s measures a city's population one typically finds that $\omega \simeq 1$. Equivalently a regression of log rank on log level of city populations yields a coefficient close to -1. This property is called Zipf's law and so we refer to ω as the Zipf coefficient. The case $\rho > 0$, which applies when TFP growth is serially correlated, has not been studied before. Simulations suggest this case behaves similarly to the $\rho = 0$ case in that it has an invariant distribution with an exponential-like upper tail. We verify below that a version of Zipf's law holds for TFP and so using the reflecting barrier process with $\rho > 0$ seems justified.

Our calibration of χ and (32) proceeds as follows. For a given χ (and θ and γ which are calibrated independently as discussed below) we measure TFP in the data following the procedure in Section 2.2, obtain its Zipf coefficient, and estimate an AR(1) in its growth rate. We then find the g, ρ and σ to match the Zipf coefficient and the serial correlation and innovation variance of the estimated AR(1) using data simulated from our model and based

[24] Coen-Pirani (2010) considers a stationary AR(2) process for the level of TFP, calibrating it to match serial correlation in net state-to-state worker flows.

on these parameters calculate the model's population Zipf coefficient. The calibrated value of χ is the one that generates a population Zipf coefficient that is as close as possible to the one we find in the data, 1.0. The best fit is at $\chi = 0.9$ with a population Zipf coefficient equal to 1.3. The corresponding values of g, ρ and σ are in Table 1.

Figure 5: Zipf's Laws for Population and TFP

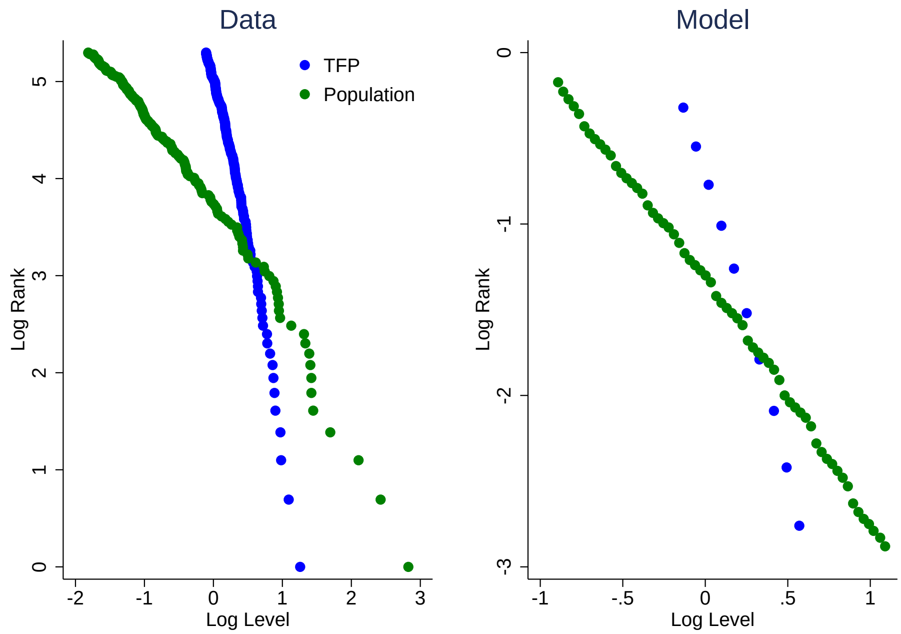


Figure 5 demonstrates the model's success at replicating the two Zipf's laws by plotting log rank versus log level for population and TFP with empirical and simulated data.[25] Log level of TFP is calculated using (4) without the first difference operator. The scales in the plots differ because we use the cumulative distribution functions to measure rank in the model and TFP's domain is narrower in the model because of the computational cost of matching the data.[26] Notice that TFP's Zipf coefficient is larger than population's in the data. This feature arises naturally in the model because population tends to be allocated away from lower toward higher TFP cities. Equivalently, the long run response of population to a

[25]The left plot is constructed using the top 200 cities by population in 1990. The right plot excludes the lower 5% of cities to be comparable and also the top 5% because of the technical necessity of limiting the upper bound of the TFP domain which introduces bunching at the top end of the population distribution.

[26]The narrower domain does not matter for our calibration. For example, the migration parameters are based on Figure 2 and KW's estimate of migration costs. In the model the former does not depend on the level of TFP and the latter depends on the distribution of TFP growth which is essentially independent of the domain. Our quantitative analysis is based on growth rates and so is similarly independent of the underlying domain.

TFP shock is larger than that of TFP. Luttmer (2007) finds a similar relationship between employment and TFP in an equilibrium model of firm size.

6.3 Labor Supply

The labor disutility parameters are calibrated to match statistics involving employment to population ratios. The multiplicative parameter ϕ is chosen to match the ratio of aggregate civilian employment to population obtained from Census Bureau data. The curvature parameter π is chosen using the first order condition for labor supply in a city. In the model's decentralization the representative household chooses labor supply to equate the disutility of putting an additional household member to work in a city with that city's wage. This implies:

$$(1-\pi)\left(\Delta\hat{n}_{it}-\Delta\hat{p}_{it}\right)+\Delta\hat{w}_{it}=0,$$

where n is the sum of n_y and n_h and the "delta" and "hat" notation is described in Section 2.2. Using the methods described in Section 2.2, we estimate the dynamic responses of $\Delta\hat{n}_{it}$, $\Delta\hat{p}_{it}$ and $\Delta\hat{w}_{it}$ to a local TFP shock and calibrate π so that this equation holds in the period of a shock. Note that this procedure does not force the model to match these variables' individual impulse responses even in the period of a shock.

