The Geography of Path Dependence*

Treb Allen
Dartmouth and NBER

Dave Donaldson
MIT and NBER

First Version: June 2017
This Version: July 2018

Preliminary and Incomplete

Abstract

How much of the spatial distribution of economic activity today is determined by history rather than by geographic fundamentals? And if history matters for the distribution, does it also matter for the total amount? This paper develops an empirically tractable theoretical framework that aims to provide answers to these questions. We derive parameter conditions, for arbitrary geographic scenarios, under which equilibrium transition paths are unique and yet steady states may nevertheless be non-unique — that is, where initial conditions ("history") may determine long-run steady-state outcomes ("path dependence"). We also derive analytical expressions, functions of the particular geography in question, that provide upper and lower bounds on the aggregate welfare level that can be sustained in any steady-state. We then estimate the model's parameters (which govern the strength of agglomeration externalities and trade and migration frictions), by focusing on moment conditions that are robust to potential equilibrium multiplicity, using spatial variation across US counties from 1800 to the present. Our simulations imply that the location of economic activity in the US today is highly sensitive to variations geographically local historical shocks, and the analytical bounds suggest the possibility of larger historical shocks mattering in the long-run.

*Allen: treb@dartmouth.edu. Donaldson: ddonald@mit.edu. We are grateful to our discussants, Jonathan Eaton and David Nagy, as well as to Rodrigo Adao, Arnaud Costinot, Jonathan Dingel, Gilles Duranton, Cecile Gaubert, Steve Redding, Stuart Rosenthal, Ivan Werning and many seminar participants, for comments that improved this paper.

1 Introduction

Economic activity in modern economies is staggeringly concentrated. For example, almost 20% of value added in the United States is currently produced in just three cities (MSAs) that take up a mere 1.5% of its land area. But perhaps even more remarkable are the historical accidents that purportedly determined the location of those same three cities—one was a Dutch fur trading post, one a *pueblo* designated by a Spanish governor for an original 22 adult and 22 children settlers, and one a river mouth known to Algonquin residents for its distinct *chicagoua* (a wild garlic).

There is no shortage of anecdotes about how the quirks of history have shaped the location of economic activity. But how widespread should we expect these and similar examples of path dependence—where initial conditions matter for long-run outcomes—to be in the real-world economies around us? If the potential for path dependence is widespread, how often did historical shocks actually matter? If initial conditions matter, do they merely reshuffle the current location of economic activity? Or does the long arm of history also impact the total amount of economic activity (and, hence, notions of aggregate welfare) by effectively concentrating modern agglomerations into fundamentally inferior locations?

In this paper we develop an empirical framework designed to shed light on these questions. We build on a rich vein of theoretical modeling (as developed in, for example, Fujita, Krugman, and Venables, 1999) that outlines stylized environments—models with two or three locations and very little heterogeneity, for example—in which strong agglomeration spillover effects can give rise to a potential multiplicity of equilibria. From there we set up a dynamic, overlapping generations model of economic geography with an arbitrary number of regions separated by arbitrary trading and migration frictions, as well as arbitrary time-varying locational fundamentals; these features allow us to map the model to empirical settings in which unobserved heterogeneity is typically substantial. To this basic setup we add agglomeration spillovers in production and consumption that, if they are sufficiently strong, can create the possibility for history to matter in determining modern outcomes.

Our main set of theoretical results aim to clarify when path dependence could potentially occur and quantify how geography constrains its impact. We first characterize a condition for dynamic equilibria—that is, the transition paths that would take this economy from any starting point to any presumptive steady-state—to be unique. This condition depends on two elasticities that promote dispersion (cross-locational elasticities of substitution in consumption and in migration decisions). It also hinges on the strength of two elasticities that govern contemporaneous agglomeration because the local attractiveness of a location at any point in time can potentially rise (due to local amenity and productivity spillovers), if these elasticities are strong, because of the presence of other workers in that location at that same point in time. As long as the agglomeration elasticities are not especially large, the dynamic paths of our economy will be unique for any path of geographic fundamentals.

However, even if these dynamic equilibria are unique, there may still exist multiple steady-state equilibria. Our second result clarifies that this can occur in a model such as ours only for sufficiently large values of a different source of agglomeration externalities—ones that work historically, whereby a location's productivity and amenity values might be functions of that location's *lagged* population level. These effects may capture both the accumulation of local knowledge (either productive or cultural), as well as the enduring payoffs from investments that previous generations may have made in a location's productivity (e.g. through improved roads, or the demarcation of property rights) or in a location's amenity appeal (e.g. through earmarking land for parks, or the discovery of an enjoyable climate).

When contemporaneous spillovers are relatively low, and yet historical spillovers are relatively high, dynamic paths will be unique but steady-states have the possibility to exhibit multiplicity. And this will be true for any arbitrary (time path of) geographic fundamentals. In this parameter range, we say that our economy could exhibit *path dependence,* because the economy's initial conditions—such as the distribution of economic activity in some early starting period, or long-irrelevant productivity shocks—may have an impact on the long-run steady-state in which the economy will end up. This parameter condition is also a necessary condition for steady-state multiplicity in a range of particular geographic scenarios, so it is the weakest geography-invariant condition for uniqueness that could be expected to hold.

In such a setting, steady-states are rankable in terms of aggregate welfare (so it is possible that unfortunate initial conditions could lead the economy to a particularly inferior steady-state). Our third theoretical result provides analytical upper and lower bounds, as a function of the underlying geography, on the aggregate welfare attainable across all possible steady states. These bounds are the product of four statistics, each of which captures one component of the underlying geography: the strength of the agglomeration forces, the cost of moving people, the cost of moving goods, the spatial variance of payoffs. Moreover, we show the ratio of the upper and lower bounds provides a limit on the extent to which history matters, which itself has a simple interpretation as the extent to which the underlying geography can sustain multiple equilibria.

In summary, this model exhibits the potential for rich and yet also well-behaved path-dependent dynamics. Whether such possibilities can obtain hinges on six elasticities: two dispersion parameters (the elasticities of trade and migration responses to payoffs), two contemporaneous spillover parameters (for production and amenities), and two historical spillover parameters (again, one for each of production and amenities).

We then set out in Section 3 to estimate these six parameters from a unique dataset on the long-run spatial history of the United States which allows us to trace local (county-level) incomes, populations, and migration flows back several centuries (to 1800) when the US Census began in full force. This estimation strategy uses trade and migration equations to infer, via market clearing conditions, the apparent attractiveness of each location-year as an origin and a destination for both trade and migration. Further, the logic of our contemporaneous and historical spillover effects suggests that these terms should be related to local contemporaneous and historical scale—and it is these expressions that provide simple 2SLS identifying moments for our estimation strategy.

Because of inevitable endogeneity in these equations we draw on an instrumental variables strategy that is based on the model's insight that other locations' geographic shifters of productivity (such as soil and elevation) or amenities (such as January temperature) should not affect any given location's own productivity or amenity values directly, after controlling for the location's own value of these shifters. We use this idea, together with the time variation driven by initial conditions and the spread of population predicted by the model, to estimate the six parameters referred to above. Importantly, this logic, and all of our moment conditions, are valid regardless of the potential for multiplicity (in dynamic paths, or steady-states) in our economy—so the usual concerns about models with non-uniqueness lacking invertibility of the mapping from data to model parameters do not arise.

Our elasticity estimates imply that the conditions for potential path dependence described above are indeed satisfied (though they are very close to the boundary identified in Propositions 1 and 2). The remaining simulation exercises in Section 4 then go beyond this qualitative result in order to assess the quantitative significance of potential path dependence. We do this by randomly reassigning the geographical incidence of various shocks to different locations—essentially, by swapping pairs of historical conditions from one location to another among pairs of locations in clusters based on similar (in a multivariate sense) geographic characteristics.

Our counterfactual simulations shock the transition path (in terms of productivity fundamentals) between 1850 and 2000 by reassigning shocks, within geographic clusters, in 1900 and 1950. Throughout, we hold initial conditions (population levels in 1800 and fundamentals in 1850) constant at their values seen in the actual data. Shocks in 2000 are held similarly constant. This implies that any differences seen in 2000 are only due to the persistent effects of long-redundant shocks, but those differences can be substantial. In particular, not only is the distribution of population in 2000 highly variable across our simulations and the spread of aggregate welfare levels in 2000 across our simulations is wide. While the simulations converge to the same steady state, they do so slowly, so that even in the year 3000 they continue to have modest impacts on aggregate welfare. And our analytical bounds suggest that no other steady state offers more than 50 percent higher welfare than that to which the economy is converging.

These results shed new light on both theoretical and empirical studies of economic geography that have aimed to speak to the phenomenon of path dependence. An important empirical literature has sought to estimate some of the ingredients of path dependence that our Propositions 1 and 2 identify. In particular, Dekle and Eaton (1999), Ellison, Glaeser, and Kerr (2010), Greenstone, Hornbeck, and Moretti (2010), Kline and Moretti (2014) and Ahlfeldt, Redding, Sturm, and Wolf (2015) all estimate the local agglomeration spillover effects of contemporaneous populations onto productivity and amenities. We are not aware, by contrast, of any papers that estimate both the historical and contemporaneous elasticities (effects which are highly correlated and hence difficult to disentangle) as we do here, and as our theoretical results highlight are independently important for the study of path dependence in these environments.

A separate empirical literature has focused on the search for direct evidence of path dependence itself. For example, Davis and Weinstein (2002) document the persistence of economic geography across locations in Japan over several millennia, including in response to the displacement and destruction of the second World War. However, evidence from destruction elsewhere suggests the empirical context may matter for whether or not path dependence can occur, as Bosker, Brakman, Garretsen, and Schramm (2007) find evidence of multiple spatial equilibria in Germany after World War II, and Michaels and Rauch (2018) in England after the fall of the Roman Empire, while persistence was confirmed in Vietnam by Miguel and Roland (2011).[1] More recently, Bleakley and Lin (2012) describe the propensity for US cities today to be located at portage sites, locations with temporarily high demand for labor (due to waterway transshipment and other services) in about 1800—and this seems to be strong evidence for path dependence in that context. Propositions 1 and 2 of our theory characterize the conditions on parameter values under which such divergent experiences with path dependence could both arise.

Just because path dependence may exist in a particular context does not necessarily imply that it is economically relevant at an aggregate level, since differing steady-states may be associated with similar or even identical welfare levels. To our knowledge, Proposition 3 offers the first analytical relationship between the welfare implications of path dependence and the underlying geography. We note that the tools used to establish this result may prove

[1] A separate literature has examined path dependence at the industry-level, with Davis and Weinstein (2008) documenting persistence at the city-industry level for manufacturing industries in Japan and Redding, Sturm, and Wolf (2011) uncovering evidence for multiple long-run steady-states in the case of airline hubs in the division and reunification of Germany.

helpful in other contexts in which multiplicity of equilibria are possible.

On the theory side, we draw on the insights of a theoretical literature that pioneered the understanding of the full dynamics of path-dependent geographic settings. Krugman (1991), Matsuyama (1991), and Rauch (1993), for example, developed models with two locations and infinitely-lived agents (with timing assumptions that meant that each agent made one locational decision at the start of her life, or at an an otherwise exogenously specified time). As fully elucidated in Ottaviano (1999), the dynamics of equilibrium paths (which includes those "sunspot"-like equilibria in which multiplicity derives purely from self-fulfilling expectations) in such settings are dauntingly complex, and would leave effectively no empirical predictability (or scope for parameter estimation) since at any point in time equilibria exhibit true multiplicity, with no mapping from parameters to data or vice versa. Counterfactual simulations in such models are similarly challenging due to indeterminacy. Herrendorf, Valentinyi, and Waldmann (2000) add agent-specific heterogeneity to such a model and reduce the range of parameter values under which extreme multiplicity does not arise, but the small number of regions and the symmetric conditions placed on those regions' fundamental conditions, in the cross-section and over time, make them unsuited to direct empirical analysis so we have endeavored to extract the core lessons of these setups and adapt them to our more empirical framework.

Finally, we build on recent work on quantitative economic geography models such as the static environments of Roback (1982), Glaeser (2008), Allen and Arkolakis (2014), Ahlfeldt, Redding, Sturm, and Wolf (2015)—summarized and synthesized in the Redding and Rossi-Hansberg (2017) review article—as well as the recent dynamic models of Desmet, Nagy, and Rossi-Hansberg (2018), Caliendo, Dvorkin, and Parro (2015), and Nagy (2017). Our advance is to extend these tools in order to facilitate the explicit study of geographic path dependence, to estimate, in the case of 200 years of US economic geography, the six key elasticities that our extended framework highlights as essential for such a theme, and then to apply the resulting estimates to counterfactual simulations about the consequentiality of path dependence for the location and aggregate efficiency of economic activity in the US today.

2 Theoretical framework

In this section we develop a dynamic economic geography model that is amenable to the empirical study of geographic path dependence throughout US history. A large set of regions possess arbitrary, time-varying fundamentals in terms of productivity and amenities. They interact in product markets that interact with one another via (costly) trade in goods, and in labor markets that interact with one another via (costly) migration. Crucially, production and consumption both potentially involve contemporary and historical non-pecuniary spillovers—the force for potential local agglomeration externalities, and hence path dependence. We now describe each of these ingredients in turn.

2.1 Setup

There are $i \in \{1, ..., N\}$ locations and time is discrete and indexed by $t \in \{0, 1, ...\}$. Each individual lives for two periods. In the first period ("childhood"), an individual is born where her parent lives and chooses where to live as an adult. In the second period ("adulthood"), an individual supplies a unit of labor inelastically to produce in the location in which she lives, consumes, and then gives birth to a child. Let L_{it} denote the number of workers (adults) residing in location i at time t, where the total number of workers $\sum_{i=1}^{N} L_{it} = \bar{L}$, is normalized to a constant in each period t.[2] The population in time $t = 0$, $\{L_{i0}\}$, is given exogenously.

2.1.1 Production

Each location i is capable of producing a unique good—the Armington (1969) assumption. A continuum of firms (indexed by ω) in location i produce this homogeneous good under perfectly competitive conditions with the following constant returns-to-scale production function

$$q_{it}(\omega) = A_{it} l_{it}(\omega)$$

where labor $l_{it}(\omega)$ is the only production input, and hence $\int l_{it}(\omega) d\omega = L_{it}$. The productivity level for the location, A_{it}, is given by

$$A_{it} = \bar{A}_{it} L_{it}^{\alpha_1} L_{it-1}^{\alpha_2} \tag{1}$$

where $\bar{A}_{it}$ is the exogenous (but unrestricted) component of this location's productivity in year t. Importantly, the two additional components of a location's productivity depend on the number of workers in that location in the current period, L_{it}, and in the previous period, L_{it-1}. We assume that firms take these aggregate labor quantities as given. Hence the parameter α_1 governs the strength of any potential (positive or negative) contemporaneous agglomeration externalities working through the size of local production. This is a simple

[2] Our model economy exhibits a form of scale-invariance that means that, for the purposes of our analysis here, the total number of workers in any time period is irrelevant.

way of capturing Marshallian externalities, external economies of scale, knowledge transfers, thick market effects in output or input markets, and the like. The presence of the term $L_{it}^{\alpha_1}$ is standard in many approaches to modeling spatial economies, albeit typically in static models that would combine effects of L_{it} and L_{it-1}.

The parameter α_2, on the other hand, governs the strength of potential *historical* agglomeration externalities. This allows for the possibility that two cities with equal fundamentals $\bar{A}_{it}$ and sizes L_{it} today might feature different productivity levels A_{it} today because they had differing sizes L_{it-1} in the past. There are many potential reasons that one might expect $\alpha_2 > 0$, and we describe two such sets of microfoundations briefly here (with complete derivations in Appendix (B.1)).

Consider first the potential persistence of local knowledge. In particular, we present a model based on Deneckere and Judd (1992), where firms can incur a fixed cost to develop a new variety, for which they earn monopolistic profits for a single period. In the subsequent period, the blueprint for the product becomes common knowledge so that the variety is produced under perfect competition, and we assume the product fully depreciates two periods after its creation. As in Krugman (1980), the equilibrium number of new varieties will be proportional to the contemporaneous local population. Given consumers' love of variety, new varieties act isomorphically to an increase in the productivity of the single Armington product, resulting in the precise form of equation (1) with $\alpha_1 \equiv \frac{\chi}{\rho-1}$ and $\alpha_2 \equiv \frac{1-\chi}{\rho-1}$, where χ is the expenditure share on all new varieties and ρ is the elasticity of substitution across individual varieties.

Second, consider the potential for durable investments in local productivity. In particular, we present a model based on Desmet and Rossi-Hansberg (2014), where firms hire workers to both produce and to innovate, where innovation increases each firm's own productivity contemporaneously and increases all firm's productivity in the subsequent period. If firms earn zero profits in equilibrium due to competitive bidding over a fixed factor (e.g. land), then as in Desmet and Rossi-Hansberg (2014), the dynamic problem of the firm simplifies to a sequence of static profit-maximizing problems. With Cobb-Douglas production functions, equilibrium productivity can be written as equation (1) with $\alpha_1 \equiv \frac{\gamma_1}{\xi} - (1-\mu)$, and $\alpha_2 \equiv \delta\frac{\gamma_1}{\xi}$, where γ_1 governs the decreasing return of innovation to productivity, ξ governs the decreasing returns of labor in innovation, δ is the depreciation of investment, and μ is the share of labor in the production function.

Of course, there may be other microfoundations that generate the productivity spillovers assumed in equation (1). In what follows, we characterize the properties of the model and estimate the strength of the spillovers without taking a stand on any particular microfoundation.

2.1.2 Consumption

Adults are the only consumers, and we assume that they care only about their own consumption.[3] They have constant elasticity of substitution (CES) preferences, with elasticity σ, across the differentiated goods that each location can produce. Letting w_{it} denote the equilibrium nominal wage, and letting P_{it} be the price index (solved for below), the deterministic component of welfare—that is, welfare up to an idiosyncratic shock that we introduce below—of any adult residing in location i at time t is given by

$$W_{it} \equiv u_{it} \cdot \left(\frac{w_{it}}{P_{it}} \right),$$

where the term u_{it} refers to a location-specific amenity shifter that is given by

$$u_{it} = \bar{u}_{it} L_{it}^{\beta_1} L_{it-1}^{\beta_2}. \tag{2}$$

The term $\bar{u}_{it}$ allows for flexible exogenous amenity offerings in any location and time period. Endogenous amenities work analogously to the production externality terms (governed by the elasticities α_1 and α_2) introduced above, with the parameters β_1 and β_2 here capturing the potential for the presence of other adults in a location to directly affect (either positively or negatively, depending on the sign of β_1 and β_2) the utility of any given resident. We assume that consumers take these terms as given, just as they take factor and goods prices as given, when making decisions.

As is well understood, a natural source of a negative value for β_1 in a model such as this one is the possibility of local congestion forces that are not directly modeled here; for example, if non-tradable goods (such as housing and land) are in fixed supply locally and are demanded in fixed (that is, Cobb-Douglas) proportions then $-\beta_1$ would equal the share of expenditure spent on such goods. These effects would work contemporaneously, so they would govern β_1.

As with α_2, the parameter β_2 captures forces by which the historical population L_{it-1} affects the utility of residents in year t directly (that is, other than through productivity, wages, prices, or current population levels). Again it seems potentially important to allow for such effects given the likelihood that previous generations of residents may leave a durable impact, positive or negative, on their former locations of residence. Positive impacts could include the construction of infrastructure (e.g. parks, sewers, or housing), and negative impacts could include environmental damage or resource depletion.

[3] If children consumed a fixed fraction of their parents' consumption amounts then allowance for consumption in childhood would simply scale up all consumption amounts in our analysis proportionally.

It is straightforward to construct a model that generates exactly the specification of equation (2) for amenities. We sketch such a microfoundation here, and present the complete set of derivations in Appendix B.2. We consider a model where agents consume both a tradable good and local housing, and each unit of land is owned by a real estate developer who bids for the rights to develop the land and then chooses the amount of housing to construct. To build housing, the developer combines local labor and the (depreciated) housing stock from the previous period. We assume the bidding process ensures developers earn zero profits, so as in Desmet and Rossi-Hansberg (2014) the dynamic problem of how much housing to construct simplifies into a series of static profit maximizing decisions. In equilibrium, the higher the contemporaneous population, the lower the utility of local residents (as the residents consume less housing), whereas the higher the population in the previous period, the higher the utility of local residents (as more workers in the previous period results in a greater housing stock today). In particular, with Cobb-Douglas production functions (with the share of old housing given by μ), and preferences (with a share $1-\lambda$ spent on housing) $\beta_1 = -\mu\frac{1-\lambda}{\lambda} < 0$ and $\beta_2 = \rho\mu\frac{1-\lambda}{\lambda} > 0$, where μ is the fraction of expenditure on the existing housing stock in the housing production function, λ is the fraction of expenditure spent on tradable goods, and ρ is the depreciation rate of the housing stock.

As with the productivity spillovers, we emphasize that there may be other microfoundations that also generate the amenity spillovers assumed in equation (2). In in what follows, there is no need to pursue any one particular microfoundation.

2.1.3 Trade

Bilateral trade from location i to location j incurs an exogenous iceberg trade cost, $\tau_{ijt} \geq 1$ (where $\tau_{ijt} = 1$ corresponds to frictionless trade). Given this, bilateral trade flows take on the well-known gravity form given by

$$X_{ijt} = \tau_{ijt}^{1-\sigma} \left(\frac{w_{it}}{A_{it}}\right)^{1-\sigma} P_{jt}^{\sigma-1} w_{jt} L_{jt}, \tag{3}$$

where $P_{it} \equiv \left(\sum_{k=1}^{N} \left(\tau_{ki}\frac{w_{kt}}{A_{kt}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ is the CES price index referred to above.

2.1.4 Migration

Recall from the discussion of timing above that L_{jt-1} adults reside in location j at time $t-1$, and they have one child each. Those children choose at the beginning of period t—as they pass into adulthood—where they want to live as adults in order to maximize their welfare.

As described above, adults who reside in a location j will enjoy a deterministic component of utility given by W_{jt} in equilibrium. Similarly to costs of trading, we allow for finite bilateral impediments to migration $\mu_{ijt} \geq 1$ (with frictionless migration denoted by $\mu_{ijt} = 1$), which act like utility-shifters conditional on migrating from i to j. This means that the deterministic utility for a migrant who moves from location i to location j is $\frac{W_{jt}}{\mu_{ijt}}$. However, we also allow for idiosyncratic unobserved heterogeneity in how each child will value living in each location j in adulthood. Letting the vector of such idiosyncratic taste differences be denoted by $\vec{\varepsilon}$, the actual welfare of a child who receives the draw $\vec{\varepsilon}$ while living in location i in time $t-1$ who chooses to move to location j as an adult is:

$$W_{ijt}(\vec{\varepsilon}) \equiv \frac{W_{jt}}{\mu_{ijt}} \varepsilon_j, \tag{4}$$

so the particular shock for location j, denoted by ε_j, simply scales up or down the deterministic component of utility, $\frac{W_{jt}}{\mu_{ijt}}$. Hence, a child chooses:

$$\max_j W_{ijt}(\vec{\varepsilon}) = \max_j \frac{W_{jt}}{\mu_{ijt}} \varepsilon_j$$

We further assume that $\vec{\varepsilon}$ is drawn independently from an extreme-value (Frechet) distribution with shape parameter θ (and a set of location parameters that we normalize to one without loss of generality). The number of children in location i in time $t-1$ who choose to move to location j in time t, L_{ijt}, is then given by:

$$L_{ijt} = \frac{(W_{jt}/\mu_{ijt})^{\theta}}{\sum_{k=1}^{N} (W_{kt}/\mu_{ikt})^{\theta}} L_{it-1}. \tag{5}$$

For future reference, we note that the expected utility of a child location in location i in time $t-1$ prior to realizing their idiosyncratic shocks $\vec{\varepsilon}$, which we denote by Π_{it}, is:

$$\Pi_{it} \equiv \mathbb{E}\left[\max_j W_{ijt}(\vec{\varepsilon})\right] = \left(\sum_{k=1}^{N} (W_{kt}/\mu_{ikt})^{\theta}\right)^{\frac{1}{\theta}}. \tag{6}$$

So, summarizing, we can write bilateral migration flows in the gravity equation form as

$$L_{ijt} = \mu_{ijt}^{-\theta} \Pi_{it}^{-\theta} L_{it-1} W_{jt}^{\theta}, \tag{7}$$

where we expect higher migration into destination locations j with high destination welfare W_{jt}, out of origin locations i that either have a lot of residents L_{it-1} or poor expected utility at birth Π_{it} or both, and among pairs for which bilateral migration costs μ_{ijt} are low.

2.2 Dynamic Equilibrium

An equilibrium in this dynamic economy is a sequence of values of prices and allocations such that goods and factor markets clear in all periods. More formally, for any initial population vector $\{L_{i0}\}$ and geography vector $\left\{\bar{A}_{it}, \bar{u}_{it}, \tau_{ijt}, \mu_{ijt}\right\}$, an equilibrium is a vector of endogenous variables $\{L_{it}, w_{it}, W_{it}, \Pi_{it}\}$ such that, for all locations i and time periods t, we have:

1. *Total sales are equal to payments to labor:* That is, a location's income is equal to the value of all locations' purchases from it, or $w_{it}L_{it} = \sum_j X_{ijt}$. Using equation (3) this can be written as

$$w_{it}^{\sigma} L_{it}^{1-\alpha(\sigma-1)} = \sum_j K_{ijt} L_{jt}^{\beta(\sigma-1)} W_{jt}^{1-\sigma} w_{jt}^{\sigma} L_{jt}, \tag{8}$$

with $K_{ijt} \equiv \left(\frac{\tau_{ij}}{\bar{A}_{it} L_{it-1}^{\alpha_2} \bar{u}_{jt} L_{jt-1}^{\beta_2}}\right)^{1-\sigma}$ defined as a collection of terms that are either exogenous, or predetermined from the perspective of period t.

2. *Trade is balanced:* That is, a location's income is fully spent on goods from all locations, or $w_{it}L_{it} = \sum_j X_{jit}$. Using equation (3) this can be written as

$$w_{it}^{1-\sigma} L_{it}^{\beta_1(1-\sigma)} W_{it}^{\sigma-1} = \sum_j K_{jit} L_{jt}^{\alpha_1(\sigma-1)} w_{jt}^{1-\sigma}. \tag{9}$$

3. *The total population is equal to the population arriving in a location:* That is, $L_{it} = \sum_j L_{jit}$. From equation (7) this implies

$$L_{it} W_{it}^{-\theta} = \sum_j \mu_{jit}^{-\theta} \Pi_{jt}^{-\theta} L_{jt-1}. \tag{10}$$

4. *The total population in the previous period is equal to the number of people exiting a location:* That is, $L_{it-1} = \sum_j L_{ijt}$. From equation (7) this can be written as

$$L_{it-1} = \sum_j \mu_{ijt}^{-\theta} \Pi_{it}^{-\theta} L_{it-1} W_{jt}^{\theta},$$

which can then be written more compactly as

$$\Pi_{it}^{\theta} \equiv \sum_j \mu_{ijt}^{-\theta} W_{jt}^{\theta}. \tag{11}$$

Summarizing, the dynamic equilibrium can be represented as the system of $4 \times N \times T$ equations (in equations 8-11) in $4 \times N \times T$ unknowns, $\{L_{it}, w_{it}, W_{it}, \Pi_{it}\}$.

This system of equations (8)-(11) comprises a high-dimensional nonlinear dynamic system whose analysis can prove challenging. But this task is facilitated by the fact that the system is a collection of additive power equations, where each of the endogenous variables $\{L_{it}, w_{it}, W_{it}, \Pi_{it}\}$ appears, on either the left-hand or right-hand side, to a particular fixed power, with weights in the system given by an exogenous "kernel" term that comprises variables that are either exogenous or pre-determined from the perspective of period t (K_{ijt} in equations 8 and 9, and $\mu_{ijt}^{-\theta}$ in equations 10 and 11). This means that the solution of each cross-sectional system for t, given values of K_{ijt} and hence solutions from the previous period $t-1$, can be solved using the methods in Allen, Arkolakis, and Li (2015). In this manner, a dynamic path can be characterized by understanding a sequence of linked dynamic problems.

Towards this goal, we define the matrix:

$$\mathbf{A}\left(\alpha_1, \beta_1\right) \equiv \begin{pmatrix} \left|\frac{\theta(1+\alpha_1\sigma+\beta_1(\sigma-1))-(\sigma-1)}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| & \left|\frac{(\sigma-1)(\alpha_1+1)}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| \\ \left|\frac{\theta/\tilde{\sigma}}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| & \left|\frac{\theta(1-(\sigma-1)\alpha_1-\beta_1\sigma)}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| \end{pmatrix}, \tag{12}$$

where $\tilde{\sigma} \equiv \frac{\sigma-1}{2\sigma-1}$. Given this definition, the following result characterizes a sufficient condition for existence and uniqueness for environments with symmetric trade costs (and unrestricted migration costs) and arbitrary positive geographic fundamentals.

Proposition 1. For any initial population $\{L_{i0}\}$ and geography $\left\{\bar{A}_{it} > 0, \bar{u}_{it} > 0, \tau_{ijt} = \tau_{jit}, \mu_{ijt} > 0\right\}$, there exists an equilibrium. The equilibrium is unique if $\rho\left(\mathbf{A}\left(\alpha_1, \beta_1\right)\right) \leq 1$, where $\rho(\cdot)$ denotes the spectral radius operator.

Proof. See Section A.1. □

We note that this sufficient condition for uniqueness will be satisfied whenever α_1 and β_1 are sufficiently small. Figure 1 illustrates this condition for two particular values of σ and θ, values at which the sufficient condition of $\rho\left(\mathbf{A}\left(\alpha_1, \beta_1\right)\right) \leq 1$ is well approximated by the simple relation of $\alpha_1 + \beta_1 \leq 0$. Finally, we note that this result concerning uniqueness of the dynamic equilibrium does not depend on the values of α_2 and β_2, since the current generation takes L_{it-1} as given.

2.3 Steady-State

Our discussion of path dependence rests on the consideration of the various potential steady-states of this model economy. Intuitively, if local agglomeration economies are strong enough then there could be multiple allocations at which the economy would be in steady-state—agents who happen to come to reside in a location could find it optimal, on average, to stay there thanks to the reinforcing logic of local positive spillovers.

To evaluate this possibility we consider a version of the above economy but for which the potentially time-varying fundamentals $\{\bar{A}_{it}\}$ and $\{\bar{u}_{it}\}$ and trade $\{\tau_{ijt}\}$ and migration $\{\mu_{ijt}\}$ costs are held constant over time at the values $\{\bar{A}_i, \bar{u}_i, \tau_{ij}, \mu_{ij}\}$. The steady-states of our economy will therefore be a set of time-invariant endogenous variables that we denote by $\{L_i, w_i, W_i, \Pi_i\}$.[4] The following result, analogous to Proposition 1, provides a sufficient condition for existence and uniqueness of the steady-state of this economy (for arbitrary geographies with symmetric trade and migration costs). It also shows how this is a "maximal domain" sufficient condition—the weakest one could impose that would be true for any geographic fundamentals.

