Tractable Dynamics in Models of Location Choice

Chase Abram*

November 5, 2024

Abstract

I lay out a dynamic model of location choice wherein the arrival of moving opportunities is random, and provide analytic results concerning equilibrium dynamics. The stationary distribution is isomorphic to standard static quantitative spatial models, and therefore can be calibrated similarly. I also show that in the baseline case of constant elasticity externalities across space, the transition path following a permanent change in fundamentals is efficient in the decentralized economy. Finally, I motivate the baseline model's potential usefulness going forward by considering a few extensions which maintain tractability while addressing more complex economic mechanisms.

*Send corrections or comments to abram@uchicago.edu. I thank Olivia Bordeu, Ignacia Cuevas, Dan Ehrlich, Olivier Kooi, Mikhail Golosov, Veronica Guerrieri, Thomas Hierons, Greg Kaplan, Xiaoyang Li, Camilla Schneier, Jordan Rosenthal-Kay, and Esteban Rossi-Hansberg for insightful comments. All errors are my own.

1 Introduction

While our understanding of the mechanisms generating the distribution of economic activity over space has grown tremendously in the past two decades, the dynamics of this distribution have remained largely understudied. Relatedly, most of the standard static models are silent on the how adjustment would occur in the event of a change in economic fundamentals, other than to say that the static equilibrium must change. Since these models may be thought of as long-run approximations, ignoring the short- to medium-run dynamics is likely benign if only concerned about long horizons or distributions which adapt quickly. However, if interested in shorter horizons, such as questions concerning within-city trends (e.g. gentrification), understanding the transition path may be key to constructing proper policy.

The starting point is essentially (Roback, 1982), where agents choose where to live in order to maximize real wages net of rents. The novelty is that agents are constrained by a mobility friction which prevents movement between locations except when an opportunity arrives. At the time of a mobility opportunity, agents also receive a preference shock with dispersion parameter θ for all locations. In the perfectly elastic limit, all agents move to the highest-valued location whenever given the opportunity. This structure will generate a stationary distribution of the population across locations that depends not only on the fundamentals in each location, but on the substitutability between locations according to the preference shock, the discount rate, and the rate at which moving opportunities arrive. This result allows for simple calibration of the stationary distribution in the nearly the same manner as static spatial models, merely requiring an additional calibrated parameter for the mobility rate.

To give analytic results and intuition for the dynamics, I solve in closed-form the transition path of the case without externalities in response to a permanent adjustment in fundamentals, and show that the transition path will only depend on the initial distribution, the stationary distribution, and the mobility rate. Additionally, I consider the limiting case $\theta \to \infty$ where preference shocks cease to matter, and agents are perfectly elastic to the value of each location, and solve for the transition paths after a permanent shock, now allowing for externalities. The path reduces to a finite number of finite length stages where agents move from suboptimal locations to the maximally-valued set of locations, and therefore the population path for each location is decreasing until it joins the maximal set, at which point it increases until the stationary distribution is reached. This same idea applies for finite θ, that population paths may be non-monotonic as agents leave earlier in the transition to move somewhere better, but later want to move back, because the exodus from their original location improved it.

Next, I consider the efficiency of the stationary distribution and transition paths. I point out that in the popular specification of constant elasticity of local wages to the local populations, if the same elasticitiy is assumed across space, the externalities "cancel out" and the decentralized allocation is efficient. Furthermore, I show this idea generalizes and that the entire transition path is also efficient. I also clarify exactly why the efficiency result holds, and show that even maintaining that the elasticity of each local externality to each local population is constant within each location, if the constant varies over space, the decentralized allocation is inefficient.

Lastly, I suggest a few extensions to the model, mainly to demonstrate the tractability of modeling migrational frictions as I have done. The simplest extension is merely to show that the HJB-KFE system still works exactly the same when fundamentals change over time. The second extension augments the baseline model with an additional moving cost for leaving the current location. The final extension adds lifecycle considerations, and while it significantly complicates the model, tractability is reasonably maintained due to the continuous-time form, and Poisson arrival of shocks.

I contribute to three strands of literature. I add to the understanding of place-based polices by considering how transitions interact with inefficienct sorting, and what policies can correct. In this area, I am closest to (Fajgelbaum and Gaubert, 2019), which considers misallocation primarily due to non-productive amenities and spillovers, but is entirely static. I abstract away from amenities, but include the spillovers and therefore my stationary distribution results are in line what they find, but I am able to add to consider the potential for dynamic inefficiency and correction. Less directly, my approach is related to (Guerrieri, Hartley, and Hurst, 2013), which considers the dynamics of gentrification. However, there are no mobility frictions in that framework, and the model is a city on a line, whereas my model has a discrete set of locations, but no geography.

To the literature on quantitative spatial economics, I provide a tractable means of adding dynamics which adds only a mobility friction without drastically changing the rest of model, or even much complicating calibration to the stationary distribution. One advantage of this approach is that it avoids the less palatable assumption that mobility costs can be recovered by reversing moves later, as in (Desmet, Nagy, and Rossi-Hansberg, 2018) and (Cruz and Rossi-Hansberg, 2021), which make the reversibility assumption entirely to maintain tractability via turning the dynamic decision problem into a static one. The key to the tractability in my approach is that Poisson arrivals of shocks in a continuous-time setting do not significantly complicate solving the relevant HJB equation, but function economically similarly to a fixed cost, creating an economically meaningful moving friction without becoming computationally unwieldy. The approach most similar to mine is (Caliendo, Dvorkin, and Parro, 2019), though they work in discrete time, so are not able to exploit the analytically clean differential equations solutions I find, nor the fact that in extensions of the baseline model, working in continuous time yields sparse HJB and KFE operators, and flows across different dimensions may be decoupled. Similarly, (Olivero and Yotov, 2012) also introduce dynamics into a gravity framework for trade a la (Anderson and van Wincoop, 2003), but they focus on asset accumulation as the mechanism as opposed to a mobility friction hindering immediate transitions after a shock to fundamentals.

To a minor degree, I add to the literature on the use of continuous-time methods for solving macroeconomic models. While (Achdou et al., 2021) and (Kaplan, Moll, and Violante, 2018) use these methods to solve incomplete markets models with heterogeneous agents, I am unaware of any applications of these methods for solving spatial models, and thus one contribution is recognizing that these tools may be applied to recasting spatial problems in a dynamic sense. The dynamic setting provides an alternative means to solving these models, compared to, for example, (Allen and Arkolakis, 2014), which recognize the static economy as boiling down to a system of two nonlinear integral equations, since in the dynamic setting (in certain cases) the HJB and KFE may be solved via backward and forward time iteration.

The remainder of the paper is structured as follows. In Section 2 I lay out the model and define the equilibrium, as well as lay out the limit case of infinite preference elasticity between locations. In Section 3 I discuss how the stationary distribution depends on the parameters governing dynamics, and consider transitional paths which are analytically simple. In Section 4 I discuss the efficiency of the stationary distribution and transition path. In Section 5 I consider a few extensions to the baseline model. In Section 6 I explain one approach for calibrating the model. Section 7 concludes.

2 Model

Time is continuous and there are n locations. Each location i offers a wage at each moment of time, $w(i,t)$. The determination of these wages has both a fundamental exogenous component and an endogenous component which will be clarified below.

2.1 Agent Choices

Agents purchase the numeraire consumption good with the wage, have log preferences for consumption, and discount the future at rate ρ. Agents choose a path of locations subject to the constraint that they may only move when a migration shock arrives. Poisson arrivals of migration shocks occur at rate λ and, when a shock arrives, an agent draws a vector ϵ of idiosyncratic preferences for all locations, chooses a location, enjoys the draw immediately, then continues to enjoy the flow wage at that location until their next move. The objective of each agent is:

$$\max_{\{i(\tau)\}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}\log w(i,t)\mathrm{d}t + \sum_\tau e^{-\rho\tau}\epsilon(i(\tau))\right]$$

subject to the random arrivals at τ of migration shocks with Poisson intensity λ:

$$F(\tau_{n+1}-\tau_n) = 1-\exp(-\lambda(\tau_{n+1}-\tau_n))$$
$$i(t) = \max_{\tau\leq t} i(\tau)$$

It will be convenient to assume the preference shocks ϵ for each location at each opportunity for migration are independently and identically distributed across locations and time and follow a Gumbel distribution with scale parameter $\frac{1}{\theta}$:

$$F(\epsilon) = \exp(-\exp(-\theta\epsilon))$$

The problem of the households may then be summarized by the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\rho V(i,t) - \partial_t V(i,t) = \log w(i,t) + \lambda\mathbb{E}\left[\max_j\{V(j,t)+\epsilon(j,t)\} - V(i,t)\right]$$

