History and the sizes of cities

Hoyt Bleakley and Jeffrey Lin*

December 31, 2014

Abstract

We discuss the tension between some of the recent evidence of path dependence in urban location with recent efforts to analyze calibrated models of city sizes. One strand of recent work, including some of our own, finds evidence of persistent city sizes following the obsolescence of historical advantages. This literature argues that such persistence cannot be understood as the medium-run effect of legacy capital, but rather as the long-run effect of equilibrium selection. In contrast, a different, recent literature uses stylized models that feature a single long-run equilibrium at each site/city. We show this disjunction in a standard model and then propose several modifications that might allow for multiplicity and thereby historical path dependence. We use mixed-integer-programming solvers to construct bounds on the degree of multiplicity of steady states across US Counties. We find, absent strong parametric restrictions, very wide bounds on the scope for multiple long-run equilibria. We then discuss a way forward for tightening these bounds using the obsolete endowments as restrictions on the model.

1 Introduction

Are city sizes uniquely determined by locational fundamentals? Or, in the presence of localized aggregate increasing returns, are city sizes instead characterized by multiple steady-states, and thus perhaps influenced by history? Because the natural advantages that first attracted economic activity to a particular location are usually persistent, separately identifying the effects of fundamentals versus history is difficult.

We attempt to deal with this identification problem by studying the role of long-obsolete endowments in determining the sizes of cities. What these natural advantages have in common—whether a convenient place to carry cargo around rapids or over rivers, a source of water power, or the confluence of two rivers—is that they were made obsolete by new technologies a century or more ago. The lasting footprint of these obsolete endowments on the spatial distribution of activity, however, suggests that city sizes are *not* uniquely determined by locational fundamentals. If they were, then cities near these features should have declined with the value of their initial natural advantages. Instead, the persistence of these cities suggests that there may be multiple steady-state spatial distributions of economic activity, with history playing in an important role in equilibrium selection. In addition, evidence of path dependence has important implications for quantitative models of city sizes.

2 The footprint of history

Many cities founded near historical portage sites persist today (Hoyt Bleakley and Jeffrey Lin, 2012*a*). During the early settlement of North America, portage sites were convenient places for carrying a boat and its cargo over land around some obstacle to navigation. These obstacles obliged traders to get out of their canoes, which made such sites focal points for commerce. But this natural advantage became obsolete a century or more ago, thanks to improvements in transportation technology. Similarly, some falls at these sites provided water power during early industrialization, an advantage that was made obsolete by the advent of other, cheaper power sources.

Yet the footprint of portage is evident, even today. Prominent examples can be found along rivers at the intersection of the *fall line*—a geomorphological feature dividing the Piedmont and the coastal plain, describing the last set of falls or rapids experienced before emptying into the Atlantic Ocean. Along rivers in the colonial era, towns tended not to form in the
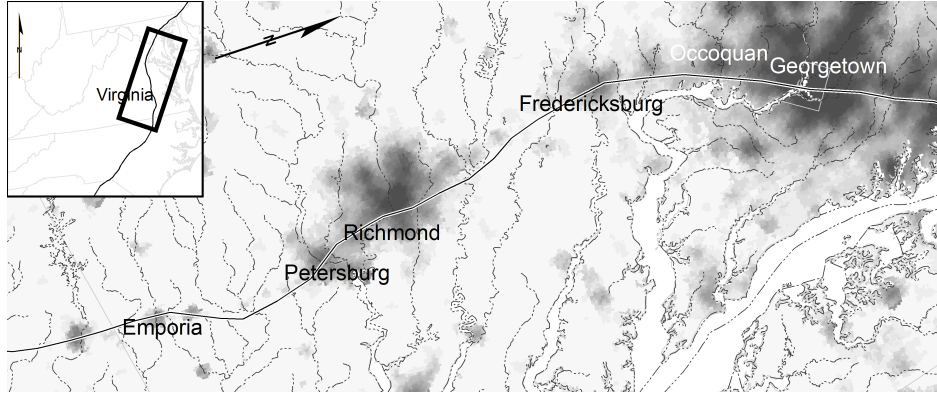
Figure 1: Fall line portage cities in contemporary Virginia