Proposition 2. For any time-invariant geography $\{\bar{A}_i > 0, \bar{u}_i > 0, \tau_{ij} = \tau_{ji}, \mu_{ij} = \mu_{ji}\}$, there exists a steady-state equilibrium and that equilibrium is unique if $\rho\left(\mathbf{A}\left(\alpha_1+\alpha_2, \beta_1+\beta_2\right)\right) \leq 1$. Moreover, if $\rho\left(\mathbf{A}\left(\alpha_1+\alpha_2, \beta_1+\beta_2\right)\right) > 1$, there exist many geographies for which there are multiple steady states.

Proof. See Section A.2. □

The condition for uniqueness of the steady-state in Proposition 2 is similar to that for uniqueness of transition paths in Proposition 1. The only difference is that the latter condition depends on the size of contemporaneous spillovers α_1 and β_1, whereas the latter condition depends on the size of total (that is, contemporaneous plus historical) spillovers $\alpha_1 + \alpha_2$ and $\beta_1 + \beta_2$. The last part of Proposition 2 demonstrates that the sufficient condition for uniqueness is indeed necessary for certain geographies, i.e. it is the weakest geography-independent sufficient condition for uniqueness one can provide.

Combining Propositions 1 and 2, we see that what matters for the potential multiplicity of steady-states (and hence the potential for initial conditions or temporary shocks to affect the steady-state that obtains – or equivalently, for path dependence to occur) is the values of α_2 and β_2. If these historical spillover parameters are large then it is likely for path dependence to occur. Further, if the values of α_1 and β_1 are low then dynamic equilibrium paths will be unique. In this range of parameters (that is, with relatively small α_1 and β_1 and yet relatively large α_2 and β_2) path dependence will both exist and be straightforward to study, since the complications of genuine multiplicity for estimation, computation, and interpretation of counterfactuals do not arise. We think of this as well-behaved path dependence.

[4]Note that while population levels at each location, L_i, will be constant in steady-state, and hence net migration flows are zero, gross migration flows will still be positive in steady-state equilibrium due to the churn induced by idiosyncratic locational preferences in equation 4.

Before discussing path dependence further, it is useful to point out two features that will obtain in any steady-state equilibrium. First, as one might suspect, a notion of welfare is equalized across locations in steady-state—otherwise, surely gross migration flows, induced by any spatial welfare arbitrage opportunities, would not be zero. The notion of welfare that is constant across locations is $\Omega \equiv \mathbb{E}\left[\max_j \left(W_j \Pi_j \varepsilon_j\right)\right]$, the expectation (across the random draws of idiosyncratic preferences in equation 4) of the maximum welfare that an agent can achieve when she has no particular attachment to any location. Recall that the welfare an agent can expect to achieve, conditional on living in location i, is $\mathbb{E}\left[\max_j \frac{W_j}{\mu_{ij}} \varepsilon_j\right]$. The essence of the steady-state version of this is to replace the term $1/\mu_{ij}$ (the relevant penalty that an agent living in i must pay to get to j and enjoy $W_j \varepsilon_j$ there) with the term Π_j (which is the appropriate weighted average of costs of getting from any location to j). A simple calculation shows that

$$\Omega = W_i \Pi_i L_i^{-\frac{1}{\theta}} \ \forall i \in \{1, ..., N\},$$

which includes the term $L_i^{-\frac{1}{\theta}}$ to account for the fact that, in equilibrium, a heavily populated location must have an average number of residents there who had relatively unfavorable idiosyncratic draws, so their average welfare is lower than otherwise.

A second feature of the steady-state is useful for fixing intuition. Algebraic manipulations of equations of the steady-state versions of equations (8)-(11) imply that the equilibrium steady state distribution of population can be written as:

$$\gamma \ln L_i = C + (1 - \tilde{\sigma}) \ln \bar{u}_i + \tilde{\sigma} \ln \bar{A}_i + (1 - \tilde{\sigma}) \ln \Pi_i - \ln P_i, \tag{13}$$

where $\gamma \equiv \frac{1}{\theta}(1 - \tilde{\sigma}) - \frac{\tilde{\sigma}}{\sigma - 1} - (\tilde{\sigma} + 1)(\beta_1 + \beta_2) + \tilde{\sigma}(\alpha_1 + \alpha_2)$. This implies that a greater density of residents can be found, in any steady-state equilibrium, in locations with high productivity $\bar{A}_i$, high amenities $\bar{u}_i$, high access to migration destinations Π_i, and high access to imported goods (low P_i), and the elasticities of these characteristics are governed by the strength of the key spillover elasticities ($\alpha_1, \alpha_2, \beta_1$ and β_2) via the combined parameter γ. Of course, while the first two of these determinants of population density, $\bar{A}_i$ and $\bar{u}_i$, are exogenous in our model, the latter two determinants, Π_i and P_i, are endogenous and involve endogenous features of all other locations. It is the self-reinforcing potential of those cross-location interactions that leads to the possibility of multiple steady-states and hence multiple vectors of $\{L_i\}$ that would satisfy the system in equation (13).

Finally, we provide a relationship between the geography of the world and steady state welfare Ω when multiple steady states (i.e. path dependence) are possible:

Proposition 3. *Consider any time-invariant geography* $\left\{\bar{A}_i > 0, \bar{u}_i > 0, \tau_{ij} = \tau_{ji}, \mu_{ij} = \mu_{ji}\right\}$ *and suppose that* $\rho \equiv \alpha_1 + \alpha_2 + \beta_1 + \beta_2 \geq 0$ *so that multiple steady states may exist. Then* the equilibrium welfare values Ω across all steady states are bounded by:

$$\underline{\Omega} \leq \Omega \leq \bar{\Omega},$$

where the upper bound is given by:

$$\bar{\Omega} \equiv c_1 \bar{\lambda}_M^{\frac{1}{\theta}} \bar{\lambda}_T^{\frac{1}{\sigma-1}} \bar{L}^{\left(\rho - \frac{1}{\theta}\right)}$$

and the lower bound is given by:

$$\underline{\Omega} \equiv c_2 \underline{\lambda}_M^{\frac{1}{\theta}} \underline{\lambda}_T^{\frac{1}{\sigma-1}} \left(\frac{\bar{L}}{N}\right)^{\rho},$$

$\bar{\lambda}_M$ and $\underline{\lambda}_M$ are the maximal and minimal eigenvalues (by moduli) of the migration matrix $\mathbf{M} \equiv \left[\mu_{ij}^{-\theta}\right]$, $\bar{\lambda}_T$ and $\underline{\lambda}_T$ are the maximal and minimal eigenvalues (by moduli) of the trade matrix $\mathbf{T} \equiv \left[\tau_{ij}^{1-\sigma} \bar{A}_i^{(\sigma-1)\tilde{\sigma}} \bar{A}_j^{\tilde{\sigma}\sigma} \bar{u}_i^{\tilde{\sigma}} \bar{u}_j^{(\sigma-1)\tilde{\sigma}}\right]$, and c_1 and c_2 are constants (defined in Section A.3) that capture the variation in location-specific welfare W_i across locations and are equal to one if $W_i = W$ for all $i \in S$. A sufficient condition for this equalization of welfare is that $\sum_l \frac{\mu_{il}^{-\theta}}{\mu_{jl}^{-\theta}}$ is constant for all $i, j \in S$).

Proof. See Section A.3. □

As the example below illustrates, when the presence of agglomeration forces results in multiple steady states, different initial conditions may lead to different steady states with different associated levels of welfare. As a result, each geography may be associated with multiple levels of potential steady-state aggregate welfare; formally, the function mapping geography to aggregate welfare Ω is multivalued. Proposition 3 deals with this by providing both lower and upper bounds to all levels of steady state welfare possible for a given geography.

The bounds offered by Proposition 3 provide an intuitive explanation for how geography can matter for welfare. The upper bound is simply the product of four terms: the largest eigenvalue (i.e. the spectral radius) of the migration matrix (scaled appropriately by the migration elasticity), the largest eigenvalue of the trade matrix (scaled appropriately by the trade elasticity), the total labor endowment (scaled by the strength of the net agglomerative forces ρ), and a term capturing the variation across locations in W_i. The spectral radius of the migration matrix will be greater the lower the migration costs, while the spectral radius of the trade matrix will be greater the lower the trade costs, the higher the fundamental amenities, or the higher the fundamental productivities. Finally, the presence of net agglomeration forces ρ means that the upper bound is increasing with the aggregate labor endowment. The lower bound includes very similar terms to the upper bound, except that the eigenvalues are now the smallest (in absolute value) of the trade and migration matrix rather than the largest, and the effect of the net agglomerative forces ρ depends on the aggregate labor endowment per location (rather than the total labor endowment).[5]

By providing bounds on the welfare of all possible steady states as a function of the underlying geography, Proposition 3 offers limits on the extent to which path dependence could matter for welfare in the long run. This is helpful when the curse of dimensionality makes it too difficult to calculate all possible steady states that could arise over the N-dimensional space of initial populations $\{L_{i0}\}$, as seems likely in many applications including ours. In particular, if we define $\hat{\Omega}^{PD}$ to be the upper bound of the welfare cost of path dependence – i.e. the best possible steady state welfare for a given geography divided by the worst possible steady state welfare for that geography – then the welfare cost of path dependence is bounded above by the ratio of the upper bound to the lower bound. Proposition 3 immediately implies:

$$\hat{\Omega}^{PD} \leq \frac{c_1}{c_2} \kappa \left(\mathbf{M}\right)^{\frac{1}{\theta}} \kappa \left(\mathbf{T}\right)^{\frac{1}{\sigma-1}} N^{\rho} \bar{L}^{-\frac{1}{\theta}}, \tag{14}$$

where $\kappa\left(\mathbf{A}\right)$ is the condition number of matrix $\mathbf{A}$. Recall that the condition number of a matrix measures how close that matrix is to being singular (where $\kappa\left(\mathbf{A}\right) = 1$ only if $\mathbf{A}$ is a scalar multiple of a linear isometry and $\kappa\left(\mathbf{A}\right) = \infty$ only if $\mathbf{A}$ is singular). Loosely speaking, equation (14) says that the welfare cost of path dependence is bounded above by the extent to which the underlying geography can support multiplicity.

We note that the upper and lower bounds provided here may not necessarily be tight for a given geography. This is clear from the nature of equation (14), which serves to decompose the sources of welfare variation across multiple steady-states into four different terms—that due to spatial dispersion in locational welfare W_i that drives $\frac{c_1}{c_2}$; that due to the geography of migration costs in $\mathbf{M}$; that due to the geography of productivity, amenities and trade costs in $\mathbf{T}$; and that due to scale economies in N^{ρ}. Steady-state welfare levels are driven by the combination of each of these four forces, and so no attempt to divide them up into separate contributions, as in equation (14), could ever provide tight bounds in general.

An illustrative special case arises if we have a set of locations with symmetric fundamen-

[5] Note too that the upper bound subtracts from the the net agglomerative force $\frac{1}{\theta}$ whereas the lower bound does not. Loosely speaking, the strongest agglomerative force occurs when all the labor endowment is in a single location, which comes at the cost of placing individuals in that location that have idiosyncratic preferences for elsewhere (hence the subtraction of $\frac{1}{\theta}$); conversely, the weakest agglomerative force occurs when the labor is equally spread across all locations, in which case individuals can sort according to their idiosyncratic preferences which is why the lower bound depends on $\frac{\bar{L}}{N}$ but the agglomerative force is not reduced by $\frac{1}{\theta}$.

tals ($\overline{A_i}\overline{u_i}$ constant across i) and both goods and labor market autarky.[6] In this geography, $\kappa(\mathbf{M}) = \kappa(\mathbf{T}) = 1$, and any allocation of population to locations would be a steady-state. With scale economies ($\rho > 0$), aggregate welfare Ω would be maximized if the population allocation happened to concentrate people in one location, and would be minimized if people were spread out evenly; and in this case the range of such possibilities is captured exactly by $\hat{\Omega}^{PD}$ in equation (14), so the bounds are tight. Equally, without scale economies ($\rho = 0$) aggregate welfare is equalized across any steady-state, so $\hat{\Omega}^{PD} = 1$ naturally.

2.4 A path dependence example

To see the logic of path dependence in this model more concretely, consider a simple example of three locations. Suppose, to begin, that these locations have identical and time-invariant geographies $\{\bar{A}_{it}, \bar{u}_{it}, \tau_{ijt}, \mu_{ijt}\}$, and trade and migration costs are symmetric across locations. Figure 2 shows the phase diagram on the two-dimensional phase space of L_{it} shares in this economy. To interpret these figures, note that each red dot has associated with it a blue ray; the direction of the ray illustrates the direction to which the system dynamics move towards the red dot, and the length of the ray conveys the speed with which those dynamics take place.

We begin in panel (a) with a setting in which the spillover parameters (α_1, β_1, α_2 and β_2) are all zero. Naturally, this symmetric economy with no spillovers has a unique steady-state, and this steady-state is located at the center of the simplex because of symmetry. Panels (b) through (f) then increase α_2 but keep all other parameters in the economy constant. At $\alpha_2 = 0.1$ this increase in α_2 has no apparent qualitative impact on the dynamics of the economy. But at $\alpha_2 = 0.2$ we see a dramatic change. The central location, a unique and stable steady-state when $\alpha_2 = 0.1$, is still a steady-state but it is no longer stable (all dynamic rays near that central point lead away from it). And, further, this steady-state is no longer unique—six additional steady-states have emerged, three stable steady states with relatively concentrated population shares (the corners of the simplex) in a single location, and three unstable steady states with equal concentration in two of the three locations (and almost no population in the third). As we increase α_2 even further this basic picture doesn't change, though speeds of convergence to steady-state do increase. One final thing to note in this example is that each steady-state will be surrounded by points that will map dynamically to it. The locus of such points around any steady-state comprise its *basin of attraction.* In this symmetric case, the three symmetric, stable steady-states have symmetric basins of attraction that partition the space of all possible starting points in the simplex. (The

[6]Note that this special case, and only this special case, violates our earlier requirement that migration costs be finite.

unstable steady-states, of course, have no basin of attraction.)

Now consider the same example but with asymmetric fundamentals. Suppose that location 2 has worse fundamental amenity value $\bar{u}_i$ than do the other two regions. When $\alpha_2 = 0$, as shown in Figure 3, the steady-state is again unique and relatively central, just as with the previous symmetric example. But the difference now is that the asymmetric fundamentals (the relative unattractiveness of location 2 in terms of fundamentals) have shifted the location of that unique steady-state—intuitively, it is shifted in the direction of the location 2 axis of the simplex, implying less population at location 2 in steady-state. As we increase α_2 from 0 to 0.5 in panels (a) to (f) we see behavior that is similar to that of the symmetric case in Figure 2. The unique steady-state under $\alpha_2 \leq 0.1$ shifts to a multiplicity of steady-states, each displaying relative concentration in the corners of the simplex, for $\alpha_2 \geq 0.2$. In this case, however, the three steady-states have different levels of aggregate welfare (in the sense of Ω described above), this example allows for an economy that might, due to a bad set of initial conditions, end up in a dominated steady-state. Reassuringly, however, the basin of attraction of a relatively good steady-state is larger than that of a dominated one. So, in the space of all possible initial conditions, good steady-states may be more likely to arise.

3 Identification and Estimation

We now describe a procedure for mapping the above model into observable features of the US economy throughout the past two centuries. The goal is to estimate the elasticity parameters (α_1, β_1, α_2, β_2, σ and θ) that are critical for assessing the likelihood and strength of path dependence, as well as the geographic fundamentals $\left\{\bar{A}_{it}, \bar{u}_{it}, \tau_{ijt}, \mu_{ijt}\right\}$ that shift the consequences of path dependence.

3.1 Data

We aim to track subnational regions throughout the period from 1800-2000. In each decennial census, information is available at the county level, but these county border definitions change over the years, so we track 25km by 25km units ("cells"), allocating each cell to the the appropriate county in each year. We then apportion uniformly the county-level information for county c in any year t into each of cell that maps to that county c in year t. In the end, our sample consists of 12,457 of such cells.

Data limitations mean that obtaining consistent time series on the long sweep of American economic (county-level) history can be challenging. Thankfully, one variable that is available throughout is a proxy for internal migration, which then corresponds to L_{ijt} in the model above. The decennial US Census tracks, from 1790-present, data on the population by county of current residence and state of birth (and age). These can be extracted from publicly available 5% samples for each year. Given that "adults" in the model are the generation that produces and consumes, we take the number of people aged 20-69 in this dataset, in each county j and year t, and use these adults' location of birth as our proxy for the origin of their adult migration journey, location i. To avoid overlaps of these cohorts of 20-69 year-olds, we then work only with the Census data for every 50 years, i.e. 1800, 1850, 1900, 1950, and 2000. Residents aged 0-19 or over-69 then play no role in our subsequent analysis. Finally, we apportion uniformly the number of people born in state s in year t equally into each of the sub-county units i contained in state s in year t. This procedure delivers our proxy for migration flows L_{ijt} (and hence also total populations $L_{jt} \equiv \sum_i L_{ijt}$).

Our second important variable is that for nominal per-capita incomes, w_{it}. The US Census did not track wage income until 1940, but an available proxy is available for the value of county-level total agricultural and manufacturing output from 1850-present. Under the assumption that local expenditure on (and hence income from) non-tradable services tracks that for agriculture and manufacturing, this data series provides a measure of $w_{it}L_{it}$ and hence w_{it}. Because this essential ingredient of our estimation procedure is only available from 1850 onwards, we treat 1800 as date 0 (and hence $\{L_{i0}\}$ comes from L_{it} in 1800).

The third data ingredient concerns intra-national trade flows, X_{ijt}. To the best of our knowledge this is only publicly available (within the 1850-2000 period) beginning in the year 1997 from the Commodity Flow Survey (CFS).

Finally, an instrumental variable estimation procedure that we describe below requires observable proxies for the geographic productivity and amenity terms, for which we collect contemporary measures of elevation, soil quality, temperature, an precipitation. For the purpose of constructing valid instruments, we treat these observed geographic characteristics as time-invariant properties of a location.

3.2 Identification and Estimation

We now describe a three-step estimation procedure designed to recover estimates of the elasticity parameters (α_1, β_1, α_2, β_2, σ and θ) and geographic fundamentals $\left\{\bar{A}_{it}, \bar{u}_{it}, \tau_{ijt}, \mu_{ijt}\right\}$ for all locations i and years t from 1850-2000 through the use of the above data on L_{ijt} from the years 1800-2000, on w_{it} from the years 1850-2000, and on X_{ijt} from one cross-section (in 1997).

In the *first step* of this procedure we assume that trade and migration costs, τ_{ijt} and μ_{ijt} are functions of observable (potential) shifters of these costs. While there are many such potential shifters, we focus on one particularly important one, for now, which is the simple distance between locations i and j (denoted $dist_{ij}$). For now we use only distance and model these costs as $\ln \tau_{ijt} = \kappa_t \ln dist_{ij}$ and $\ln \mu_{ijt} = \lambda_t \ln dist_{ij}$. Substituting these expressions into the gravity equations for trade and migration flows, equations (3) and (7) respectively, we obtain

$$\ln X_{ijt} = (1-\sigma)\,\kappa_t \ln dist_{ij} + \gamma_{it} + \delta_{jt} + \varepsilon_{ijt} \tag{15}$$

$$\ln L_{ijt} = -\theta \lambda_t \ln dist_{ij} + \rho_{it} + \pi_{jt} + \nu_{ijt}, \tag{16}$$

where the terms γ_{it}, δ_{jt}, ρ_{it}, and π_{jt} represent fixed effects in these gravity estimation equations, and we interpret ε_{ijt} and ν_{ijt} as potential measurement error (that is uncorrelated with distance) in trade and migration flows respectively.

In principle, one could estimate a separate κ_t for any year t in which data on trade flows X_{ijt} are available. However, as described above, we only have access to such data for one year, 1997. So we assume that κ is constant over time, as is broadly consistent with the patterns in international trade data surveyed by Disdier and Head (2008). By contrast, data on migration flows L_{ijt} are available for all decades from 1850 onwards so we estimate corresponding λ_t separately for each year. The result of this first step is an estimate of the composite parameters $(1-\sigma)\kappa_t$ and $\theta\lambda_t$.[7]

Turning to our *second step,* we define $T_{ijt} \equiv \hat{\tau}_{ijt}^{1-\sigma} = dist_{ij}^{(1-\sigma)\kappa_t}$, $M_{ijt} \equiv \hat{\mu}_{ijt}^{-\theta} = dist_{ij}^{-\theta\lambda_t}$, which are identified in step one (as they are a function of observables and the identified composite parameters only). For notational ease, further define $p_{it} \equiv \frac{w_{it}}{A_{it}}$, and $Y_{it} \equiv w_{it} L_{it}$. Then the system of equations (8)-(11) defining equilibrium for each period can be written as

$$p_{it}^{\sigma-1} = \sum_j T_{ijt} \left(\frac{Y_{jt}}{Y_{it}} \right) P_{jt}^{\sigma-1} \tag{17}$$

$$P_{it}^{\sigma-1} = \sum_j T_{jit} \left(p_{jt}^{\sigma-1} \right)^{-1} \tag{18}$$

$$\left(W_{it}^{\theta} \right)^{-1} = \sum_j M_{jit} \frac{L_{jt-1}}{L_{it}} \left(\Pi_{jt}^{\theta} \right)^{-1} \tag{19}$$

$$\Pi_{it}^{\theta} = \sum_i M_{ijt} W_{jt}^{\theta}. \tag{20}$$

Noting that we have data on Y_{it} and L_{it} for all locations i and periods t and that the values T_{ijt} and M_{ijt} were identified in step 1, the following proposition shows that the four remaining variables in equations (8)-(11) are identified because this system of equations has

[7] Given data limitations, we estimate the trade gravity regression on bilateral state to state trade flows, whereas we estimate the migration gravity regression on state (of birth) to county (of residence) population flows.

a unique solution given (Y_{it}, L_{it}, T_{ijt} and M_{ijt}).

Proposition 4. Given observed data on $\{Y_{it}, L_{it}, L_{it-1}\}$ and identified values of $\{T_{ijt}, M_{ijt}\}$ from step two there exists unique (up to-scale) values of $\left\{p_{it}^{\sigma-1}, P_{it}^{\sigma-1}, W_{it}^{\theta}, \Pi_{it}^{\theta}\right\}$ that satisfy equations *(8)-(11)*.

Proof. See Section A.4. □

An important feature of this second step result is that it does not depend on the values of the trade or migration elasticities, σ and θ, only on the composite parameters that were recovered in step 1. The basic intuition of this recovery procedure is that we are recovering the analogs of the exporter and importer fixed effects of the trade gravity equation (which allow recovery, by standard arguments, of $p_{it}^{\sigma-1}$ and $P_{it}^{\sigma-1}$, respectively) and the origin and destination fixed effects of the migration gravity equation (W_{it}^{θ} and Π_{it}^{θ}, respectively) from non-bilateral data (unlike the usual approach to recovery of such fixed effects) by making use of the goods and labor market clearing equations for equilibrium.

Finally, we turn to the *third step* of our estimation procedure. By definition, $p_{it} \equiv \frac{w_{it}}{A_{it}}$; therefore, given the definition of $A_{it} = \overline{A}_{it} L_{it}^{\alpha_1} L_{it-1}^{\alpha_2}$ we take logs of $p_{it}^{\sigma-1}$ to obtain:

$$\begin{aligned} \ln\left(p_{it}^{\sigma-1}\right) = & (\sigma-1)\ln w_{it} + \alpha_1(1-\sigma)\ln L_{it} + \alpha_2(1-\sigma)\ln L_{it-1} \\ & + (1-\sigma)\ln \bar{A}_{it}. \end{aligned} \tag{21}$$

Recall that the value of $p_{it}^{\sigma-1}$ was identified (up to an irrelevant scale parameter) in step two. Therefore, equation (21) represents an equation that can be used as a simple regression estimating equation given data on the right-hand side variables, w_{it}, L_{it} and L_{it-1}. Consistent estimates of this equation can therefore identify σ, $\alpha_1$and α_2. However, the unobservable term in equation (21), $(1-\sigma)\ln \bar{A}_{it}$, the error term in this estimating equation, would be correlated with the regressors w_{it} and L_{it}—indeed, the migration behavior in equation (7) suggests that migrants would move to locations with exceptional values of this residual, $\bar{A}_{it}$. We come back to our (instrumental variables) strategy to deal with this endogeneity problem below.

Analogous manipulations on the migration side imply

$$\ln\left(W_{it}^{\theta}\right) = \theta \ln w_{it} + \left(\frac{\theta}{\sigma-1}\right)\ln\left(P_{it}^{1-\sigma}\right) + \beta_1\theta\ln L_{it} + \beta_2\theta\ln L_{it-1} + \theta\ln\bar{u}_{it}, \tag{22}$$

which is again an equation that relates a variable recovered from step two, the migration equation origin fixed-effect W_{it}^{θ}, on observables (w_{it}, L_{it} and L_{it-1}) as well as another variable recovered from step two, the trade destination fixed-effect $P_{it}^{1-\sigma}$. Again, this regression specification allows the opportunity to estimate three key elasticities (θ, β_1 and β_2), but the logic of migration suggests that there is an unavoidable endogeneity problem due to the correlation between the unobserved amenity shifter $\bar{u}_{it}$, the regression residual in equation (22), and regressors such as $\ln w_{it}$ and $\ln L_{it}$. Finally, we note that equations (21) and (22) together over-identify the parameter σ, so there are opportunities for testing this restriction.

To construct instrumental variables (IVs) for the endogenous regressors $\{\ln w_{it}, \ln L_{it}, \ln L_{it-1}, \ln P_{it}\}$ in equations (21) and (22), we draw on model-based simulations of these variables along the lines of Allen, Arkolakis, and Takahashi (2014) and Adao, Arkolakis, and Esposito (2018). This proceeds as follows. First, we begin with a candidate guess of the elasticity parameters, at values motivated by the existing literature.[8] We denote these candidate values as ($\alpha_1^{(IV)}$, $\beta_1^{(IV)}$, $\alpha_2^{(IV)}$, $\beta_2^{(IV)}$, $\sigma^{(IV)}$ and $\theta^{(IV)}$). Second, we assume that, for the purposes of constructing IVs, we can model the productivity shifter that will enter our IV, $\overline{A}_{it}^{(IV)}$, as a function of a vector of certain time-invariant observable geographic characteristics of location i that we denote $\mathbf{z}_i$; in particular, we let $\ln \overline{A}_{it}^{(IV)} = \gamma_A \cdot \mathbf{z}_i$ in any year t. Similarly, we model the amenity shifter $\overline{u}_{it}$ as $\ln \overline{u}_{it}^{(IV)} = \gamma_u \cdot \mathbf{z}_i$.[9] We note that these assumptions (and in particular that $\overline{A}_{it}^{(IV)}$ and $\overline{u}_{it}^{(IV)}$ are not time-varying) refer to the construction of the IVs, not the model that is used for counterfactuals below. Third, in order to estimate the values of γ_A and γ_u, we use the estimates from step two (at the distance elasticities estimated in step one) above to estimate, along with the candidate guess of our six elasticity parameters, candidate values of $\overline{A}_{it}$ and $\overline{u}_{it}$ in the year $t = 2000$. We then use OLS to project $\ln \overline{A}_{i,t=2000}$ and $\ln \overline{u}_{i,t=2000}$ on $\mathbf{z}_i$ in order to estimate what could be thought of as the "zeroth-stage" parameters, $\widehat{\gamma}_A$ and $\widehat{\gamma}_u$.[10] Fourth, starting from the observed initial population shares in 1800 as $\{L_{i0}\}$, we simulate the IV-generating model forwards in all years from 1850 onwards, using the candidate elasticities ($\alpha_1^{(IV)}$, $\beta_1^{(IV)}$, $\alpha_2^{(IV)}$, $\beta_2^{(IV)}$, $\sigma^{(IV)}$ and $\theta^{(IV)}$) and with the productivity and amenity values for each location and year set to $\ln \overline{A}_i^{(IV)} = \widehat{\gamma}_A \cdot \mathbf{z}_i$ and $\ln \overline{u}_i^{(IV)} = \widehat{\gamma}_u \cdot \mathbf{z}_i$. This procedure generates

[8] We set $\beta_1 = -0.3$ to match the interpretation of this parameter as minus the housing share in consumption, as discussed in Section 2, the remaining spillover terms to $\alpha_1 = \alpha_2 = \beta_2 = 0.1$ as is roughly in line with common estimates of agglomeration externalities, $\sigma - 1 = 8$ so that the trade elasticity is in the range estimated by Donaldson and Hornbeck (2016) for the 19th Century US at country-level spatial resolution, and $\theta = 8$ so that the trade and migration elasticities are equal. In practice our eventual 2SLS parameter estimates are not much affected by these choices.

[9] In practice, the geographic variables are a combination of climatic observables (average January temperature and precipitation), the soil quality variables (the net primary productivity and soil nutrient availability), and topographic variables (elevation and ruggedness).

[10] Figure 15 in the Appendix depicts the relationship between the inverted $\ln \overline{A}_{i,t=2000}$ and $\ln \overline{u}_{i,t=2000}$ (i.e. the productivities and amenities consistent with the observed data given our gravity estimates and candidate elasticity values) and $\ln \overline{A}_i^{(IV)} = \widehat{\gamma}_A \cdot \mathbf{z}_i$ and $\ln \overline{u}_i^{(IV)} = \widehat{\gamma}_u \cdot \mathbf{z}_i$ (i.e. the productivities and amenities that depend only on observed geographic variables. As can be seen, the correlation between the two is reasonably high (0.27 for productivities and 0.51 for amenities), although the variation in the inverted values necessary to match the observed data is substantially larger than the variation predicted by geography alone.

predicted values of the model's endogenous variables, including for the endogenous regressors $\{\ln w_{it}, \ln L_{it}, \ln L_{it-1}, \ln P_{it}\}$ in equations (21) and (22). We denote those predictions, from the IV-generating model, as $\left\{\ln w_{it}^{(IV)}, \ln L_{it}^{(IV)}, \ln L_{it-1}^{(IV)}, \ln P_{it}^{(IV)}\right\}$. Fifth, we use these variables as IVs when estimating equations (21) and (22) vis 2SLS. Sixth, when estimating these equations we control directly for $\mathbf{z}_i$ and L_{i0} so that the excluded geographical component of the IV is the effect of these geographical characteristics and initial populations in a *other* locations. Finally, we also divide our 25km×25km "cells" into 500km×500km geographically contiguous "boxes" and include box-year fixed effects in our preferred specification to control for spatially coarse regional variation in fundamental productivities or amenities that is time-varying but unobserve.