The first term on the left accounts for discounting, and the second term accounts for changes in wages which affect the value of each location at each instant. The first term on the right is the flow log wage (which equals log consumption, so utility) for the current location, and the second term accounts for the λ-rate migration opportunities. The HJB can be rearranged, and the Gumbel shock distribution allows for a closed-form expression for the max problem, to yield the following expression[1].

$$(\rho+\lambda)V(i,t) - \partial_t V(i,t) = \log w(i,t) + \lambda[\gamma\theta + \log\sum_j \exp(\theta V(j,t))]$$

In the stationary equilibrium, the $\partial_t V(i,t)$ is zero, and the far right side term is constant across i, so the value function may be explicitly solved in terms of wages.

$$V(i) = \frac{\log w(i)}{\rho+\lambda} + \frac{\lambda}{\rho+\lambda(1-\theta)}[\gamma\theta + \log\sum_j \exp\left(\frac{\theta\log w(j)}{\rho+\lambda}\right)]$$

[1] Details for this section can be found in Appendix A.1

Additionally, the Gumbel preference shock distribution "smooths out" the choice of location, when migration opportunities arrive, so that $a(i,t)$ fraction of agents choose to migrate to i at time t.

$$a(i,t) = \frac{\exp(\theta V(i,t))}{\sum_j \exp(\theta V(j,t))}$$

This choice only depends on the destination, not the origin, because whenever a shock arrives the agent's choice set and value moving forward do not depend on their current location.

2.2 Wage Determination

Wages are the product of a fundamental component which varies by location, and an agglomeration externality which has constant elasticity β with respect to local residence.

$$w(i,t) = \overline{w}(i,t)R(i,t)^{\beta}$$
$$\log w(i,t) = \log \overline{w}(i,t) + \beta \log R(i,t)$$

Note that in the case $\beta = 0$, wages are invariant to the distribution of agents across space.

2.3 Population Evolution

The population flows can then be summarized by the following Kolmogorov Forward Equation (KFE).

$$\dot{R}(j,t) = \lambda[-R(j,t) + \sum_i R(i,t)a(j,t)] = \lambda[-R(j,t) + a(j,t)]$$

The first term on the right accounts for the flows out of every location at rate λ, and the second term accounts for the flows into each location. Note that, because agents' choices do not depend on their current state, the flow rate into each state is the same across all other states.

In the stationary equilibrium, $\dot{R} = 0$, so $R(j) = a(j)$, and the solution may be found explicitly, where $\eta = \frac{\rho+\lambda}{\theta}$.

$$\begin{aligned} R(j) &= \frac{\exp(\theta V(j,t)}{\sum_i \exp(\theta V(i,t)} \\ &= \frac{\overline{w}(j)^{\frac{1}{\eta-\beta}}}{\sum_i \overline{w}(i)^{\frac{1}{\eta-\beta}}} \end{aligned}$$

2.4 Equilibrium

The equilibrium of the model is a pair of agent choices and aggregates which are consistent with the each other and the path of the exogenous component of wages.

Definition 2.1. An equilibrium is a pair of functions $V : \{1, \ldots, n\} \times [0, \infty) \to \mathbb{R}$, $R : \{1, \ldots, n\} \times [0, \infty) \to [0, 1]$ such that

(i) Given wages implied by the population paths, agents optimally choose where to migrate, when given the opportunity (the HJB equation is satisfied)

$$(\rho+\lambda)V(i,t)-\partial_t V(i,t)=\log\overline{w}(i,t)+\beta\log R(i,t)+\lambda[\gamma\theta+\log\sum_j\exp(\theta V(j,t))]$$

(ii) Population path dynamics evolve according to agent choices (the KFE equation is satisfied)

$$\dot{R}(j,t)=\lambda[-R(j,t)+\frac{\exp(\theta V(j,t)}{\sum_i\exp(\theta V(j,t)}]$$

2.5 Perfectly Elastic Limit

In order to obtain further analytic results and intuition, I consider the limiting case without preference shocks: $\theta\to\infty$. This provides a helpful benchmark because agents fully select locations based on their wage values going forward, and removes distributional questions that must be addressed when agents are receiving idiosyncratic shocks[2].

Since $\eta>0$ was previously providing a dispersion force, and $\theta\to\infty$ implies $\eta\to 0$, the only remaining agglomeration or dispersion force is due to local externalities. If $\beta\geq 0$, then many equilibria exist, and below I lay out the explicit set of equilibria. On the other hand, when $\beta<0$, a unique equilibrium exists.

For clarity I restate the relevant expressions for this limiting case. The objective of each agent is

$$\max_{\{i(\tau)\}}\mathbb{E}\left[\int_0^\infty e^{-\rho t}\log w(i,t)\mathrm{d}t\right]$$

subject to the random arrivals τ of migration shocks with Poisson intensity λ

$$F(\tau_{n+1}-\tau_n)=1-\exp(-\lambda(\tau_{n+1}-\tau_n))$$
$$i(t)=\max_{\tau\leq t}i(\tau)$$

The problem of the agent may then be summarized by the HJB

$$\begin{aligned}\rho V(i,t)-\partial_t V(i,t)&=\log w(i,t)+\lambda\mathbb{E}\left[\max_j\{V(j,t)\}-V(i,t)\right]\\&=\log\overline{w}(i,t)+\beta\log R(i,t)+\lambda\mathbb{E}\left[\max_j\{V(j,t)\}-V(i,t)\right]\end{aligned}$$

The KFE is

[2] When considering population welfare (say for a planner problem) is the correct notion ex-ante or ex-post welfare? Should agents with better preference draws simply be luckier in their outcomes, or should some wages be redistributed away from them? These issues are addressed in (Davis and Gregory, 2021), but may be sidestepped entirely in the limit case.

$$\dot{R}(j,t) = \lambda[-R(j,t) + a(j,t)]$$

where $a(j,t)$ is the share of the population choosing to flow into j. Note that if there is a unique maximum $V(j,t)$, then $a(j,t) = 1$, otherwise the flows will potentially be split between all locations sharing the maximum value.[3] In Appendix A.2 I show the stationary distribution is

$$R(i) = \frac{\overline{w}(i)^{-\frac{1}{\beta}}}{\sum_j \overline{w}(j)^{-\frac{1}{\beta}}}$$

2.5.1 Classification of Equilibria

I now turn to classifying how the equilibria depend on β, still maintaining $\theta \to \infty$. The gist of this section is that when $\beta \geq 0$, a mess of equilibria generally exist, but when $\beta < 0$, a unique equilibrium exists. This motivates the focus on $\beta < 0$ in the $\theta \to \infty$ limit.

Theorem 2.1. If $\beta > 0$, there exist up to $2^n - 1$ stationary equilibria, where the exact number depends on how agents select a location when a tie in maximal value occurs. Let E be the set of singletons in $\{1, \ldots, n\}$. Then for any set S of combinations of $\{1, \ldots, n\}$, each combination with at least 2 elements, there exists a selection criteria such that the equilibria are exactly the combinations $E \cup S$ having positive mass, and no other equilibria exist. If $\beta = 0$, and $m = |\{i : \overline{w}(i) = \max_j \overline{w}(j)\}|$, then the set of distributions which form an equilibrium is the m-dimensional probability simplex.

See Appendix A.3 for the proof. The model is not well-behaved in this case because there are no dispersion forces, and the externalities are providing an agglomeration force (or no force when $\beta = 0$). Therefore agents want to agglomerate and full agglomeration at any location is consistent with equilibrium. The additional stationary equilibria are knife-edge cases which hinge upon the selection criteria to exist or not. In sum, if $\beta \geq 0$, it is unclear how one would select the "relevant" equilibrium for a problem. The case with externalities providing a dispersion force is clearer.

Theorem 2.2. If $\beta < 0$, there exists a unique equilibrium.

See Appendix A.3 for the proof. As mentioned above, this is the standard result that when dispersion forces dominate agglomeration forces, then there will exist a unique equilibrium (see (Allen and Arkolakis, 2014), (Ahlfeldt et al., 2015), (Redding and Rossi-Hansberg, 2017)). Below, I show that this equilibrium is also stable in the sense that any initial distribution will converge to the stationary distribution.

For the remainder of the paper, whenever $\theta \to \infty$, I will focus on the case where $\beta < 0$, since this simplifies analysis by ruling out a multiplicity of equilibria, and is likely more empirically relevant in many settings (e.g. if supply of housing is limited, then increasing the density of agents in a location will increase rents). In Appendix A.4 I provide a simple microfoundation supporting the assumption $\beta < 0$.

[3] Unlike in finite θ case, the churn behavior within locations sharing the current maximal value is indeterminate, as any churn process which does not move agents away from their maximal value, and which has the same implied path for the distribution, will be indistinguishable.