6.4 Remaining Parameters

Our strategy for calibrating the remaining parameters borrows from studies based on the neo-classical growth model. Several calibration targets involve GDP and we measure this in the model as

$$GDP = Y + I, \tag{33}$$

where Y is output of non-construction final goods and I is residential investment. Residential investment is measured as the value in contemporaneous consumption units of the total additions to local housing in a year. Specifically,

$$I=\int\left[\beta\int q_h(z')dQ(s';s,s_{-1})\right]n_h(z)^{\alpha}k_h(z)^{\vartheta}b_h(z)^{1-\alpha-\vartheta}d\mu$$

where q_h denotes the price of residential structures. This price is obtained as the solution to the following no arbitrage condition

$$q_h(z) = r_h(z) + (1 - \delta_h)\beta \int q_h(z')\, dQ(s'; s, s_{-1})$$

where the rental price of residential structures, r_h, equals the marginal product of structures in the provision of housing services. The National Income and Product Accounts (NIPA) measure of private residential investment is the empirical counterpart to I. Our empirical measure of Y is the sum of personal consumption expenditures less housing services, equipment investment and private business inventory investment. Because our model does not include non-residential structures investment, government expenditures and net exports we exclude these from our empirical concept of GDP.

Our measurement of model GDP and wages excludes the value of guided trip services, which might be problematic. For example, workers produce guided trips and in principle they should be compensated for this. Using the decentralization discussed in Section 3.5, we calculate the total value of guided trips in our baseline calibration to be 1.8% of model GDP. Recall that we interpret guided trips as encompassing many market and non-market activities. Some of these activities appear in the national accounts as business services and therefore count as intermediate inputs that do not end up directly in measured GDP. Others do not appear anywhere in the national accounts because they are essentially home production or are impossible to measure. Fortunately, given its small size including the total value of guided trips in our model-based measures of GDP and wages does not change our baseline calibration.

Measuring employment also is complicated by the fact that all household members participate in generating guided trips. We count those agents engaged in intermediate good production, n_y, and residential construction, n_h, as employed and measure their wages by their marginal products excluding the value of guided trips. The non-employed who also produce guided trips are assumed to be engaged in home production and so are not included in our accounting of employment. In Table 1 the labor share parameters are chosen to match total labor compensation as a share of GDP (the target is borrowed from traditional real business cycle studies) and our estimate of the share of residential construction employment in total private non farm employment.[27]

[27] We estimate residential construction employment by multiplying total construction employment by the average over our sample period of the nominal share of residential investment expenditures in total structures investment expenditures.

We fix the discount rate so the model's real interest rate is 4%. Combined with this target the equipment-output ratio in the non-construction sector, K_y/Y, identifies equipments's share in that sector's production. Our empirical measure of equipment for this calculation is the Bureau of Economic Analysis' (BEA) measure of the stock of equipment capital. Equipment's depreciation rate is identified using the investment to GDP ratio, where we measure investment using the NIPA estimate of equipment investment. Equipment's share in residential construction is identified by the ratio of capital employed in the residential construction sector, K_h, to GDP where the empirical counterpart to capital in this ratio is the BEA measure of equipment employed in residential construction. The depreciation rate of residential structures is identified using the residential investment to GDP ratio.[28]

We identify the housing service parameters as follows. First the housing coefficient H is chosen to match the residential capital to GDP ratio, where the measurement of residential capital is consistent with our measure of residential investment described above. Land's share in housing services, ζ, is chosen to match the estimate of land's share of the total value of housing in Davis and Heathcote (2007). To measure this object in the model we need the price of land, q_b. We obtain this variable as the solution to the no-arbitrage condition

$$q_b\left(z\right) = r_b\left(z\right) + \beta \int q_b\left(z'\right) dQ(s'; s, s_{-1}),$$

where r_b denotes the rental price of land which equals the marginal product of land in the provision of housing services. Land's share of the economy-wide value of housing is then given by $\int q_b b_r d\mu / \left[\int q_h h d\mu + \int q_b b_r d\mu\right]$.

7 Quantitative Analysis

We now consider the model's empirical predictions. First, we confirm that the model accounts for population's slow adjustment and that this success comes with generally accurate predictions for gross migration and the behavior of local labor and housing markets. Next, we study the model's predictions for unconditional dynamics and find the model is similarly successful at accounting for the data even though TFP shocks are the only source of city level fluctuations in the model. So, despite choosing parameters to match evidence not directly

[28] The depreciation rate for residential structures obtained this way is close to the mean value of the (current cost) depreciation-stock ratio for residential structures obtained from the BEA publication "Fixed Assets and Consumer Durable Goods," once output and population growth are taken into account. Calibrating to this alternative depreciation rate has virtually no impact on our quantitative findings.

related to the dynamics of interest our model nonetheless excels in replicating them.

After establishing the empirical credibility of our framework, we investigate how migration and housing influence slow population adjustments. We find that costly directed migration through the model's guided trip technology is the principle source of slow population adjustments. We interpret this finding as demonstrating that the myriad ways individuals get informed about desirable locations to live and work represent significant barriers to rapid population (and worker) reallocation. The fact that we identify the model's migration parameters without consideration of within-city dynamic responses to TFP shocks lends substantial credibility to this interpretation. Interestingly, we find that housing plays only a small role slowing population adjustments once migration costs are taken into account.