This map shows major cities at the intersections of the fall line (solid line) and rivers (dot-dash), and the present-day distribution of economic activity using 2003 nighttime lights from NationalAtlas.gov.

coastal plain, but instead at the fall line where the obstacle to water transport required the offloading of goods sourced upstream. Thomas Jefferson (1781) noted this phenomenon in *Notes on the State of Virginia*:

> [The Tidewater area] being much intersected with navigable waters, and trade brought generally to our doors, instead of our being obliged to go in quest of it, has probably been one of the causes why we have no towns of any consequence.

Today, these rivers are no longer used for commercial transportation, yet the major cities of Virginia continue to persist at fall line portage sites. Figure 1 shows several portage cities—among them Richmond (at the falls of the James River), Petersburg (Appomattox), Fredericksburg (Rappahannock)—in present-day Virginia.

Importantly, portage cities did not shrink over time compared to either the average location or locations that were similarly dense historically. Nor does any specific legacy capital (such as infrastructure, housing, sectoral composition, or literacy) explain the persistence of portage city sizes. Finally, the relatively smooth landscapes of the coastal plain and Piedmont suggests an absence of persistent geographic factors that might explain persistence in other historical contexts (Donald R. Davis and David E. Weinstein., 2002; Sanghoon Lee and Jeffrey Lin, 2013).

A recent literature finds other important effects of temporary historical factors on the sizes and types of cities. For example, German division
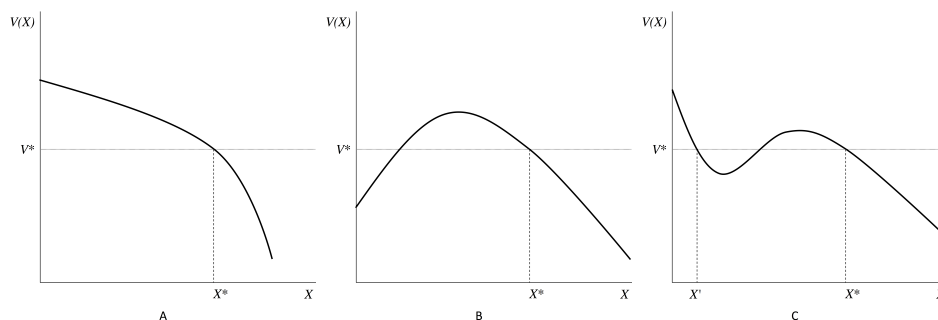
Figure 2: Utility paths with unique (A,B) and multiple (C) equilibrium city sizes

resulted in a permanent diversion of air traffic from Berlin to Frankfurt (Stephen J. Redding, Daniel M. Sturm and Nikolaus Wolf, 2011). Dramatic but temporary reductions in the supply of raw cotton to the British textile industry during the U.S. Civil War had a long-run impact on English towns where cotton production had been concentrated before the war (Walker Hanlon, 2014). Within Manhattan, historical marshes affect housing prices even today, despite sewers having rendered their initial disadvantages moot (Carlos Villareal, 2012). And rail lines that are subsequently scuttled appear to have permanent effects on the spatial distribution of activity, both across cities (Remi Jedwab and Alexander Moradi, 2014) and within cities (Leah Brooks and Byron Lutz, 2014). Intriguingly, the location of Roman towns predicts later city locations in France, but not Britain—suggesting that there may be significant consequences to path dependence in city sizes (Guy Michaels and Ferdinand Rauch, 2013).

3 History and theory

What, then, can explain persistence in city sizes at these sites, if natural advantages went obsolete and historical legacy capital has long ago depreciated? Localized aggregate increasing returns are a natural explanation for path dependence in city sizes. Importantly, in the presence of increasing returns, a particular location may feature multiple equilibrium city sizes. Then, persistent differences in city size can be rationalized even in the absence of differences in natural advantages or sunk capital.