To summarize, our instrumental variables are simply functions of observed 1800 populations, the full matrix of bilateral distances, and the stated geographical characteristics of $\mathbf{z}_i$ (i.e. climate, soil quality, and topography) of each location. But since we control for any location's own 1800 population and geographic characteristics, the *excluded* instruments are these variables for all other regions. Our identifying assumption is that those excluded instruments are uncorrelated with the error terms (unobserved productivity and amenity terms, $\overline{A}_{it}$ and $\overline{u}_{it}$) in equations (21) and (22). Formally, this means that any function of these variables would deliver consistent 2SLS estimates of the parameters (α_1, β_1, α_2, β_2, σ and θ). But given the many such potential functions of this long set of variables, we use the logic of our model to achieve dimensionality-reduction and effective instrument selection.

Finally, we note that, conditional on obtaining consistent estimates of the elasticity parameters (α_1, β_1, α_2, β_2, σ and θ), equations (21) and (22) allow recovery of the geographic fundamentals $\left\{\bar{A}_{it}, \bar{u}_{it}\right\}$ as well. Combined with the earlier estimates of $\{T_{ijt}, M_{ijt}\}$ from step two all model parameters are thereby identified.

3.3 Estimation Results

We begin with estimates from the trade and migration gravity equations in step one. As is standard, our estimate of the elasticity of trade flows with respect to distance (using the 1997 CFS data to estimate equation 15) is close to minus one: in particular, we estimate $\kappa(1-\sigma) = -1.20$ ($SE = 0.24$). Perhaps surprisingly, the migration-distance elasticity is also close to minus one in all periods that we investigate. These estimates are shown in Figure 4, where the combination of parameters $\theta\lambda_t$ appears to range from -1.5 to -0.8, with no clear trend over the 150 years for which we display decadal estimates.[11] This is similar to the

[11] While we explore in Figure 4 these estimates of $\theta\lambda_t$ across each decade, our simulations below apply the estimates (in step two of our procedure, for the purposes of constructing M_{ijt}) only in the years 1850, 1900, 1950 and 2000.

persistence of the trade-distance elasticity over time, as discussed above.

Turning to step three, the parameter values implied by our 2SLS estimates of the coefficients in equations (21) and (22) are reported in Table 1.[12] These estimates have some noteworthy features. First, our estimates of σ do differ substantially across columns (1) and (2), with the more precise estimates from column (1) being closer to standard values in the literature (if perhaps on the high end, as might be expected given our study of trade among small, intranational spatial units).[13] Second, our estimates of productivity spillovers, α_1 and α_2, are both positive, as prior work (on static estimation settings, which lump these two parameters together) might suggest. The contemporaneous amenity spillover, $\beta_1 = -0.341$, is negative and strikingly close to the value predicted by a model of Cobb-Douglas preferences (with expenditure share on housing of about one-third) and fixed local housing supply. The estimate of the historical amenity spillovers parameter, β_2, is also negative but the actual point value of -0.004 is very close to zero in practice.

Because the productivity spillovers are positive and the amenity spillovers negative, it is not clear whether these estimates are in the range for uniqueness of equilibria, and of steady-states, implied by Propositions 1 and 2. Figure 4 plots these estimates in the ranges implied by these propositions (which also depend on the estimates of σ and θ). From the fact that the red star is inside the yellow region we see that the equilibria will be unique at these parameter estimates; similarly, from the fact that the green star is (just) inside the blue region we see that multiple steady-states are a possibility (that is, the sufficient condition for uniqueness of steady-states identified in Proposition 2 is not satisfied by these estimates). These two findings suggest that path dependent outcomes are indeed a possibility in this model economy, and that computation of equilibrium paths, given any starting point, is guaranteed to be straightforward given the equilibrium uniqueness result of Proposition 1.

4 The Geography of Path Dependence

Having estimated all of the ingredients of the model introduced above on data from the history of US economic geography from 1800 onwards, we now use the estimated model to

[12] The corresponding first-stage estimates for this 2SLS system are all of the expected signs—implying that the IV-model predictions are in line, at least qualitatively, with the actual data in these first-stage moments—and statistically significant (with t-statistics in excess of 10 and hence implied univariate F-statistics in excess of 100).

[13] In our counterfactual simulations below we use the value of σ implied by column (1) because of its greater precision and its greater proximity to values in the existing literature. We also note that the spirit of using trade data to estimate the trade elasticity, $\sigma - 1$, as we do in column (1) is common in the literature, whereas the method of column (2), which infers the trade elasticity from the responsiveness of migration flows to the implied price of tradable goods from the destination fixed-effect of the trade gravity equation, has no parallel in existing literature to our knowledge.

obtain a quantitative understanding of its path-dependent features. To do so, we assess both how different historical conditions would affect the distribution of economic activity and welfare today (i.e. in the year 2000) and in the long-run (i.e. in the steady state).

4.1 The Effect of History on the Distribution of Economic Activity Today

To assess the affect of history on the distribution of economic activity today, we pursue a computational approach where we simulate alternative paths of the distribution of US economic activity under differing historical conditions, holding constant contemporary geography. Many such historical conditions could be studied in the framework developed and estimated above. But for now we focus on the question of how sensitive are long-run outcomes to the spatial distribution of historical location-specific shocks to locational fundamentals ($\left\{\bar{A}_{it}\right\}$ in our model). Clearly, in a setting without path dependence long-run (i.e. steady-state) outcomes would not depend on such historical shocks, but in our model, at our estimated parameter values, these shocks may matter and we aim to quantify the extent to which they do.

Our simulations proceed as follows. First, we fix initial population levels $\{L_{i0}\}$ to those seen in the data. Second, we fix the values of $\left\{\bar{A}_{it}\right\}$ for the years 1850 and 2000 to those we have backed out from the estimation procedure above—this means that the starting point and ending point of each location's path of fundamentals remain fixed across simulations. We also fix the amenity values $\{\bar{u}_{it}\}$ at their factual values in all time periods. Finally, we perturb the path of the historical shocks to $\left\{\bar{A}_{it}\right\}$, in the years 1900 and 1950, in a spatially clustered manner. To do that, we define a set of geographically similar clusters $c \in \mathcal{C}$—where $\mathcal{C}$ defines a partition of the space of all sub-county locations i—according to a procedure that we describe below. Within a given cluster c, and within the year of $t = 1900$, say, there is a set of observed values of $\left\{\bar{A}_{it=1900}\right\}_{i \in c}$ backed out from the factual data above. Our simulations then assign counterfactual values of $\bar{A}'_{it=1900}$ to every location $i \in c$ by redrawing without replacement from the set $\left\{\bar{A}_{it=1900}\right\}_{i \in c}$. This essentially randomly reshuffles the values of $\bar{A}_{it=1900}$ in a geographically clustered manner, and allows us to evaluate whether geographically local changes in fundamentals could have had a lasting impact on the economy. Finally, we emphasize that because the current (year 2000) productivity and amenity shifters remain unchanged, the only effect of the path of productivity shocks that we simulate on current outcomes is through their effects on $\{L_{it=2000}\}$ and $\{L_{it=1950}\}$.

To define the geographic clusters c, we use a "k-mean clustering" algorithm that effectively finds the construction of partitions into k different groups of locations that minimizes differences on geographic observables within groups. The particular geographic variables we use for this are the same as those used to define our instrumental variables above (soil quality, elevation, climate, and water access).

Finally, in practice, to reduce computational burdens (resulting from the large number of simulations performed below), we work for now with gridded spatial units that are larger than (i.e. spatial aggregates of) the sub-county units tracked in the data above. There are 570 such spatial grid cells in the simulations reported below, each approximately 125km by 125km. We set $k = 57$, so that there are 10 grid cells per cluster c. A map of the resulting clusters is shown in Figure 6, where each of 57 colors refers to a different cluster, but the absolute color scale is irrelevant.

Figures 7-9 begin to convey a sense of what happens in our simulations. These maps illustrate the distribution of equilibrium 1900-2000 population for each of our 570 gridded simulation locations, among the first three of our 200 simulations. Also shown are the actual population distributions in these years. While these are only three random simulations, a clear impression emerges of the disruption that even our local geographic reshuffling shocks can cause.

Consistent with this evidence from just three simulations, Figure 10 demonstrates that something similar is at work, on average, in all 200 simulations that we have run. In this figure we plot, for each year from 1850-2000, the spread of population in each location across all simulations. For each simulation we calculate the rank of each location in the nationwide population size ranking, and the y-axis reports the tendencies of this ranking across the simulations—the thin blue bar indicates the max and min, the thicker blue bar the interquartile range, and the black dot the mean. The locations are then ordered along the x-axis by their median ranking across simulations. Naturally, the picture from 1850 shows no variance since we are not perturbing any shocks in that year. But, as suggested by the maps in Figures 7-9, for 1900 and 1950 the counterfactual productivity shocks are evidently disruptive in terms of generating a wide spread of alternative histories, simply from reordering local (that is, geographically similar) productivity values. By 2000 the variance is reduced, but it is still substantial.

These counterfactual re-shuffling of productivity shocks provide insight into how robust a particular location is to historical shocks. Figure 11 depicts the relationship between the observed population of each cell in the year 2000 and the variance of its (log) population across all of the simulations. For interpretability, we have labeled each cell with its largest city if that city has a population of more than 200,000 people.[14] As can be seen, there

[14]Note, however, that this mapping is imperfect, as some cities (like New York) span multiple cells, in which case the population of a cell is only part of the population of the city.

is little systematic relationship between population size and how robust a location is to productivity shocks: some large cities like Los Angeles remain large across most simulations (i.e. they are robust to historical shocks), whereas others like Miami vary substantially in size depending on the particular counterfactual (i.e. they are less robust). Figure 12 shows that the degree of robustness to historical shocks does exhibit systematic spatial variation, with Florida, the upper Midwest, and the Rocky Mountain states showing large amounts of variation of population across different historical shocks, whereas California and Arizona exhibiting substantially greater robustness to historical shocks.

Consistent with this, Figure 13 shows that the spread of possible aggregate welfare outcomes across simulations here can be significant, with a 45 log point difference across the max-min range in these 200 simulations. Another feature of Figure 13 is the fact that the actual 2000 aggregate welfare in the US (shown with a yellow star) is right in the middle of the distribution of possible welfare levels across our simulations. So the actual path of productivity shocks $\bar{A}_{it}$ that occurred in 1900-1950 was evidently not that out of line with what was likely, on average, across our simulations.

4.2 The Long-run Impact of Path Dependence

To assess the long-run impact of path dependence, we pursue complementary analytical and computational approaches. Computationally, we simulate the model forward in time for both the observed distribution of economic activity today and each of the 200 historical simulations. Figure 14 shows, in blue, the path of aggregate welfare levels along which the US economy would travel, according to our estimates, if all geographic fundamentals were held at their year 2000 levels for ever more. It also shows, in in green, the corresponding trajectory for each of our 200 simulations. We find that all 200 simulations converge to the same steady state as the observed distribution of economic activity, but the speed of convergence is slow: even after 1,000 years, there exists variation of about 5 log points in the welfare across different historical shocks.

Of course, the fact that all 200 simulations we consider converge to the same steady state does not imply that history will not matter for the steady state distribution of economic activity: it could simply be that the shocks we considered were of insufficient size to push the economy out of its current basin of attraction. To consider the scope for large welfare impacts of all possible historical paths, we calculate bounds on the welfare impacts of path dependence, according to the geography that obtains in our setting, using Proposition 3.

The upper bound is given by:

$$\underbrace{\bar{\Omega}}_{1.16} \equiv \underbrace{c_1}_{2.40} \times \underbrace{\bar{\lambda}_M^{\frac{1}{\theta}}}_{0.68} \times \underbrace{\bar{\lambda}_T^{\frac{1}{\sigma-1}}}_{1.22} \times \underbrace{\bar{L}^{\left(\rho-\frac{1}{\theta}\right)}}_{0.58},$$

whereas the lower bound is given by:

$$\underbrace{\underline{\Omega}}_{0.05} \equiv \underbrace{c_2}_{0.66} \times \underbrace{\underline{\lambda}_M^{\frac{1}{\theta}}}_{0.37} \times \underbrace{\underline{\lambda}_T^{\frac{1}{\sigma-1}}}_{0.22} \times \underbrace{\left(\frac{\bar{L}}{N}\right)^{\rho}}_{1}.$$

Given that the steady state welfare of the steady state we are converging toward is has an associated welfare of $\Omega = 0.76$, this implies that no historical path could have increased the steady state welfare of the current path by more than about 50 percent. As a basis of comparison, this is roughly equal to the range of short-run welfare impacts of the historical shocks we simulate.

Finally, recall that the ratio of the upper and lower bounds provides a convenient decomposition of geography of path dependence. From equation (14) we have:

$$\underbrace{\hat{\Omega}^{PD}}_{21.57} \leq \underbrace{\frac{c_1}{c_2}}_{3.61} \times \underbrace{\kappa\left(\mathbf{M}\right)^{\frac{1}{\theta}}}_{1.83} \times \underbrace{\kappa\left(\mathbf{T}\right)^{\frac{1}{\sigma-1}}}_{5.63} \times \underbrace{N^{\rho}\bar{L}^{-\frac{1}{\theta}}}_{0.58}.$$

Hence, the contribution of the trade matrix (which, recall, includes variation in productivities and amenities across locations) is roughly three times as great as that of the migration matrix to the geography of path dependence, although both suggest there exists substantial latitude for history to matter in the long run.

5 Conclusion

It is not hard to look at the geographic patterns of economic activity around us and believe both that agglomeration forces are important, and that they are strong enough to be the source of self-reinforcing, stable clustering of economic activity. This opens up the possibility that there are many such locations at which economic activity could settle in steady-state—some good, some bad—and the potential for historical accidents, such as initial conditions or long-defunct technological shocks, to play an outsized role in governing both where economic activity occurs and how efficiently it occurs overall.

This paper has sought to develop a theoretical framework that can be combined with historical data to characterize and quantify such path-dependent spatial phenomena. Six elasticities matter for geographic path dependence, according to our theory: two dispersion elasticities coming from the desire for goods and migrants to seek substitute locations, two elasticities governing the the strength of contemporaneous local productivity and amenity agglomeration externalities, and two elasticities capturing the propensity for lagged agglomeration spillovers to matters. In the steady state, the extent to which historical accidents affect steady-state welfare (i.e. path dependence) is bounded by simple statistics of these elasticities and the underlying geography.

When applied to US Census data from 1800 onwards, we estimate values of these elasticities that imply the potential for path dependence in this context. Our simulations of randomly chosen, spatially local permutations in initial conditions and historical shocks suggest that the location of economic activity in the US today is highly sensitive to the variations in historical shocks that we consider and the analytical bounds suggest the possibility of larger historical shocks mattering in the long-run.

While we have developed these empirical and theoretical tools in the hopes of an improved understanding of inter-city economic geography, these techniques could be applied to other areas of economics in which increasing returns and coordination failures, and hence multiplicity and path dependence, have long appeared as objects of theoretical interest that lack a corresponding amount of empirical estimation, quantification, and simulation. Potential areas of application could include: intra-city issues such as residential segregation, sorting, and so-called "tipping" dynamics (Schelling, 1971; and Card, Mas, and Rothstein, 2008); traditional "big push" models of economic development (Rosenstein-Rodan, 1943; Murphy, Shleifer, and Vishny, 1989; and Krugman and Venables, 1995); policy questions surrounding technology adoption and competition in the presence of network effects and switching costs (David, 1985; and Farrell and Klemperer, 2007); and the study of dynamic questions of political economy such as those surveyed in Acemoglu and Robinson (2005).

References

Acemoglu, D., and J. A. Robinson (2005): *Economic Origins of Dictatorship and Democracy*. MIT press.

Adao, R., C. Arkolakis, and F. Esposito (2018): "Spatial Linkages, Global Shocks, and Local Labor Markets: Theory and Evidence," mimeo Chicago Booth.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83(6), 2127–2189.

Allen, T., and C. Arkolakis (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*.

Allen, T., C. Arkolakis, and X. Li (2015): "On the existence and uniqueness of trade equilibria," *mimeo, Dartmouth and Yale Universities*.

Allen, T., C. Arkolakis, and Y. Takahashi (2014): "Universal Gravity," *NBER Working Paper 20787*.

Armington, P. S. (1969): "A Theory of Demand for Products Distinguished by Place of Production," *International Monetary Fund Staff Papers*, 16, 159–178.

Bleakley, H., and J. Lin (2012): "Portage and path dependence," *Quarterly Journal of Economics*, 127(2), 587–644.

Bosker, M., S. Brakman, H. Garretsen, and M. Schramm (2007): "Looking for multiple equilibria when geography matters: German city growth and the WWII shock," *Journal of Urban Economics*, 61(1), 152–169.

Caliendo, L., M. Dvorkin, and F. Parro (2015): "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," NBER working paper number 21149.

Card, D., A. Mas, and J. Rothstein (2008): "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics*, 133(1), 177–218.

David, P. A. (1985): "Clio and the Economics of QWERTY," *American Economic Review, Papers and Proceedings*, 75(2), 332–337.

Davis, D., and D. Weinstein (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92(5), 1269–1289.

——— (2008): "A Search for Multiple Equilibria in Urban Industrial Structure," *Journal of Regional Science*, 48(1), 29–65.

Dekle, R., and J. Eaton (1999): "Agglomeration and Land Rents: Evidence from the Prefectures," *Journal of Urban Economics*, 46(2), 200–2014.

Dekle, R., J. Eaton, and S. Kortum (2008): "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55(3), 511–540.

Deneckere, R. J., and K. L. Judd (1992): "Cyclical and Chaotic Behavior in a Dynamic Equilibrium Model, with Implications for Fiscal Policy," *Cycles and Chaos in Economic Equilibrium*, p. 308.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018): "The Geography of Development," *Journal of Political Economy*, 126(3), 903–983.

Desmet, K., and E. Rossi-Hansberg (2014): "Spatial development," *American Economic Review*, 104(4), 1211–43.

Disdier, A.-C., and K. Head (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade," *Review of Economics and Statistics*, 90(1), 37–48.

Donaldson, D., and R. Hornbeck (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *Quarterly Journal of Economics*, 131(2), 799–858.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010): "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100(3), 1195–1213.

Farrell, J., and P. Klemperer (2007): "Coordination and Lock-In: Competition with Switching Costs and Network Effects," in *Handbook of Industrial Organization*, ed. by M. Armstrong, and R. Porter, vol. 3, pp. 1967–2072. Elsevier.

Fujita, M., P. Krugman, and A. J. Venables (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Boston, Massachussetts.

Glaeser, E. L. (2008): *Cities, Agglomeration and Spatial Equilibrium*. Oxford University Press.

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

Herrendorf, B., A. Valentinyi, and R. Waldmann (2000): "Ruling out multiplicity and indeterminacy: the role of heterogeneity," *Review of Economic Studies*, 67(2), 295–307.

Kline, P., and E. Moretti (2014): "Local economic development, agglomeration economies and the big push: 100 years of evidence from the tennessee valley authority," *Quarterly Journal of Economics*, 129, 275–331.

Krugman, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70(5), 950–959.

——— (1991): "History versus expectations," *Quarterly Journal of Economics*, 106(2), 651–667.

Krugman, P., and A. J. Venables (1995): "Globalization and the Inequality of Nations," *Quarterly Journal of Economics*, 110(4), 857–880.

Matsuyama, K. (1991): "Increasing returns, industrialization, and indeterminacy of equilibrium," *Quarterly Journal of Economics*, 106(2), 617–650.

Michaels, G., and F. Rauch (2018): "Resetting the Urban Network: 117-2012," *Economic Journal*, 128(608), 378–412.

Miguel, E., and G. Roland (2011): "The long-run impact of bombing Vietnam," *Journal of Development Economics*, 96(1), 1–15.

Mitrinovic, D. S., and P. M. Vasic (1970): *Analytic inequalities*, vol. 1. Springer.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1989): "Industrialization and the Big Push," *Journal of Political Economy*, 97(5), 1003–1026.

Nagy, D. K. (2017): "City location and economic development," mimeo CREi.

Ottaviano, G. I. (1999): "Integration, geography and the burden of history," *Regional Science and Urban Economics*, 29(2), 245–256.

Rauch, J. E. (1993): "Does History Matter Only When it Matters Little? The Case of City-Indu try Location," *Quarterly Journal of Economics*, 108(3), 843–867.

Redding, S., D. Sturm, and N. Wolf (2011): "History and Industrial Location: Evidence from German Airports," *Review of Economics and Statistics*, 93(3), 814–831.

Redding, S. J., and E. A. Rossi-Hansberg (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9(1).

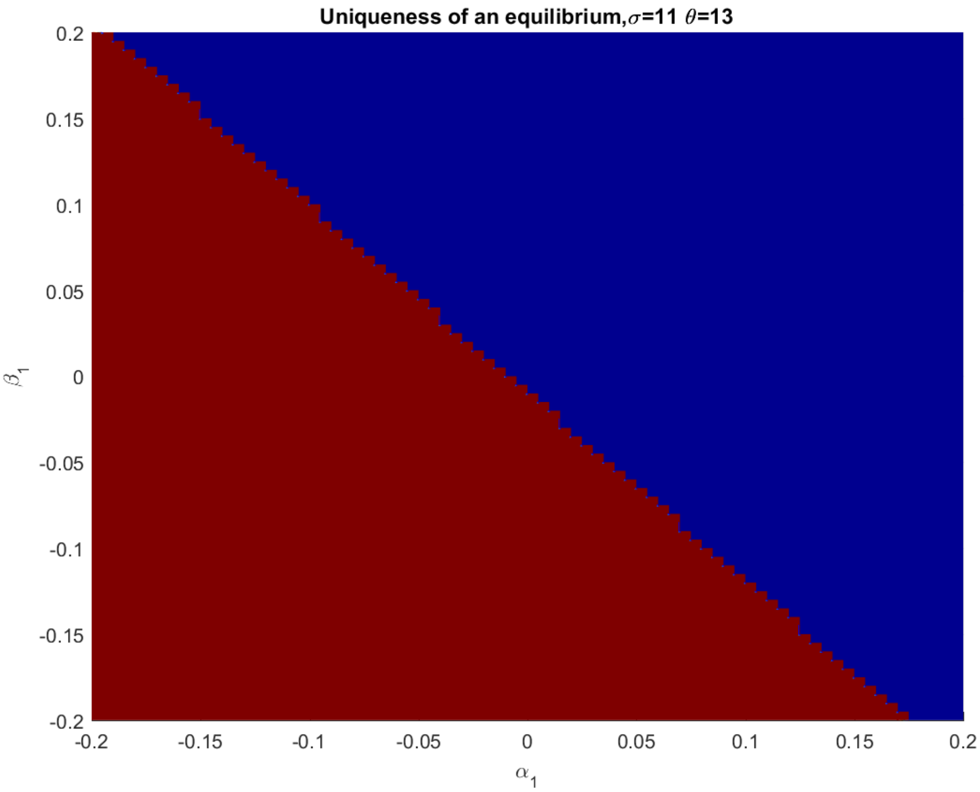
Roback, J. (1982): "Wages, rents, and the quality of life," *The Journal of Political Economy*, pp. 1257–1278.

Rosenstein-Rodan, P. N. (1943): "Problems of Industrialisation of Eastern and South-eastern Europe," *Economic Journal*, June-September, 202–11.

Schelling, T. C. (1971): "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1(July), 143–186.

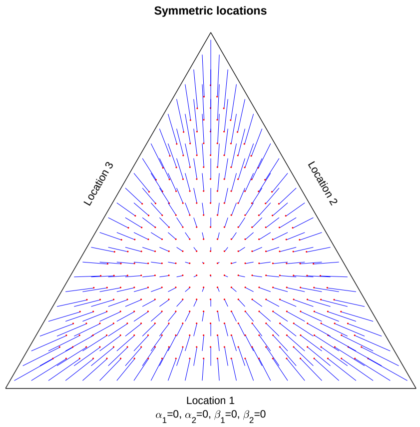
Specht, W. (1960): "Zur theorie der elementaren mittel," *Mathematische Zeitschrift*, 74(1), 91–98.

Figure 1: Illustration of Proposition 1

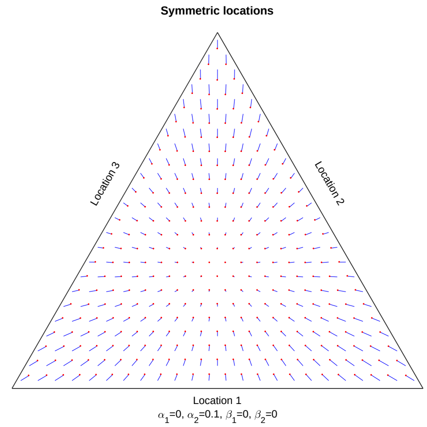


Notes: This figure illustrates the regions of the parameter range (in the space of α_1 and β_2, holding σ and θ constant at the example values shown above) that satisfy the condition for uniqueness of equilibrium, as per Proposition 1.

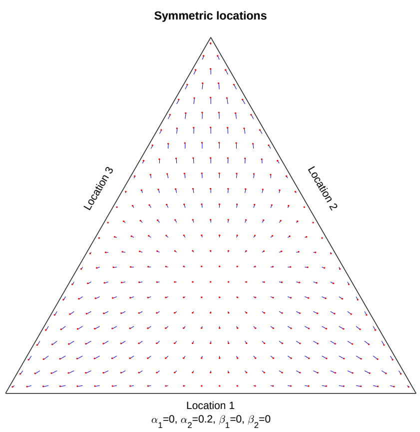
Figure 2: Phase diagrams for 3-region symmetric example



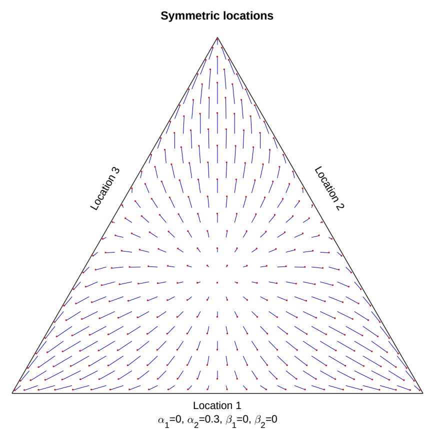
(a) $\alpha_2 = 0$



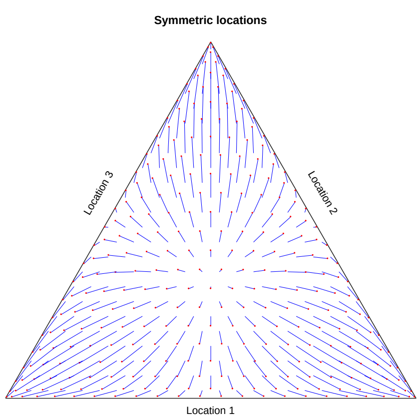
(b) $\alpha_2 = 0.1$



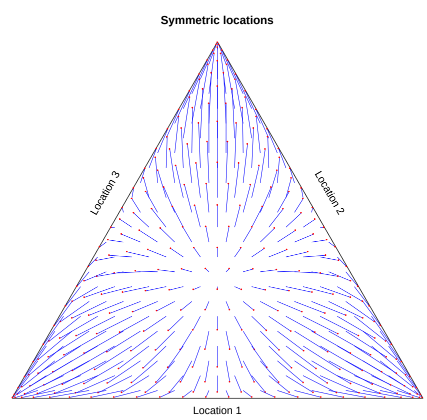
(c) $\alpha_2 = 0.2$



(d) $\alpha_2 = 0.3$



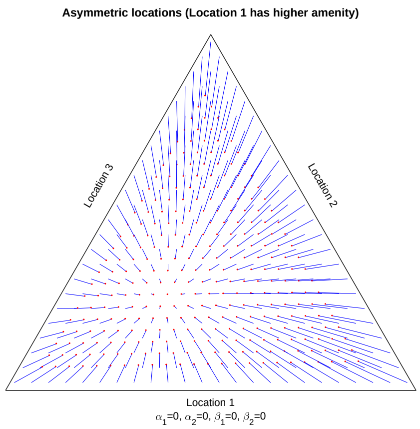
(e) $\alpha_2 = 0.4$



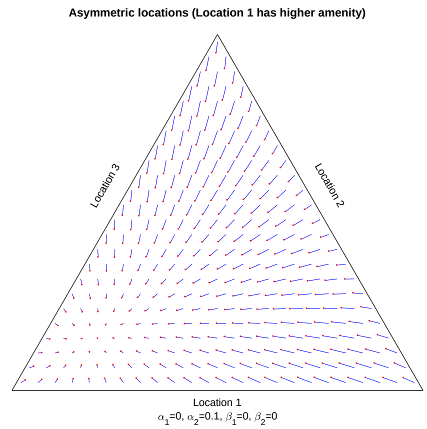
(f) $\alpha_2 = 0.5$

Notes: This figure illustrates phase diagrams for an asymmetric three-region example economy. The parameters $\alpha_1, \beta_1, \beta_2, \sigma$ and θ are held constant as α_2 varies.