Finally, we investigate the implications of our model's successful accounting of slow short run population adjustments for persistent urban decline. There are many cities in our data that experience declining population throughout the sample period. These cities also experience declining TFP suggesting our model might account for the persistence of urban decline. To investigate this possibility, we study the average experience of the 15 cities with the largest population declines. Simulating our model using the empirical path of TFP for these cities shows that our model accounts for essentially all of the average population decline. This finding suggests that costly migration, in particular the costs of finding new locations to live and work, is a major factor determining the persistence of urban decline.

7.1 Model Validation with Conditional Correlations

Comparisons of model and empirical impulse response functions is a model validation tool common in macroeconomics, see for example Christiano, Eichenbaum, and Evans (2005). Its key advantage over studying unconditional statistics, is that in principle it is robust to the presence of other shocks. We estimate the responses to TFP shocks of population, gross migration, employment, wages, residential investment and house prices in both the model and the data using the identical procedure described in Section 2.2, basing our model responses on the simulation of a large panel of cities over a long time period. TFP in the model is measured as we do in the data.

Figure 6 displays model and estimated responses of population to a one standard deviation positive innovation to TFP. Here and for similar figures below the vertical lines with hash marks indicate plus and minus 2 standard error bands for the estimates.[29] Figure 6

[29]These standard errors do not take into account the sampling uncertainty in our estimates of the under-

Figure 6: Responses of Population

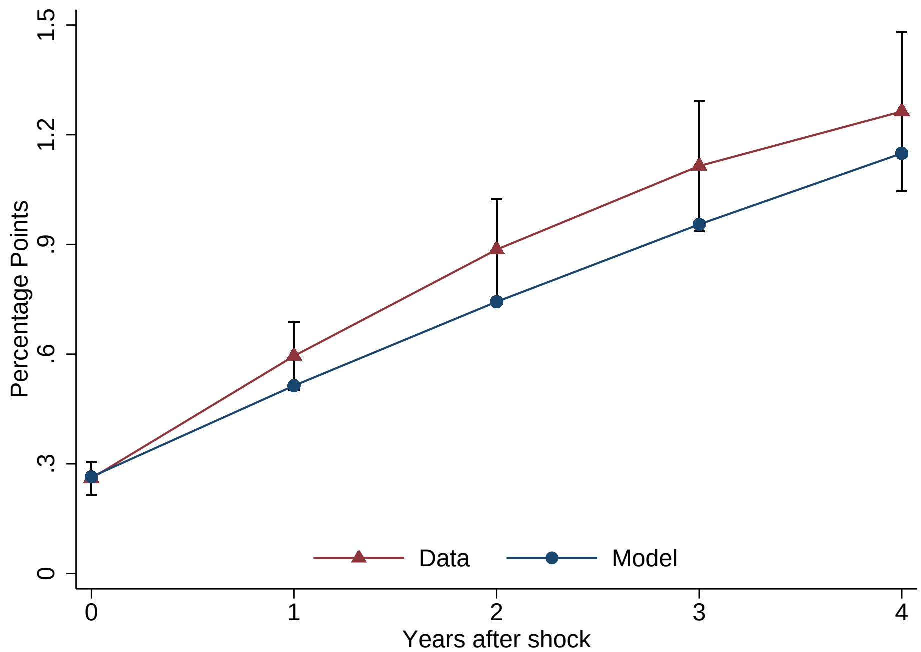
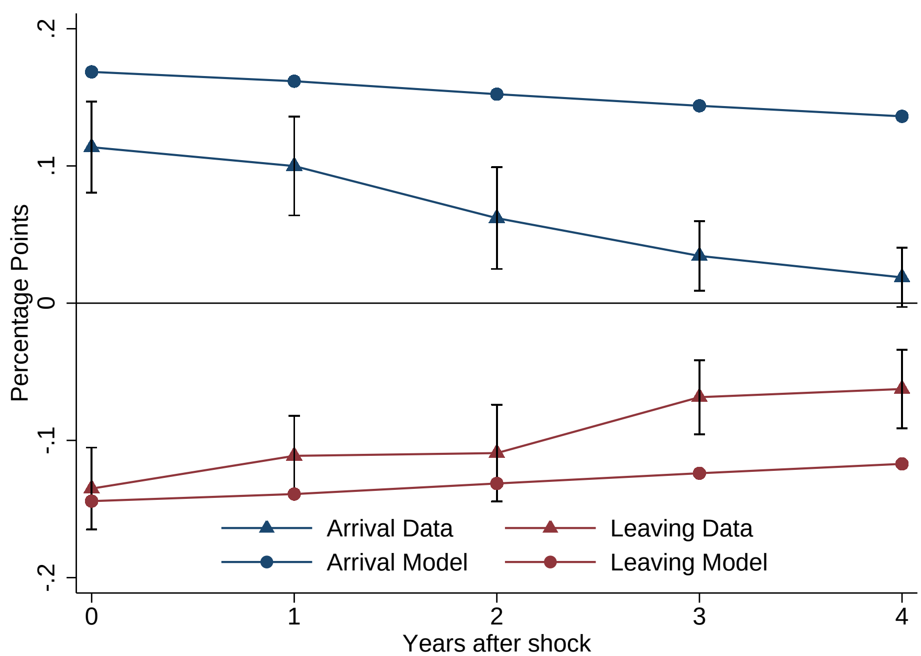


Figure 7: Responses of Arrival and Leaving Rates



---

lying TFP process.

demonstrates that the model's population response is statistically and economically close to the one for the data; our model accounts for the slow response of city populations to local TFP shocks. It may appear that the model's slow population response is inconsistent with population's variance exceeding that for TFP as indicated by Figure 5. The model is able to account for the unconditional population distribution with a slow conditional response of population to a TFP shock because ultimately the long run response of population exceeds that for TFP.