To see this, note that equilibrium models of city sizes typically include

both agglomeration forces and congestion forces. Larger cities trade off superior agglomeration benefits against higher congestion costs. For a particular location, a convenient way to describe equilibrium is to derive indirect utility $V(X)$ as a function of total city population or employment X. In long-run spatial equilibrium, city size is determined so that the marginal mobile household receives the same utility in that city as a reservation level of utility (which may be endogenous) available in other cities.

The shapes of these utility paths vary with the number and types of centripetal and centrifugal forces considered. Figure 2 shows these utility paths as a function of total city size X, under three different assumptions about the number and types of agglomeration and congestion forces.

The discussion that follows is relevant to a recent literature that has attempted to quantify equilibrium models of city sizes.[1] Many assume functional forms or parameter values to guarantee unique equilibrium city sizes. But if we want to take path dependence seriously, it may be useful and important to work with a more flexible model that admits more features, including multiple equilibria.[2]

A model featuring a single congestion force is the easiest to analyze. For example, each location might feature a fixed land endowment that is diluted with increasing population.[3] As there are no countervailing agglomeration economies, the essential feature of such a model is a downward sloping pseudo-demand curve for labor at a location (Panel A). Then, there is a unique long-run equilibrium city size X^*, shown where the utility path intersects V^*, the reservation level of utility in spatial equilibrium. In this model, differences in city sizes can be rationalized by innate variation in natural production or consumption amenities, which correspond to vertical translations of the utility path. However, in the (very) long run, if labor and households are mobile, then persistent differences in city sizes are difficult to explain absent differences in fundamentals. In this simple model, there can be no long-run history dependence.

Even the addition of a standard agglomerative force may not yield predictions consistent with path dependence. For example, consider a pro-

[1]Some recent examples include David Albouy and Bryan Stuart (2014); Satyajit Chatterjee (2006); Rebecca Diamond (2013); Klaus Desmet and Esteban Rossi-Hansberg (2013); Andrew F. Haughwout and Robert P. Inman (2001); Sanghoon Lee and Qiang Li (2013); and Jordan Rappaport (2008a, 2008b).

[2]Elhanan Helpman (1998) shows that the welfare properties of equilibrium are very different in models featuring unique versus multiple equilibria.

[3]Alternatively, a congestion externality with a Poisson arrival process might be used to justify an exponential functional form..

duction function of the form $Y = \bar{X}^\delta f(X, L, K)$, where $f()$ is a firm-level constant returns to scale production function in labor, land, and capital, and $0 < \delta < 1$ is the degree of (external) increasing returns to scale in the city-level employment $(\bar{X})$.[4] For δ large enough, a typical utility path under these assumptions features a single-peaked hump shape (Panel B).

Though there are two points where the utility path crosses the reservation utility level, only the larger city size is a stable equilibrium. (At the smaller size, utility is rising with city size, inducing factor movements into the city.) Thus, these assumptions yield the result of a unique equilibrium city size for each location.[5] Two cities may be of different size because of variation in locational fundamentals or the degree of increasing returns. But again, it is difficult in this framework to explain path dependence in city sizes.

Notice, too, that under this specification for agglomeration economies, the model has very strong predictions about the pattern of city sizes. In particular, such a model has difficulty explaining the preponderance of small, non-empty cities without requiring them to be in an unstable equilibrium.[6] This difficulty arises because of how agglomeration economies are modeled: one typology of agglomeration economies might distinguish the effects of city size on marginal product at very small scales. Modeling spillover benefits as X^δ implies an infinite marginal agglomeration benefit or spillovers near zero. At city size increases, the marginal spillover benefits strictly decline.

Some models of agglomeration have this property—for example, ones based on matching. Peter Diamond's (1982) coconut model is a leading case, but see also Kevin M. Murphy (1986) and Bleakley and Lin (2012b). But agglomeration economies might have the opposite pattern: negligible effects at very small scales that don't kick in until some important threshold is crossed. For example, the presence of fixed costs implies variation across industries in minimum efficient scale. Models of a "big push" have this flavor (e.g., Murphy, Schleifer, and Vishny, 1993).