3 Results Concerning Dynamics

I now turn to explaining how the stationary equilibrium depends on fundamental parameters which govern dynamics, and the economic interpretation. The closed-form solution for the stationary distribution allows me to consider comparative statics. Along transition paths the same exercises can generally not be completed, but I consider a special class of transition paths for finite θ in which the entire path, from any initial distribution, has a simple solution. I also consider a slightly more general class of transition paths for the perfectly elastic limit case.

3.1 Parameters Governing Dynamics

As already discussed above, the stationary equilibrium has the following closed-form solution.

$$R(j) = \frac{\overline{w}(j)^{\frac{1}{\eta-\beta}}}{\sum_i \overline{w}(i)^{\frac{1}{\eta-\beta}}}$$

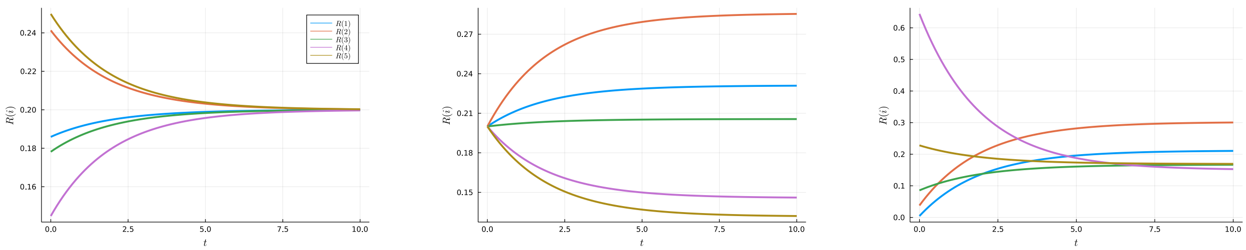
Recall $\eta = \frac{\rho+\lambda}{\theta}$, so changes in both the discount rate and the rate of migration opportunity may be summarized by changes in η, as well as changes in the dispersion of the preference shocks. These three parameters govern substitution choices for agents, since they determine how much they care about the current preference draws versus the flow wages, how soon another preference draw will occur, and how dispersed the preference draws are expected to be at each arrival. Then a marginal change in η tells how the stationary equilibrium changes when the core dynamic parameters of the model change, and it may be found explicitly in terms of fundamentals and the stationary equilibrium

$$\frac{\partial R(j)}{\partial \eta} = -\frac{1}{(\eta-\beta)^2} R(j) \sum_i [\log \overline{w}(j) - \log \overline{w}(i)] R(i)$$

Therefore increasing η will increase $R(j)$ only if the difference between the log wage and all other wages is negative in the weighted average sense using the stationary distribution for weighting. The intuition is that $\eta \equiv \frac{\rho+\lambda}{\theta}$ is a parameter governing how much agents will value the tradeoff between preference draws and flow wages at any given location. If agents sufficiently discount the future (high ρ), if moving opportunities are sufficiently frequent (high λ), or if the dispersion in preference draws is sufficiently high (low θ), then in aggregate agents tend to care more about idiosyncratic draws in the way they make decisions (high η) compared to if η were lower. Therefore as η marginally increases, agents care less about the wages and more about the shocks, and η acts as a dispersion force for determining the population distribution. Whether this dispersion force increases or decreases the population at a location will depend on whether the local log wage is "on average" high or low, where the weighting in the above expression makes this point precise. When the wage is "low" (the sum term is negative) then the dispersion force increases the population, since agents would not select the location for its wage, but will select it when they care more about the preference draws.

3.2 Dynamic Special Case

While it is comparatively simple to characterize the stationary distribution and its dependence on model parameters, it is generally difficult to characterize the transition path analytically. However,



(a) Random Initial R (b) Random $\overline{w}$ (c) Random Initial R and $\overline{w}$

Figure 1: Transition response to permanent change in $\overline{w}$ when $\beta = 0$, $\lambda = \frac{1}{2}$

in the special case without local externalities and without dynamics in the fundamental components of the wages, the equilibrium path starting from any initial distribution is simply exponential convergence to the stationary distribution at rate λ.

Theorem 3.1. When local externalities are not present ($\beta = 0$) and local wages are constant ($w(i,t) = w(i)$), there is a unique stationary distribution $\overline{R}$, and from any initial distribution $\hat{R}$, the time paths of the populations follow

$$R(i,t) = \overline{R}(i) + [\hat{R}(i) - \overline{R}(i)] \exp(-\lambda t)$$

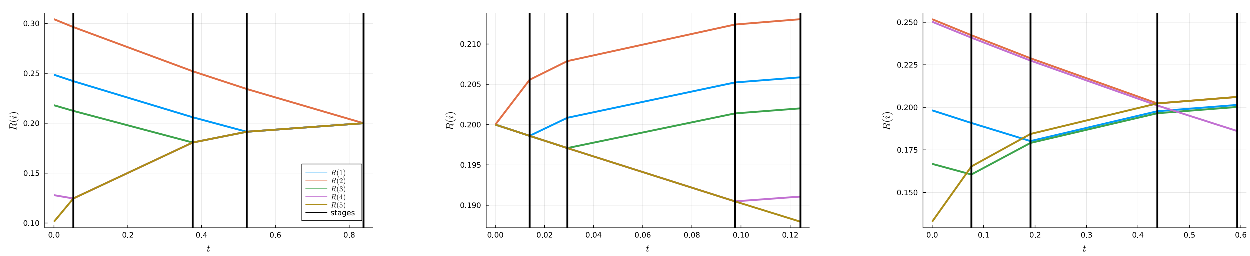
See Appendix A.5 for the proof. The key to the result is that when $\beta = 0$ and $\dot{\overline{w}} = 0$, agents' choice probabilities are constant over time and equal to the stationary distribution, and moving opportunities occur at rate λ, so mass churns at rate λ towards the stationary distribution. It may be tempting to guess that this rate should instead depend on η, or some other combination involving ρ and θ, but ρ merely determines the degree to which agents want to move, so irrelevant to choices when $\overline{w}$ is time-invariant, and θ will affect how individual agents locate, and the rate of churn along the transition path, but in aggregate will not affect choice probabilities. An even simpler intuition for why λ should matter for the rate, but ρ and θ should not, is that λ is a material constraint limiting the mobility of agents, so there is no sense in which this friction can wash out in the aggregate, whereas ρ and θ are preference parameters, which can wash out (θ) or be irrelevant to determining sorting patterns (ρ).

Three examples of transition paths are shown in Figure 1. The thought experiment is that the economy was in a stationary distribution, then received an unanticipated and permanent shock to fundamentals $\overline{w}$. In Figure 1a, the initial stationary distribution had heterogeneity in $\overline{w}$, but the shock equalized $\overline{w}$, so all paths converge to the same populations level. In Figure 1b, there is no heterogeneity in the initial $\overline{w}$, but after the shock the $\overline{w}$ are dispersed. In Figure 1c, there is heterogeneity in the initial $\overline{w}$, and the shock randomly alters the $\overline{w}$, so each individual path still displays exponential convergence, but has unique starting and ending points.

3.3 Transition Paths when $\theta \to \infty$

Perfect substitutability across locations means that if any locations differ in their values, all agents will want to migrate to the highest-valued location. This idea is made formal in the following theorem.

Theorem 3.2. For any given initial distribution $\hat{R}$ and time-invariant fundamental wages $\overline{w}$, the transition path to the stationary distribution may be broken into finitely many stages, each of a finite time horizon.



(a) Random Initial R (b) Random $\overline{w}$ (c) Random Initial R and $\overline{w}$

Figure 2: Transition response to permanent change in $\overline{w}$ when $\beta < 0$, $\lambda = \frac{1}{2}$

See Appendix A.6 for the proof. It is worth further elucidating how the stages work. At the beginning of each stage, all locations are either in the set M, of locations sharing the maximal value, or its complement, M^c. Agents living in $i \in M^c$ want to move to a location $j \in M$ as soon as they are able, so flows are entirely from M^c to M. This will mean that the population for all $i \in M^c$ declines at exponential rate λ during the stage. The populations for $j \in M$ are slightly trickier, because merely sending all the incoming mass to locations in M arbitrarily will generally mean that some location values will rise faster than others. This cannot be an equilibrium because agents have perfect foresight, and thus no agent would move to a location with any value but the highest. Therefore the flows into M must be apportioned exactly such that all locations in M continue to share the maximal value. The final element to describe the stage is that its ending is determined by the time needed for the maximal value (shared by all $i \in M$) to equal the maximum value of all locations in M^c. Since $\beta < 0$, the intermediate value theorem guarantees that there exists a t at which the population in some location $i \in M^c$ will be low enough that its value will have risen to equal the value of all locations in M, which will be falling.