Figure 7 shows this accounting for slow population adjustments involves replicating quite closely the dynamic responses of the arrival and leaving rates.[30] Crucially the model is consistent with the negative conditional correlation between the gross migration rates. The intuition for this finding is simple. Having multiple margins to respond to the increase in productivity, the city planner takes advantage of all of them. It can raise employment per person and bring more workers to the city. For the latter it can cut back on the fraction of the initial population that leaves for other cities, that is reduce the leaving rate, and attract more workers to the city with more guided trips. The goodness-of-fit is weaker for gross migration than it is for population, for example both responses are more persistent than in the data and the arrival rate's initial response is a little too strong. Nevertheless given its simplicity the model does surprisingly well.

Figure 8 shows the dynamic responses of employment, wages, residential investment and house prices. We define house prices, q_{sf}, as the total value of structures and land used to produce housing services per unit of housing services provided:

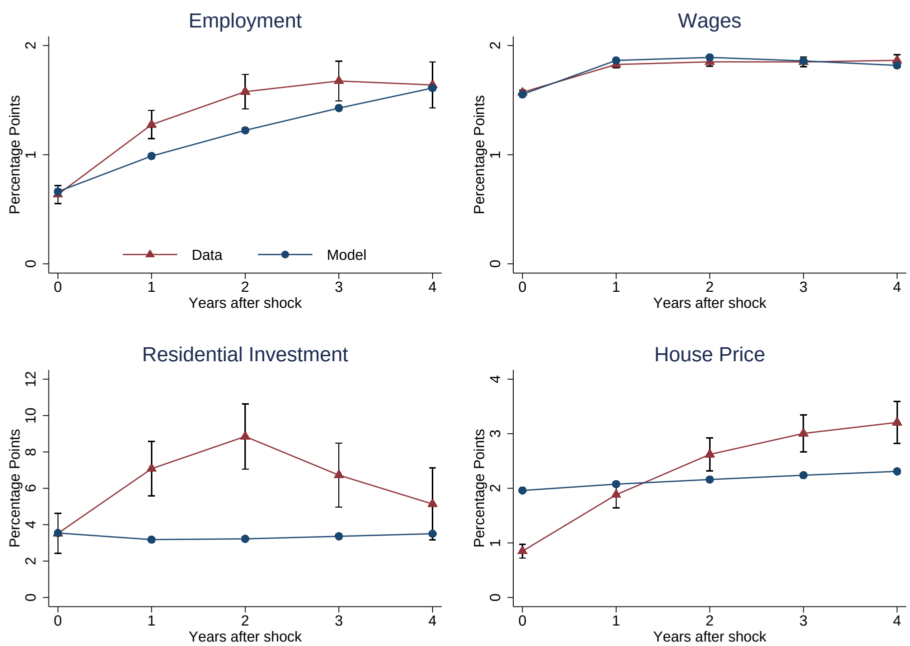
$$q_{sf}\left(z\right) = \frac{q_h\left(z\right) h\left(z\right) + q_b\left(z\right) b_n(z)}{h(z)^{1-\varsigma} b_r(z)^{\varsigma}}.$$

The price q_{sf} corresponds to the price of housing per square foot under the assumption that every square foot of built housing yields the same quantity of housing services.

The labor market responses are a very good fit. Observe that the employment response in the model, as in the data, is stronger than the population response. That is, the employment to population ratio rises after a positive TFP shock indicating that the labor supply margin is indeed exploited in both the data and the model. The qualitative responses of construction and housing also are consistent with the data. These findings derive from a higher population

[30] The difference between the model's migration rates in the the first period do not correspond exactly to the response of population which in principle it should according to equation (24). The discrepancy is due to using logarithmic first differences to approximate the net rate of population growth.

Figure 8: Responses of Labor and Housing Markets



desiring additional housing and that local factor inputs with alternative uses are used in construction thereby creating an imperfectly elastic supply of new housing. The model is less successful accounting for the quantitative responses of housing. Residential investment misses the hump shape in the data and the house price response is too fast. However in both cases the order of magnitude of the responses are about right. One explanation for housing's discrepancy with the data is that our model does not include search frictions in the local housing market. Lloyd-Ellis et al. (2014) demonstrate that search frictions show promise in generating serially correlated responses of construction and house price growth to productivity shocks.

7.2 The Model's Predictions for Unconditional Dynamics

A city's response to TFP shocks is in principle robust to the presence of other shocks and is therefore informative about model validity even with such shocks. However it is likely that there are shocks to local taxes, amenities and intermediate good demand and so it is worth knowing the extent to which TFP shocks alone account for unconditional moments of the data. Tables 2 and 3 display unconditional standard deviations, contemporaneous correlations, and serial correlations of the same variables discussed above in the model and in our data. Except for population, standard deviations are expressed relative to population and the contemporaneous correlations are all with population. The statistics are based on the levels of the gross migration rates and on growth rates for the other variables. All variables have been transformed as described in Section 2 prior to the analysis.

Table 2: Volatility and Co-movement Within Cities

Variable	Standard Deviation Data	Standard Deviation Model	Correlations Data	Correlations Model
Population	1.33	0.87	–	–
Arrival Rate	0.68	0.53	0.59	1.00
Leaving Rate	0.57	0.48	-0.43	-1.00
Employment	1.58	1.23	0.56	0.93
Wages	1.23	1.81	0.16	0.32
Res. Investment	19.7	4.27	0.14	0.40
House Prices	3.77	2.32	0.29	0.47

Note: The statistics are based levels of the gross migration rates and on the growth rates of the other variables. The latter variables have been transformed as described in Section 2.2 prior to calculating growth rates. Standard deviations of all variables except population are expressed relative to the standard deviation for population. Correlations are with population.