For these reasons, it seems desirable to choose a different set of assumptions about agglomeration and congestion forces. Helpman's (1998) model addresses many of these issues, using fixed costs, transport costs, and a fixed

---

[4]J.V. Henderson's (1974) canonical model includes a similar specification for increasing returns to scale at the city level. His congestion force is an exponential commuting cost that increases with the size of the city. Several of the above-cited studies use this $\bar{X}^\delta$ form of agglomeration.

[5]There may also exist an equilibrium city size of zero, but this is unstable by the same argument.

[6]This observation was made by Chatterjee (2006).

endowment of housing to generate utility paths as in Panel C. In this formulation, utility paths are S-shaped, and the S-shape reflects different ranges of city size where congestion or agglomeration forces dominate. There are now two stable, nonzero equilibrium city sizes, labeled X' and X^*. Note that locations that are otherwise identical in terms of locational fundamentals might have very different long-run city sizes. Across locations, which equilibria are selected might depend on history (Paul Krugman, 1991). Thus, a temporary shock to fundamentals might lead to persistent differences. The additional curvature in the utility paths yields multiplicity in equilibrium city sizes, providing an intriguing and attractive way to rationalize path dependence in city sizes.

However, note that such a model would have now have difficulty in explaining *intermediate* city sizes, corresponding to the range where utility is upward sloping, without requiring them to be in unstable equilibrium. (Note that the empirical distribution of city sizes does not exhibit bimodality around X' and X^*.) This is because of the way locational fundamentals enter the model: natural production or consumption amenities shift utility paths across locations vertically, up or down. A positive amenity is neutral with respect to density, thus yielding the same proportional benefit to a large or small city.

Reflecting on differences across places, this neutrality assumption seems at best incomplete. One can think of a whole host of investments or endowments that might complement density. Infrastructure investments, for example, are a way to reduce the negative externalities from congestion at a given density; so, too, might congestion pricing complement density (Jeffrey Brinkman, 2013). Geographic barriers to development might also affect how congestion responds to city size (Mariaflavia Harari, 2014). Or, echoing Henderson's (1974) explanation, cities may vary in industrial composition and therefore the degree of increasing returns.

The non-neutrality of fundamentals echoes the time-honored concept of "economic base" from regional economics. Economic base matters because—roughly speaking—products that can be exported are not subject to a sharply downward sloping local demand curve. Locations endowed with better transport access to the rest of the world face, in effect, a more elastic demand curve for their exports. Market access complements density because sites with better access can get larger without depressing prices received for their output.

4   Model

We outline a static, multi-region, general equilibrium model of city sizes that can rationalize persistent differences in population densities across locations, even absent differences in natural advantages. The model is a simplified, multi-region version of Helpman (1998). Our main departure is to model agglomeration economies and congestion costs in a reduced-form way, versus from micro-foundations as in Helpman. This is helpful for quantitatively matching the data (notably, by not restricting agglomeration and congestion forces to be identical across locations), while (1) accommodating multiple equilibria and (2) keeping the model parsimonious.

There are two kinds of agglomerating forces: heterogeneity in the natural amenity or productivity value of locations and scale economies to the geographic concentration of economic activity. These centripetal forces are balanced with congestion costs and scarce fixed factors that discourage all economic activity from concentrating in a single black-hole location.

We show that the model, given the observed data and a handful of calibrated parameters, is log linear in unobserved parameters (productivity shifters, e.g.). This allows us to use linear programming and mixed-integer linear programming to solve the model efficiently.

4.1   Setup

There are i distinct geographic areas ($i = 1, 2, 3, ..., I$), in which i is a large number covering the entire spatial economy. Household location choice is the central endogenous decision in the model, determining the size of regions.