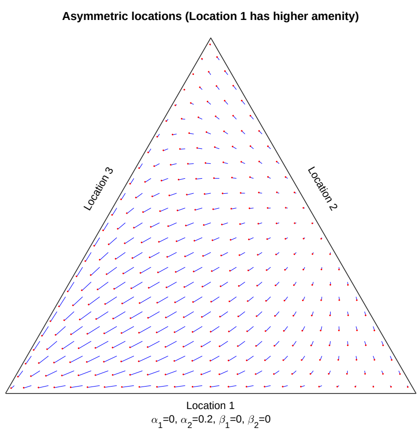
Figure 3: Phase diagrams for 3-region asymmetric example



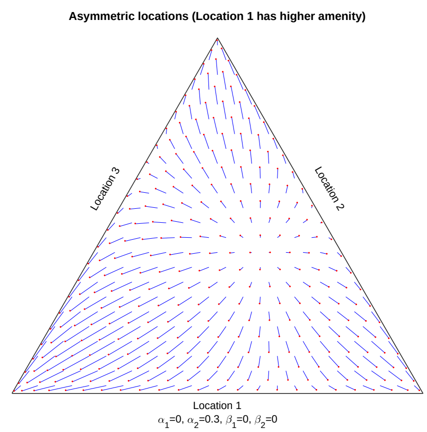
(a) $\alpha_2 = 0$



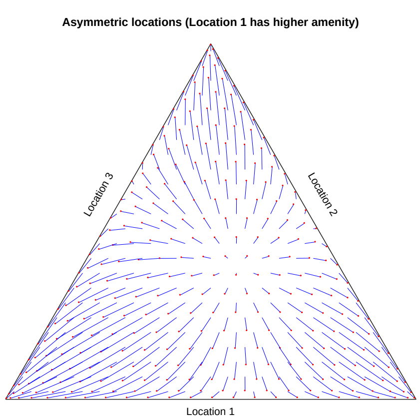
(b) $\alpha_2 = 0.1$



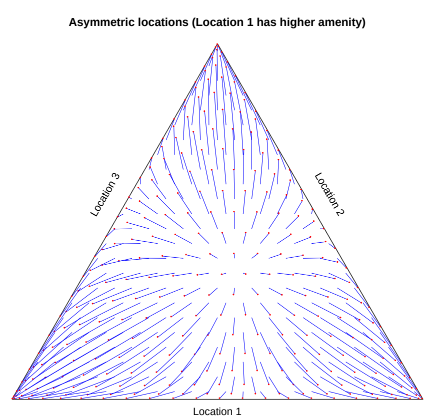
(c) $\alpha_2 = 0.2$



(d) $\alpha_2 = 0.3$



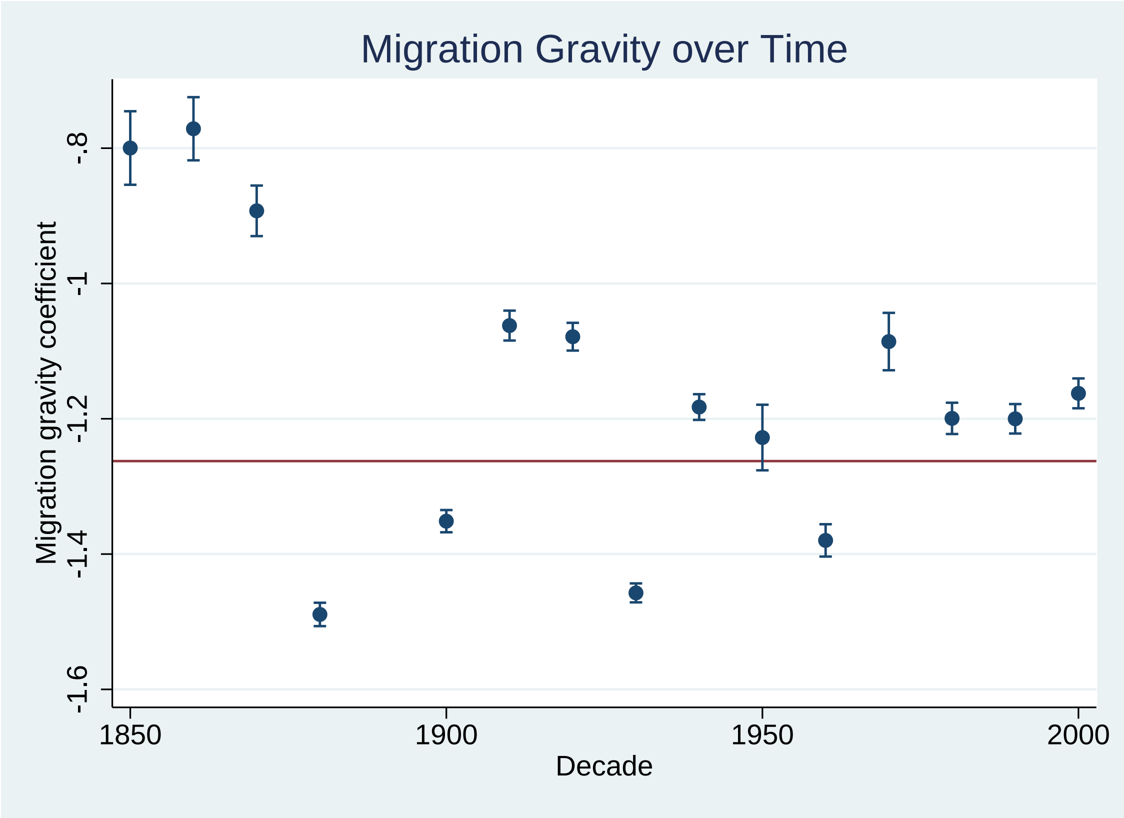
(e) $\alpha_2 = 0.4$



(f) $\alpha_2 = 0.5$

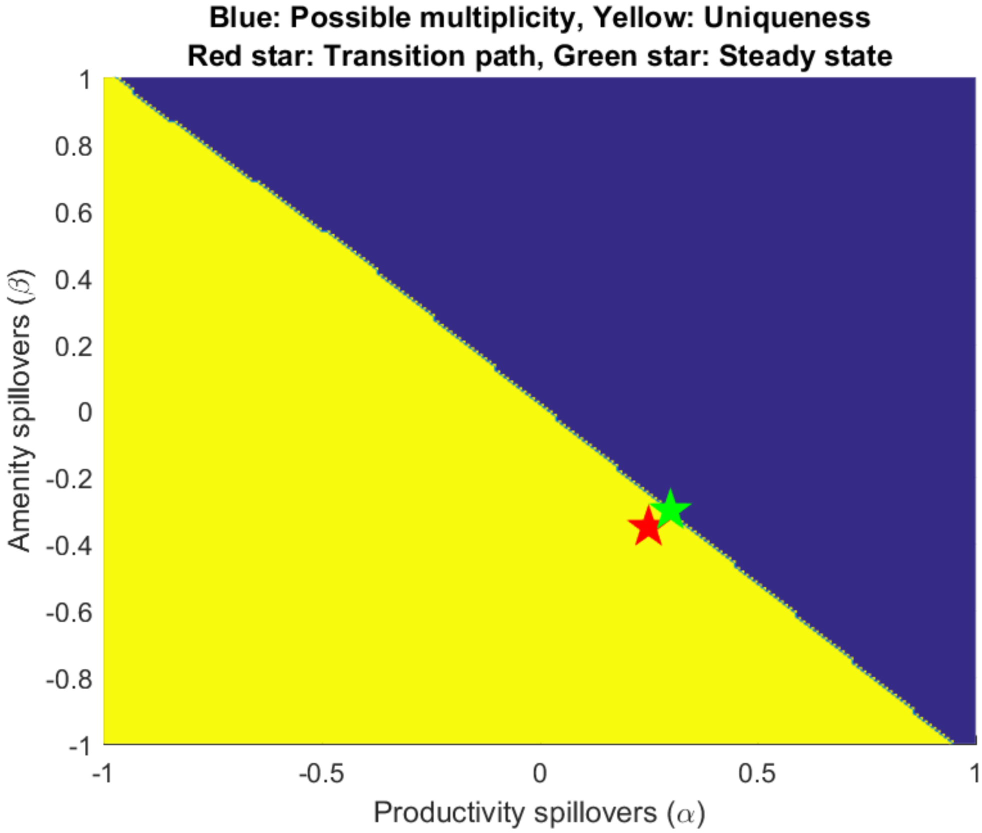
Notes: This figure illustrates phase diagrams for an asymmetric three-region example economy. The parameters $\alpha_1, \beta_1, \beta_2, \sigma$ and θ are held constant as α_2 varies.

Figure 4: Migration gravity parameter estimates



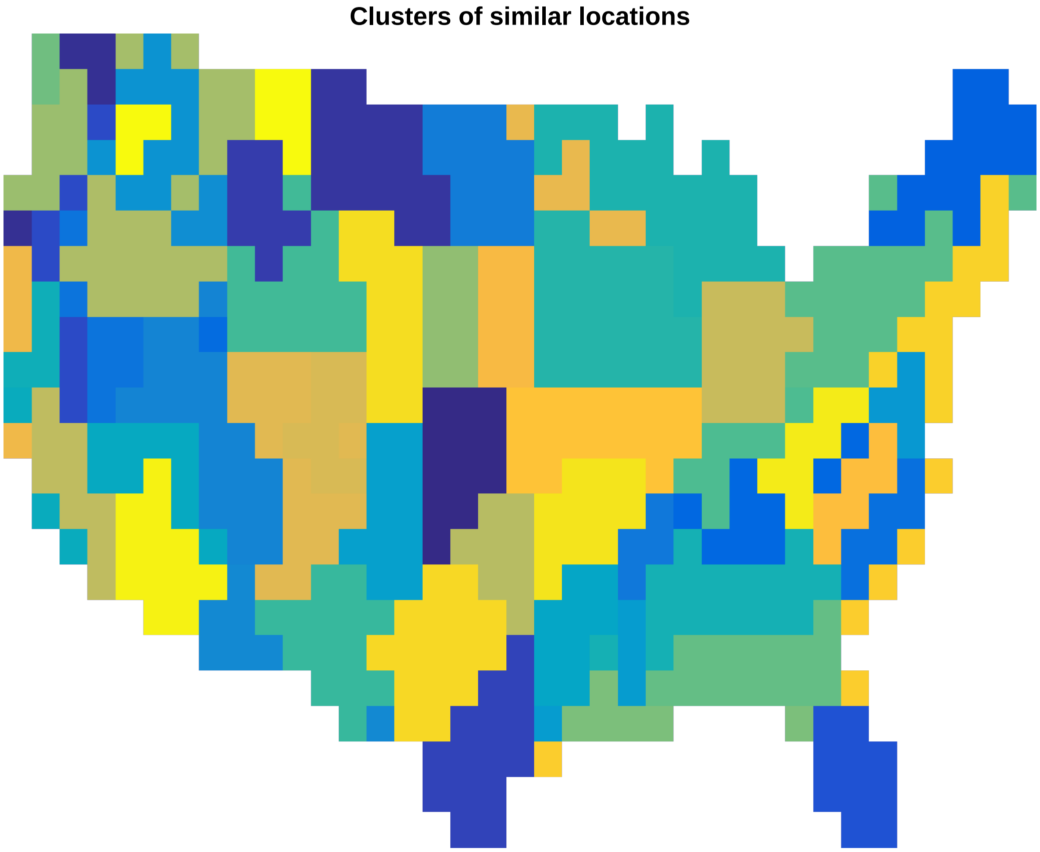
Notes: This figure illustrates estimates (and associated 95% confidence intervals) for the combined parameter $\theta\lambda_t$ from the migration gravity equation (16) estimated in various cross-section decades, t.

Figure 5: Parameter Estimates



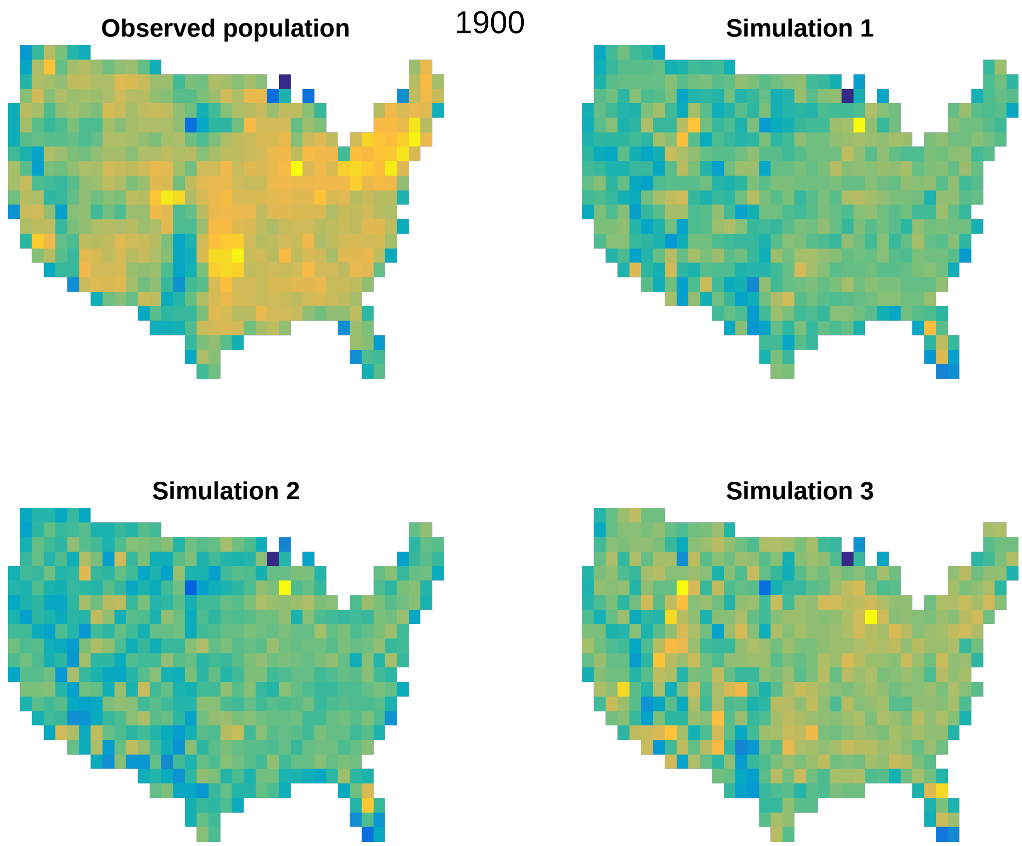
Notes: This figure illustrates the regions of the parameter range (in the space of $\alpha_1+\alpha_2$, along the x-axis, and $\beta_1 + \beta_2$, along the y-axis, holding σ and θ constant at the values estimated in Section 3.3) that satisfy the condition for uniqueness of equilibrium, as per Proposition 1, and uniqueness of steady-states, as per Proposition 2.

Figure 6: Map of geographic clusters



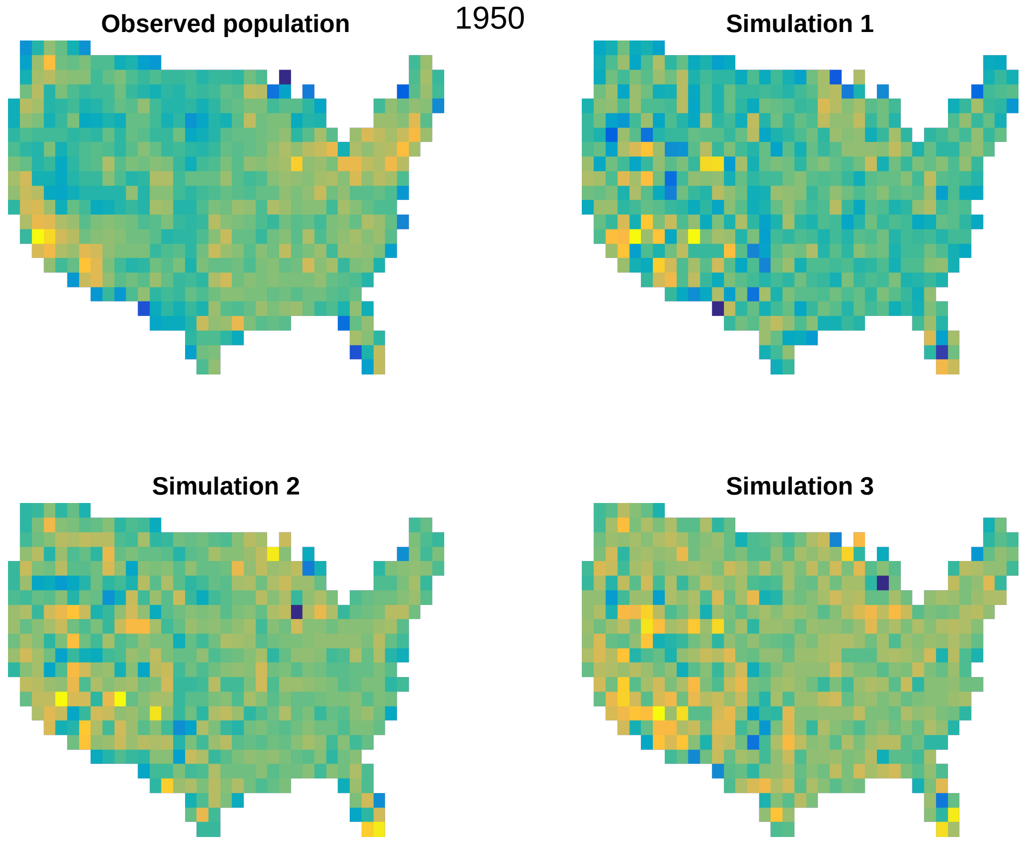
Notes: This figure illustrates a map of the 570 locations in our simulations, as well as how they are grouped into 57 clusters designed to minimize the within-cluster variation in geographic characteristics (according to a k-mean clustering algorithm).

Figure 7: Map of 3 example simulations of random (within-cluster) productivity values in 1900 and 1950



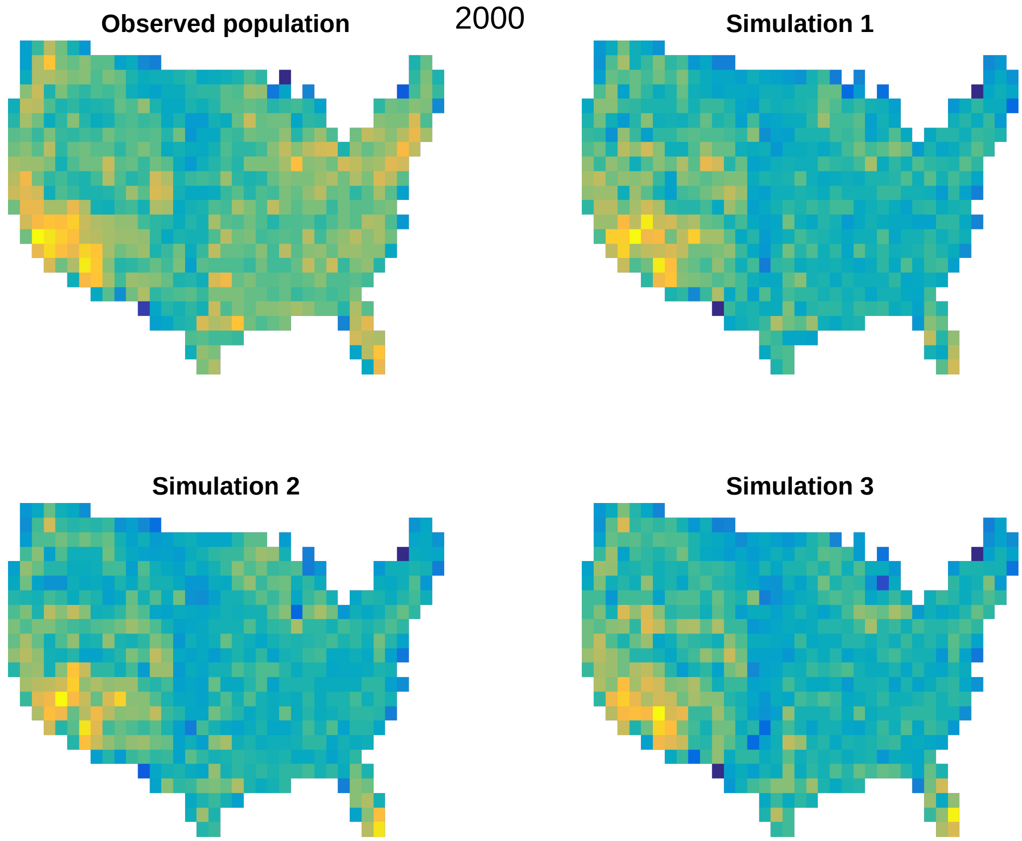
Notes: This figure illustrates (for 1900) the results of the first three of our 200 simulations of randomized (within the geographically similar cluster regions of Figure 6, drawn without replacement) productivity values in 1900 and 1950. Also show, at the top left, is the map of the actual population distribution.

Figure 8: Map of 3 example simulations of random (within-cluster) productivity values in 1900 and 1950



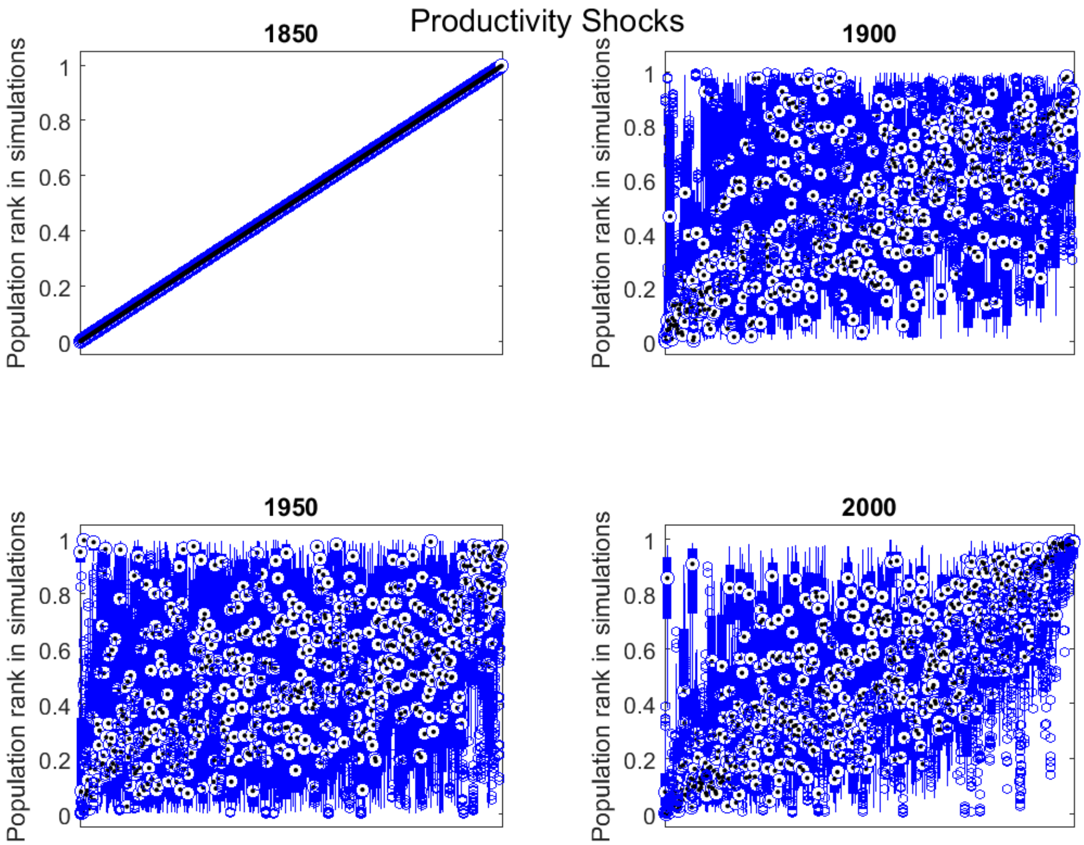
Notes: This figure illustrates (for 1950) the results of the first three of our 200 simulations of randomized (within the geographically similar cluster regions of Figure 6, drawn without replacement) productivity values in 1900 and 1950. Also show, at the top left, is the map of the actual population distribution.

Figure 9: Map of 3 example simulations of random (within-cluster) productivity values in 1900 and 1950



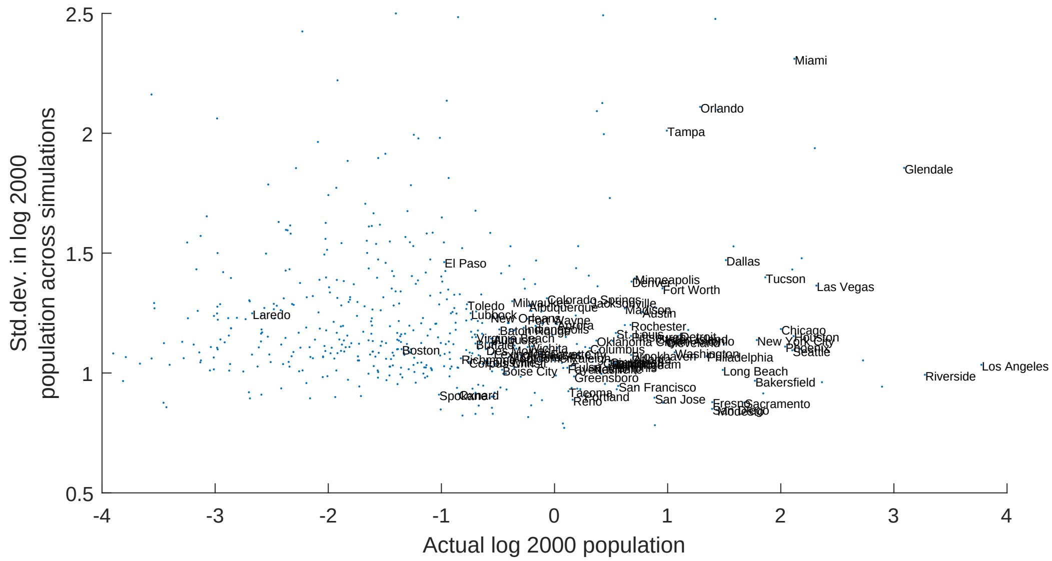
Notes: This figure illustrates (for 2000) the results of the first three of our 200 simulations of randomized (within the geographically similar cluster regions of Figure 6, drawn without replacement) productivity values in 1900 and 1950. Also show, at the top left, is the map of the actual population distribution.

Figure 10: The distribution of population, by year, across 200 simulations of random productivity values in 1900 and 1950



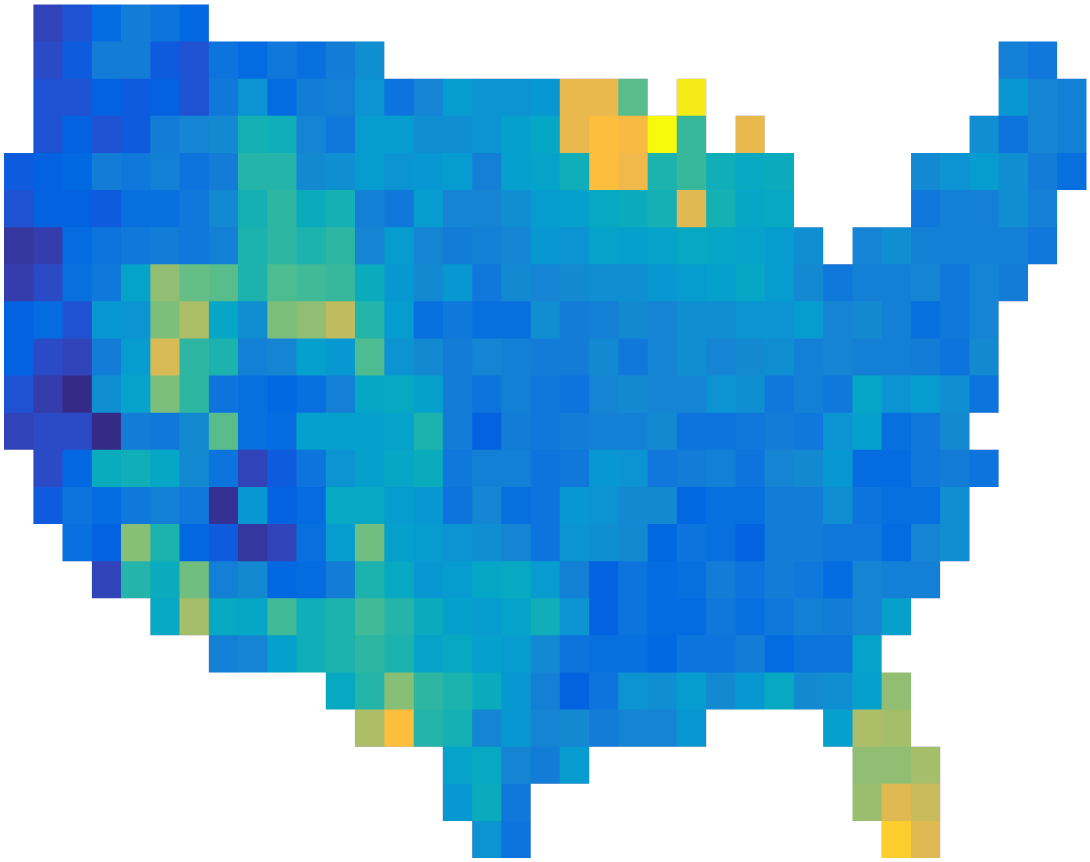
Notes: This figure illustrates how a location's rank in the nationwide population distribution (within any given year, as shown) can vary across 200 simulations of random (drawn, within-the geographically similar clusters of Figure 6, without replacement) productivity shocks $\overline{A}_{it}$ in years 1900 and 1950 (but not 1850 or 2000). The x-axis refers to each location, ordered according to its across-200 simulations median nationwide population rank within the year shown. Then, for each location, the figure contains box plots (with the max-min range in narrow blue, the interquartile range in wide blue, and the mean shown with a black circle) of that location's cross-simulations distribution of nationwide population rank.

Figure 11: The robustness of cities to historical shocks



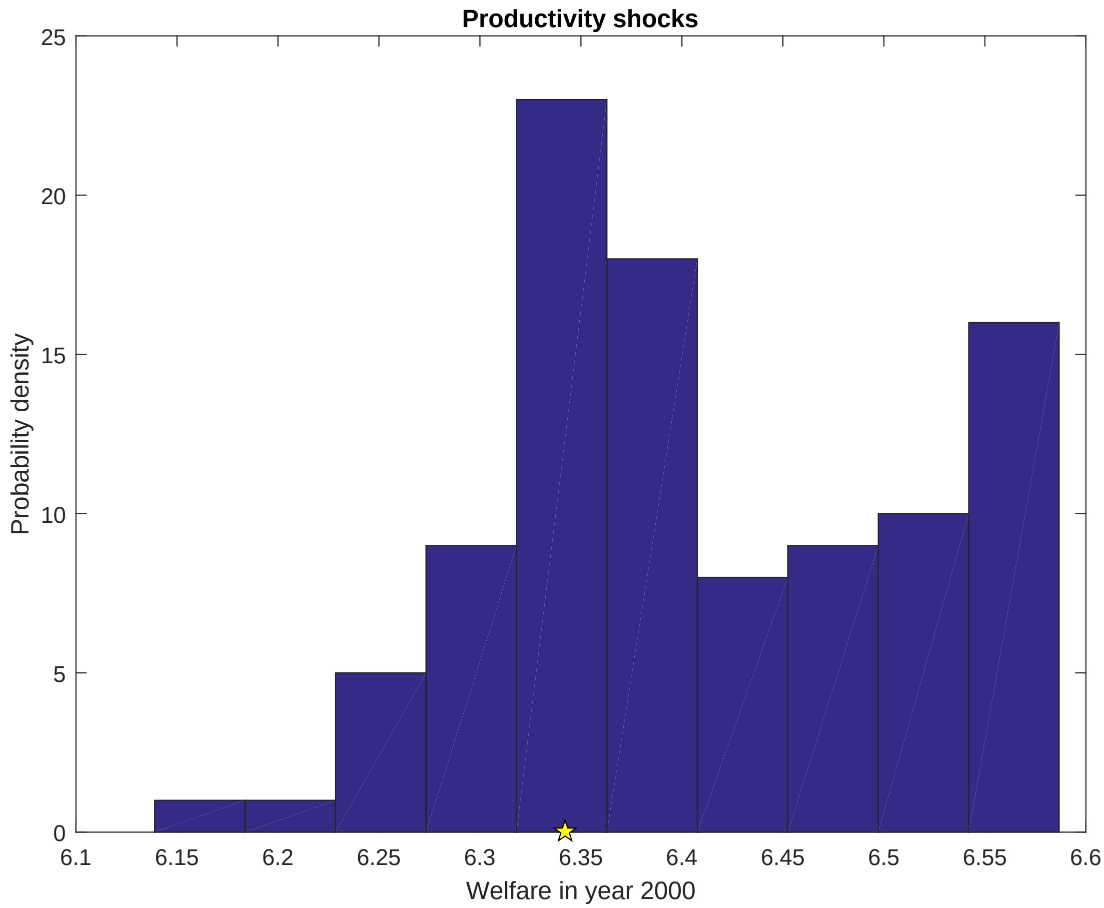
Notes: This figure illustrates the relationship between "robustness" of a cell to historical shocks and its actual year 2000 population. Robustness is measured as the standard deviation of the year 2000 population of a given cell across 200 simulations of random random (drawn, within-the geographically similar clusters of Figure 6, without replacement) productivity shocks $\overline{A}_{it}$ in years 1900 and 1950 (but not 1850 or 2000). For interpretability, we identify each cell with its largest city if the largest city has a population of more than 200,000. Note that cells without such cities are unlabeled, and some cities (e.g. New York) span multiple cells.