The first thing to notice from Table 2 is that TFP shocks generate about two thirds of the overall variation in population – they are a quantitatively important source of local variation. The model is strikingly successful at replicating the qualitative pattern of relative volatilities and only somewhat less successful quantitatively. Gross migration is less volatile than population and the labor and housing market variables are all more volatile than population, just as in the data. The relative volatility among the variables other than population also mostly match the data. Only the labor market variables miss, with wages a little too volatile compared to employment. The model is consistent with residential construction being the most volatile variable, but it fluctuates much less in the model than in the data. House prices in the model are more than twice as volatile as population, but not quite as volatile as in the data. The high volatility of house prices is a direct consequence of local land and labor that have alternative uses being factor inputs in construction.

Table 2 shows that the model is qualitatively consistent with all the correlations with population growth. The largest discrepancies involve the arrival and leaving rates being per-

Table 3: Serial Correlation Within Cities

	Lag			
Variable	1	2	3	4
Population				
Data	0.81	0.74	0.67	0.63
Model	0.93	0.87	0.81	0.75
Arrival Rate				
Data	0.81	0.69	0.56	0.45
Model	0.93	0.87	0.81	0.75
Leaving Rate				
Data	0.80	0.76	0.70	0.63
Model	0.93	0.87	0.81	0.75
Employment				
Data	0.52	0.29	0.21	0.15
Model	0.73	0.63	0.58	0.54
Wages				
Data	0.15	0.04	0.05	0.07
Model	0.20	0.02	-0.02	-0.02
Res. Investment				
Data	0.12	0.10	-0.02	-0.11
Model	-0.09	0.02	0.04	0.04
House Prices				
Data	0.73	0.31	-0.06	-0.25
Model	0.07	0.06	0.05	0.05

Note: The variables are have been transformed as described in Section 2.2 prior to calculating the statistics. The gross migration rates are levels and all other variables are growth rates.

fectly positively and negatively correlated with population growth. Including a mechanism to reproduce KW's finding that out-migration is relatively high for recent in-migrants could move the model in the direction of the data.[31]

From Table 3 we see that population, gross migration, employment and wages all display similar persistence to that in the data, although the model's variables are more persistent. Construction in the model and data are similarly random-walk like, although this feature of the unconditional moments clearly is due to the effects of other shocks given the serially correlated growth rate of construction in response to TFP shocks we find in the data. House prices display the greatest differences with house price growth displaying substantial serial

[31]See Coen-Pirani (2010) for one such mechanism.

correlation in the data while in the model house prices are more like a random-walk.

7.3 The Source of Slow Population Adjustments

We now address the sources of slow population adjustments in our model. Figure 9 displays impulse responses to TFP shocks implied by several different versions of the model. The different versions consist of perturbations relative to the baseline, calibrated version of the model, holding parameters not involved in the perturbation fixed at their baseline values. These perturbations are as follows:

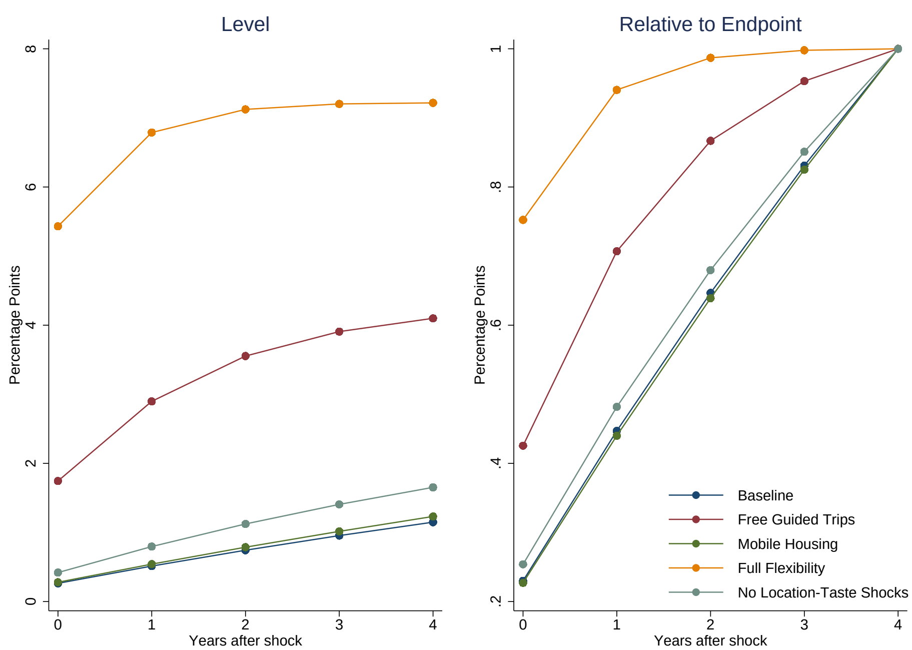
- The "Free Guided Trips" case sets $A = 0$. This case is identical to assuming all the migration parameters are set to zero, because when guided trips are free the city-planner sets the leaving rate in each city to the constant value that minimizes leaving costs and adjusts population by changing the arrival rate at zero cost. So in this case only the housing frictions are operative.
- "No Location-Taste Shocks" is the case where $\psi_1 = \psi_2 = 0$ so that costly guided trips are the only migration friction.
- "Mobile Housing" corresponds to the case discussed in Section 4 in which housing can be rented at a fixed price from any city; housing is perfectly mobile. In this case a city's dynamics are not influenced by the durability or the size of the local housing stock nor the city's ability to produce houses to accommodate new workers.
- "Full Flexibility" combines all the perturbations so there are no mobility costs and housing is perfectly mobile.