The dynamics found in this solution also match the way that tatonnement is discussed in the context of solving static models. In the static framework, there is some sense that if one location becomes more productive, then agents will move in and lower the value of that location (raise rents in the microfoundation provided in Appendix A.4) until it equalizes with the other locations. The dynamic setting here plays out the tatonnement process over a finite time horizon, and indeed locations which start with low valuations (either due to high R or low $\overline{w}$) see residents leaving until values enter the maximal set, at which point they rise and continue to do so until all locations are equalized. However, to my knowledge, this is the first result making explicit the non-monotonicity of the population paths, however, as in informal discussions of tatonnement in static problems there may be the (here shown to be wrong) sense that all locations will fall or rise monotonically until hitting their stationary distribution at the exact same time. While this non-monotonicity may be surprising, it is natural given how agents will sort towards the highest-valued location(s), and below I show this path is also efficient.

3.4 Example: Many Improvements

Figure 2 displays some example transition paths and clarifies how the staging process works. Each vertical black bar is the end of a stage. In Figure 2a, the initial distribution is random, but the $\overline{w}$ is constant across locations, so the stationary distribution has the same population everywhere. This symmetry means that for two locations to have the same value, they need to have the same population, and this can be seen in the figure where each location loses mass at rate λ until their population matches the set of locations with the lowest mass, then this mass rises more slowly. In

Figure 2b, the initial distribution has the same population everywhere, but the $\overline{w}$ is random, so the stationary distribution is not uniform. Since the initial distribution is uniform, all locations' populations fall along the same path until their value rises enough that they enter the maximal value set, at which point they start rising. Figure 2c has both a random initial distribution and random $\overline{w}$. In all three figures, it is immediate to see whether any given location is in M for any given stage, since the path is falling when in M^c, and rising when in M.

4 Efficiency

I now consider efficiency concerns regarding the stationary equilibrium and dynamics. I find that when β is constant across locations, not only is the stationary equilibrium efficient, but so will be the transition path from any initial distribution to the stationary distribution.

4.1 Stationary

I first consider the stationary distribution. Since the stationary distribution is isomorphic to the some of the models considered in (Fajgelbaum and Gaubert, 2019), except with amenities constant across locations, the stationary equilibrium will be efficient. The problem of the planner is

$$\max_R \sum_i R(i)w(i)$$
$$1 = \sum_i R(i)$$

To clarify exactly what delivers the efficiency result, in this section I allow for the elasticity of wages to to the local population to vary by location, call it β_i. Then the planner will select $R(i)$ such that

$$\overline{w}(i)(1+\beta_i)R(i)^{\beta_i} = \overline{w}(j)(1+\beta_j)R(j)^{\beta_j}$$

for all i, j. The details for solving the planner's problem can be found in Appendix A.7.

Therefore the equilibrium is generally inefficient, except for the case of interest, $\beta_i \equiv \beta$ in which case externalities "cancel out". Otherwise over or under allocation depends on relative elasticities and fundamentals. If $|\beta_i| > |\beta_j|$, then $\frac{1+\beta_j}{1+\beta_i} > 1$, so the planner will want to allocate less to i than the decentralized allocation.

A planner with access to taxes and transfers by location can achieve the first-best by setting transfers such that $w(i) = \overline{w}(i)(1+\beta_i)R^*(i)^{\beta_i}$. Then agents will face the same relative valuations across locations as the planner. This is a transfer from areas with high externalities to lower ones, such that when agents only take into account the externality for themselves, they choose the planner solution. In locations where the competitive equilibrium yields $R(i) > R^*(i)$, the planner then taxes wages, and transfers them via subsidies to locations where $R(i) < R^*(i)$.

It is worth noting that the existence of externalities alone is not what what generates the inefficiency, but the potential for some locations to have externalities with larger elasticities than others. Total population is always unity, so inefficiency can only ever be rooted in misallocation, not uniform under- or overallocation. To see this point particularly clearly, not that when $\beta_i = \beta$, the planner's solution has a closed form and it matches the decentralized allocation. In this special case, agents are not internalizing their effect on any location, but the the way in which they are doing so is uniform across locations, so the externalities "cancel out".

4.2 Dynamics

The transitional dynamics presented here are, to my knowledge, new. Similarly to the stationary distribution, the planner is subject to the same material constraints as in the stationary distribution, so when $\beta_i \equiv \beta$, the planner takes into account that the externalities uniformly affect the marginal value of sending flows to any given location, and thus the planner delivers the same allocation along the entire transition path as the decentralized outcome.

The planner solves the following problem, given an initial distribution $\hat{R}$ and subject to churn rate λ.

$$\max_a \int_0^\infty e^{-\rho t} \sum_i R(i,t) w(i,t)$$

$$\dot{R}(i,t) = \lambda[-R(i,t) + \sum_j a(j,i,t) R(j,t)]$$

$$\sum_j a(i,j,t) = 1$$

The problem will generally not be analytically solvable, but in the special case of a permanent and unexpected change in $\overline{w}$, the following theorem applies.

Theorem 4.1. If the wage elasticity to local population is constant and constant across locations, then the decentralized allocation path is efficient after a permanent change in fundamentals $\overline{w}$.

See Appendix A.8 for the proof. As in the stationary case, the key is that with constant β_i, the planner accounts for externalities across time and directs flows accordingly, but the externalities have a uniform effect on valuations, compared to the way decentralized outcome generated by agents' valuations, so the flows are exactly the same.

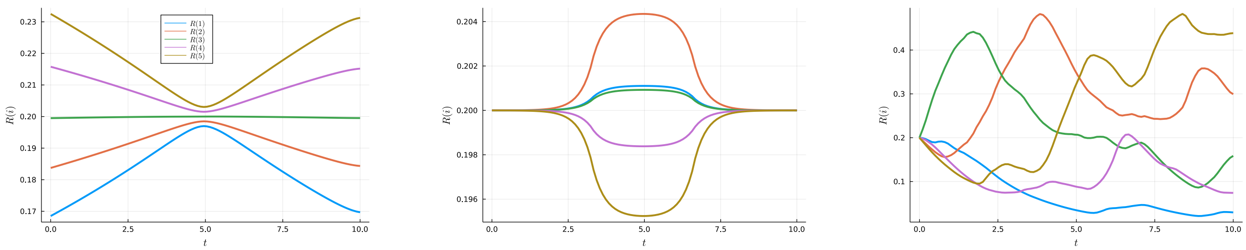
5 Extensions

The way dynamics have been introduced in the simple spatial model above is generalizable to more involved models without losing the core tractability. To make this point more concretely, I now consider three brief extensions of the baseline model to point towards further extensions that maintain tractability for potentially otherwise unwieldy problems.

5.1 Dynamic Fundamental Component of Wages

I have thus far only considered permanent and immediate changes in $\overline{w}$ because in that case the analytics are simpler and even closed-form. However, the general theory boils down to the HJB-KFE system of differential equations, which makes no requirements about the paths of exogenous parameters, and thus the system can be solved numerically. To this point, I now consider some more exotic paths of $\overline{w}$ and solve the system. The computational details are the same as for the permanent change in $\overline{w}$, and are in Appendix B.1.

In Figure 3, I display the time paths of the population distribution for three different paths of $\overline{w}$. In Figure 3a, the paths of $\overline{w}$ fall until the midpoint of the time horizon, then rise, but each location falls and rises at a different rate. The locations which have $\overline{w}$ falling most quickly lose the most mass in the middle of the path, and those with the smallest decline in $\overline{w}$ even gain population. Figure 3b is something of a reverse story, where during the the middle third of the time, each



(a) $\overline{w}$ fall then rise (b) $\overline{w}$ increase in middle (c) $\overline{w}$ follows AR(1)

Figure 3: Transition response to variety of paths for $\overline{w}$

location receives a random shock to $\overline{w}$. The perfect foresight of agents will lead to anticipatory moving towards the locations with the better shocks even before the shock is realized. Finally in Figure 3c, the $\overline{w}$ of each location follows an independent AR(1) process. Agents anticipate the paths and recognize the tradeoff that some locations will be better today, but worse tomorrow, and there is the risk of becoming stuck at any location longer than desired because of the infrequent migration opportunities.

5.2 Movement Costs

The tractability of introducing moving frictions through Poisson shocks opens the door to more involved phenomena. For example, in the baseline model above (regardless of the value of θ), the only friction to moving is the arrival of a moving opportunity. This simplification meant that individuals in different locations would value moving to any given location in the same way. In reality, there will also be costs to moving from the current location. With this aspect in mind, the value of moving to each location will depend on the current location. I now show how to incorporate this friction into the model.