Figure 12: The geography of robustness to historical shocks



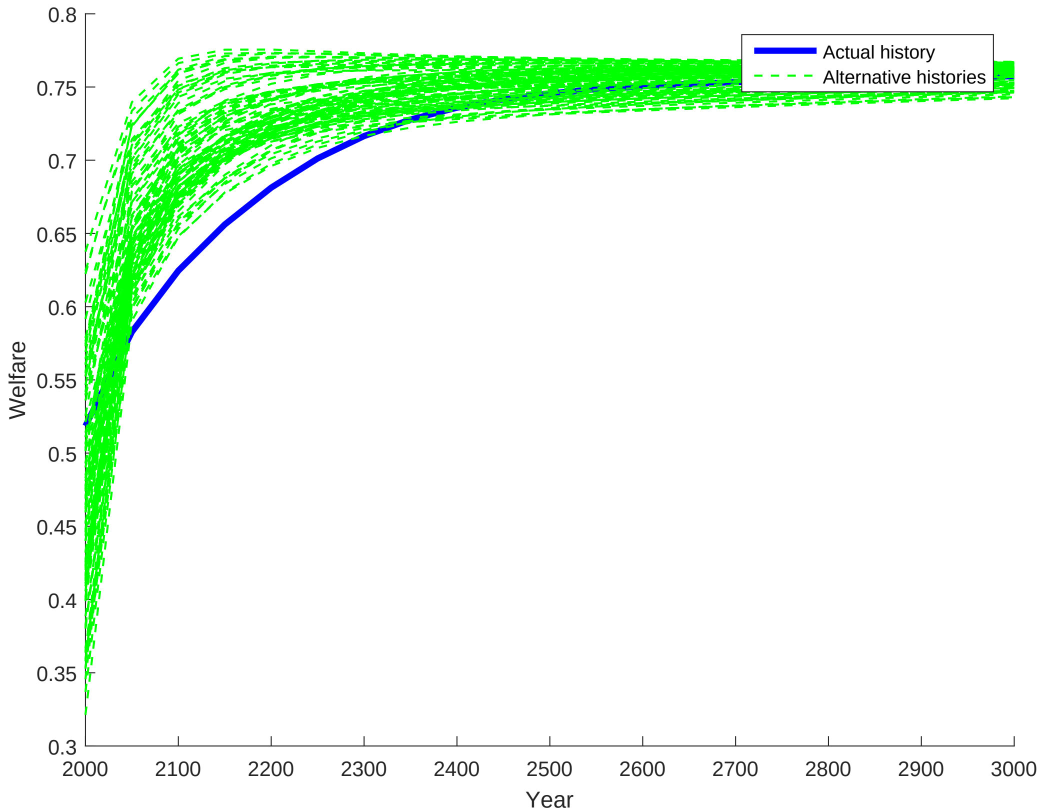
Notes: This figure illustrates the spatial variation in "robustness" to historical shocks, Robustness is measured as the standard deviation of the year 2000 population of a given cell across 200 simulations of random random (drawn, within-the geographically similar clusters of Figure 6, without replacement) productivity shocks $\overline{A}_{it}$ in years 1900 and 1950 (but not 1850 or 2000). Yellow indicates less robustness (i.e. a greater variation in year 2000 population across simulations), whereas blue indicates more robustness (i.e. less variation in year 2000 population across simulations).

Figure 13: The distribution of aggregate welfare in 2000 across 200 simulations of random productivity values in 1900 and 1950



Notes: This figure illustrates how aggregate (population-weighted average) welfare, in logs, in 2000 varies across all 200 simulations of random (drawn, within-the geographically similar clusters of Figure 6, without replacement) productivity shocks $\overline{A}_{it}$ in years 1900 and 1950 (but not 1850 or 2000). The yellow star indicates the factual value from 2000.

Figure 14: The path to steady-states in 200 simulations of random productivity in 1900 and 1950



Notes: This figure illustrates, in green, the level of aggregate (population-weighted average) welfare, in logs, across all 200 simulations of random (drawn, within-the geographically similar clusters of Figure 6, without replacement) productivity shocks $\overline{A}_{it}$ in years 1900 and 1950 (but not 1850 or 2000). The blue line indicates the corresponding path given actual data.

Table 1: Estimating elasticities and spillovers

	(1) Trade orig. FE	(2) Migr. dest. FE
Elasticity of	13.676***	49.821
substitution (σ)	(1.913)	(36.513)
Migration elasticity		11.736***
(θ)		(1.621)
Contemporaneous	0.239***	
productivity spillover (α_1)	(0.010)	
Lagged productivity	0.028***	
spillover (α_2)	(0.003)	
Contemporaneous		-0.341***
amenity spillover (β_1)		(0.018)
Lagged amenity		-0.004*
spillover (β_2)		(0.002)
R-squared	0.432	0.628
Observations	44,408	44,408

Notes: Column (1) reports estimates of the parameter estimates implied by the 2SLS estimation of equation (21). The instruments used are model-implied predictions of the three endogenous variables (the wage, the population level, and the lagged population level), where the model predictions are formed on the basis of geographical characteristics (soil quality, elevation, climate, and water access), observed initial population in 1800, and candidate parameter values (described in Section 3.2) in line with existing work. Column (2) is analogous but for equation (22). The sample includes the years 1850, 1900, 1950 and 2000. Regressions control for each location's own value of initial population and geographical characteristics.

A Proofs

The following three proofs are special cases of Theorem 3 (parts (i) and (ii)) of Allen, Arkolakis, and Li (2015), which we restate here for convenience:

Consider the following system of $N \times K$ system of equations

$$\prod_{h=1}^{K}\left(x_i^h\right)^{\beta_{kh}} = \sum_{j=1}^{K} K_{ij}^k \left[\prod_{h=1}^{H}\left(x_j^h\right)^{\gamma_{kh}}\right],$$

where $\{\beta_{kh,}\gamma_{kh}\}$ are known elasticities and $\left\{K_{ij}^k > 0\right\}$ are known bilateral frictions. Let $\mathbf{B} \equiv [\beta_{kh}]$ and $\mathbf{\Gamma} \equiv [\gamma_{kh}]$ be the $K \times K$ matrices of the known elasticities. Define $\mathbf{A} \equiv \mathbf{\Gamma}\mathbf{B}^{-1}$ and the absolute value (element by element) of $\mathbf{A}$ as $\mathbf{A}^p$. Then there exists a strictly positive set of $\left\{x_i^h > 0\right\}_{i\in\{1,\ldots,N\}}^{h\in\{1,\ldots,K\}}$ and that solution is unique if the spectral radius (i.e. the absolute value of the largest eigenvalue, denote $\rho\left(\cdot\right)$) of $\mathbf{A}^p$ is weakly less than one, i.e. $\rho\left(\mathbf{A}^p\right) \leq 1$.

A.1 Proof of Proposition 1

When trade costs are symmetric, Allen and Arkolakis (2014) show that the origin and destination fixed effects of the gravity trade equation are equal up to scale. That is if $X_{ij} = K_{ij}\gamma_i\delta_j$, $K_{ij} = K_{ji}$, and $\sum_j X_{ij} = \sum_j X_{ji}$, then we have:

$$\gamma_i \propto \delta_i.$$

[15]From equation (3), this implies:

$$\begin{aligned}
w_i^{1-\sigma}A_i^{\sigma-1} &\propto P_i^{\sigma-1}w_iL_i \iff \\
w_i^{1-\sigma}A_i^{\sigma-1} &\propto \left(\frac{w_iu_i}{W_i}\right)^{\sigma-1} w_iL_i \iff \\
w_i &\propto W_i^{\tilde{\sigma}}u_i^{-\tilde{\sigma}}A_i^{\tilde{\sigma}}L_i^{\frac{1}{1-2\sigma}} \iff \\
w_i &\propto W_i^{\tilde{\sigma}}\bar{u}_i^{-\tilde{\sigma}}\bar{A}_i^{\tilde{\sigma}}L_i^{\left(\alpha_1-\beta_1+\frac{1}{1-\sigma}\right)\tilde{\sigma}}\left(L_i^{lag}\right)^{(\alpha_2-\beta_2)\tilde{\sigma}}
\end{aligned}$$

where $\tilde{\sigma} \equiv \frac{\sigma-1}{2\sigma-1}$.

We can use this to simplify our equilibrium equations:

$$\left(W_i^{\tilde{\sigma}}u_i^{-\tilde{\sigma}}A_i^{\tilde{\sigma}}L_i^{\frac{\tilde{\sigma}}{1-\sigma}}\right)^{\sigma}L_i = \sum_j \tau_{ij}^{1-\sigma}A_i^{\sigma-1}u_j^{\sigma-1}W_j^{1-\sigma}\left(W_j^{\tilde{\sigma}}u_j^{-\tilde{\sigma}}A_j^{\tilde{\sigma}}L_j^{\frac{\tilde{\sigma}}{1-\sigma}}\right)^{\sigma}L_j$$

$$\Pi_i^{\theta} = \sum_j \mu_{ij}^{-\theta}W_j^{\theta}$$

[15]The exact scale will be determined by the aggregate labor market clearing condition. However, the scale can be ignored by first solving for the "scaled" labor (i.e. imposing the scalar is equal to one) and then recovering the scale by imposing the labor market clearing condition. Note that this does not affect any of the other equilibrium equations below, as they are all homogeneous of degree 0 with respect to labor.

$$L_i = \sum_j \mu_{ji}^{-\theta} W_i^{\theta} \Pi_j^{-\theta} L_j^{lag}$$

or equivalently:

$$W_i^{\tilde{\sigma}\sigma} L_i^{1+\frac{\sigma}{1-\sigma}\tilde{\sigma}} = \sum_j \tau_{ij}^{1-\sigma} A_i^{\sigma-1-\sigma\tilde{\sigma}} u_i^{\tilde{\sigma}\sigma} u_j^{\sigma-1-\tilde{\sigma}\sigma} A_j^{\tilde{\sigma}\sigma} W_j^{1-\sigma+\sigma\tilde{\sigma}} L_j^{1+\frac{\sigma}{1-\sigma}\tilde{\sigma}}$$

$$\Pi_i^{\theta} = \sum_j \mu_{ij}^{-\theta} W_j^{\theta}$$

$$L_i = \sum_j \mu_{ji}^{-\theta} W_i^{\theta} \Pi_j^{-\theta} L_j^{lag}$$

Let us then use our spillover equations:

$$A_i = \bar{A}_i L_i^{\alpha_1} \left(L_i^{lag}\right)^{\alpha_2}$$

$$u_i = \bar{u}_i L_i^{\beta_1} \left(L_i^{lag}\right)^{\beta_2}$$

to get:

$$\begin{aligned} W_i^{\tilde{\sigma}\sigma} L_i^{1+\frac{\sigma}{\sigma-1}\tilde{\sigma}-\alpha_1(\sigma-1-\sigma\tilde{\sigma})-\beta_1\sigma\tilde{\sigma}} &= \sum_j \tau_{ij}^{1-\sigma} \bar{A}_i^{\sigma-1-\sigma\tilde{\sigma}} \bar{u}_i^{\sigma\tilde{\sigma}} \beta_j^{\sigma-1-\tilde{\sigma}\sigma} \bar{A}_j^{\tilde{\sigma}\sigma} \left(L_i^{lag}\right)^{\alpha_2(\sigma-1-\sigma\tilde{\sigma})+\beta_2\tilde{\sigma}\sigma} \left(L_j^{lag}\right)^{\beta_2(\sigma-1-\tilde{\sigma}\sigma)+} \\ &\quad \times W_j^{1-\sigma+\sigma\tilde{\sigma}} L_j^{1+\frac{\sigma}{\sigma-1}\tilde{\sigma}+\alpha_1(\tilde{\sigma}\sigma)+\beta_1(\sigma-1-\tilde{\sigma}\sigma)} \end{aligned}$$

$$\Pi_i^{\theta} = \sum_j \mu_{ij}^{-\theta} W_j^{\theta}$$

$$L_i = \sum_j \mu_{ji}^{-\theta} W_i^{\theta} \Pi_j^{-\theta} L_j^{lag}.$$

With a little algebra, this simplifies to:

$$\begin{aligned} W_i^{\tilde{\sigma}\sigma} L_i^{\tilde{\sigma}(1-\alpha_1(\sigma-1)-\beta_1\sigma)} &= \sum_j \tau_{ij}^{1-\sigma} \bar{A}_i^{(\sigma-1)\tilde{\sigma}} \bar{u}_i^{\tilde{\sigma}} \beta_j^{(\sigma-1)\tilde{\sigma}} \bar{A}_j^{\tilde{\sigma}\sigma} \left(L_i^{lag}\right)^{\tilde{\sigma}(\alpha_2(\sigma-1)+\beta_2\sigma)} \\ &\quad \times \left(L_j^{lag}\right)^{\tilde{\sigma}(\alpha_2\sigma+\beta_2(\sigma-1))} W_j^{-(\sigma-1)\tilde{\sigma}} L_j^{\tilde{\sigma}(1+\alpha_1\sigma+\beta_1(\sigma-1))} \end{aligned}$$

$$\Pi_i^{\theta} = \sum_j \mu_{ij}^{-\theta} W_j^{\theta}$$

$$L_i W_i^{-\theta} = \sum_j \mu_{ji}^{-\theta} \Pi_j^{-\theta} L_j^{lag}.$$

If we order the endogenous variables as L, W, Π, then the matrix of LHS coefficients becomes:

$$\boldsymbol{B} \equiv \begin{pmatrix} \tilde{\sigma}\left(1-\alpha_1\left(\sigma-1\right)-\beta_1\sigma\right) & \tilde{\sigma}\sigma & 0 \\ 0 & 0 & \theta \\ 1 & -\theta & 0 \end{pmatrix}$$

and the matrix on the RHS coefficients becomes:

$$\boldsymbol{\Gamma} \equiv \begin{pmatrix} \tilde{\sigma}\left(1+\alpha_1\sigma+\beta_1\left(\sigma-1\right)\right) & -\left(\sigma-1\right)\tilde{\sigma} & 0 \\ 0 & \theta & 0 \\ 0 & 0 & -\theta \end{pmatrix}.$$

Hence, we have:

$$\mathbf{A} \equiv \boldsymbol{\Gamma}\mathbf{B}^{-1} = \begin{pmatrix} \frac{\theta-\sigma-\beta_1\theta+\alpha_1\sigma\theta+\beta_1\sigma\theta+1}{\sigma+\theta+\alpha_1\theta-\alpha_1\sigma\theta-\beta_1\sigma\theta} & 0 & \frac{\tilde{\sigma}(2\sigma-1)(\alpha_1+1)}{\sigma+\theta+\alpha_1\theta-\alpha_1\sigma\theta-\beta_1\sigma\theta} \\ \frac{\theta/\tilde{\sigma}}{\sigma+\theta+\alpha_1\theta-\alpha_1\sigma\theta-\beta_1\sigma\theta} & 0 & \frac{-\theta(\alpha_1-\alpha_1\sigma)-\beta_1\sigma+1}{\sigma+\theta+\alpha_1\theta-\alpha_1\sigma\theta-\beta_1\sigma\theta} \\ 0 & -1 & 0 \end{pmatrix}.$$

Note that the spectral radius of the absolute value will be equal to no less than one given the -1 in the third row and second column. Hence the uniqueness condition requires the absolute remainder of the matrix (removing the third row and second column) to feature a spectral radius no greater than one, i.e.:

$$\rho\begin{pmatrix} \left|\frac{\theta(1+\alpha_1\sigma+\beta_1(\sigma-1))-(\sigma-1)}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| & \left|\frac{(\sigma-1)(\alpha_1+1)}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| \\ \left|\frac{\theta/\tilde{\sigma}}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| & \left|\frac{\theta(1-(\sigma-1)\alpha_1-\beta_1\sigma)}{\sigma+\theta(1+(1-\sigma)\alpha_1-\beta_1\sigma)}\right| \end{pmatrix} \leq 1,$$

as required.

A.2 Proof of Proposition 2

The proof proceeds similarly to the proof of Proposition 1. If migration costs are symmetric and we are in the steady state, we have: $\sum_i L_{ij} = \sum_j L_{ji}$, $L_{ij} = M_{ij} g_i d_j$, and $M_{ij} = M_{ji}$. So then it will be the case that:

$$g_i \propto d_i.$$

In our case, this implies:

$$W_i \Pi_i L_i^{\frac{1}{\theta}} = \Omega,$$

which recall is our measure of steady state welfare.

This simplifies our system of equations as follows:

$$W_i^{\tilde{\sigma}\sigma} L_i^{\tilde{\sigma}(1-(\alpha_1+\alpha_2)(\sigma-1)-\sigma(\beta_1+\beta_2))} = \sum_j \tau_{ij}^{1-\sigma} \bar{A}_i^{(\sigma-1)\tilde{\sigma}} \bar{u}_i^{\tilde{\sigma}} u_j^{(\sigma-1)\tilde{\sigma}} \bar{A}_j^{\tilde{\sigma}\sigma} W_j^{-(\sigma-1)\tilde{\sigma}} L_j^{\tilde{\sigma}(1+(\alpha_1+\alpha_2)\sigma+(\beta_1+\beta_2)(\sigma-1))}$$

$$L_i W_i^{-\theta} = \Omega^{-\theta} \sum_j \mu_{ij}^{-\theta} W_j^{\theta}.$$

Let us order the endogenous variables as L, W. Define $\tilde{\alpha} \equiv \alpha_1 + \alpha_2$ and $\tilde{\beta} \equiv \beta_1 + \beta_2$ Then the matrix of LHS coefficients becomes:

$$\boldsymbol{B} \equiv \begin{pmatrix} \tilde{\sigma}\left(1 - \tilde{\alpha}\left(\sigma - 1\right) - \tilde{\beta}\sigma\right) & \tilde{\sigma}\sigma \\ 1 & -\theta \end{pmatrix}$$

and the matrix on the RHS coefficients becomes:

$$\boldsymbol{\Gamma} \equiv \begin{pmatrix} \tilde{\sigma}\left(1 + \tilde{\alpha}\sigma + \tilde{\beta}\left(\sigma - 1\right)\right) & -\left(\sigma - 1\right)\tilde{\sigma} \\ 0 & \theta \end{pmatrix}.$$

Hence, we have:

$$\mathbf{A} \equiv \boldsymbol{\Gamma}\mathbf{B}^{-1} = \begin{pmatrix} \frac{\theta - \sigma - \tilde{\beta}\theta + \tilde{\alpha}\sigma\theta + 1}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} & \frac{-(\sigma-1)(\tilde{\alpha}+1)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \\ \frac{\theta/\tilde{\sigma}}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} & \frac{-\theta\left(\tilde{\alpha}(1-\sigma) - \tilde{\beta}\sigma + 1\right)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{pmatrix}.$$

As a result, the condition for uniqueness is identical to that above, where we simply replace α_1 and β_1 with $\tilde{\alpha} \equiv (\alpha_1 + \alpha_2)$ and $\tilde{\beta} \equiv (\beta_1 + \beta_2)$, as required:

$$\rho\left(\begin{vmatrix} \frac{\theta\left(1 + \tilde{\alpha}\sigma + \tilde{\beta}(\sigma-1)\right) - (\sigma-1)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix} \quad \begin{vmatrix} \frac{(\sigma-1)(\tilde{\alpha}+1)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix} \atop \begin{vmatrix} \frac{\theta/\tilde{\sigma}}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix} \quad \begin{vmatrix} \frac{\theta\left(1 - (\sigma-1)\tilde{\alpha} - \tilde{\beta}\sigma\right)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix}\right) \leq 1.$$

The final part of the proof claims that there exists a geography for which if

$$\rho\left(\begin{vmatrix} \frac{\theta\left(1 + \tilde{\alpha}\sigma + \tilde{\beta}(\sigma-1)\right) - (\sigma-1)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix} \quad \begin{vmatrix} \frac{(\sigma-1)(\tilde{\alpha}+1)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix} \atop \begin{vmatrix} \frac{\theta/\tilde{\sigma}}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix} \quad \begin{vmatrix} \frac{\theta\left(1 - (\sigma-1)\tilde{\alpha} - \tilde{\beta}\sigma\right)}{\sigma + \theta\left(1 + (1-\sigma)\tilde{\alpha} - \tilde{\beta}\sigma\right)} \end{vmatrix}\right) > 1,$$

then there exist multiple equilibria. For readability, we present it this result as a general theorem, under which our model clearly falls:

Theorem 1. *Consider the following mathematical system:*

$$x_{i,1} = \lambda_1 \sum_{j=1}^{N} K_{ij,1} x_{j,1}^{a_{11}} x_{j,2}^{a_{12}} \tag{23}$$

$$x_{i,2} = \lambda_2 \sum_{j=1}^{N} K_{ij,2} x_{j,1}^{a_{21}} x_{j,2}^{a_{22}}, \tag{24}$$

where $\{K_{ij,k}\}_{i,j\in\{1,\ldots,N\}}^{l\in\{1,2\}}$ *are the "kernels" of (exogenous) bilateral frictions,* $\{a_{lk}\}_{l,k\in\{1,2\}}$ *are (exogenous) elasticities,* $\{x_{i,k}\}_{i\in\{1,\ldots,N\}}^{k\in\{1,2\}}$ *are (endogenous) strictly positive vectors and* $\{\lambda_k\}_{k\in\{1,2\}}$ *are either endogenous scalars determined by additional constraints or are exogenous. If the* *spectral radius of the* 2×2 *matrix* $\mathbf{A}^p \equiv [|a_{kl}|]$ *is greater than one, then there exists kernels* $\{K_{ij,k}\}_{i,j\in\{1,\ldots,N\}}^{l\in\{1,2\}}$ *such that there are multiple solutions to equations (23) and (24).*

Proof. The proof proceeds by construction. We begin by performing two transformations of the problem that simplifies the setup. First, we absorb the scalars into the endogenous variables. To do so, define $y_{i,k} = \left(\lambda_1^{d_{k,1}} \lambda_2^{d_{k,2}}\right) x_{i,k}$, where $\mathbf{D} = [d_{kl}] \equiv -\left(\mathbf{I} - \mathbf{A}\right)^{-1}$. Note that this is well defined as long as the spectral radius of $\mathbf{A}$ is not equal to one. It is straightforward to then show that the following equations:

$$y_{i,1} = \sum_j K_{ij,1} y_{j,1}^{a_{11}} y_{j,2}^{a_{12}}$$
$$y_{i,2} = \sum_j K_{ij,2} y_{j,1}^{a_{21}} y_{j,2}^{a_{22}}$$

are equivalent to equations (23) and (24). To see this, substitute in the definition of $y_{i,k}$, yielding:

$$\left(\lambda_1^{d_{11}} \lambda_2^{d_{12}}\right) x_{i,1} = \sum_j K_{ij,1} x_{j,1}^{a_{11}} \left(\lambda_1^{d_{11}} \lambda_2^{d_{12}}\right)^{a_{11}} x_{j,2}^{a_{12}} \left(\lambda_1^{d_{21}} \lambda_2^{d_{22}}\right)^{a_{12}}$$
$$\left(\lambda_1^{d_{21}} \lambda_2^{d_{22}}\right) x_{i,2} = \sum_j K_{ij,2} y_{j,1}^{a_{21}} \left(\lambda_1^{d_{11}} \lambda_2^{d_{12}}\right)^{a_{21}} y_{j,2}^{a_{22}} \left(\lambda_1^{d_{21}} \lambda_2^{d_{22}}\right)^{a_{22}}$$

which rearranging yields:

$$x_{i,1} = \lambda_1^{-d_{11}+a_{11}d_{11}+a_{12}d_{21}} \lambda_2^{-d_{12}+a_{11}d_{12}+a_{12}d_{22}} \sum_j K_{ij,1} x_{j,1}^{a_{11}} x_{j,2}^{a_{12}}$$
$$x_{i,2} = \lambda_1^{-d_{21}+a_{21}d_{11}+a_{22}d_{21}} \lambda_2^{-d_{22}+a_{21}d_{12}+a_{22}d_{22}} \sum_j K_{ij,2} y_{j,1}^{a_{21}} y_{j,2}^{a_{22}}$$

Note that the lambda equations can be written as:

$$\begin{aligned} \exp\left(\left(-\mathbf{D} + \mathbf{AD}\right) \ln \boldsymbol{\lambda}\right) &= \exp\left(\left(-\left(\mathbf{I} - \mathbf{A}\right) \mathbf{D}\right) \ln \boldsymbol{\lambda}\right) \iff \\ &= \exp\left(\left(\left(\mathbf{I} - \mathbf{A}\right) \left(\mathbf{I} - \mathbf{A}\right)^{-1}\right) \ln \boldsymbol{\lambda}\right) \iff \\ &= \exp\left(\mathbf{I} \ln \boldsymbol{\lambda}\right) \iff \\ &= \lambda, \end{aligned}$$

as claimed.

The second transformation is closely related to the "exact hat" algebra pioneered by Dekle, Eaton, and Kortum (2008) in the field of trade and considers a "normalized" system of equations around an observed equilibrium. Suppose we observe a steady state solution

$\{y_{i,k}\}_{i\in S,k\in\{1,.2\}}$ that satisfies:

$$y_{i,1} = \sum_j K_{ij,1} y_{j,1}^{a_{11}} y_{j,2}^{a_{12}}$$
$$y_{i,2} = \sum_j K_{ij,2} y_{j,1}^{a_{21}} y_{j,2}^{a_{22}}$$

We are interested in knowing whether there exists a different steady state solution $\{x_{i,k}\}_{i\in S,k\in\{1,.2\}}$ that also satisfies the same equations: [Note that these x's are different than the $x's$ with the scalar above]

$$x_{i,1} = \sum_j K_{ij,1} x_{j,1}^{a_{11}} x_{j,2}^{a_{12}}$$
$$x_{i,2} = \sum_j K_{ij,2} x_{j,1}^{a_{21}} x_{j,2}^{a_{22}}$$

Define $z_{i,k} \equiv \frac{x_{i,k}}{y_{i,k}}$ and note that the previous equations can be written as:

$$z_{i,1} = \sum_j F_{ij,1} z_{j,1}^{a_{11}} z_{j,2}^{a_{12}} \tag{25}$$
$$z_{i,2} = \sum_j F_{ij,2} z_{j,1}^{a_{21}} z_{j,2}^{a_{22}}, \tag{26}$$

where $F_{ij,k} \equiv \left(\frac{K_{ij,k}}{y_{i,k}} y_{j,1}^{a_{k1}} y_{j,2}^{a_{k2}}\right)$. By construction, note that we have $z_{i,k} = 1$ as a solution to this system of equations. Moreover, the matrices $\mathbf{F}_k$ are stochastic, i.e.:

$$\sum_j F_{ij,k} = 1 \ \forall i \in S, k \in \{1,2\}.$$

In what follows, we will search for stochastic matrices $\mathbf{F}_k$ that have two solutions: one in which $z_{i,k} = 1$ for all $i \in \{1,..,N\}$ and $k \in \{1,2\}$ and another in which there exists a $z_{i,k} \neq 1$.

It turns out to do this requires $N = 4$. Choose any $m_k < 1 < M_k$ for $k \in \{1,2\}$. Then we will construct a set of kernels that have the following solution:

$$\begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{2,1} & z_{2,2} \\ z_{3,1} & z_{3,2} \\ z_{4,1} & z_{4,2} \end{pmatrix} = \begin{pmatrix} \tilde{m}_1^A & \tilde{m}_2^A \\ \tilde{m}_1^B & \tilde{m}_2^B \\ \tilde{m}_1^C & \tilde{m}_2^C \\ \tilde{m}_1^D & \tilde{m}_2^D \end{pmatrix} = \begin{pmatrix} m_1^{\mathbf{1}\{a_{11}>0\}} M_1^{\mathbf{1}\{a_{11}\leq 0\}}; & m_2^{\mathbf{1}\{a_{12}>0\}} M_2^{\mathbf{1}\{a_{12}\leq 0\}} \\ m_1^{\mathbf{1}\{a_{21}>0\}} M_1^{\mathbf{1}\{a_{21}\leq 0\}}; & m_2^{\mathbf{1}\{a_{22}>0\}} M_2^{\mathbf{1}\{a_{22}\leq 0\}} \\ m_1^{\mathbf{1}\{a_{11}\leq 0\}} M_1^{\mathbf{1}\{a_{11}>0\}}; & m_2^{\mathbf{1}\{a_{12}\leq 0\}} M_2^{\mathbf{1}\{a_{12}>0\}} \\ m_1^{\mathbf{1}\{a_{21}\leq 0\}} M_1^{\mathbf{1}\{a_{21}>0\}}; & m_2^{\mathbf{1}\{a_{22}\leq 0\}} M_2^{\mathbf{1}\{a_{22}>0\}} \end{pmatrix}. \tag{27}$$

Before constructing the kernel, we note the following helpful properties.