The left plot in Figure 9 displays the levels of the responses and the right one shows the responses after first dividing them by the value attained in the last (fifth) period of the response to more clearly show the speed of adjustment. Figure 9 shows that in the Full Flexibility case the population dynamics essentially follow the path of TFP with roughly 90% of the long run (five year) adjustment occurring after 2 years compared to 85% for TFP (see Figure 3) – absent migration and housing frictions the model has essentially no internal mechanism to propagate TFP shocks.

The No Location-Taste Shocks and Mobile Housing cases are very close to the baseline, with the latter being almost the same. In other words removing from the model costly out-migration or immobile housing, leaving costly guided trips as the only model friction, leaves the population response essentially as slow as it is in the baseline economy and hence the data. In the Free Guided Trip case the population response is closer to the full-flexibility case and does not take the model all the way to the data. In the Free Guided Trip case the only friction is that housing is immobile, suggesting some role for housing in slowing population adjustments.

Despite this last result, we still conclude that costly guided trips are the main source of slow population adjustments. The discrepancy of Free Guided Trips with Full Flexibility arises from a property of adjustment costs highlighted by Abel and Eberly (1994). The first adjustment cost introduced to an otherwise frictionless model always has a relatively large impact on dynamics. So, introducing immobile housing into an otherwise frictionless model has seemingly large effects. However immobile housing on its own is not sufficient to deliver the amplitude and persistence of the population response in the data. Yet, the population dynamics with migration costs and mobile housing, the Mobile Housing case, are essentially the same as the baseline. This suggests that the prime driver of slow population adjustment in the model is the costly guided trip technology.

Figure 9: Impact of Model Features on Population Adjustment



The finding of slow population adjustment driven mostly by costly migration confirms and reinforces results in KW. Using the parameters of a migration choice problem estimated with data on the frequency of inter-state moves taken from the National Longitudinal Survey of Youth, KW calculate optimizing responses of individuals in all states to a one-time permanent change in wages of one particular state (they consider changes in California, Illinois and New York). From these choices they obtain a matrix of transition probabilities which they simulate to trace out the response of population in the state to the permanent change in wages. Strikingly we find roughly the same five year elasticity of population with respect to the wage, about .5.[32] KW find that about 30% of the five period response occurs in the first period (see their Figure 1), whereas we find a response closer to 20%.

These similarities are quite striking given the very different methodologies used to generate the responses. The slower initial response we obtain is consistent with the fact that in our analysis wages take a few periods to reach their long run level due to the nature of the TFP process we estimate and our identification takes into account feedback to future migration from lower wages induced by greater net in-migration. The fact that our model includes housing does not appear to be an important source of the difference. Overall our results establish that KW's findings are robust to the presence of housing and equilibrium interactions as well as considering migration between cities instead of states.

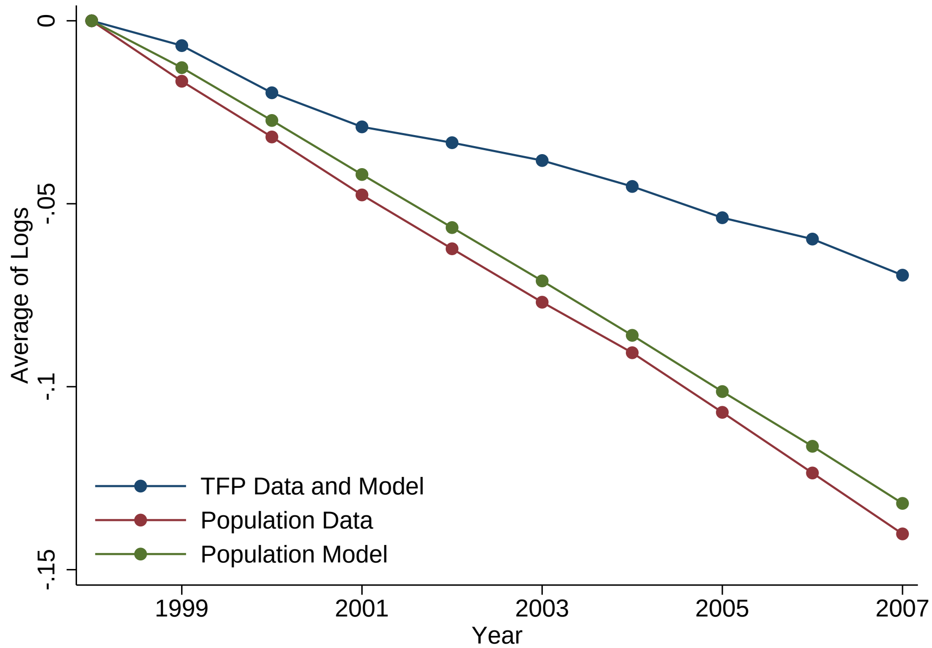
7.4 Migration and Urban Decline

There are many cities which experience declining populations (relative to the aggregate) over the sample period 1985-2007. This is evidence of the persistent urban decline studied by Glaeser and Gyourko (2005). Interestingly, the cities with declining populations also have TFP declining for most of the sample. Our model's ability to reproduce the *short run* response of population to TFP shocks then suggests it might account for population dynamics over the *long run* and in particular persistent urban decline. We now discuss a simple experiment that demonstrates that indeed our calibrated model does account for persistent urban decline.