For simplicity, I will assume the moving costs κ are paid immediately, so are similar to the preference shocks. Additionally, I will assume true fixed costs, so that the same cost is required to move to any location, from any location, with the exception that agents may choose to not move, in which case they still enjoy the new preference shock from their own location, but do not have to pay the moving cost. The HJB is then

$$\rho V(i,t) - \partial_t V(i,t) = \log w(i,t) + \lambda \mathbb{E}\left[\max_j \{V(j,t) + \epsilon(j,t) - \kappa \mathbf{1}\{i \neq j\}\} - V(i,t)\right]$$

Again, the expectation on the right may be computed, the equation can be rearranged, and the form for w may be substituted to yield the following equation for the stationary HJB.

$$(\rho + \lambda)V(i) = \log \overline{w}(i) + \beta \log R(i) + \lambda[\gamma\theta + \log \sum_j \exp\left(\theta(V(j) - \kappa \mathbf{1}\{i \neq j\})\right)]$$

The stationary distribution no longer has such a simple closed-form solution, since the term on the far right depends on i through the moving cost. The KFE only changes through the changes in choice probabilties. Unsurprisingly, the effect of the moving cost is to reduce the amount of flows out of any location, since every location other than the current location is devalued by κ.

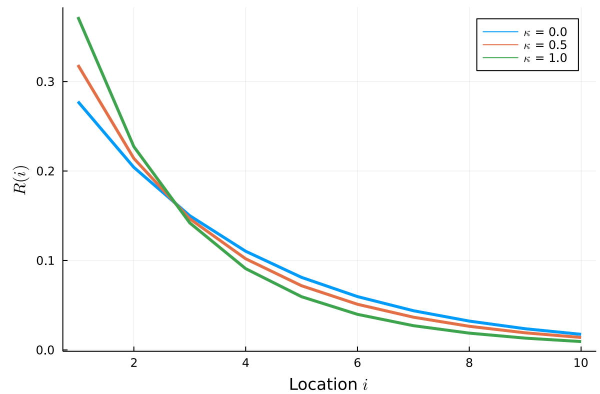


Figure 4: Dependence of stationary distribution on moving costs κ

$$\dot{R}(i) = \lambda[-R(i) + \sum_j R(j)a(j,i)]$$
$$a(j,i) = \frac{\exp(\theta(V(i) - \kappa \mathbf{1}\{i = j\})}{\sum_k \exp(\theta(V(k) - \kappa \mathbf{1}\{k = j\})}$$

I now consider a simple concrete example to show how the stationary distribution varies with the moving cost. I consider a set of $n = 10$ locations, ordered by $\overline{w}$ (1 is highest, 10 is lowest), and set $\beta = -0.2, \theta = 2, \lambda = 0.2$. I consider three cases: free moving costs, moderate moving costs, and high moving costs. In Figure 4, I show the implied stationary distribution for each case. The additional mobility cost imposed by κ can be thought of as skewing the preference shocks towards home bias, and since this creates an agglomeration force, the population concentrates more at the locations with the higher $\overline{w}$.

5.3 Lifecycle Considerations

Incorporating lifecycle aspects may also be of interest for properly modeling individual location decisions over the lifetime, and for capturing how the distribution of agents across space varies with age. For example, (De La Roca and Puga, 2017) find that workers gain experience from living in big cities, and that this experience continues to be useful even after emigrating from these cities. To incorporate these effects, I augment the state space to include age A and human capital k, but assume that aggregates are invariant and focus on the stationary distribution. Instead of an infinite horizon, agents live for a finite horizon $[0, T]$, and $\frac{1}{T}$ agents die every instant to be reborn at a random location over some exogenous distribution. Additionally, I assume human capital acquisition depends on how effective each location is at giving experience, so

$$\dot{k} = l(i)k^\alpha - \delta k$$

where $l(i)$ is a local learning coefficient, and δ is experience depreciation (constant across locations). After simplifying using the preference shock distribution, the stationary HJB is

$$(\rho + \lambda)V(i, A, k) = \log w(i) + \partial_A V(i, A, k) + \partial_k V(i, A, k)[l(i)k^\alpha - \delta k] + \lambda[\gamma\theta + \log \sum_j \exp\left(\theta(V(j, A, k))\right)]$$

The KFE will also be more involved.

$$\dot{R}(i,A,k) = \underbrace{\lambda[-R(i,A,k) + \sum_j \frac{\exp(\theta V(j,A,k)}{\sum_{j'} \exp(\theta V(k,A,j'))}]}_{\text{location flow}} - \underbrace{\partial_A R(i,A,k)}_{\text{age flow}} - \underbrace{\partial_k [a(i,A,k) R(i,A,k)}_{\text{human capital flow}}$$

Note that the model is now a system of partial differential equations. This will increase the difficulty of computation, but since the model is cast in continuous time, the difficulties of simultaneous decisions are sidestepped, and flows in each direction may be considered separately, as seen in both the HJB and KFE above. Additionally, only one dimension of choice exists, since both age and human capital evolve exogenously at each location. Then finite difference methods as in (Barles and Sougandis, 1991) can be used to solve the system, as detailed in (Achdou et al., 2021).

6 Quantification

The model is simple to take to data, if one wants to calibrate the stationary distribution to a given moment in time. It is also flexible to scale, in that agents choosing locations may be choosing between cities, counties, or even areas within a city. The core idea is that the churning parameter λ may be estimated from microdata on how frequently agents move locations at the desired scale. For example, using the 2007 American Community Survey (ACS) public use microdata sample, it is estimated that the average individual moves around 12 times in their lifetime. Given a current life expectancy of around 80 years, and allowing one unit of time to be one year in the model, a back-of-the-envelope calculation yields $\lambda = \frac{12}{80} = 0.15$. This calculation is overly simple and misses, for example, the lifecycle considerations above, but is fine for gaining an understanding of the basic calibration approach. [4].

The remaining parameters of ρ, θ, and β which govern, respectively, substitution across time, substitution across space, and local externalities, are less straightforward to calculate. Typically, a set of plausible estimates for each parameter is chosen, and results are reports for the range of estimates. Alternatively, perhaps one or two parameters will be selected, such as β and θ, and the third parameter (ρ) is then calibrated to some other moment. From the solution for the stationary distribution, however, it should be clear that for this model all that will matter will be η and β, in terms of determining the stationary distribution, and thus it would be wise to consider a range of these parameters in order to check the robustness of any given calibration.

Finally, to estimate the fundamental components of wages for each location, assume that at a given time the equilibrium is stationary and simply use the log of the stationary equilibrium solution to see

$$\log R(j) = \frac{\log \overline{w}(j)}{\eta - \beta} - \log \sum_i \exp(\frac{\log \overline{w}(i)}{\eta})^{\frac{\eta}{\eta - \beta}}$$

Note that log wages are only defined up to an additive constant, so proper normalization yields

$$\log \overline{w}(j) = (\eta - \beta) \log R(j)$$
$$\Rightarrow \overline{w}(j) = R(j)^{\eta - \beta}$$

[4]Individuals with different wealth, income, skill levels etc. may all have different movement rates, which may generate selection effects across locations after a change in fundamentals. Since this paper is merely concerned with laying out the foundations of understanding the transitional dynamics, these concerns are left for future work.

One advantage of the way I have chosen to model migrational frictions is that flow rates are bounded above by λ. Thus, if flow rates greater than λ are needed to justify the data, then it must be the case that the migrational friction alone cannot explain the transitional dynamics in the spatial distribution, and this may motivate the inclusion of the moving costs or lifecycle elements considered above.

7 Conclusion

I have exposited a tractable way to introduce dynamics into models of location choice. Two key ingredients are the Poisson arrival of migration opportunities, which makes arrival opportunity memoryless, and the use of continuous-time methods, which allows the problem to be reduced to a coupled system of differential equations (ordinary for the baseline model, but partial for extensions with additional states). I solve for the stationary distribution and for transition paths in certain cases, and show that the baseline model is efficient not only in the stationary distribution, but also along the transition path. I give a few examples of how the baseline model may be extended, to show that the tractability generalizes to more involved applications. Lastly, I suggest one means for calibrating the model. My hope is that the techniques explored here can be used to introduce dynamics into quantitative spatial models in a way that is both theoretically appealing and intuitive, and that remains tractable.

References

Achdou, Yves et al. (2021). "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach". In: *Review of Economics Studies*.

Ahlfeldt, Gabriel M. et al. (2015). "The Economics of Density: Evidence from the Berlin Wall". In: *Econometrica* 83.6, pp. 2127–2189.

Allen, Treb and Costas Arkolakis (2014). "Trade and the Topography of the Spatial Economy". In: *The Quarterly Journal of Economics*, pp. 1085–1139.

— (2020). "The Welfare Effects of Transportation Infrastructure Improvements". In: *Working Paper*.

Anderson, James E. and Eric van Wincoop (2003). "Gravity with Gravitas: A Solution to the Border Puzzle". In: *The American Economic Review* 93.1, pp. 170–192.

Barles, G. and P.E. Sougandis (1991). "Convergence of approximation schemes for fully nonlinear second order equations". In: *Asymptotic Analysis* 4, pp. 271–283.