First, define $\ln \mathbf{m} \equiv \begin{pmatrix} \ln m_1 \\ \ln m_2 \end{pmatrix}$, $\ln \mathbf{M} \equiv \begin{pmatrix} \ln M_1 \\ \ln M_2 \end{pmatrix}$, and the indicator matrix

$$\mathbf{P} \equiv \begin{pmatrix} \mathbf{1}\{a_{11} > 0\} & \mathbf{1}\{a_{12} > 0\} \\ \mathbf{1}\{a_{21} > 0\} & \mathbf{1}\{a_{22} > 0\} \end{pmatrix}$$

(for "positive"); and $\mathbf{1} \equiv \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Then we can write the bounds as follows:

$$\begin{aligned}
(\mathbf{A} \circ \mathbf{P}) \ln \mathbf{m} + (\mathbf{A} \circ (\mathbf{1} - \mathbf{P})) \ln \mathbf{M} \leq \ln \mathbf{m} \leq \ln \mathbf{M} \leq (\mathbf{A} \circ (\mathbf{1} - \mathbf{P})) \ln \mathbf{m} + (\mathbf{A} \circ \mathbf{P}) \ln \mathbf{M} &\iff \\
(\mathbf{A} \circ \mathbf{P}) \ln \mathbf{m} + (\mathbf{A} - (\mathbf{A} \circ \mathbf{P})) \ln \mathbf{M} \leq \ln \mathbf{m} \leq \ln \mathbf{M} \leq (\mathbf{A} - (\mathbf{A} \circ \mathbf{P})) \ln \mathbf{m} + (\mathbf{A} \circ \mathbf{P}) \ln \mathbf{M} &\iff \\
\mathbf{A} \ln \mathbf{M} - (\mathbf{A} \circ \mathbf{P}) (\ln \mathbf{M} - \ln \mathbf{m}) \leq \ln \mathbf{m} \leq \ln \mathbf{M} \leq \mathbf{A} \ln \mathbf{m} + (\mathbf{A} \circ \mathbf{P}) (\ln \mathbf{M} - \ln \mathbf{m}) &\iff \\
\ln \mathbf{B} - (\mathbf{A} \circ \mathbf{P}) (\ln \mathbf{M} - \ln \mathbf{m}) \leq \ln \mathbf{m} \leq \ln \mathbf{M} \leq \ln \mathbf{b} + (\mathbf{A} \circ \mathbf{P}) (\ln \mathbf{M} - \ln \mathbf{m}) &\iff \\
\ln \mathbf{B} - \ln \mathbf{D} \leq \ln \mathbf{m} \leq \ln \mathbf{M} \leq \ln \mathbf{b} + \ln \mathbf{D}, &
\end{aligned} \tag{28}$$

where $\ln \mathbf{D} \equiv (\mathbf{A} \circ \mathbf{P}) (\ln \mathbf{M} - \ln \mathbf{m}) = \begin{pmatrix} \ln \left(\frac{M_1}{m_1}\right)^{a_{11}\mathbf{1}\{a_{11}>0\}} & \left(\frac{M_2}{m_2}\right)^{a_{12}\mathbf{1}\{a_{12}>0\}} \\ \ln \left(\frac{M_1}{m_1}\right)^{a_{21}\mathbf{1}\{a_{21}>0\}} & \left(\frac{M_2}{m_2}\right)^{a_{22}\mathbf{1}\{a_{22}>0\}} \end{pmatrix}$ and $D_k \equiv \exp\left((\ln \mathbf{D})_k\right)$.

Second, we note the existence and uniqueness of weights that can be used to relate the $\tilde{m}_k^n$ $(n \in \{A, B, C, D\})$ variables to other endogenous objects. In what follows, we define those weights for $\tilde{m}_k^A$, but the corresponding results also hold for $\tilde{m}_k^B$, $\tilde{m}_k^C$, and $\tilde{m}_k^D$. Since $m_k \leq \tilde{m}_k^A \leq M_k$, then there exists a weight $C_k^A \in [0, 1]$ such that:

$$\tilde{m}_k^A = C_k^A m_k + \left(1 - C_k^A\right) M_k$$

and there exists a weight $c_k^A \in [0, 1]$ such that:

$$\begin{aligned}
\ln \tilde{m}_k^A &= c_k^A \ln m_k + \left(1 - c_k^A\right) \ln M_k \iff \\
\tilde{m}_k^A &= m_k^{c_k^A} M_k^{1 - c_k^A} \iff \\
\tilde{m}_k^A &= M_k \left(\frac{M_k}{m_k}\right)^{-c_k^A}
\end{aligned} \tag{29}$$

or conversely:

$$\tilde{m}_k^A = m_k \left(\frac{M_k}{m_k}\right)^{1 - c_k^A} \tag{30}$$

Note that because $\tilde{m}_k^A = M_k \left(\frac{M_k}{m_k}\right)^{-c_k^A}$ from equation (29) we can write:

$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = \left(M_1 \left(\frac{M_1}{m_1}\right)^{-c_1^A}\right)^{a_{11}} \left(M_2 \left(\frac{M_2}{m_2}\right)^{-c_2^A}\right)^{a_{12}} \iff$$
$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = M_1^{a_{11}} M_2^{a_{12}} \left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{-c_1^A} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{-c_2^A} \iff$$
$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = \frac{B_1}{D_1} D_1 \left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{-c_1^A} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{-c_2^A} \iff$$
$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = \frac{B_1}{D_1} \left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-c_1^A} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-c_2^A} \tag{31}$$

Similarly:

$$\left(\tilde{m}_1^A\right)^{a_{21}} \left(\tilde{m}_2^A\right)^{a_{22}} = \frac{B_2}{D_2} \left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\mathbf{1}\{a_{21}>0\}-c_1^A} \left(\left(\frac{M_2}{m_2}\right)^{a_{22}}\right)^{\mathbf{1}\{a_{22}>0\}-c_2^A} \tag{32}$$

Similarly, because $\tilde{m}_k^A = m_k \left(\frac{M_k}{m_k}\right)^{1-c_k^A}$ from equation (30) we can write::

$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = \left(m_1 \left(\frac{M_1}{m_1}\right)^{\left(1-c_1^A\right)}\right)^{a_{11}} \left(m_2 \left(\frac{M_2}{m_2}\right)^{\left(1-c_2^A\right)}\right)^{a_{12}} \iff$$
$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = m_1^{a_{11}} m_2^{a_{12}} \left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\left(1-c_1^A\right)} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\left(1-c_2^A\right)} \iff$$
$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = b_1 D_1 \frac{\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\left(1-c_1^A\right)} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\left(1-c_2^A\right)}}{D_1} \iff$$
$$\left(\tilde{m}_1^A\right)^{a_{11}} \left(\tilde{m}_2^A\right)^{a_{12}} = b_1 D_1 \left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\left(1-c_1^A\right)-\mathbf{1}\{a_{11}>0\}} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\left(1-c_2^A\right)-\mathbf{1}\{a_{12}>0\}} \tag{33}$$

Similarly,:

$$\left(\tilde{m}_1^A\right)^{a_{21}} \left(\tilde{m}_2^A\right)^{a_{22}} = b_2 D_2 \left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\left(1-c_1^A\right)-\mathbf{1}\{a_{11}>0\}} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\left(1-c_2^A\right)-\mathbf{1}\{a_{12}>0\}} \tag{34}$$

As a result, the system of equations (25) and (26) become:

$$\tilde{m}_1^A = F_{11,1}\frac{B_1}{D_1} + F_{12,1}\frac{B_1}{D_1}\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-c_1^B} \left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-c_2^B} + F_{13,1} b_1 D_1 + F_{14,1} b_1 D_1 \left(\left(\frac{M_1}{m_1}\right.\right.$$

$$\tilde{m}_1^B = F_{21,1}\frac{B_1}{D_1} + F_{22,1}\frac{B_1}{D_1}\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-c_1^B}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-c_2^B} + F_{23,1}b_1D_1 + F_{24,1}b_1D_1\left(\left(\frac{M_1}{m_1}\right)\right.$$

$$\tilde{m}_1^C = F_{31,1}\frac{B_1}{D_1} + F_{32,1}\frac{B_1}{D_1}\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-c_1^B}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-c_2^B} + F_{33,1}b_1D_1 + F_{34,1}b_1D_1\left(\left(\frac{M_1}{m_1}\right)\right.$$

$$\tilde{m}_1^D = F_{41,1}\frac{B_1}{D_1} + F_{42,1}\frac{B_1}{D_1}\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-c_1^B}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-c_2^B} + F_{43,1}b_1D_1 + F_{44,1}b_1D_1\left(\left(\frac{M_1}{m_1}\right)\right.$$

$$\tilde{m}_2^A = F_{11,2}\frac{B_2}{D_2}\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\mathbf{1}\{a_{21}>0\}-c_1^A}\left(\left(\frac{M_2}{m_2}\right)^{a_{22}}\right)^{\mathbf{1}\{a_{22}>0\}-c_2^A} + F_{12,2}\frac{B_2}{D_2} + F_{13,2}b_2D_2\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\left(1-c_1^C\right)-\mathbf{1}}$$

$$\tilde{m}_2^B = F_{21,2}\frac{B_2}{D_2}\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\mathbf{1}\{a_{21}>0\}-c_1^A}\left(\left(\frac{M_2}{m_2}\right)^{a_{22}}\right)^{\mathbf{1}\{a_{22}>0\}-c_2^A} + F_{22,2}\frac{B_2}{D_2} + F_{23,2}b_2D_2\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\left(1-c_1^C\right)-1}$$

$$\tilde{m}_2^C = F_{31,2}\frac{B_2}{D_2}\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\mathbf{1}\{a_{21}>0\}-c_1^A}\left(\left(\frac{M_2}{m_2}\right)^{a_{22}}\right)^{\mathbf{1}\{a_{22}>0\}-c_2^A} + F_{32,2}\frac{B_2}{D_2} + F_{33,2}b_2D_2\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\left(1-c_1^C\right)-\mathbf{1}}$$

$$\tilde{m}_2^D = F_{41,2}\frac{B_2}{D_2}\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\mathbf{1}\{a_{21}>0\}-c_1^A}\left(\left(\frac{M_2}{m_2}\right)^{a_{22}}\right)^{\mathbf{1}\{a_{22}>0\}-c_2^A} + F_{42,2}\frac{B_2}{D_2} + F_{43,2}b_2D_2\left(\left(\frac{M_1}{m_1}\right)^{a_{21}}\right)^{\left(1-c_1^C\right)-1}$$

We now move on to constructing the kernel. Note that given the inequality (28), there exists constants $P_k \in (0,1)$ and $Q_k \in (0,1)$ such that:

$$m_k = P_k\frac{B_k}{D_k} + (1-P_k)\,b_kD_k \tag{35}$$

$$M_k = Q_k\frac{B_k}{D_k} + (1-Q_k)\,b_kD_k \tag{36}$$

Combining the last two results (where again we focus on $\tilde{m}_k^A$, but the following holds for $\tilde{m}_k^B$, $\tilde{m}_k^C$, and $\tilde{m}_k^D$ as well) note that:

$$\tilde{m}_k^A = C_k^A m_k + \left(1-C_k^A\right)M_k$$

and

$$m_k = P_k\frac{B_k}{D_k} + (1-P_k)\,b_kD_k$$

$$M_k = Q_k\frac{B_k}{D_k} + (1-Q_k)\,b_kD_k$$

so that:

$$\tilde{m}_k^A = C_k^A \left(P_k \frac{B_k}{D_k} + (1 - P_k)\, b_k D_k \right) + \left(1 - C_k^A\right) \left(Q_k \frac{B_k}{D_k} + (1 - Q_k)\, b_k D_k \right) \iff$$
$$\tilde{m}_k^A = \left(\left(C_k^A P_k + \left(1 - C_k^A\right) Q_k\right)\right) \frac{B_k}{D_k} + \left(C_k^A (1 - P_k) + \left(1 - C_k^A\right)(1 - Q_k)\right) b_k D_k$$

Moreover, note that:

$$\begin{aligned} \left(\left(C_k^A P_k + \left(1 - C_k^A\right) Q_k\right)\right) + \left(C_k^A (1 - P_k) + \left(1 - C_k^A\right)(1 - Q_k)\right) &= C_k^A P_k + \left(1 - C_k^A\right) Q_k + C_k^A (1 - P_k) + (\\ &= C_k^A \left(P_k + (1 - P_k)\right) + \left(1 - C_k^A\right)(Q_k + (1 \\ &= C_k^A + \left(1 - C_k^A\right) \\ &= 1 \end{aligned}$$

This tells us that $\tilde{m}_k^A$ can also be written as weighted average of $\frac{B_k}{D_k}$ and $b_k D_k$, with the weight being $\omega_k^A \equiv \left(\left(C_k^A P_k + \left(1 - C_k^A\right) Q_k\right)\right)$.

With all of these properties established, we have enough information to define our kernels:

$$\mathbf{F}_1 = \begin{pmatrix} \omega_1^A & 0 & 1 - \omega_1^A & 0 \\ \omega_1^B & 0 & 1 - \omega_1^B & 0 \\ \omega_1^C & 0 & 1 - \omega_1^C & 0 \\ \omega_1^D & 0 & 1 - \omega_1^D & 0 \end{pmatrix}$$
$$\mathbf{F}_2 = \begin{pmatrix} 0 & \omega_2^A & 0 & 1 - \omega_2^A \\ 0 & \omega_2^B & 0 & 1 - \omega_2^B \\ 0 & \omega_2^C & 0 & 1 - \omega_2^C \\ 0 & \omega_2^D & 0 & 1 - \omega_2^D \end{pmatrix}.$$

Note that the $z_{i,k} = 1$ for all $i \in \{1, .., 4\}$ and $k \in \{1, 2\}$ trivially satisfies the equilibrium system. But it is also straightforward to confirm that the proposed solution (27) is also an equilibrium. This is because every equation has a term of $\left(\frac{B_k}{D_k}\right)$ and $(b_k D_k)$, which we know every endogenous variable is a weighted average of (see equations (35) and (36)).

Finally, we mention that there are many geographies that deliver this multiplicity for two reasons. First, the argument above holds for any choice of $m_k < 1 < M_k$. Second, it is straightforward to show that perturbations of the above kernel also generate multiple equilibria. Suppose we considered the perturbed system of equations:

$$\mathbf{F}_1 = \begin{pmatrix} \omega_1^A - \kappa\varepsilon & \delta\varepsilon & 1 - \omega_1^A - (1 - \kappa)\,\varepsilon & (1 - \delta)\,\varepsilon \\ \omega_1^B & 0 & 1 - \omega_1^B & 0 \\ \omega_1^C & 0 & 1 - \omega_1^C & 0 \\ \omega_1^D & 0 & 1 - \omega_1^D & 0 \end{pmatrix},$$

where $\varepsilon > 0$, $\kappa \in [0, 1]$ and $\delta \in [0, 1]$. The only restriction we place is that $\omega_1^A - \kappa\varepsilon > 0 \iff \kappa\varepsilon < \omega_1^A$ and $\left(1 - \omega_1^A - (1 - \kappa)\,\varepsilon\right) > 0 \iff \varepsilon\,(1 - \kappa) < 1 - \omega_1^A$. Note that both of these equations will hold for sufficiently small ε, as $\omega_k^A = \left(C_k^l P_k + \left(1 - C_k^l\right) Q_k\right)$ and $P_k \in (0, 1)$ and $Q_k \in (0, 1)$. In what follows, we show for any choice of $\varepsilon > 0$ (that is sufficiently small to satisfy these inequalities) and any choice of $\delta \in [0,1]$, there exists a $\kappa \in [0,1]$ that ensures the multiplicity still holds.

Then the relevant equation becomes:

$$\tilde{m}_1^A = \omega_1^A \frac{B_1}{D_1} - \kappa\varepsilon\left(\frac{B_1}{D_1}\right) + \delta\varepsilon\left(\frac{B_1}{D_1}\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-c_1^B}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-c_2^B}\right) +$$

$$\kappa\varepsilon\frac{B_1}{D_1} + (1-\varepsilon)\, b_1 D_1 = \delta\varepsilon\frac{B_1}{D_1}\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-\mathbf{1}\{a_{21}>0\}}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-\mathbf{1}\{a_{22}>0\}} + (1-\delta)\,\varepsilon b_1 D_1\,($$

$$\kappa\varepsilon\frac{B_1}{D_1} + (1-\kappa)\,\varepsilon b_1 D_1 = \delta\varepsilon\left(\frac{B_1}{D_1}\right)G + (1-\delta)\,\varepsilon\frac{1}{G}\,(b_1 D_1) \iff$$

$$\kappa\frac{B_1}{D_1} + (1-\kappa)\, b_1 D_1 = \delta\left(\frac{B_1}{D_1}\right)G + (1-\delta)\,\frac{1}{G}\,(b_1 D_1)$$

where $G \equiv \left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-\mathbf{1}\{a_{21}>0\}}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-\mathbf{1}\{a_{22}>0\}}$. Recall that

$$\frac{B_1}{D_1} = \frac{M_1^{a_{11}} M_2^{a_{12}}}{\left(\frac{M_1}{m_1}\right)^{a_{11}\mathbf{1}\{a_{11}>0\}}\left(\frac{M_2}{m_2}\right)^{a_{12}\mathbf{1}\{a_{12}>0\}}}$$

$$b_1 D_1 = m_1^{a_{11}} m_2^{a_{12}}\left(\frac{M_1}{m_1}\right)^{a_{11}\mathbf{1}\{a_{11}>0\}}\left(\frac{M_2}{m_2}\right)^{a_{12}\mathbf{1}\{a_{12}>0\}},$$

i.e. B_1/D_1 is always the lowest that can be achieved given the signs of the exponents, and $b_1 D_1$ is the highest that can be achieved given the sign of the exponents. As a result, we have:

$$G\left(\frac{B_1}{D_1}\right) = \left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-\mathbf{1}\{a_{21}>0\}}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-\mathbf{1}\{a_{22}>0\}} \times \frac{M_1^{a_{11}} M_2^{a_{12}}}{\left(\frac{M_1}{m_1}\right)^{a_{11}\mathbf{1}\{a_{11}>0\}}\left(\frac{M_2}{m_2}\right)^{a_{12}\mathbf{1}\{a_{12}>0\}}}$$

$$G\left(\frac{B_1}{D_1}\right) = \frac{M_1^{a_{11}} M_2^{a_{12}}}{\left(\frac{M_1}{m_1}\right)^{a_{11}\mathbf{1}\{a_{21}>0\}}\left(\frac{M_2}{m_2}\right)^{a_{12}\mathbf{1}\{a_{22}>0\}}}$$

$$\frac{1}{G}\,(b_1 D_1) = \frac{m_1^{a_{11}} m_2^{a_{12}}\left(\frac{M_1}{m_1}\right)^{a_{11}\mathbf{1}\{a_{11}>0\}}\left(\frac{M_2}{m_2}\right)^{a_{12}\mathbf{1}\{a_{12}>0\}}}{\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{11}>0\}-\mathbf{1}\{a_{21}>0\}}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{12}>0\}-\mathbf{1}\{a_{22}>0\}}} \iff$$

$$\frac{1}{G}\,(b_1 D_1) = m_1^{a_{11}} m_2^{a_{12}}\left(\left(\frac{M_1}{m_1}\right)^{a_{11}}\right)^{\mathbf{1}\{a_{21}>0\}}\left(\left(\frac{M_2}{m_2}\right)^{a_{12}}\right)^{\mathbf{1}\{a_{22}>0\}}$$

Together this implies that:

$$\frac{B_1}{D_1} \le G\left(\frac{B_1}{D_1}\right), \frac{1}{G}\,(b_1 D_1) \le b_1 D_1$$

since B_1/D_1 and b_1D_1 are designed to be the lowest and highest (respectively) given the signs of the exponents. As a result, there exists constants (weights) $\lambda_1 \in [0,1]$ and $\lambda_2 \in [0,1]$ such that:

$$G\left(\frac{B_1}{D_1}\right) = \lambda_1 \frac{B_1}{D_1} + (1-\lambda_1)\, b_1 D_1$$
$$\frac{1}{G}(b_1 D_1) = \lambda_2 \frac{B_1}{D_1} + (1-\lambda_2)\, b_1 D_1$$

We now return to the above equation:

$$\begin{aligned}
&\kappa \frac{B_1}{D_1} + (1-\kappa)\, b_1 D_1 = \delta\left(\frac{B_1}{D_1}\right) G + (1-\delta)\frac{1}{G}(b_1 D_1) \iff \\
&\kappa \frac{B_1}{D_1} + (1-\kappa)\, b_1 D_1 = \delta\left(\lambda_1 \frac{B_1}{D_1} + (1-\lambda_1)\, b_1 D_1\right) + (1-\delta)\left(\lambda_2 \frac{B_1}{D_1} + (1-\lambda_2)\, b_1 D_1\right) \iff \\
&\kappa \frac{B_1}{D_1} + (1-\kappa)\, b_1 D_1 = (\delta\lambda_1 + (1-\delta)\,\lambda_2)\frac{B_1}{D_1} + (\delta(1-\lambda_1) + (1-\delta)(1-\lambda_2))\, b_1 D_1 \qquad (37)
\end{aligned}$$

Choose $\kappa \equiv \delta\lambda_1 + (1-\delta)\,\lambda_2$. Then

$$\begin{aligned}
1-\kappa &= 1 - \delta\lambda_1 - (1-\delta)\,\lambda_2 \iff \\
1-\kappa &= 1 + \delta - \delta - \delta\lambda_1 - (1-\delta)\,\lambda_2 \iff \\
1-\kappa &= \delta(1-\lambda_1) + (1-\delta)(1-\lambda_2)\,,
\end{aligned}$$

so that equation (37) holds. Hence, for any choice of δ, we can find a κ that ensures the equilibrium still holds. Note that there is nothing in this argument that is particular to $\tilde{m}_1^A$. As a result, we can construct examples of multiple equilibria of the form:

$$\mathbf{F}_1 = \begin{pmatrix}
\omega_1^A - \kappa_1^A \varepsilon_1^A; & \delta_1^A \varepsilon_1^A; & 1-\omega_1^A - \left(1-\kappa_1^A\right)\varepsilon_1^A; & \left(1-\delta_1^A\right)\varepsilon_1^A \\
\omega_1^B - \kappa_1^B \varepsilon_1^B; & \delta_1^B \varepsilon_1^B; & 1-\omega_1^B - \left(1-\kappa_1^B\right)\varepsilon_1^B; & \left(1-\delta_1^B\right)\varepsilon_1^B \\
\omega_1^C - \kappa_1^C \varepsilon_1^C; & \delta_1^C \varepsilon_1^C; & 1-\omega_1^C - \left(1-\kappa_1^C\right)\varepsilon_1^C; & \left(1-\delta_1^C\right)\varepsilon_1^C \\
\omega_1^D - \kappa_1^D \varepsilon_1^D; & \delta_1^D \varepsilon_1^D; & 1-\omega_1^D - \left(1-\kappa_1^D\right)\varepsilon_1^D; & \left(1-\delta_1^D\right)\varepsilon_1^D
\end{pmatrix},$$

$$\mathbf{F}_2 = \begin{pmatrix}
\delta_2^A \varepsilon_2^A; & \omega_2^A - \kappa_2^A \varepsilon_2^A; & \left(1-\delta_2^A\right)\varepsilon_2^A & 1-\omega_2^A - \left(1-\kappa_2^A\right)\varepsilon_2^A \\
\delta_2^B \varepsilon_2^B; & \omega_2^B - \kappa_2^B \varepsilon_2^B; & \left(1-\delta_2^B\right)\varepsilon_2^B & 1-\omega_2^B - \left(1-\kappa_2^B\right)\varepsilon_2^B \\
\delta_2^C \varepsilon_2^C; & \omega_2^C - \kappa_2^C \varepsilon_2^C; & \left(1-\delta_2^C\right)\varepsilon_2^C & 1-\omega_2^C - \left(1-\kappa_2^C\right)\varepsilon_2^C \\
\delta_2^D \varepsilon_2^D; & \omega_2^D - \kappa_2^D \varepsilon_2^D; & \left(1-\delta_2^D\right)\varepsilon_2^D & 1-\omega_2^D - \left(1-\kappa_2^D\right)\varepsilon_2^D
\end{pmatrix}$$

for many different chosen values of $\left\{\varepsilon_k^l\right\}$ and $\left\{\delta_k^l\right\}$. □

A.3 Proof of Proposition 3

As a reminder, the steady state system of equations we would like to examine can be written as:

$$L_i W_i^{-\theta} = \Omega^{-\theta} \sum_j M_{ij} W_j^{\theta}$$

$$W_i^{\tilde{\sigma}\sigma}\left(L_i^{\frac{1}{\rho}}\right)=\sum_{j\in S}T_{ij}W_j^{-(\sigma-1)\tilde{\sigma}}\left(L_j^{\frac{1}{\rho}}\right)^a$$

where $T_{ij}\equiv\tau_{ij}^{1-\sigma}\bar{A}_i^{(\sigma-1)\tilde{\sigma}}\bar{A}_j^{\tilde{\sigma}\sigma}\bar{u}_i^{\tilde{\sigma}}\bar{u}_j^{(\sigma-1)\tilde{\sigma}}$ and $M_{ij}\equiv\mu_{ij}^{-\theta}$, $p\equiv\left(\tilde{\sigma}\left(1-\left(\alpha_1+\alpha_2\right)\left(\sigma-1\right)-\sigma\left(\beta_1+\beta_2\right)\right)\right)^{-1}$, and $a\equiv\frac{(1+(\alpha_1+\alpha_2)\sigma+(\beta_1+\beta_2)(\sigma-1))}{(1-(\alpha_1+\alpha_2)(\sigma-1)-\sigma(\beta_1+\beta_2))}$. In what follows, we will assume $p>0$ and $a>0$. In addition, we have the labor market clearing constraint $\sum_{i\in S}L_i=\bar{L}$. Our goal is to provide bounds on Ω.

It proves helpful to define $x_i\equiv\left(\frac{L_i}{\bar{L}}\right)^{\frac{1}{\rho}}$ so that the system of equations become:

$$L_iW_i^{-\theta}=\Omega^{-\theta}\sum_j M_{ij}W_j^{\theta}\tag{38}$$

$$W_i^{\tilde{\sigma}\sigma}x_i{}^{\frac{1}{\rho}}=\bar{L}^{\frac{a-1}{\rho}}\sum_{j\in S}T_{ij}W_j^{-(\sigma-1)\tilde{\sigma}}x_j^a\tag{39}$$

where note that the labor market constraint now becomes $\sum_{i\in S}x_i^p=1$. In what follows, we refer to equation (38) as the "migration equation" and equation (39) as the "trade equation".

Before continuing with the proof, we remind the reader of a number of helpful mathematical properties. Define $\|\mathbf{x}\|_p\equiv\left(\sum_{i\in S}x_i^p\right)^{\frac{1}{p}}$. (With some abuse of notation, we refer to $\|\mathbf{x}\|_p$ as the "p-norm of $\mathbf{x}$", even though it is formally a norm only if $p\geq 1$). First, we remind the reader of the relationship between different p norms. For any $0<p<q$, we have the convenient relationship:

$$\|\mathbf{x}\|_q\leq\|\mathbf{x}\|_p\,.\tag{40}$$

More generally, for any $p<q$, we have:

$$N^{\frac{1}{q}-\frac{1}{p}}\|\mathbf{x}\|_p\leq\|\mathbf{x}\|_q\leq C\left(p,q\right)N^{\frac{1}{q}-\frac{1}{p}}\|\mathbf{x}\|_p\tag{41}$$

where $N=|S|$, $C\left(p,q\right)\equiv\left(\frac{p(\mu^q-\mu^p)}{(q-p)(\mu^p-1)}\right)^{\frac{1}{q}}\left(\frac{q(\mu^p-\mu^q)}{(p-q)(\mu^q-1)}\right)^{-\frac{1}{p}}$, and $\mu\geq\left(\frac{\max_i x_i}{\min_i x_i}\right)$. Note that if $\mu=1$, $C\left(p,q\right)=1$. The first inequality is the well known generalized mean inequality, whereas the second inequality is due to the less known result originally due to Specht (1960) and reprinted (in English) in the textbook by Mitrinovic and Vasic (1970) (see Theorem 1 on p.79).

Second, recall the Cauchy–Schwarz inequality that for any $N\times 1$ vectors $\mathbf{x}\equiv[x_i]$ and $\mathbf{y}\equiv[y_i]$, we have:

$$\sum_{i\in S}|x_iy_i|\leq\left(\sum_{i\in S}x_i^2\right)^{\frac{1}{2}}\left(\sum_{i\in S}y_i^2\right)^{\frac{1}{2}}\iff\|\{x_iy_i\}\|_1\leq\|\{x_i\}\|_2\|\{y_i\}\|_2\,.$$

Third, recall that the matrix norm induced by the vector p-norm for square matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_p\equiv\sup\left\{\frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p}|\mathbf{x}\neq 0\right\}$, which immediately implies that $\|\mathbf{Ax}\|_p\leq\|\mathbf{A}\|_p\|\mathbf{x}\|_p$ (this is known as the sub-multiplicative property of the matrix norm). Moreover, we have

$\|\mathbf{A}\|_2 = \sigma(\mathbf{A})$, i.e. the matrix norm induced by the Euclidean vector norm is equal to the largest singular value of matrix $\mathbf{A}$. (If $\mathbf{A}$ is normal, then $\sigma(\mathbf{A})$ is simply the absolute value of the largest eigenvalue. If $\mathbf{A}$ is real and positive, this is the spectral radius (Perron root) of $\mathbf{A}$. More generally, we have).

Fourth, recall that if an $N \times N$ matrix $\mathbf{A}$ is real and symmetric, there exists an eigenvalue-decomposition such that:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T,$$

where $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues of $\mathbf{A}$ and $\mathbf{Q} = [\mathbf{q}_i]$ is a matrix of the orthonormal eigenvectors, i.e. that $\mathbf{q}_i'\mathbf{q}_i = 1$ for all $i \in S$ and $\mathbf{q}_i'\mathbf{q}_j = 0$ for all $i \neq j$. An implication of this decomposition is that the quadratic form of $\mathbf{A}$ – i.e. $\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i\in S}\sum_{j\in S} A_{ij}x_i x_j$ – is bounded above by $\bar{\lambda}_A \|\mathbf{x}\|_2$, where $\bar{\lambda}_A$ is the absolute value of the largest eigenvalue of matrix $\mathbf{A}$.

A.3.1 Lemma

We now offer a lemmas which provides a bound on the maximum of the ratio to highest welfare and lowest welfare across locations within a given equilibrium.

Lemma 1. *In any steady state equilibrium, we can bound the ratio of the maximum to minimum period welfare* $W^* \equiv \frac{\max_{i\in S} W_i}{\min_{i\in S} W_i}$ *by:*

$$1 \leq W^* \leq \mu, \tag{42}$$

where $\mu \equiv (M^*)^{\left|\frac{1}{\theta - \frac{y_1}{y_2}}\right|}$, $M^* \equiv \max_{i,j}\sum_l \frac{M_{il}}{M_{jl}}$, *and* $[y_k]$ *are the eigenvectors associated with the largest eigenvalue of matrix* $\mathbf{A} = \begin{pmatrix} \left|\frac{(1+(\alpha_1+\alpha_2)\sigma+(\beta_1+\beta_2)(\sigma-1))}{(1-(\alpha_1+\alpha_2)(\sigma-1)-\sigma(\beta_1+\beta_2))}\right| & \left|\frac{2\sigma-1}{(1-(\alpha_1+\alpha_2)(\sigma-1)-\sigma(\beta_1+\beta_2))}\right| \\ \frac{1}{\theta} & 1 \end{pmatrix}$.
We remark that if $M^* = 1$, *then welfare will be equalized in all locations, i.e.* $W^* = 1$.