The traded good Y is produced at the plant level according to $y = \lambda \phi_i \sigma(X_i) f(x, k, h)$, in which λ is an economy-wide productivity shifter, ϕ_i is a local natural advantage that affects productivity, and x, k, and h_p are labor, capital, and land inputs at the plant level. Capital is freely traded at a national rental rate r; local wages w and land prices p are determined in equilibrium. Land is inelastically supplied and non-traded.

Each plant's production function is constant returns to scale and Cobb Douglas: $f \equiv x^\beta k^\omega (h_p)^{1-\beta-\omega}$. Also, a reduced-form function of total population X_i, $\sigma(X_i)$ is modeled as

$$\log \sigma(X_i) \equiv \left[\delta_1 \log X_i + \delta_2 (\log X_i)^2 \right],$$

and accounts for local scale economies not internalized by plants.[7] While an individual plant may face constant returns to scale in its own factor inputs

---

[7]This parameterization is useful for quantitatively matching the model to data; see

(i.e., in f), its choice of factors can create positive externalities to other plants in the same location—for instance, by raising the marginal product of other factors in that location. Goods are costlessly traded across locations. Traded goods plants behave competitively, taking populations and factor prices as given.

People receive income from inelastically supplying one unit of labor at wage w.[8] Utility is $U = y^{1-\theta}(h_c)^{\theta}\psi\tau(X_i)$, where h_c is consumption of land, ψ_i is the value of local natural amenities in consumption, and $\tau(X_i) \equiv e^{-\gamma X_i}$ is a reduced-form function that captures disutility from crowdedness.

4.2 Equilibrium

In equilibrium, local factor prices w_i, p_i, local shifters ϕ_i, ψ_i and factors X_i, K_i are such that marginal products equal factor prices; factor, goods, and land markets clear; and utility is equal across locations. Equations 1–4.2 define an equilibrium at each site. (Here local subscripts i are suppressed.) These conditions must hold for each site and a global population constraint must also be satisfied.

Aggregate production in each location is

$$Y = \lambda\phi\sigma(X)X^{\beta}K^{\omega}\left[(1-\alpha)H\right]^{1-\beta-\omega} \tag{1}$$

$$\sigma(X) = \exp\left[\delta_1 \log X + \delta_2 (\log X)^2\right], \tag{2}$$

where $(1-\alpha)$ is the share of land H devoted to production (with the balance left for consumption), determined in equilibrium. Equilibrium $\alpha = \frac{\theta\beta}{1-\beta-\omega+\theta\beta}$.

Incorporating the budget constraint and land market clearing, indirect utility is:

$$\overline{V} = V \equiv [w(1-\theta)]^{(1-\theta)}\left(\frac{\alpha H}{X}\right)^{\theta}\psi e^{-\gamma X}, \tag{3}$$

in which $\overline{V}$ is equilibrium utility at every location.

---

below. It also has the attractive features of no spillovers for a local employment of one person ($\sigma(X = 1) = 1$) and that agglomeration spillovers are non-declining for $X = \geq 1$ and $\delta_1, \delta_2 \geq 0$. Note that, for $\delta_2 = 0$, this collapses to a commonly used X^{δ} specification for agglomeration spillovers.

[8]Payments to land and capital are made to absentee landlords. Helpman (1998) uses these payments rebated lump-sum to workers and the consumption of rival non-traded goods (e.g., land for housing) to generate disutility from crowdedness from micro-foundations. Instead, we assume absentee landlords and introduce a reduced-form function of population τ to achieve a similar effect.