Caliendo, Lorenzo, Maximilliano Dvorkin, and Fernando Parro (2019). "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock". In: *Econometrica* 87.3, pp. 741–835.

Cruz, Jose-Luis and Esteban Rossi-Hansberg (2021). "The Economic Geography of Global Warming". In: *Working Paper*.

Davis, Morris A. and Jesse Gregory (2021). "Place-Based Redistribution in Location Choice Models". In: *Working Paper*.

De La Roca, Jorge and Diego Puga (2017). "Learning by Working in Big Cities". In: *Review of Economic Studies* 84, pp. 106–142.

Desmet, Klaus, David Krisztian Nagy, and Esteban Rossi-Hansberg (2018). "The Geography of Development". In: *Journal of Political Economy* 126.3.

Eaton, Jonathan and Samuel Kortum (2002). "Technology, Goegraphy, and Trade". In: *Econometrica* 70.5, pp. 1741–1779.

Fajgelbaum, Pablo D. and Cecile Gaubert (2019). "Optimal Spatial Policies, Geography and Sorting". In: *Working Paper*.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst (2013). "Endogenous gentrification and housing price dynamics". In: *Journal of Public Economics* 100, pp. 45–60.

Kaplan, Greg, Benjamin Moll, and Gianluca Violante (2018). "Monetary Policy According to HANK". In: *American Economic Review* 108.3, pp. 697–743.

Olivero, Maria Pia and Yoto T. Yotov (2012). "Dynamic gravity: endogenous country size and asset accumulation". In: *The Canadian Journal of Economics* 45.1, pp. 64–92.

Redding, Stephen J. and Esteban Rossi-Hansberg (2017). "Quantitative Spatial Economics". In: *Annual Reviews of Economics* 9, pp. 21–58.

Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life". In: *Journal of Political Economy* 90.6, pp. 1257–1278.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation, Second Edition*. New York, NY: Cambridge University Press.

A Mathematical

In this appendix I provide proofs and missing details for the analytic results in the main text. The finite θ and $\theta \to \infty$ cases are generally treated separately since at the limit slightly different arguments are needed. The core economics do not change, however, as the finite θ case may be thought of as a "smoothing approximation" to the limit case (see (Eaton and Kortum, 2002) for this idea in a trade context, or (Allen and Arkolakis, 2020) for this idea in the context of commuting over a network of linked locations).

A.1 Stationary distribution for finite θ

The HJB is

$$\rho V(i,t) - \partial_t V(i,t) = \log w(i,t) + \lambda \mathbb{E}\left[\max_j \{V(j,t) + \epsilon(j,t)\} - V(i,t)\right]$$

Each preference shock is Gumbel$(0,\theta)$, so $V(j,t) + \epsilon(j,t)$ has distribution

$$\begin{aligned} F_j(x) &= Pr[V(j,t) + \epsilon(j,t) \leq x] \\ &= Pr[\epsilon(j,t) \leq x - V(j,t)] \\ &= \exp(-\exp(-\theta(x - V(j,t)))) \end{aligned}$$

and $V(j,t) + \epsilon(j,t) \sim$ Gumbel$(V(j,t), \theta)$. Since the preference shocks are independent and identically distributed, their maximum is simply the product of their cdfs

$$\begin{aligned} F(x) &= \prod_j F_j(x) \\ &= \exp(-\sum_j \exp(-\theta(x - V(j,t)))) \\ &= \exp(-\exp(-\theta(x - \left(\log \sum_j \exp(\theta V(j,t))\right)))) \end{aligned}$$

So $\max_j V(j,t) + \epsilon(j,t) \sim$ Gumbel$(\log \sum_j \exp(\theta V(j,t)), \theta)$. This distribution has mean $\gamma\theta + \log \sum_j \exp(\theta V(j,t))$, where γ is the Euler-Mascheroni constant. The HJB may then be written

$$(\rho + \lambda)V(i,t) - \partial_t V(i,t) = \log w(i,t) + \lambda[\gamma\theta + \log \sum_j \exp(\theta V(j,t))]$$

Considering the stationary distribution ($V(i,t) \equiv V(i)$), and noting the farthest right term is a constant, call it K, I have

$$(\rho+\lambda)V(i) = \log w(i) + K$$
$$\Rightarrow V(i) = \frac{\log w(i) + K}{\rho+\lambda}$$

Then I can solve for K

$$\begin{aligned}
K &= \lambda[\gamma\theta + \log \sum_j \exp(\theta V(j))] \\
&= \lambda[\gamma\theta + \log \sum_j \exp(\theta \frac{\log w(j) + K}{\rho+\lambda})] \\
&= \lambda[\gamma\theta + \frac{\theta K}{\rho+\lambda} + \log \sum_j \exp(\theta \frac{\log w(j)}{\rho+\lambda})] \\
\Rightarrow K &= \frac{\rho+\lambda}{\rho+\lambda(1-\theta)} \lambda[\gamma\theta + \log \sum_j \exp(\theta \frac{\log w(j)}{\rho+\lambda})]
\end{aligned}$$

Which solves for the values

$$V(i) = \frac{\log w(i)}{\rho+\lambda} + \frac{\lambda}{\rho+\lambda(1-\theta)} [\gamma\theta + \log \sum_j \exp(\theta \frac{\log w(j)}{\rho+\lambda})]$$

To solve for the stationary population distribution, I may calculate the probability that location j is the max at an opportunity for migration, $V(j) + \epsilon(j) = \max_i V(i) + \epsilon(i)$, using the choice probabilities implied by the Gumbel distribution (see (Train, 2009)).

$$a(j) = \frac{\exp(\theta V(j))}{\sum_i \exp(\theta V(i))}$$

The KFE will imply that the choice probabilities in the stationary distribution must match the stationary distribution

$$0 = \dot{R}(i) = \lambda[-R(i) + a(i)]$$
$$\Rightarrow a(i) = R(i)$$

Using the solution for V above, that $w(i) = \overline{w}(i) + \beta \log R(i)$, and that the choice probabilities must equal the stationary distribution, I solve

$$R(j) = \frac{\exp(\frac{\theta(\overline{\log w}(j)+\beta\log R(j))}{\rho+\lambda})}{\sum_i \exp(\frac{\theta(\log\overline{w}(i)+\beta\log R(i))}{\rho+\lambda})}$$

$$\Rightarrow R(j) = \frac{\exp(\frac{\theta}{\rho+\lambda}\log\overline{w}(j))^{\frac{\rho+\lambda}{\rho+\lambda-\beta\theta}}}{\sum_i \exp(\frac{\theta}{\rho+\lambda}\log\overline{w}(i))^{\frac{\rho+\lambda}{\rho+\lambda-\beta\theta}}}$$

$$= \frac{\exp(\frac{\log\overline{w}(j)}{\eta})^{\frac{\eta}{\eta-\beta}}}{\sum_i \exp(\frac{\log\overline{w}(i)}{\eta})^{\frac{\eta}{\eta-\beta}}}$$

$$= \frac{\overline{w}(j)^{\frac{1}{\eta-\beta}}}{\sum_i \overline{w}(i)^{\frac{1}{\eta-\beta}}}$$

A.2 Stationary distribution for $\theta \to \infty$ case

The HJB is

$$\rho V(i,t) - \partial_t V(i,t) = \log\overline{w}(i,t) + \beta\log R(i,t) + \lambda\mathbb{E}\left[\max_j\{V(j,t)\} - V(i,t)\right]$$

Impose stationarity and that values are equalized across space $V(i,t) \equiv V$ (else there would be flows from low value into high value locations).

$$\rho V = \log\overline{w}(i) + \beta\log R(i)$$

Exponentiate and equate the i and j versions of the HJB, since values are equalized

$$\overline{w}(i)R(i)^{\beta} = \overline{w}(j)R(j)^{\beta}$$

$$R(i) = \frac{\overline{w}(j)^{\frac{1}{\beta}}}{\overline{w}(i)^{\frac{1}{\beta}}}R(j)$$

Use that total labor supply is unity

$$1 = \sum_i R(i) = R(j)\overline{w}(j)^{-\frac{1}{\beta}}\sum_i \overline{w}(i)^{-\frac{1}{\beta}}$$

$$R(j) = \frac{\overline{w}(j)^{-\frac{1}{\beta}}}{\sum_i \overline{w}(i)^{-\frac{1}{\beta}}}$$

A.3 Classifying equilibria

Theorem A.1. If $\beta > 0$, there exist up to $2^n - 1$ stationary equilibria, where the exact number depends on how agents select a location when a tie in maximal value occurs. Let E be the set of singletons in $\{1,\ldots,n\}$. Then for any set S of combinations of $\{1,\ldots,n\}$, each combination with at least 2 elements, there exists a selection criteria such that the equilibria are exactly the combinations $E \cup S$ having positive mass, and no other equilibria exist. If $\beta = 0$, and $m = |\{i : \overline{w}(i) = \max_j \overline{w}(j)\}|$, then the set of distributions which form an equilibrium is the m-dimensional probability simplex.