Proof. We begin by writing the steady state equilibrium system of equations more compactly as:

$$L_i^{\gamma_{11}} W_i^{\gamma_{12}} = c_1 \sum_j K_{ij,1} L_j^{\beta_{11}} W_j^{\beta_{12}}$$

$$L_i^{\gamma_{21}} W_i^{\gamma_{22}} = c_2 \sum_j K_{ij,1} L_j^{\beta_{21}} W_j^{\beta_{22}},$$

where $\mathbf{\Gamma} \equiv [\gamma_{kl}] = \begin{pmatrix} \tilde{\sigma}\left(1 - (\alpha_1 + \alpha_2)(\sigma - 1) - \sigma(\beta_1 + \beta_2)\right) & \tilde{\sigma}\sigma \\ 1 & -\theta \end{pmatrix}$, $\mathbf{B} \equiv [\beta_{kl}] = \begin{pmatrix} \tilde{\sigma}\left(1 + (\alpha_1 + \alpha_2)\sigma + (\beta_1 + \right. \\ 0 \end{pmatrix}$
$c_1 = 1$, $c_2 = \Omega^{-\theta}$, $K_{ij,1} = \tau_{ij}^{1-\sigma} \bar{A}_i^{(\sigma-1)\tilde{\sigma}} \bar{u}_i^{\sigma\tilde{\sigma}} u_j^{(\sigma-1)\tilde{\sigma}} \bar{A}_j^{\tilde{\sigma}\sigma}$, and $K_{ij,2} = \mu_{ij}^{-\theta}$.

We can re-write this system of equations as:

$$L_i = c_1 W_i^{-\frac{\gamma_{12}}{\gamma_{11}}} \left(\sum_j K_{ij,1} L_j^{\beta_{11}} W_j^{\beta_{12}} \right)^{\frac{1}{\gamma_{11}}}$$

$$W_i = c_2 L_i^{-\frac{\gamma_{21}}{\gamma_{22}}} \left(\sum_j K_{ij,1} L_j^{\beta_{21}} W_j^{\beta_{22}} \right)^{\frac{1}{\gamma_{22}}}$$

Define $L^{max} \equiv \max_i L_i$, $L^{min} \equiv \min_i L_i$, $W^{max} \equiv \max_i W_i$, $W^{min} \equiv \min_i W_i$.

$$\begin{aligned} L^{max} \leq & c_1 \left(\left(W^{min}\right)^{\mathbf{1}\left\{\frac{\gamma_{12}}{\gamma_{11}}>0\right\}} \left(W^{max}\right)^{\mathbf{1}\left\{\frac{\gamma_{12}}{\gamma_{11}}<0\right\}} \right)^{-\frac{\gamma_{12}}{\gamma_{11}}} \times \\ & \left(\sum_j K_{ij,1} \left(\left(L^{max}\right)^{\mathbf{1}(\beta_{11}>0)} \left(L^{min}\right)^{\mathbf{1}(\beta_{11}<0)} \right)^{\beta_{11}} \left(\left(W^{max}\right)^{\mathbf{1}(\beta_{12}>0)} \left(W^{min}\right)^{\mathbf{1}(\beta_{12}<0)} \right)^{\beta_{12}} \right)^{\frac{1}{\gamma_{11}}\mathbf{1}\{\gamma_{11}>0\}} \times \\ & \left(\sum_j K_{ij,1} \left(\left(L^{max}\right)^{\mathbf{1}(\beta_{11}<0)} \left(L^{min}\right)^{\mathbf{1}(\beta_{11}>0)} \right)^{\beta_{11}} \left(\left(W^{max}\right)^{\mathbf{1}(\beta_{12}<0)} \left(W^{min}\right)^{\mathbf{1}(\beta_{12}>0)} \right)^{\beta_{12}} \right)^{\frac{1}{\gamma_{11}}\mathbf{1}\{\gamma_{11}<0\}} \end{aligned}$$

Similarly, we have:

$$\begin{aligned} L^{min} \geq & c_1 \left(\left(W^{max}\right)^{\mathbf{1}\left\{\frac{\gamma_{12}}{\gamma_{11}}>0\right\}} \left(W^{min}\right)^{\mathbf{1}\left\{\frac{\gamma_{12}}{\gamma_{11}}<0\right\}} \right)^{-\frac{\gamma_{12}}{\gamma_{11}}} \times \\ & \left(\sum_j K_{ij,1} \left(\left(L^{min}\right)^{\mathbf{1}(\beta_{11}>0)} \left(L^{max}\right)^{\mathbf{1}(\beta_{11}<0)} \right)^{\beta_{11}} \left(\left(W^{min}\right)^{\mathbf{1}(\beta_{12}>0)} \left(W^{max}\right)^{\mathbf{1}(\beta_{12}<0)} \right)^{\beta_{12}} \right)^{\frac{1}{\gamma_{11}}\mathbf{1}\{\gamma_{11}>0\}} \times \\ & \left(\sum_j K_{ij,1} \left(\left(L^{min}\right)^{\mathbf{1}(\beta_{11}<0)} \left(L^{max}\right)^{\mathbf{1}(\beta_{11}>0)} \right)^{\beta_{11}} \left(\left(W^{min}\right)^{\mathbf{1}(\beta_{12}<0)} \left(W^{max}\right)^{\mathbf{1}(\beta_{12}>0)} \right)^{\beta_{12}} \right)^{\frac{1}{\gamma_{11}}\mathbf{1}\{\gamma_{11}<0\}} \end{aligned}$$

Define $L^* \equiv \frac{\max_i L_i}{\min_i L_i} = \frac{L^{max}}{L^{min}}$ and $W^* \equiv \frac{\max_i W_i}{\min_i W_i} = \frac{W^{max}}{W^{min}}$ Then combining the previous two equations we have:

$$L^* \leq \left(L^*\right)^{\left|\frac{\beta_{11}}{\gamma_{11}}\right|} \left(W^*\right)^{\left|\frac{\gamma_{12}}{\gamma_{11}}\right| + \left|\frac{\beta_{12}}{\gamma_{11}}\right|}$$

We can proceed similarly for the second equation, which yields:

$$W^* \leq \left(L^*\right)^{\left|\frac{\gamma_{21}}{\gamma_{22}}\right| + \left|\frac{\beta_{21}}{\gamma_{22}}\right|} \left(W^*\right)^{\left|\frac{\beta_{22}}{\gamma_{22}}\right|}$$

Define: $\mathbf{A} \equiv \begin{pmatrix} \left|\frac{\beta_{11}}{\gamma_{11}}\right| & \left|\frac{\gamma_{12}}{\gamma_{11}}\right| + \left|\frac{\beta_{12}}{\gamma_{11}}\right| \\ \left|\frac{\gamma_{21}}{\gamma_{22}}\right| + \left|\frac{\beta_{21}}{\gamma_{22}}\right| & \left|\frac{\beta_{22}}{\gamma_{22}}\right| \end{pmatrix} = \begin{pmatrix} \left|\frac{(1+(\alpha_1+\alpha_2)\sigma+(\beta_1+\beta_2)(\sigma-1))}{(1-(\alpha_1+\alpha_2)(\sigma-1)-\sigma(\beta_1+\beta_2))}\right| & \left|\frac{2\sigma-1}{(1-(\alpha_1+\alpha_2)(\sigma-1)-\sigma(\beta_1+\beta_2))}\right| \\ \frac{1}{\theta} & 1 \end{pmatrix}$

so that our inequalities in matrix notation become:

$$\begin{pmatrix} \ln L^* \\ \ln W^* \end{pmatrix} \leq \mathbf{A} \begin{pmatrix} \ln L^* \\ \ln W^* \end{pmatrix} \iff$$
$$\lambda \begin{pmatrix} \ln L^* \\ \ln W^* \end{pmatrix} = \mathbf{A} \begin{pmatrix} \ln L^* \\ \ln W^* \end{pmatrix},$$

for some $\lambda > 1$. As long as the spectral radius of $\mathbf{A}$ is greater than one, the unique solution to this inequality is that:

$$\begin{pmatrix} \ln L^* \\ \ln W^* \end{pmatrix} = \begin{pmatrix} cy_1 \\ cy_2 \end{pmatrix},$$

where $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} > 0$ is the eigenvector associated with the largest eigenvalue of matrix $\mathbf{A}$ and $c > 0$ is some scalar. As a result, we have the following equation relating the spatial variation of welfare and population density across locations:

$$\ln W^* = \frac{y_2}{y_1} \ln L^*, \tag{43}$$

We continue by recalling that in the steady state we have the following equality:

$$W_i^\theta \Pi_i^\theta = \Omega^\theta L_i$$

which in turn implies for any equilibrium we have:

$$L^* = \frac{W^\theta_{i_L^{max}} \Pi^\theta_{i_L^{max}}}{W^\theta_{i_L^{min}} \Pi^\theta_{i^{min}L}},$$

where $i_L^{max} \equiv \arg\max_{i \in S} L_i$ and $i_L^{min} \equiv \arg\min_{i \in S} L_i$. As a result, we have:

$$\ln L^* \leq \theta \ln W^* + \ln \frac{\Pi^\theta_{i_L^{max}}}{\Pi^\theta_{i_L^{min}}}$$

Note that for any $i \in S$ and $j \in S$ we have:

$$\frac{\Pi_i^\theta}{\Pi_j^\theta} = \frac{\sum_k M_{ik} W_k^\theta}{\sum_k M_{jk} W_k^\theta} = \sum_k \frac{M_{ik} W_k^\theta}{\sum_l M_{jl} W_l^\theta} \leq \sum_k \frac{M_{ik} W_k^\theta}{M_{jk} W_k^\theta} = \sum_k \frac{M_{ik}}{M_{jk}}$$

Let $M^* \equiv \max_{i,j} \sum_k \left(\frac{M_{ik}}{M_{jk}}\right)$. We then have:

$$\ln \frac{\Pi^\theta_{i^{max}}}{\Pi^\theta_{i^{min}}} \leq \ln M^*,$$

so that:

$$\ln L^* \leq \theta \ln W^* + \ln M^* \tag{44}$$

Combining equations (43) and (44) yields:

$$\begin{aligned}\ln W^* \leq \left(\frac{y_2}{y_1}\right)(\theta \ln W^* + \ln M^*) \iff \\ \left(1 - \theta \frac{y_2}{y_1}\right) \ln W^* \leq \left(\frac{y_2}{y_1}\right) \ln M^*.\end{aligned}$$

If $\theta \frac{y_2}{y_1} < 1$ we then have:

$$\ln W^* \leq \frac{\left(\frac{y_2}{y_1}\right)}{1 - \theta \frac{y_2}{y_1}} \ln M^* \tag{45}$$

Similarly, we have:

$$(W^*)^\theta = \frac{L_{i_W^{max}}}{L_{i_W^{min}}} \times \frac{\Pi^\theta_{i_W^{min}}}{\Pi^\theta_{i_W^{max}}},$$

where $i_W^{max} \equiv \arg\max_{i \in S} W_i$ and $i_W^{min} \equiv \arg\min_{i \in S} W_i$ so that:

$$\begin{aligned}\theta \ln W^* \leq \ln L^* + \ln \frac{\Pi^\theta_{i_W^{min}}}{\Pi^\theta_{i_W^{max}}} \leq \ln L^* + \ln M^* \iff \\ \ln L^* \geq \theta \ln W^* - \ln M^*. \end{aligned} \tag{46}$$

Combining equations (43) and (46) yields:

$$\begin{aligned}\ln W^* \geq \left(\frac{y_2}{y_1}\right)(\theta \ln W^* - \ln M^*) \iff \\ \left(\left(\frac{y_2}{y_1}\right)\theta - 1\right) \ln W^* \leq \left(\frac{y_2}{y_1}\right) \ln M^*\end{aligned}$$

If $\theta \frac{y_2}{y_1} > 1$ we then have:

$$\ln W^* \leq \frac{\left(\frac{y_2}{y_1}\right)}{\left(\left(\frac{y_2}{y_1}\right)\theta - 1\right)} \ln M^* \tag{47}$$

Combining equations (45) and (47) then yields:

$$
\begin{aligned}
\ln W^* &\leq \left| \frac{\left(\frac{y_2}{y_1}\right)}{\left(\frac{y_2}{y_1}\right)\theta - 1} \right| \ln M^* \iff \\
\ln W^* &\leq \left| \frac{1}{\theta - \frac{y_1}{y_2}} \right| \ln M^* \iff \\
W^* &\leq (M^*)^{\left|\frac{1}{\theta - \frac{y_1}{y_2}}\right|},
\end{aligned}
$$

as required. □

A.3.2 The upper bound

We now proceed by constructing the upper bound. The proof proceeds by first constructing an upper bound for steady state welfare as a function of the norm of the period welfare using the migration equation. The proof then constructs an upper bound for the norm of period welfare using the trade equation.

The migration equation We first examine the migration equation (38). Define $\omega_i \equiv W_i^\theta$. Recall that $\sum_{i \in S} L_i = \bar{L} \iff \|\{L_i\}\|_1 = \bar{L}$. Then:

$$
\begin{aligned}
L_i W_i^{-\theta} &= \Omega^{-\theta} \sum_j M_{ij} W_j^\theta \iff \\
L_i &= \Omega^{-\theta} \omega_i \sum_j M_{ij} \omega_j \implies \\
\|\{L_i\}\|_1 &= \Omega^{-\theta} \left\| \left\{ \omega_i \sum_j M_{ij} \omega_j \right\} \right\|_1 \iff \\
\bar{L}\Omega^\theta &= \left\| \left\{ \omega_i \sum_j M_{ij} \omega_j \right\} \right\|_1 \iff \\
\bar{L}\Omega^\theta &= \boldsymbol{\omega}' \mathbf{M} \omega \implies \\
\bar{L}\Omega^\theta &\leq \bar{\lambda}_M \|\boldsymbol{\omega}\|_2 \iff \\
\bar{L}\Omega^\theta &\leq \bar{\lambda}_M \left(\sum_{i \in S} \left(W_i^\theta\right)^2 \right)^{\frac{1}{2}} \iff \\
\Omega &\leq \bar{L}^{-\frac{1}{\theta}} \bar{\lambda}_M^{\frac{1}{\theta}} \|\boldsymbol{W}\|_{2\theta}
\end{aligned} \tag{48}
$$

where $\bar{\lambda}_M$ is the largest eigenvalue of $\mathbf{M}$. Because $\mathbf{M}$ is positive, this is also the spectral radius, Perron-root, and largest singular value of $\mathbf{M}$.

The trade equation We now turn to the trade equation (39), which if we define $y_i \equiv W_i^{\tilde{\sigma}\sigma}\left(L_i^{\frac{1}{\rho}}\right)$ can be written as follows:

Defining $y_i \equiv W_i^{\tilde{\sigma}\sigma}\left(L_i^{\frac{1}{\rho}}\right)$ so that:

$$
\begin{aligned}
y_i &= \sum_j T_{ij} y_j \frac{W_j^{-(\sigma-1)\tilde{\sigma}}\left(L_j^{\frac{1}{\rho}}\right)^a}{W_j^{\tilde{\sigma}\sigma}\left(L_j^{\frac{1}{\rho}}\right)} \iff \\
y_i &= \sum_j T_{ij} y_j W_j^{1-\sigma} L_j^{\frac{a-1}{\rho}}
\end{aligned}
$$

We then sum both sides over $i \in S$:

$$
\begin{aligned}
y_i &= \sum_j T_{ij} y_j W_j^{1-\sigma} L_j^{\frac{a-1}{\rho}} \implies \\
\|\{y_i\}\|_1 &= \left\|\left\{\sum_j T_{ij} y_j W_j^{1-\sigma} L_j^{\frac{a-1}{\rho}}\right\}\right\|_1 \implies \\
\|\{y_i\}\|_1 &\leq \left\|\left\{\sum_j T_{ij} y_j\right\}\right\|_2 \left\|\left\{W_j^{1-\sigma} L_j^{\frac{a-1}{\rho}}\right\}\right\|_2 \implies \\
\|\{y_i\}\|_1 &\leq \|\mathbf{T}\|_2 \|\{y_i\}\|_2 \left\|\left\{W_j^{1-\sigma} L_j^{\frac{a-1}{\rho}}\right\}\right\|_2 \implies \\
\|\{y_i\}\|_1 &\leq \bar{L}^{\frac{a-1}{\rho}} \|\mathbf{T}\|_2 \|\{y_i\}\|_2 \left\|\left\{W_i^{1-\sigma}\right\}\right\|_1 \left\|\left\{\frac{L_i}{\bar{L}}\right\}\right\|_1 \implies \\
\|\{y_i\}\|_1 &\leq \bar{L}^{\frac{a-1}{\rho}} \|\mathbf{T}\|_2 \|\{y_i\}\|_1 \left\|\left\{W_i^{1-\sigma}\right\}\right\|_1 \iff \\
\left(\sum_{i\in S} W_j^{1-\sigma}\right)^{-1} &\leq \bar{L}^{\frac{a-1}{\rho}} \|\mathbf{T}\|_2 \iff \\
\|\mathbf{W}\|_{1-\sigma} &\leq \bar{L}^{\frac{a-1}{\rho}\frac{1}{\sigma-1}} \lambda_T^{\frac{1}{\sigma-1}},
\end{aligned}
\tag{49}
$$

where the third line uses the Cauchy-Schwartz inequality, the fourth line uses the property of matrix norms induced by the vector norm, the fifth line uses the fact that $\frac{a-1}{p} > 1$ and $\left(\frac{L_i}{\bar{L}}\right) \in [0,1]$ then $\left(\frac{L_i}{\bar{L}}\right)^{\frac{a-1}{p}} < \frac{L_i}{\bar{L}}$, the sixth lines uses the fact that $|y_i|_2 \leq |y_i|_1$, and the seventh the fact that $|\mathbf{A}|_2 \leq \bar{\lambda}_A$, i.e. the matrix p-norm with $p = 2$ is bounded above by the Perron root), and $\bar{\lambda}_T$ is the largest eigenvalue of $\mathbf{T}$ (because $\mathbf{T}$ is strictly positive, by the Perron-Frobenius theorem, the largest eigenvalue is positive) and we used the fact that $\frac{1}{\sigma-1}\frac{a-1}{p} = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)$.

The Bound Recall from equation (41) that because $(1-\sigma) < 2\theta$, we have:

$$\|\boldsymbol{W}\|_{2\theta} \leq c\,\|\boldsymbol{W}\|_{1-\sigma}\,, \tag{50}$$

where

$$c_1 \equiv \left(\frac{(1-\sigma)\left(\mu^{2\theta}-\mu^{(1-\sigma)}\right)}{(2\theta+\sigma-1)\left(\mu^{(1-\sigma)}-1\right)}\right)^{\frac{1}{2\theta}}\left(\frac{2\theta\left(\mu^{(1-\sigma)}-\mu^{2\theta}\right)}{((1-\sigma)-2\theta)\left(\mu^{2\theta}-1\right)}\right)^{-\frac{1}{(1-\sigma)}} N^{\frac{1}{2\theta}+\frac{1}{\sigma-1}}$$

from equation (28) and μ is defined above in equation (42) from Lemma 1. Combining equation (50) with the migration bound from equation (48) and the trade bound from equation (49) then yields:

$$\Omega \leq c_1 \bar{\lambda}_M^{\frac{1}{\theta}} \bar{\lambda}_T^{\frac{1}{\sigma-1}} \bar{L}^{\left(\rho-\frac{1}{\theta}\right)}, \tag{51}$$

where $\rho \equiv (\alpha_1+\alpha_2)+(\beta_1+\beta_2)$, as claimed.

A.3.3 The lower bound

We now proceed to prove the lower bound. As above, we first consider the migration equation and then consider the trade equation.

The migration equation We first examine the migration equation. Define $\mu_i \equiv L_i W_i^{-\theta}$ and, with some abuse of notation, $M_{ij}^{-1} \equiv \mathbf{M}_{ij}^{-1}$. Then:

$$\begin{aligned}
L_i W_i^{-\theta} = \Omega^{-\theta} \sum_j M_{ij} W_j^{\theta} &\iff \\
\sum_j M_{ij}^{-1} L_j W_j^{-\theta} = \Omega^{-\theta} W_i^{\theta} &\iff \\
\sum_j M_{ij}^{-1}\left(W_i^{-\theta} L_i\right)\left(L_j W_j^{-\theta}\right) = L_i \Omega^{-\theta} &\iff \\
\sum_j M_{ij}^{-1} \mu_i \mu_j = L_i \Omega^{-\theta} &\implies \\
\mu' \mathbf{M}^{-1} \mu = \bar{L} \Omega^{-\theta} &\implies \\
\bar{L}\Omega^{-\theta} \leq \bar{\lambda}_{M^{-1}} \|\boldsymbol{\mu}\|_2 &\iff \\
\bar{L}\Omega^{-\theta} \leq \bar{\lambda}_{M^{-1}} \left(\sum_{i\in S} L_i^2 W_i^{-2\theta}\right)^{\frac{1}{2}} &\implies \\
\bar{L}\Omega^{-\theta} \leq \bar{\lambda}_{M^{-1}} \left(\sum_{i\in S} L_i^4\right)^{\frac{1}{4}} \left(\sum_{i\in S} W_i^{-4\theta}\right)^{\frac{1}{4}} &\implies \\
\bar{L}\Omega^{-\theta} \leq \bar{\lambda}_{M^{-1}} \bar{L}\left(\left\|\left\{(W_i)^{-1}\right\}\right\|_{4\theta}\right)^{\theta} &\iff \\
(\underline{\lambda}_M)^{\frac{1}{\theta}}\left(\|\{W_i\}\|_{-4\theta}\right) \leq \Omega,
\end{aligned} \tag{52}$$

where $\bar{\lambda}_{M^{-1}}$ is the largest eigenvalue (in absolute value) of $\mathbf{M}^{-1}$ and the third to last line applied the Cauchy–Schwarz inequality, the second to last line used equation (40) to note that $\|\{L_i\}\|_4 \leq \|\{L_i\}\|_1 = \bar{L}$, and the last lined used the fact that $\bar{\lambda}_{M^{-1}} = (\underline{\lambda}_M)^{-1}$, i.e. the largest eigenvalue (in absolute value) of $\mathbf{M}^{-1}$ is the inverse of the smallest eigenvalue of $\mathbf{M}$.

The trade equation We now turn to the trade equation (39). Again, with some abuse of notation, define $T_{ij}^{-1} \equiv \mathbf{T}_{ij}^{-1}$. As a result, we can write equation (39) as follows:

$$\begin{aligned}
W_i^{\tilde{\sigma}\sigma} L_i^{\frac{1}{\rho}} &= \sum_{j\in S} T_{ij} W_j^{-(\sigma-1)\tilde{\sigma}} L_j^{\frac{a}{\rho}} \iff \\
W_i^{-(\sigma-1)\tilde{\sigma}} L_i^{\frac{a}{\rho}} &= \sum_{j\in S} T_{ij}^{-1} W_j^{\tilde{\sigma}\sigma} L_j^{\frac{1}{\rho}} \iff \\
y_i L_i^{\frac{a-1}{\rho}} &= \sum_{j\in S} T_{ij}^{-1} W_j^{\tilde{\sigma}\sigma} L_j^{\frac{1}{\rho}} \frac{y_j}{W_j^{-(\sigma-1)\tilde{\sigma}} L_j^{\frac{1}{\rho}}} \iff \\
y_i &= \sum_{j\in S} T_{ij}^{-1} W_j^{\sigma-1} y_j L_i^{-\left(\frac{a-1}{\rho}\right)},
\end{aligned}$$

where $y_i \equiv W_i^{-(\sigma-1)\tilde{\sigma}} L_i^{\frac{1}{\rho}}$. Taking the Euclidean norm of both sides yields:

$$\begin{aligned}
\|\{y_i\}\|_2 &= \left\| \left\{ \sum_{j\in S} T_{ij}^{-1} W_j^{\sigma-1} L_i^{-\frac{a-1}{\rho}} y_j \right\} \right\|_2 \implies \\
\|\{y_i\}\|_2 &\leq \max_{\mathbf{L}} \left\| \left\{ \sum_{j\in S} T_{ij}^{-1} W_j^{\sigma-1} L_i^{-\frac{a-1}{\rho}} y_j \right\} \right\|_2 \implies \\
\|\{y_i\}\|_2 &\leq \left(\frac{\bar{L}}{N}\right)^{-\frac{a-1}{p}} \left\| \left\{ \sum_{j\in S} T_{ij}^{-1} W_j^{\sigma-1} y_j \right\} \right\|_2 \implies \\
\|\{y_i\}\|_2 &\leq \left(\frac{\bar{L}}{N}\right)^{-\frac{a-1}{p}} \left\|\mathbf{T}^{-1}\right\|_2 \left\|\{W_j^{\sigma-1} y_j\}\right\|_2 \implies \\
\|\{y_i\}\|_1 &\leq \left(\frac{\bar{L}}{N}\right)^{-\frac{a-1}{p}} \left\|\mathbf{T}^{-1}\right\|_2 \left\|\{W_j^{\sigma-1}\}\right\|_4 \|\{y_j\}\|_4 \implies \\
\|\{y_i\}\|_1 &\leq \left(\frac{\bar{L}}{N}\right)^{-\frac{a-1}{p}} \left\|\mathbf{T}^{-1}\right\|_2 \left\|\{W_j^{\sigma-1}\}\right\|_4 \|\{y_j\}\|_1 \implies \\
1 &\leq \left(\frac{\bar{L}}{N}\right)^{-\frac{a-1}{p}} \left\|\mathbf{T}^{-1}\right\|_2 \left\|\{W_j^{\sigma-1}\}\right\|_4 \iff \\
(\underline{\lambda}_T)^{\frac{1}{\sigma-1}} \left(\left(\frac{\bar{L}}{N}\right)^{\frac{1}{\sigma-1}\frac{a-1}{p}} \right) &\leq \|\{W_j\}\|_{4(\sigma-1)}, \qquad (53)
\end{aligned}$$

where the second line took a maximum over all possible distributions of labor, the third line used the fact that the solution to that maximization is $L_i = \frac{\bar{L}}{N}$, fourth line used a property of matrix norms, the fifth line used Cauchy-Schwartz, the sixth line used the fact that $\|\{y_j\}\|_4 \leq \|\{y_j\}\|_1$, seventh line used the fact that $\underline{\lambda}_T = \bar{\lambda}_{T^{-1}}^{-1}$.

The Bound Recall from equation (41) that because $-4\theta < 4(\sigma - 1)$, we have:

$$c_2 \|\boldsymbol{W}\|_{4(\sigma-1)} \leq \|\boldsymbol{W}\|_{-4\theta}, \tag{54}$$

where

$$c_2 \equiv \left(\left(\frac{(-4\theta)\left(\mu^{4(\sigma-1)} - \mu^{(-4\theta)}\right)}{(4(\sigma-1)+4\theta)\left(\mu^{(-4\theta)} - 1\right)} \right)^{\frac{1}{4(\sigma-1)}} \left(\frac{4(\sigma-1)\left(\mu^{(-4\theta)} - \mu^{4(\sigma-1)}\right)}{((-4\theta) - 4(\sigma-1))\left(\mu^{4(\sigma-1)} - 1\right)} \right)^{\frac{1}{4\theta}} N^{\frac{1}{4(\sigma-1)} + \frac{1}{4\theta}} \right)^{-1}$$

from equation (28) and μ is defined above in equation (42) from Lemma 1. Combining equation (52) from the migration bound with equation (53) from the trade bound then yields:

$$\Omega \geq c_2 \left(\underline{\lambda}_T\right)^{\frac{1}{\sigma-1}} \left(\underline{\lambda}_M\right)^{\frac{1}{\theta}} \left(\frac{\bar{L}}{N}\right)^{\rho}, \tag{55}$$

where $\rho \equiv (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)$, as claimed.

A.4 Proof of Proposition 3

The two systems of equations are:

$$p_{it}^{\sigma-1} = \sum_j T_{ijt} \left(\frac{Y_{jt}}{Y_{it}}\right) P_{jt}^{\sigma-1}$$
$$P_{it}^{\sigma-1} = \sum_j T_{jit} \left(p_{jt}^{\sigma-1}\right)^{-1}$$

and:

$$\left(W_{it}^{\theta}\right)^{-1} = \sum_j M_{jit} \frac{L_{jt-1}}{L_{it}} \left(\Pi_{jt}^{\theta}\right)^{-1}$$
$$\Pi_{it}^{\theta} = \sum_i M_{ijt} W_{jt}^{\theta}.$$

Both systems of equations can be written as:

$$x_i = \sum_j K^A_{ij} y_j$$
$$y_i = \sum_j K^B_{ij} x_j^{-1}$$

which has a corresponding LHS matrix of coefficients:

$$\boldsymbol{B} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the matrix on the RHS coefficients becomes:

$$\boldsymbol{\Gamma} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Hence, we have:

$$\mathbf{A} \equiv \boldsymbol{\Gamma}\mathbf{B}^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

so that $\mathbf{A}^p = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. It is straightforward to check that $\rho\left(\mathbf{A}^p\right) = 1$, as required.

B Possible Microfoundations for Spillovers

B.1 Productivity spillovers

B.1.1 The Accumulation of ideas

We follow Deneckere and Judd (1992). Suppose that firms can pay a fixed cost f_i (in terms of local labor) to create a new variety, over which they have monopoly rights over for one period (the period in which they introduce the variety). In the subsequent period, the new variety exists but is produced under conditions of perfect competition. In the following period (two periods after its introduction), we assume the variety no longer exists (i.e. it fully depreciates). Finally, we assume that consumers have Cobb-Douglas preferences (within location) over the the new varieties and the old varieties, and CES preferences across respectively.

Setup

Demand Let Ω^{new}_{it} be the set of varieties created by monopolistically competitive firms in period t in location $i \in S$ and $\Omega^{old}_{i,t}$ be the set of varieties created in the previous period that are now produced under perfect competition. We assume that consumers have Cobb-Douglas preferences over CES aggregates of the two types of goods within location and then

CES aggregates of the Cobb-Douglas combinations across locations, i.e.:

$$C_{jt} = \left(\sum_{i \in S} \left(\left(\left(\left(\int_{\Omega_{it}^{new}} q_{ijt}(\omega)^{\frac{\rho-1}{\rho}} d\omega \right)^{\frac{\rho}{\rho-1}} \right)^{\chi} \left(\left(\int_{\Omega_{it}^{old}} q_{ijt}(\omega)^{\frac{\rho-1}{\rho}} d\omega \right)^{\frac{\rho}{\rho-1}} \right)^{1-\chi} \right)^{\frac{\sigma-1}{\sigma}} \right) \right)^{\frac{\sigma}{\sigma-1}},$$

where $q_{ijt}(\omega)$ is the quantity consumed in country j of variety ω from location i, ρ is the elasticity of substitution between varieties of a given type from a given location, χ is the Cobb-Douglas share of the CES composite of new varieties from a given location, and σ is the elasticity of substitution of the Cobb-Douglas aggregate across locations.