Finally, marginal products equal factor prices:

$$w = \frac{\beta Y}{X}; r = \frac{\omega Y}{K}; p = \frac{(1 - \beta - \omega)Y}{(1 - \alpha)H}$$

Following Helpman, equilibrium can be solved numerically. A convenient way to describe possible equilibrium configurations at a single site i is to derive indirect utility $V(X_i)$ as a function of a single endogenous variable, population X_i, and parameters:

$$\log \bar{V} = \log V(X_i) \tag{4}$$

$$= C + \left(\frac{1 - \theta}{1 - \omega}\right) \left[\delta_1 \log X_i + \delta_2 (\log X_i)^2\right]$$

$$- (1 - \theta) \left(\frac{1 - \omega - \beta}{1 - \omega}\right) \log X_i$$

$$- \theta \log X_i - \gamma X_i \tag{5}$$

$$C \equiv \left(\frac{1 - \theta}{1 - \omega}\right) \log \phi_i + \log \psi_i$$

$$+ (1 - \theta) \left(\frac{1 - \omega - \beta}{1 - \omega}\right) \log H_i + \theta \log H_i$$

$$+ F(\theta, \beta, \omega, \lambda, r), \tag{6}$$

where $F()$ is a function of global parameters. (The endogenous variables for wages, land prices, and the capital stock were solved out.) Note that in equation 4, δ_1 and δ_2 govern the strength of agglomeration economies (conditional on X_i in the first term after C. The subsequent terms capture decreasing marginal product of labor, crowding of the fixed factor (land) in consumption, and general congestion costs, respectively. Equation 6 describes how the constant term C is composed of site-specific natural advantages and land area, as well as global parameters.

Depending on parameters, a site may feature one of two types of equilibrium configurations. These two cases are shown in figure 3. Each panel shows $V(X_i)$ as a function of population X_i for a particular location. Equilibrium population densities can be seen by the point where the $V(X_i)$ curve intersects the equilibrium utility across locations, $\bar{V}$.

In case 1 (left panel), if agglomeration economies are weak, then $V(X_i)$ is monotonically decreasing in X_i. Thus, conditioned on these parameters there is a unique equilibrium population at this site.

Note that if *every* site features parameters ensuring a unique equilibrium size, then differences in city sizes must be rationalized by differences in land

10

area and/or natural amenities in production or consumption. According to equation 6, these differences correspond to vertical translations of the (log) indirect utility curve. When equilibrium population is unique, then city size is determined by the point where congestion costs exhaust natural advantages.

Alternatively, in case 2 (right panel), $V(X_i)$ is declining in X_i except for some range of X_i where agglomeration economies dominate congestion costs. This is the case where agglomeration economies are strong and multiple equilibria in city size are possible. In this figure, δ_1 and δ_2 are such that agglomeration economies are greater than congestion costs over an intermediate range of population—the region corresponding to the upward-sloping portion of $V(X_i)$.

A key implication of strong increasing returns is the possibility of multiple stable equilibrium city sizes. Note that this feature allows for differences in long-run populations between two locations—say, along the same river—that are nearly identical in terms of natural advantages.
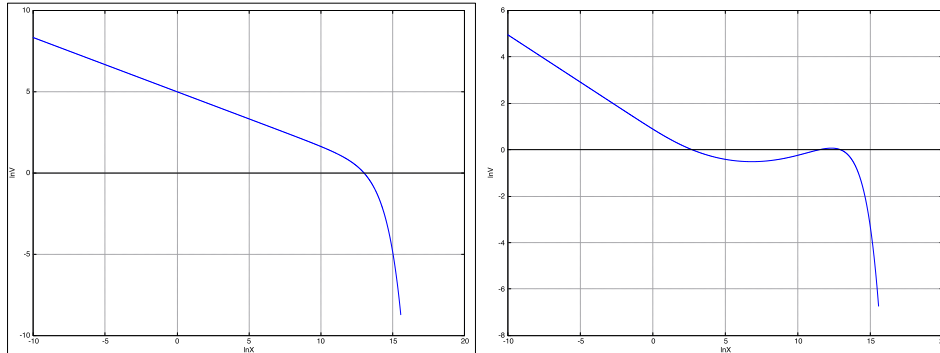
Figure 3: $V(X_i)$ with unique (left) and multiple equilibrium (right) sizes

In the figure there are two "stable" and one "unstable" equilibrium city size, in the Krugman (1991) sense. Note that the middle equilibrium is "unstable" in the sense that small increases in city size increase utility, which in turn might attract more workers. In contrast, $V'(X) < 0$ in the two outer "stable" equilibria.