Proof. Let $\beta > 0$, and S be some combinations of $\{1, \ldots, n\}$ with at least 2 elements. Then the selection criteria such that, if the distribution is such that the values for locations $s \in S$ are equalized, and zero mass is elsewhere, then agents do not move, will maintain the desired combination as an equilibrium. Then set this selection rule for all $s \in S$, and for $s' \notin S$, instead set a lexicographic selection rule for locations by their index. This will rule out all other combinations with 2 or more elements, since if those potential equilibria arose, agents would move to the lower indexed value, increasing its value due to the positive externality, and then the selection rule would not longer be relevant, because all agents will move to the location with the highest value, again further increasing its value due to the the positive externality. The remaining singleton equilibria always persist, however, because if all the mass is at a single location, then that location has positive value, but all other locations have zero value, so no one will move.

Let $\beta = 0$. Then $w(i) = \overline{w}(i)$ for all i, and the population distribution is irrelevant for determining wages. Any agent at location with $\overline{w}$ less than the maximal $\overline{w}$ will then flow towards one of the maximal locations, so there must be zero mass in all locations without the maximal $\overline{w}$. Now allow for an arbitrary distribution of the population over the support of locations with the maximal $\overline{w}$. No agent will have an incentive to move, and thus this is an equilibrium (both the HJB and KFE are satisfied and stationary). Then the result follows, since the set of distributions with support over the m maximal locations is the m-dimensional probability simplex. □

Theorem A.2. If $\beta < 0$, there exists a unique equilibrium.

Proof. If any location has zero mass, then since $\beta < 0$ the effective wage will be unbounded, and hence provide greater value than any location that has positive mass. This cannot be a stationary equilibrium, because agents would flow from every location with positive mass to the one with zero mass. Thus, every location has positive mass, and for the same reasoning must have the same value.

Consider two distinct distributions, R^1 and R^2, of the population across locations. Since they are distinct, at least one location j must have a lower population in R^2 than R^1, since not all locations can have the same population or higher and still satisfy $\sum_i R(i) = 1$. Then the wage at j is higher in R^2. Similarly, at least one location k must have a higher population in R^2 than R^1, and therefore must have a lower wage. So $V^2(j) > V^1(j)$ and $V^2(k) < V^1(k)$. If R^1 is a stationary equilibrium, then $V^2(j) > V^1(j) = V^1(k) > V^2(k)$, so R^2 is not a stationary equilibrium, and if R^2 is a stationary equilibrium, then $V^1(k) > V^2(k) = V^2(j) > V^1(j)$, so R^1 is not a stationary equilibrium. Hence the stationary equilibrium is unique, and exists by the direct argument above in Appendix A.2 which found the allocation.

□

A.4 Microfoundation for $\beta < 0$

One microfoundation for $\beta < 0$ simply relies on the fact that land is a scarce resource, and thus housing a larger population in a location will increase rents.

First, assume that the consumption good in the main text (call it C here) is comprised of a Cobb-Douglas aggregation of housing h and all other consumption goods c, with weight α on housing. Then flow utility follows

$$\begin{aligned}\log C &= \log\left(\left(\frac{h}{\alpha}\right)^{\alpha}\left(\frac{c}{1-\alpha}\right)^{1-\alpha}\right)\\ &= \alpha\log h + (1-\alpha)\log c - \alpha\log\alpha - (1-\alpha)\log(1-\alpha)\end{aligned}$$

Letting c now be the numeraire and denoting rents r, the agent's budget is

$$c + rh = \overline{w}$$

They then solve

$$\max_{c,h} \alpha \log h + (1-\alpha) \log c$$
$$c + rh = \overline{w}$$

and find

$$c = (1-\alpha)\overline{w}$$
$$h = \frac{\alpha \overline{w}}{r}$$

with indirect flow utility

$$v = \log \overline{w} - \alpha \log r$$

Now suppose that each location has a competitive housing sector which builds housing subject to convex costs

$$\mathcal{C}(h) = h^{1+\nu}, \quad \nu > 0$$

Since the sector is competitive, there will be zero profits, and rents for any given demand of housing are

$$r = h^{\nu}$$

Using this price in the indirect flow utility of agents I find

$$v = \log \overline{w} - \alpha\nu \log h$$

If $\beta = -\alpha\nu$, then the reduced form indirect flow utility is recovered when $R = h$ agents live in a location, and $\beta < 0$, as long as the consumption share in housing is positive and building costs are convex.

A.5 Dynamic Special Case with Finite θ

Theorem A.3. When local externalities are not present ($\beta = 0$) and local wages are constant ($w(i,t) = w(i)$), there is a unique stationary distribution $\overline{R}$, and from any initial distribution $\hat{R}$, the time paths of the populations follow

$$R(i,t) = \overline{R}(i) + [\hat{R}(i) - \overline{R}(i)] \exp(-\lambda t)$$

Proof. Since externalities are not present and local wages are constant, there is no time dependence in the HJB, so it may be solved explicitly as in the stationary distribution case to yield the following choice probabilities.

$$a(i,j) = \frac{\overline{w}(j)^{\frac{1}{\eta}}}{\sum_k \overline{w}(k)^{\frac{1}{\eta}}}$$

Consider A as the matrix with $a(i,j)$ in its (i,j)-th entry. Each row is identical and sums to 1. Now recall that A' governs the population evolutions via the KFE

$$\dot{R} = \lambda(A' - I)R$$

where $R \equiv (R(1), \ldots, R(n))$. Note that A' idempotent, and therefore

$$(A' - I)^2 = (A')^2 - 2A' + I = -(A' - I)$$

So $(A' - I)$ is skew-idempotent, and its eigenvalues are $\{0, -1\}$, by the following line, where v is an arbitrary eigenvector, and ξ the corresponding eigenvalue

$$\xi v = (A' - I)v = -(A' - I)^2 v = -(A' - I)\xi v = -\xi^2 v$$

Now note that A' is an irreducible and aperiodic (hence ergodic) Markov matrix[5], so there is a unique distribution $\overline{R}$ such that $A'\overline{R} = \overline{R}$. Then this distribution yields $\dot{R} = 0$, and no other distribution does, for then that other distribution would also be a stationary distribution of A', violating uniqueness. The remainder of the eigenbasis may be constructed of linearly independent vectors of the form $\tilde{R}_i = (-1, \underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{n-i-1})$ for $i \in \{2, \ldots, n\}$, which all have eigenvalue -1. To verify this, note that the $\tilde{R}_i$ are a linearly independent set, because if any linear combination $\sum_{i=2}^n c_i \tilde{R}_i$ has a $c_j \neq 0$ for some i, then $(\sum_{i=2}^n c_i \tilde{R}_i)_j = c_j \neq 0$. To see that $\overline{R}$ is not in the span of the set of $\tilde{R}$, note that if $\overline{R}_j = (\sum_{i=2}^n c_i \tilde{R}_i)_j$ for $j = \{2, \ldots, n\}$, then $c_j = \overline{R}_j$, which then implies $(\sum_{i=2}^n c_i \tilde{R}_i)_1 = -\sum_{i=2}^n c_i < 0 < \overline{R}_1$.

The initial distribution may then be decomposed as $\hat{R} = c_1 \overline{R} + \sum_{i=2}^n c_i \tilde{R}_i$ for some $\{c_i\}$. The KFE may be solved for some constants $\{b_i\}$

$$R(j,t) = b_1 \overline{R}(j) + \sum_{i=2}^{n} b_i \tilde{R}_i(j) \exp(-\lambda t)$$

The initial condition determines the $\{b_i\}$, which yields the stated result.

□

[5] Transitions for the matrix in the Markov interpretation are notationally $x^{t+1} = A'x^t$.

A.6 Transition Path for $\theta \to \infty$

Theorem A.4. For any given initial distribution $\hat{R}$ and time-invariant fundamental wages $\overline{w}$, the transition path to the stationary distribution may be broken into finitely many stages, each of a finite time horizon.