The total quantity a consumer in country $j \in S$ in period t will demand from firm ω in location i can be written as:

$$q_{ij,t}(\omega) = \begin{cases} \chi p_{ij,t}(\omega)^{-\rho} \left(P_{i,t}^{new}\right)^{\rho-1} \times \dfrac{\tau_{ij}^{1-\sigma} \left(\left(P_{i,t}^{new}\right)^{\chi} \left(P_{i,t}^{old}\right)^{1-\chi} \right)^{1-\sigma}}{\sum_{k \in S} \tau_{ij}^{1-\sigma} \left(\left(P_{k,t}^{new}\right)^{\chi} \left(P_{k,t}^{old}\right)^{1-\chi} \right)^{1-\sigma}} Y_{j,t} & \text{if } \omega \in \Omega_{it}^{new} \\ (1-\chi) p_{ij,t}(\omega)^{-\rho} \left(P_{i,t}^{old}\right)^{\rho-1} \times \dfrac{\tau_{ij}^{1-\sigma} \left(\left(P_{i,t}^{new}\right)^{\chi} \left(P_{i,t}^{old}\right)^{1-\chi} \right)^{1-\sigma}}{\sum_{k \in S} \tau_{ij}^{1-\sigma} \left(\left(P_{k,t}^{new}\right)^{\chi} \left(P_{k,t}^{old}\right)^{1-\chi} \right)^{1-\sigma}} Y_{j,t} & \text{if } \omega \in \Omega_{it}^{old} \end{cases} \tag{56}$$

where:

$$\left(P_{i,t}^{new}\right)^{1-\rho} \equiv \int_{\Omega_{it}^{new}} p_{ij,t}(\omega)^{1-\rho} d\omega \tag{57}$$

$$\left(P_{i,t}^{old}\right)^{1-\rho} \equiv \int_{\Omega_{it}^{old}} p_{ij,t}(\omega)^{1-\rho} d\omega \tag{58}$$

is the Dixit-Stiglitz price index of the inner CES nest.

Supply Let $c_{i,t} \equiv \frac{w_{i,t}}{\bar{A}_{i,t}}$ denote the marginal cost of production by a firm., where $\bar{A}_{i,t}$ is the (exogenous) productivity. The optimization problem faced by firm ω is:

$$\max_{\{q_{ij}(\omega)\}_{j \in S}} \sum_{j \in S} \left(p_{ij,t}(\omega) q_{ij}(\omega) - c_{i,t} \tau_{ij} q_{ij,t}(\omega) \right) - w_{i,t} f_i$$

subject to consumer demand given by equation (56).

As a result, conditional on positive production (more on that below), the first order conditions imply:

$$p_{ij,t}(\omega) = \frac{\rho}{\rho-1} c_{i,t} \tau_{ij} \tag{59}$$

so that the price index across new varieties within a location is:

$$P_{i,t}^{new} \equiv \left(M_{it}^{new}\right)^{\frac{1}{1-\rho}} \left(\frac{\rho}{\rho-1} c_{i,t} \right) \tag{60}$$

Profits of monopolistically competitive firms Firms profits of a firm $\omega \in \Omega_{i,t}^{new}$ are:

$$\pi_{i,t}(\omega) \equiv \sum_j \left(p_{ij,t}(\omega) - c_{i,t}\tau_{ij}\right) q_{ij,t}(\omega) - w_{i,t} f_i \tag{61}$$

Substituting the consumer demand expression (56) and the price expression (59) into equation (61) yields:

$$\pi_{i,t}(\omega) = \chi \frac{1}{\rho} \left(\frac{\rho}{\rho-1}\right)^{1-\rho} \sum_j \left(c_{i,t}\tau_{ij}\right)^{1-\rho} \left(P_{i,t}^{new}\right)^{\rho-1} \frac{\tau_{ij}^{1-\sigma} \left(\left(P_{i,t}^{new}\right)^{\chi} \left(P_{i,t}^{old}\right)^{1-\chi}\right)^{1-\sigma}}{\sum_{k\in S} \tau_{ij}^{1-\sigma} \left(\left(P_{k,t}^{new}\right)^{\chi} \left(P_{k,t}^{old}\right)^{1-\chi}\right)^{1-\sigma}} Y_{j,t} - w_{i,t} f_i$$

It turns out that in this framework, the profits of a firm have a simple relationship to the quantity the firm produces, which greatly simplifies the equilibrium. To see this, we first relate the profits a firm to its revenues. Note that from the consumer demand equation (56) and the price expression (59) that the revenue a producer receives is:

$$r_i(\omega) \equiv \sum_{j\in S} p_{ij,t}(\omega)\, q_{ij,t}(\omega) \iff$$

$$r_i(\omega) \left(\frac{\rho}{\rho-1}\right)^{\rho-1} \frac{1}{\chi} = \sum_j \left(c_{i,t}\tau_{ij}\right)^{1-\rho} \left(P_{i,t}^{new}\right)^{\rho-1} \frac{\tau_{ij}^{1-\sigma} \left(\left(P_{i,t}^{new}\right)^{\chi} \left(P_{i,t}^{old}\right)^{1-\chi}\right)^{1-\sigma}}{\sum_{k\in S} \tau_{ij}^{1-\sigma} \left(\left(P_{k,t}^{new}\right)^{\chi} \left(P_{k,t}^{old}\right)^{1-\chi}\right)^{1-\sigma}} Y_{j,t} \tag{62}$$

so that variable profits are simply equal to revenue divided by the elasticity of substitution, i.e.:

$$\pi_{i,t}(\omega) + w_{i,t} f_i = \frac{1}{\rho} r_i(\omega). \tag{63}$$

Free entry From the free entry condition, total profits of a firm are zero, i.e. $\pi_{i,t}(\omega) = 0$. Applying the free entry condition to equation (63) yields:

$$w_{i,t} f_i = \frac{1}{\rho} r_i(\omega) \tag{64}$$

Substituting equation (64) into equation (62) yields:

$$\sum_j \tau_{ij}^{1-\rho} w_{i,t}^{-\rho} A_{i,t}^{\rho-1} \left(P_{i,t}^{new}\right)^{\rho-1} \frac{\tau_{ij}^{1-\sigma} \left(\left(P_{i,t}^{new}\right)^{\chi} \left(P_{i,t}^{old}\right)^{1-\chi}\right)^{1-\sigma}}{\sum_{k\in S} \tau_{ij}^{1-\sigma} \left(\left(P_{k,t}^{new}\right)^{\chi} \left(P_{k,t}^{old}\right)^{1-\chi}\right)^{1-\sigma}} Y_{j,t} = \frac{1}{\chi} \left(\frac{\rho}{\rho-1}\right)^{\rho-1} \rho f_i, \tag{65}$$

where we use the fact that $c_{i,t} = w_{i,t}/A_{i,t}$.

Perfectly competitive varieties The price charged for the perfectly competitive varieties $\omega \in \Omega_{i,t}^{new}$ is simply the marginal cost:

$$p_{ij,t}(\omega) = \tau_{ij} c_{i,t} \ \forall \omega \in \Omega_{i,t}^{new}$$

so that:

$$P_{i,t}^{old} = \left(M_{it}^{old}\right)^{\frac{1}{1-\rho}} c_{i,t} \tag{66}$$

Labor market clearing Let $M_{i,t}^{new} \equiv \left|\Omega_{i,t}^{new}\right|$ and $M_{i,t}^{old} \equiv \left|\Omega_{i,t}^{old}\right|$ denote the measure of new and existing varieties, respectively.

Labor market clearing requires that the total labor used by all firms (for entry and production of the new varieties as well as production of the existing varieties) must equal to the total number of workers in the location, $L_{i,t}$.

The total amount of labor required by new varieties is:

$$\begin{aligned} L_{i,t}^{new} &= \int_{\Omega_{i,t}^{new}} \left(\sum_{j \in S} \tau_{ij} \frac{q_{ij}(\omega)}{\bar{A}_{i,t}} + f_i \right) d\omega \iff \\ L_{i,t}^{new} &= \rho f_i M_{i,t}^{new}, \end{aligned}$$

where the last line used the free entry equation (65).

The total amount of labor required by old varieties is:

$$\begin{aligned} L_{i,t}^{old} &= \int_{\Omega_{i,t}^{old}} \left(\sum_{j \in S} \tau_{ij} \frac{q_{ij}(\omega)}{\bar{A}_{i,t}} \right) d\omega \iff \\ L_{i,t}^{old} &= M_{it}^{new} \frac{1-\chi}{\chi} \rho f_i, \end{aligned}$$

where the second to last line used the equations for the old and new variety price indices from equations (60) and (66).

Total labor used by all firms is hence:

$$\begin{aligned} L_{i,t}^{new} + L_{i,t}^{old} &= L_{i,t} \iff \\ M_{i,t}^{new} &= \chi \frac{L_{i,t}}{\rho f_i}, \end{aligned} \tag{67}$$

so that the measure of new firms is proportional to the labor supply.

The micro-foundation Combining the old and new variety price indices from equations (60) and (66) yields:

$$\left(\left(P_{i,t}^{new}\right)^{\chi} \left(P_{i,t}^{old}\right)^{1-\chi} \right)^{1-\sigma} = \left(c_{i,t}\right)^{1-\sigma} \frac{\rho}{\rho-1}^{(1-\sigma)\chi} \left(M_{it}^{new}\right)^{\chi\left(\frac{1-\sigma}{1-\rho}\right)} \left(M_{it}^{old}\right)^{(1-\chi)\left(\frac{1-\sigma}{1-\rho}\right)}$$

Total trade flows from $i \in S$ to $j \in S$ in time t is determined by simply aggregating across all firms of both types. The total trade of new varieties is:

$$X_{ijt}^{new} = \int_{\Omega_{i,t}^{new}} p_{ij,t}(\omega) \, q_{ij,t}(\omega) \, d\omega \iff$$

$$X_{ijt}^{new} = \chi \frac{\left(\tau_{ij} c_{i,t}\right)^{1-\sigma} \left(M_{it}^{new}\right)^{\chi\left(\frac{1-\sigma}{1-\rho}\right)} \left(M_{it}^{old}\right)^{(1-\chi)\left(\frac{1-\sigma}{1-\rho}\right)}}{\sum_{k \in S} \left(\tau_{kj} c_{k,t}\right)^{1-\sigma} \left(M_{kt}^{new}\right)^{\chi\left(\frac{1-\sigma}{1-\rho}\right)} \left(M_{kt}^{old}\right)^{(1-\chi)\left(\frac{1-\sigma}{1-\rho}\right)}} Y_{j,t}$$

Similarly, the total trade of existing varieties is:

$$X_{ijt}^{old} = \int_{\Omega_{i,t}^{old}} p_{ij,t}(\omega) \, q_{ij,t}(\omega) \, d\omega \iff$$

$$X_{ijt}^{old} = (1-\chi) \frac{\left(\tau_{ij} c_{i,t}\right)^{1-\sigma} \left(M_{it}^{new}\right)^{\chi\left(\frac{1-\sigma}{1-\rho}\right)} \left(M_{it}^{old}\right)^{(1-\chi)\left(\frac{1-\sigma}{1-\rho}\right)}}{\sum_{k \in S} \left(\tau_{kj} c_{k,t}\right)^{1-\sigma} \left(M_{kt}^{new}\right)^{\chi\left(\frac{1-\sigma}{1-\rho}\right)} \left(M_{kt}^{old}\right)^{(1-\chi)\left(\frac{1-\sigma}{1-\rho}\right)}} Y_{j,t}$$

so that the total trade flows are:

$$X_{ij,t} = X_{ij,t}^{new} + X_{ij,t}^{old} \iff$$
$$X_{ij,t} = \tau_{ij}^{1-\sigma} w_{i,t}^{1-\sigma} A_{i,t}^{\sigma-1} P_{j,t}^{\sigma-1} Y_{j,t},$$

where:

$$P_{j,t}^{1-\sigma} \equiv \sum_{k \in S} \tau_{kj}^{1-\sigma} w_{k,t}^{1-\sigma} A_{k,t}^{\sigma-1}$$

and:

$$A_{i,t} \equiv \bar{A}_{i,t} f_i^{\frac{1}{\rho-1}} \times L_{i,t}^{\alpha_1} \times L_{i,t-1}^{\alpha_2}$$

and $\alpha_1 \equiv \frac{\chi}{\rho-1}$ and $\alpha_2 \equiv \frac{1-\chi}{\rho-1}$, as claimed.

B.1.2 Durable Investment

Setup In each location $i \in S$, there There is a measure of firms in all locations, each endowed with that compete a la Bertrand. Firms can hire workers either to produce or to innovate, where the total quantity produced at location $i \in S$ and time t depends on the amount of labor used in the production $L_{i,t}$, the amount of land $H_{i,t}$, the amount of innovation $\phi_{i,t}$ and some productivity shifter $B_{i,t}$:

$$Q_{i,t} = \phi_{i,t}^{\gamma_1} B_{i,t} L_{i,t}^{\mu} H_{i,t}^{1-\mu} \iff$$
$$q_{i,t} = \phi_{i,t}^{\gamma_1} B_{i,t} l_{i,t}^{\mu},$$

where in what follows we focus on the output per unit land $q_{i,t}$ and the labor per unit land $l_{i,t}$. We assume the parameters $\mu < 1$ (due to the diminishing marginal product of labor per unit land) and $\gamma_1 < 1$ (due to the diminishing marginal product of innovation).

To employ a level of innovation $\phi_{i,t}$, a firm must hire $\nu \phi_{i,t}^{\xi}$ additional units of labor, where

$\xi < \gamma_1/(1-\mu)$. We assume that innovation today has an affect on the level of productivity tomorrow so that:

$$B_{i,t} = \phi_{i,t-1}^{\delta\gamma_1} \bar{B}_{i,t}, \tag{68}$$

where $\bar{B}_{i,t}$ is an exogenous shock and $\delta < 1$ indicates the extent to which innovation decays from one period to the next.

We assume the cost per unit of land $r_{i,t}$ is determined by a competitive auction, so that firms obtain zero profits.

Profit maximization Even though innovations today affect innovations in future periods, because firms earn zero profits in the future, the dynamic problem reduces to a sequence of static profit maximizing problems (see Desment and Rossi-Hansberg '14).

As a result the firms profit maximization problem becomes:

$$\max_{l_{i,t},\phi_{i,t}} p_{i,t} B_{i,t} \left(\phi_{i,t}^{\gamma_1}\right) \times \left(l_{i,t}^{\mu}\right) - w_{i,t} \underbrace{l_{i,t}}_{\#\text{ of production workers}} - w_{i,t} \underbrace{\left(\nu\phi_{i,t}^{\xi}\right)}_{\#\text{ of innovation workers}} - r_{i,t}$$

which has the following first order conditions:

$$\begin{aligned} \gamma_1 B_{i,t} p_{it} \phi_{i,t}^{\gamma_1 - 1} l_{i,t}^{\mu} &= \xi \nu w_{i,t} \phi_{i,t}^{\xi-1} \\ \mu B_{i,t} p_{i,t} \phi_{i,t}^{\gamma_1} l_{i,t}^{\mu-1} &= w_{i,t} \end{aligned}$$

which combined yields:

$$\begin{aligned} \frac{\gamma_1}{\mu} l_{i,t} &= \xi\nu\phi_{i,t}^{\xi} \iff \\ \left(\frac{\gamma_1}{\mu\xi\nu} l_{i,t}\right)^{\frac{1}{\xi}} &= \phi_{i,t} \end{aligned} \tag{69}$$

Total employment $\tilde{l}_{i,t}$ per unit land is equal to the sum of the production workers and the innovation workers:

$$\begin{aligned} \tilde{l}_{i,t} &= l_{i,t} + \nu\phi_{i,t}^{\xi} \iff \\ \tilde{l}_{i,t} &= \left(1 + \frac{\gamma_1}{\mu\xi}\right) l_{i,t} \end{aligned}$$

Rent and income Equilibrium rent ensures profits per unit land are equal to zero:

$$\begin{aligned} r_{i,t} &= B_{i,t} p_{i,t} \phi_{i,t}^{\gamma_1} l_{i,t}^{\mu} + w_{i,t} l_{i,t} + \nu w_{i,t} \phi_{i,t}^{\xi} \iff \\ r_{i,t} &= \left(\frac{1}{\mu} + 1 + \frac{\gamma_1}{\mu\xi}\right) w_{i,t} l_{i,t} \end{aligned}$$

Note that total income per unit labor in a location is:

$$\begin{aligned} Y_{i,t} &= r_{i,t}H_{i,t} + w_{i,t}\tilde{L}_{i,t} \iff \\ \frac{Y_{i,t}}{\tilde{L}_{i,t}} &= \left(\frac{\frac{1}{\mu} + 1 + \frac{\gamma_1}{\mu\xi}}{\left(1 + \frac{\gamma_1}{\mu\xi}\right)} + 1\right) w_{i,t} \end{aligned}$$

The productivity microfoundation The output price is:

$$\begin{aligned} \mu B_{i,t} p_{i,t} \phi_{i,t}^{\gamma_1} L_{i,t}^{\mu-1} = w_{i,t} \iff \\ p_{i,t} = \frac{1}{B_{i,t}} \left(\frac{1}{\mu}\left(\frac{\xi\nu\mu}{\gamma_1}\right)^{\frac{\gamma_1}{\xi}}\right) w_{i,t} l_{i,t}^{1-\mu-\frac{\gamma_1}{\xi}} \end{aligned}$$

total output is:

$$\begin{aligned} q_{i,t} &= \phi_{i,t}^{\gamma_1} B_{i,t} l_{i,t}^{\mu} \iff \\ Q_{i,t} &= \left(\frac{\gamma_1}{\mu\xi\nu}\right)^{\frac{\gamma_1}{\xi}} B_{i,t} \tilde{L}_{i,t}^{\mu+\frac{\gamma_1}{\xi}} H_{i,t}^{1-\mu-\frac{\gamma_1}{\xi}}, \end{aligned}$$

where $\tilde{L}_{i,t}$ is total employment in location i at time t. Combining equations (68) and (69) yields:

$$\begin{aligned} B_{i,t} &= \phi_{i,t-1}^{\delta\gamma_1} \bar{B}_{i,t} \iff \\ B_{i,t} &= \left(\frac{\frac{\gamma_1}{\mu\xi\nu}}{\left(1 + \frac{\gamma_1}{\mu\xi}\right)} \frac{\tilde{L}_{i,t-1}}{H_{i,t-1}}\right)^{\delta\frac{\gamma_1}{\xi}} \bar{B}_{i,t} \end{aligned}$$

so that in total we have:

$$\begin{aligned} Q_{i,t} &= \left(\frac{\gamma_1}{\mu\xi\nu}\right)^{\frac{\gamma_1}{\xi}} \left(\left(\frac{\frac{\gamma_1}{\mu\xi\nu}}{\left(1 + \frac{\gamma_1}{\mu\xi}\right)} \frac{\tilde{L}_{i,t-1}}{H_{i,t-1}}\right)^{\delta\frac{\gamma_1}{\xi}} \bar{B}_{i,t}\right) \tilde{L}_{i,t}^{\mu+\frac{\gamma_1}{\xi}} H_{i,t}^{1-\mu-\frac{\gamma_1}{\xi}} \iff \\ Q_{i,t} &= \bar{A}_{i,t} \tilde{L}_{i,t}^{\alpha_1} \tilde{L}_{i,t-1}^{\alpha_2} \tilde{L}_{i,t}, \end{aligned}$$

where $\bar{A}_{i,t} \equiv \left(\frac{\gamma_1}{\mu\xi\nu}\right)^{\frac{\gamma_1}{\xi}} \left(\left(\frac{\frac{\gamma_1}{\mu\xi\nu}}{\left(1+\frac{\gamma_1}{\mu\xi}\right)} \frac{\tilde{L}_{i,t-1}}{H_{i,t-1}}\right)^{\frac{\gamma_1}{\xi}} \bar{B}_{i,t}\right) H_{i,t}^{1-\mu-\frac{\gamma_1}{\xi}}$, $\alpha_1 \equiv \frac{\gamma_1}{\xi} - (1-\mu)$, and $\alpha_2 \equiv \delta\frac{\gamma_1}{\xi}$, as required.

B.2 Amenity spillover

B.2.1 Setup

Demand Suppose that consumers have Cobb-Douglas preferences over land and a consumption good, so that their indirect utility function can be written as:

$$W_{i,t} = \frac{(Y_{i,t}/L_{i,t})}{(P_{i,t})^{\lambda} \left(r^H_{i,t}\right)^{1-\lambda}},$$

where $r^H_{i,t}$ is the rental cost of housing. Let $H_{i,t}$ denote the (equilibrium quantity) of housing and let K_i denote the (exogenous) quantity of land in a location, so that $h_{i,t} \equiv H_{i,t}/K_{i,t}$ is the housing density (e.g. square feet of housing per acre of land).

Given the Cobb-Douglas preferences (and, from balanced trade, that income equals expenditure, $Y_{i,t} = E_{i,t}$), we have:

$$r^H_{i,t} H_{i,t} = (1-\lambda) Y_{i,t}$$

$$w_{i,t} L_{i,t} = \lambda Y_{i,t}$$

so that we can write the payment to housing as a function of the payment to labor:

$$r^H_{i,t} = \left(\frac{1-\lambda}{\lambda}\right) \frac{1}{H_{i,t}} w_{i,t} L_{i,t}$$

Note then that we can write:

$$W_{i,t} = \frac{(Y_{i,t}/L_{i,t})}{(P_{i,t})^{\lambda} \left(r^H_{i,t}\right)^{1-\lambda}} \iff$$

$$\tilde{W}_{i,t} = \frac{1}{\lambda (1-\lambda)^{\frac{1-\lambda}{\lambda}}} \frac{w_{i,t}}{P_{i,t}} \left(\frac{H_{i,t}}{L_{i,t}}\right)^{\frac{1-\lambda}{\lambda}}, \tag{70}$$

where $\tilde{W}_{i,t} \equiv W_{i,t}^{\frac{1}{\lambda}}$ is a positive monotonic transform of $W_{i,t}$ and hence can be our measure of welfare.

Supply We now determine the equilibrium stock of housing $H_{i,t}$. Suppose that each unit of land is owned by a representative developer, who decides how much to upgrade the housing tract. The amount of housing per unit land ($h_{it} \equiv \frac{H_{i,t}}{K_i}$) is a function of the housing stock that has survived from the previous period $\left(h^{existing}_{i,t} \equiv \frac{H^{existing}_{i,t}}{K_i}\right)$ and the amount of labor that the firm chooses to hire to rebuild it:

$$h_{i,t} = \left(h^{existing}_{i,t}\right)^{\mu} \left(l^d_{i,t}\right)^{1-\mu} \iff$$
$$H_{i,t} = \left(H^{existing}_{i,t}\right)^{\mu} \left(L^d_{i,t}\right)^{1-\mu}$$

In what follows, we assume for simplicity that the existing housing stock from period $t-1$ in period t is some fraction of the development in the previous period:

$$H_{i,t}^{existing} = \bar{C}_{i,t} \left(L_{i,t-1}^{d}\right)^{\rho}, \tag{71}$$

where $\bar{C}_{i,t}$ is an (exogenous) shock.

B.2.2 Profit maximization

A developer solves:

$$\begin{aligned} \max_{l_{i,t}^{d}} r_{i,t}^{H} h_{i,t} - w_{i,t} l_{i,t}^{d} - f_{i,t} &\iff \\ \max_{l_{i,t}^{d}} r_{i,t}^{H} \left(h_{i,t}^{existing}\right)^{\mu} \left(l_{i,t}^{d}\right)^{1-\mu} - w_{i,t} l_{i,t}^{d} - f_{i,t}&, \end{aligned}$$

where $f_{i,t}$ is a fixed cost (a "permit cost") that is remitted back to local residents and is set via a competitive bid, ensuring that the firm earns zero profits (and hence the dynamic problem simplifies into a series of static profit maximization problems, as above).

First order conditions are:

$$\begin{aligned} (1-\mu) r_{i,t}^{H} \left(h_{i,t}^{existing}\right)^{\mu} \left(l_{i,t}^{d}\right)^{-\mu} &= w_{i,t} \iff \\ \left(h_{i,t}^{existing}\right)^{\mu} \left(l_{i,t}^{d}\right)^{1-\mu} &= \frac{1}{1-\mu} \frac{1}{r_{i,t}^{H}} w_{i,t} l_{i,t}^{d} \end{aligned}$$

Note that the fixed "permit costs" are then:

$$\begin{aligned} f_{i,t} &= r_{i,t}^{H} \left(h_{i,t}^{existing}\right)^{\mu} \left(l_{i,t}^{d}\right)^{1-\mu} - w_{i,t} l_{i,t}^{d} \iff \\ f_{i,t} &= \left(\frac{\mu}{1-\mu}\right) w_{i,t} l_{i,t}^{d}, \end{aligned}$$

which recall are remitted to workers and ensure profits are zero.

We can combine this with the rental rate above to calculate the fraction of workers hired in the development of the land:

$$\begin{aligned} h_{i,t} = \left(h_{i,t}^{existing}\right)^{\mu} \left(l_{i,t}^{d}\right)^{1-\mu} &\iff \\ (1-\mu)\left(\frac{1-\lambda}{\lambda}\right) L_{i,t} = L_{i,t}^{d}&, \end{aligned}$$

so we require as a parametric restriction (so that only a fraction of workers are hired as local developers):

$$\begin{aligned} (1-\mu)\left(\frac{1-\lambda}{\lambda}\right) < 1 &\iff \\ 1 < \mu + \lambda + (1-\mu)\lambda& \end{aligned}$$

Since a constant fraction of local workers are hired, we can express the housing density solely as a function of the local population, the local land area, and the

$$\begin{aligned}
h_{i,t} &= \left(h_{i,t}^{existing}\right)^{\mu} \left(l_{i,t}^{d}\right)^{1-\mu} \iff \\
H_{i,t} &= \left((1-\mu)\left(\frac{1-\lambda}{\lambda}\right)\right)^{(1-\mu)+\rho\mu} \bar{C}_{i,t}^{\mu} \left(L_{i,t-1}\right)^{\rho\mu} \left(L_{i,t}\right)^{1-\mu} \qquad (72)
\end{aligned}$$

B.2.3 Microfoundation

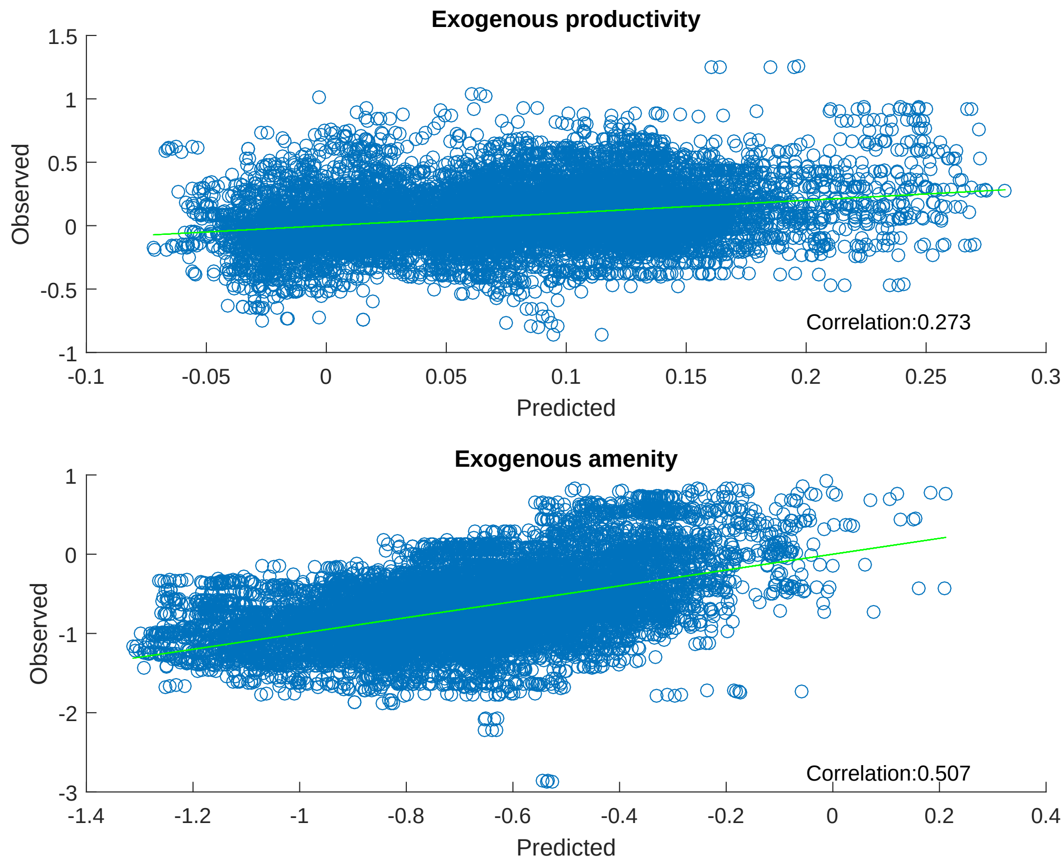
We substitute equation (72) for the equilibrium stock of housing into the welfare equation (70) to yield:

$$\begin{aligned}
\tilde{W}_{i,t} &= \frac{1}{\lambda\left(1-\lambda\right)^{\frac{1-\lambda}{\lambda}}} \frac{w_{i,t}}{P_{i,t}} \left(\frac{H_{i,t}}{L_{i,t}}\right)^{\frac{1-\lambda}{\lambda}} \iff \\
\tilde{W}_{i,t} &= \frac{w_{i,t}}{P_{i,t}} \bar{u}_{i,t} L_{i,t}^{\beta_1} L_{i,t-1}^{\beta_2},
\end{aligned}$$

where $\bar{u}_{it} \equiv \frac{1}{\lambda(1-\lambda)^{\frac{1-\lambda}{\lambda}}} \left((1-\mu)\left(\frac{1-\lambda}{\lambda}\right)\right)^{\frac{1-\lambda}{\lambda}((1-\mu)+\rho\mu)} \bar{C}_{i,t}^{\frac{1-\lambda}{\lambda}}$, $\beta_1 \equiv -\mu\frac{1-\lambda}{\lambda}$, and $\beta_2 \equiv \rho\mu\frac{1-\lambda}{\lambda}$.

C Additional Tables and Figures

Figure 15: The "0th stage": Predicting productivities and amenities from geographic observables



Notes: This figure shows the relationship between the observed productivities and amenities in the year 2000 (given estimates of migration and trade frictions and candidate elasticity values) and the predicted productivities and amenities using variation in observed geographic variables, namely climatic variables average January temperature and precipitation), soil quality variables (the net primary productivity and soil nutrient availability), and topographic variables (elevation and ruggedness). The (time-invariant) predicted productivities and amenities are then used to construct the model-based instruments used in the 2SLS procedure detailed in Section 3.2.