We focus on the 15 cities of the 365 total that experience the greatest population declines in our sample. The corresponding TFP paths are fed into the model from the common initial condition that takes the mid-point of our TFP grid and assumes TFP stays at that level for a long time. We use the first 12 years of our sample, 1985 to 1997, to simulate unique initial conditions for each city based on each city's empirical TFP path. This procedure builds in

[32]KW consider a 10% increase in wages and find that population is 5% higher after five years. We find a 1.1% response of population to a 2% (roughly) permanent increase in the wage.

Figure 10: Persistent Urban Decline



the possibility that past declines in TFP show through into future population declines. 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.[33]

Figure 10 shows the average log paths for TFP and population for the data and the model. TFP falls by .06 from 1998-2007 and population falls by twice as much.[34] Strikingly, the model's predicted path for population lies very close to its empirical counterpart. Obviously the fit is not as perfect for the individual cities, but the general impression is similar. There are two key factors driving the model's success: persistent declines in TFP taken from the data and the slow response of population to past declines in TFP predicted by the model. 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 innovation is much smaller than TFP's, but over longer horizons it responds by much more.

Since the dominant source of slow population adjustment in the model is the cost of finding new cities to live and work, we conclude that these costs are integral to our model's

[33]For this experiment we (incorrectly) equate our empirical measure of TFP to model TFP, s. We do this for computational tractability but the small differences involved should not affect our conclusions.

[34]The much larger drop in population is a reflection of the forces driving our model's reproduction of Zipf's law discussed above.

explanation of persistent urban decline. Durable and immobile housing is not important at all in the sense that migration frictions essentially account for slow population adjustments on their own. This contrasts with Glaeser and Gyourko (2005) who argue that durable and immobile housing explains persistent urban decline.[35] These authors do not integrate costly migration into their empirical analysis. We consider both housing and migration costs in a unified framework, but housing turns out to be unimportant.

8 Conclusion

This paper documents that population adjusts slowly to near random-walk TFP shocks and proposes an explanation for why: the incentive to reallocate population after a TFP shock is limited by costs of finding new cities to live and work. We show that these same costs can also account for the persistence of urban decline. The framework that delivers these results is not arbitrary, but is dictated by the nature of the relationship between gross and net population flows in cities that we uncover in our panel of 365 cities from 1985 to 2007 and microeconomic estimates of migration costs from KW.

Our model has left out interesting features that are undoubtedly important for understanding the full range of adjustment to shocks within and across cities. Chief among these omissions are search frictions in local labor and housing markets. We think it would be interesting to add these features to our framework. Doing so would help disentangle the contributions to labor reallocation of traditional search from the migration frictions we introduce. For example, the local housing and labor search frictions considered by Karahan and Rhee (2012) and others might play a role similar to the migration frictions in our framework. Nevertheless it remains to be determined whether these local frictions can account for the empirical relationship between gross and net migration and the slow response of population to TFP shocks that we uncover in this paper.

[35] They show that irreversible housing in cities with declining populations has several empirical predictions which they verify in the data. Our model does not share these predictions since the irreversibility constraint is never binding. It is never binding because of the relatively small variance of TFP innovations compared to the depreciation rate for housing. This constraint appears not to bind in the panel as new building permits are always strictly positive. It presumably binds for neighborhoods within a city and this may play a role in explaining Glaeser and Gyourko (2005)'s findings.

References

Abel, A. B. and J. C. Eberly (1994). A unified model of investment under uncertainty. *American Economic Review 84*(1), 1369–1384.

Albouy, D. (2009). The unequal geographic burden of federal taxation. *Journal of Political Economy 117*, 635–667.

Alvarez, F. and R. Shimer (2011). Search and rest unemployment. *Econometrica 79*(1), 75–122.

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

Christiano, L., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*(1), 1–45.

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 (2011). The role of housing in labor reallocation. Revision to Chicago Fed working paper no. 2010-18.

Davis, M., J. Fisher, and T. Whited (2014). Macroeconomic implications of agglomeration. *Econometrica 82*(2), 731–764.

Davis, M. and J. Heathcote (2007). The price and quantity of residential land in the United States. *Journal of Monetary Economics 54*(8), 2595–2620.

Davis, M. and F. Ortalo-Magné (2011). Household expenditures, wages, rents. *Review of Economic Dynamics 14*(2), 248–261.

Davis, S., J. Faberman, and J. Haltiwanger (2006). The flow approach to labor markets: New data sources and micro-macro links. *Journal of Economic Perspectives 20*(3).

Davis, S., J. Faberman, and J. Haltiwanger (2011). Labor market flows in the cross section and over time. *Journal of Monetary Economics 59*, 1–18.

Diamond, R. (2012). The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. Harvard University working paper.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review 94*(5), 1429–1451.

Ferreira, F., J. Gyourko, and J. Tracy (2011). Housing busts and household mobility: An update. NBER Working Paper No. 17405.

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.

Greenwood, M. (1997). Internal migration in developed countries. In M. Rosenzweig and O. Stark (Eds.), *Handbook of Population and Family Economics*, Volume 1B. North Holland.

Kaplan, G. and S. Schulfofer-Wohl (2012). Understanding the long-run decline in interstate migration.

Karahan, F. and S. Rhee (2012). Geographical reallocation and unemployment during the great recession: The role of the housing bust. University of Pennsylvania working paper.

Karahan, F. and S. Rhee (2013). Population aging, migration spillovers, and the decline in interstate migration. Univeristy of Pennsylvania manuscript.

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 indivisual 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*.