Proof. Suppose $\overline{w}$ are fixed and $\hat{R}$ is an arbitrary distribution of the population at $t = 0$. Let $M(t)$ be the set of locations which share the maximal value at t, and $M^c(t)$ the complement. Then for $i \in M^c(t)$, all agents want to leave, so

$$\begin{aligned} \dot{R}(i,t) &= -\lambda R(i,t) \\ \Rightarrow R(i,t) &= R(i,0)\exp(-\lambda t) \end{aligned}$$

All the flows exiting M^c enter M, but they must do so as to preserve the value equalization within M. If value equalization were not preserved, then some agents would be flowing into a location with less than the maximal value at some instant, which would be suboptimal. Also, I note

$$V(i,t) = \frac{\log \overline{w}(i,t) + \beta \log R(i,t)}{\rho + \lambda} + K$$

where K does not depend on i. Thus

$$\dot{V}(i,t) = \frac{\beta \dot{R}(i,t)}{(\rho + \lambda)R(i,t)} + \dot{K}$$

Since values are equalized in M, so are their paths, and it must be the case that $\frac{\dot{R}}{R}$ is equalized across locations. Then I require

$$\sum_{i \in M} \dot{R}(i,t) = -\sum_{j \in M^c} \dot{R}(j,t), \quad \forall t \qquad \text{(Flow Preservation)}$$

$$\dot{R}(j,t) = -\lambda R(j,t) \qquad \text{(Flows out of } M^c\text{)}$$

$$\frac{\dot{R}(i,t)}{R(i,t)} = \frac{\dot{R}(i',t)}{R(i',t)}, \quad \forall i, i' \in M \qquad \text{(Value Equalization in } M\text{)}$$

This system has the following solution for $i \in M$ and any time interval where no location moves from M^c to M

$$R(i,t) = R(i,0)[1 + \frac{\sum_{j \in M^c} R(j,0)}{\sum_{j \in M} R(j,0)}(1 - \exp(-\lambda t))]$$

In order to find the length of time before some location in M^c joins M, I find the t that equates the maximal value in M^c with the value in M.

$$\max_{i\in M^c}\{\log \overline{w}(i) + \beta \log R(i,t)\} = \log \overline{w}(j) + \beta \log R(j,t)$$

where $j \in M$ may be arbitrary. The value in M is strictly falling with t, and the values for all $i \in M^c$ are strictly rising with t, so the max is also, and the equation will have a solution by the intermediate value theorem. The strictness guarantees the uniqueness.

□

A.7 Planner's problem for stationary equilibrium

The planner solves

$$\max_R \sum_i R(i)w(i)$$
$$1 = \sum_i R(i)$$

Their Lagrangian is then

$$\mathcal{L} = \sum_i \overline{w}(i)R(i)^{1+\beta_i} + \mu[1 - \sum_i R(i)]$$

The first-order condition for $R(i)$ is

$$\overline{w}(i)(1+\beta_i)R(i)^{\beta_i} = \mu$$

And μ satisfies the feasibility constraint

$$1 = \sum_i [\overline{w}(i)(1+\beta_i)]^{-\frac{1}{\beta_i}} \mu^{-\frac{1}{\beta_i}}$$

A.8 Planner's problem for transition path

Theorem A.5. If the wage elasticity to local population is constant and constant across locations, then the decentralized allocation path is efficient after a permanent change in fundamentals $\overline{w}$.

Proof. Since the change in $\overline{w}$ is permanent, the planner's problem's solution is simply to direct flows towards locations with the highest marginal contribution to $\sum_i R(i)w(i)$. Directing flows to locations without the highest marginal contribution would be suboptimal because this will mean that $\sum_i R(i)w(i)$ will be lower for the entire remaineder of the transition, and thus the present discounted value of this path will be dominated by the path which always directs flows to the location(s) with the highest marginal value. Now note that the marginal value of additional mass at location j is

$$\frac{\partial}{\partial R(j)} \sum_i w(i)R(i) = \frac{\partial}{\partial R(j)} \sum_i \overline{w}(i)R(i)^{1+\beta}$$
$$= (1+\beta_j)\overline{w}(j)R(j)^{\beta_j}$$

The planner then directs flows to the location(s) with the highest marginal value, and $\beta < 0$ guarantees that the marginal value is falling as $R(j)$ rises. Additionally, the planner will direct flows such that all locations with the maximal marginal value continue to share that value, for otherwise some flows would for some instant be towards a location with less than the maximal marginal value.

To see that the planner solution is the same as the decentralized outcome when $\beta_i \equiv \beta$ is constant across locations, recognize that the planner uses marginal values $(1+\beta)\overline{w}(j)R(j)^{\beta}$ to direct flows, and the decentralized allocation directs flows in the exact same manner, but using marginal values $\overline{w}(j)R(j)^{\beta}$. So while the planner accounts for the externality, it has the same effect on marginal values across all locations, and does not affect the allocation, in comparison to the decentralized outcome.

□

B Computation

In this appendix I outline the numerical algorithms for solving the transition paths in both the finite θ and $\theta \to \infty$ cases. The finite θ algorithm applies to more general paths of fundamentals, but a separate algorithm is required for the limit case since choices exhibit "bang-bang" type behavior by flowing entirely into locations with the maximal value.

B.1 Finite θ Case

Recall that the system to be solved is

$$(\rho+\lambda)V(i,t) - \partial_t V(i,t) = \log \overline{w}(i,t) + \beta \log R(i,t) + \lambda[\gamma\theta + \log \sum_j \exp(\theta V(j,t))]$$

$$\dot{R}(j,t) = \lambda[-R(j,t) + \frac{\exp(\theta V(j,t)}{\sum_i \exp(\theta V(j,t)}]$$

where some initial distribution $\hat{R}$ is given. The problem may be discretized and recast as solving a system of nonlinear equations via the following steps.

(i) Pick a time horizon T which is sufficiently long so that the economy will have reached the stationary distribution by time T. Then discretize $[0, T]$ into $\{t_0, t_1, t_2, \ldots, t_N\}$ for some N number of time steps.

(ii) For $k \in \{0, \ldots, N-1\}$, the discretized HJB is

$$(\rho+\lambda)V(i,t_{k+1}) - \frac{V(i,t_{k+1}) - V(i,t_k)}{t_{k+1}-t_k} = \log \overline{w}(i,t_{k+1}) + \beta \log R(i,t_{k+1}) + \lambda[\gamma\theta + \log \sum_j \exp(\theta V(j,t_{k+1})]$$

(iii) For $k \in \{0, \ldots, N-1\}$, the discretized KFE is

$$\frac{R(j,t_{k+1}) - R(j,t_k)}{t_{k+1}-t_k} = \lambda[-R(j,t_k) + \frac{\exp(\theta V(j,t_k))}{\sum_i \exp(\theta V(j,t_k))}$$

(iv) The relevant boundary condition for the HJB is that the final value matches the stationary distribution (which has a closed form)

$$V(i,t_N) = V(i)$$

(v) The relevant boundary condition for the KFE is that the initial distribution matches the given initial distribution

$$R(i,0) = \hat{R}(i)$$

(vi) The above equations and boundary conditions comprise $n \times (N+1)$ equations in $n \times (N+1)$ unknowns, and can be solved via a nonlinear equation solver (e.g using a trust-region method, or even a (quasi-)Newton routine, since automatic differentiation is possible for this system)

B.2 $\theta \to \infty$ Case

In the limit case, solving the ODE system above will generally not work because of the "bang-bang" nature of the flows only going to the highest-valued location. Discretizing the time steps as above then will provide a poor approximation because between some time steps there will need to be a qualitative shift in flows, but the discretization will not allow this. Additionally, for the case of a permanent change in $\overline{w}$, as considered in the text, the analytical properties can be leveraged so that the length of each stage may be found explicitly. Then the discretization will exactly coincide with the timing of the stages during the transition. The algorithm for solving is then

(i) Initialize the initial distribution $\hat{R}$ and value function. Since $\overline{w}$ is time-invariant, the ordering of V is preserved by using $V(i,t) = \log \overline{w}(i) + \beta \log R(i,t)$ throughout.

(ii) Partition the set of locations into M and M^c according to whether their initial value is maximal or not.

(iii) While $M^c \neq \varnothing$ (this loop is for each stage)

  (I) Populations evolve as

$$R(i,t) = R(i,0)[1 + \frac{\sum_{j \in M^c} \beta R(j,0)}{\sum_{j \in M} \beta R(j,0)}(1 - \exp(-\lambda t))] \qquad (i \in M)$$

$$R(j,t) = R(j,0)\exp(-\lambda t) \qquad (j \in M^c)$$

  (II) Find the T such that the maximal value is equalized with the maximum value of all locations $j \in M^c$

$$\max_{i \in M^c}\{\log \overline{w}(i) + \beta \log R(i,T)\} = \log \overline{w}(j) + \beta \log R(j,T)$$

  (III) Update all $R(\cdot, 0)$ to match their levels at the end of the stage. Move all $j \in M^c$ that matched the maximal value at the end of the stage to M.

Note that in the inner loop a T solving the equation to end each stage always exists by the intermediate value theorem (using $\beta < 0$), and at at least one location moves from M^c to M every stage. Since there are finitely many locations, there are finitely many stages, and the transition path terminates in finite time at the new stationary distribution.