Lloyd-Ellis, H., A. Head, and H. Sun (2014). Search, liquidity, and the dynamics of house prices and construction. *American Economic Review 104*(4), 1172–1210.

Lucas, R. (1997). Internal migration in developing countries. In M. Rosenzweig and O. Stark (Eds.), *Handbook of Population and Family Economics*, Volume 1B. North Holland.

Luttmer, E. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics 122*(3), 1103–1144.

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

Modestino, A. and J. Dennett (2012). Are american homeowners locked into their houses? the impact of housing market conditions on state-to-state migration. Federal Reserve Bank of Boston Working Paper No. 12-1.

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. (2012). Regional mismatch and labor reallocation in an equilibrium model of migration. Norwegian Business School working paper.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy 90*(6), 1257–1278.

Rosen, S. (1979). Wagebased indexes of urban quality of life. In P. N. Miezkowski and M. R. Straszheim (Eds.), *Current Issues in Urban Economics*, pp. 74–104. Baltimore, MD: Johns Hopkins University Press.

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.

Schulhofer-Wohl, S. (2012). Negative equity does not reduce homeowners' mobility. NBER Working Paper No. 16701.

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.

Van Nieuwerburgh, S. and P.-O. Weil (2010). Why has house price dispersion gone up? *Review of Economic Studies 77*, 1567–1606.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Appendix

This appendix describes how we calculate gross migration flows using IRS data; how we assess the bias in using migration costs based on inter-state migration in a model of inter-city migration; and the methods used to solve the quantitative model.

A Calculating Gross Migration Flows with IRS Data

We construct data on gross MSA-level population inflows and outflows using county-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 are annual and 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.

For each of the years, 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.[36] 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 as the number of new entrants to an MSA during a given year, l as the number of people exiting the MSA during the year, and $\bar{p}$ as all the people that did not move into or out of the MSA during the year. We compute the beginning of year population x and end of year population p as

$$\begin{aligned} x &= \bar{p} + l \\ p &= \bar{p} + a \end{aligned}$$

Net migration is therefore $p - x = a - l$. Note that due to births and deaths and foreign migration, p in any given year is typically less than x in the subsequent year.

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

B 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 KW'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 KW 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.[37] 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

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

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 KW's finding that moving costs increase with distance moved.[38]

Following KW 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

[37] For simplicity we abstract from idiosyncratic wage shocks included by KW. KW 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.

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

MSA in 1990 (the year KW 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.[39]

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 KW (see footnote 23 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 KW of 2.9 percent.

We compute all three moments analytically. The probability agents migrate to location j given their current location of i, $\gamma(j,i)$ has the straightforward expression

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

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

[39] 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.

for all moves from j to i is therefore

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

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

$$\frac{\sum\limits_{j\neq i}\gamma\left(j,i\right)\beta\left(V_j-V_i\right)}{\sum\limits_{k\neq i}\gamma\left(k,i\right)} = \frac{\sum\limits_{j\neq i}\gamma\left(j,i\right)\beta\left(V_j-V_i\right)}{1-\gamma\left(i,i\right)}.$$

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

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

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

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 KW 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, 0.01 higher than the average simulated benefits accruing to all movers. The bias is therefore 5%.[40] We find that a bias of this size does not affect our conclusions.

[40] 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%.

C Solving the Quantitative 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 first difficulty is that the value function of the city planner's problem has two 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. To make the task of computing the value function manageable we used spline approximations.

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

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

As already mentioned, this spline approximation preserves monotonicity and concavity of the original function f (e.g. Lyche and Morken (2011), Section 5.2). The definition of variation diminishing splines is easily generalized to functions of more than one variable using tensor products (e.g. Lyche and Morken (2011), Section 7.2.1). These properties greatly simplify the value function iterations of the city planner's problem and they should prove useful in a variety of other settings. In our actual computations we worked with an approximation of degree $p = 3$.

An additional complication involves the return function of the city planner's problem.

Conditional on the current states (h, x, s, s_{-1}) and future states (h', x'), evaluating the one period return function of the city planner requires solving a nonlinear system of equations in $(n_y, n_h, k_y, k_h, b_r, b_h)$ allowing for the possibility that the constraint $n_y + n_h \leq x'$ may bind. This is not a hard task. However, doing this for every combination of (h, x, s, s_{-1}) and (h', x') considered in solving the maximization problem at each value function iteration would slow down computations quite considerably. For this reason, we chose to construct a cubic variation-diminishing spline approximation to the return function $R(h, x, s, , s_{-1}, h', x')$ once, before starting the value function iterations, and use this approximation instead. In practice, for each value of (h, x, s, s_{-1}) we used a different knot vector for h' and x' to gain accuracy of the return function over the relevant range.

Performing the maximization over (h', x') for each value of (h, x, s, s_{-1}) at each value function iteration is a well behaved problem given the concavity of the spline approximations to the return function and the next period value function. There are different ways of climbing such a nice hill in an efficient way. In our case, given that we could offload computations into two Tesla C2075 graphic cards (with a total of 896 cores), we used the massively parallel capabilities of the system to implement a very simple generalized bisection method. Essentially for each value of (h, x, s, s_{-1}) we used a block of 16×16 threads to simultaneously evaluate 16×16 combinations of (h', x') over a predefined square. We then zoom to the smallest square area surrounding the highest value and repeat. In practice, a maximum would be found after only three or four passes.

Statistics under the invariant distribution were computed using Monte Carlo simulations. This part of the computations was also offloaded to the 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 all decision rules.

Speeding up the solution to the city planner's problem and Monte Carlo simulations was crucial since finding solutions (Y, 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.