4.3 Flexibility in agglomeration and congestion forces

Allowing for multiple equilibria introduces some problems when we attempt to quantitatively match the model to data. First note that parameterizations

restricting the model to a unique spatial equilibrium have no trouble fitting the model to data. Given observed population, there is some combination of natural advantages (ϕ, ψ) that can rationalize the employment data. Visually, in figure 3, left panel, variation in population along the horizontal axis is taken care of via vertical translations of the $V(X)$ curve.

However, things are not so simple when multiple equilibria are possible.[9] First, consider the case where agglomeration and congestion forces are identical across locations (i.e., each site's indirect utility curve is a vertical translation of the curve in Figure 3, right panel). Then, medium-sized cities (i.e., along the upward-sloping portion of $V(X)$) are predicted by the model to be in an unstable equilibrium. Thus, this restricted version of the model would have difficulty accounting for the large number of cities featuring intermediate levels of population.

A key feature of our quantitative work is to allow δ_1, δ_2, and/or γ to vary across locations, thus allowing the model to explain the data without requiring any site to be in an "unstable" equilibrium. Reasons why agglomeration and congestion forces may vary across locations include variation in industrial composition and spillovers. In Henderson's (1974) model, city-industry specialization and varying spillovers are used in a similar way to justify differences in city sizes. On the congestion side, differences in natural geography affecting transportation and housing supply, or differences in the productivity of a non-traded sector, may affect the elasticity of congestion forces across locations.

4.4   Data and Calibration

We use 1990 data for US counties on employment X_i, wages w_i, and land area H_i, reported in the 1994 and 1998 county data books (Michael R. Haines and the Inter-university Consortium for Political and Social Research, 2004). For wages, we use total BEA earnings, all industries (ea10090d) divided by total BEA employment (ge50090d). 3,109 counties (out of 3,142) have nonzero employment and earnings reported. Total annual earnings per employee (in thousands) range from under 8 (Mercer County, MO) to over 43 (New York, NY). (An outlier is North Slope, AK, where average annual earnings top 57.) Employment ranges from 108 (Loving, TX) to 5.3 million (Los Angeles, CA). Land area is reported in square miles.

The model contains many parameters, some of which we calibrate. Global

---

[9]These issues were recognized by Chatterjee (2006), who dealt with them by (1) fixing some population, instead of fixing land as we do, and (2) setting agglomeration economies to zero below some threshold, versus introducing a new congestion force.

β	2/3	cost share of labor
ω	1/4	cost share of capital
θ	1/3	expenditure share on land
r	0.04	national rental rate
$\alpha^{\dagger}$	$8/11 \approx 0.73$	consumption share of land
λ	1	output normalization
$\overline{V}$	1	utility normalization

Table 1: Calibrated and normalized global parameters

†—Implied by other parameters.

parameters include β, ω, θ, r, λ, and $\overline{V}$. We normalize units of output and utility so that λ and $\overline{V}$ are 1. Other parameters are calibrated as described in Table 1. Local shifters ϕ_i and ψ_i are determined in determined (perhaps uniquely) by the model, data, and calibrated parameters. Endogenous local variables σ, Y, K, V', V'', and p are similarly determined. The parameters governing the strength of agglomeration and congestion forces, δ_1, δ_2, and γ, are either fixed, restricted to be equal across locations, or locally flexible, as explained later.

4.5 Solution method

In general, there may be many solutions for equilibrium factor prices, shifters, and factor allocations, even conditioned on the calibrated parameters and the data on X, w, and H. Nevertheless, conditional on these data and a few calibrated parameters, the model is log-linear in the remaining parameters. This form makes the model amenable to linear programming, which is suited to optimizing possibly underdetermined systems. We use a binary variable to measure multiplicity, thus necessitating an integer-constraint on the otherwise linear program. We use the model to formulate a mixed-integer-linear program (MIP) to construct bounds on various concepts defined below.

The integer aspects of this program would render impossible a simple brute-force approach to the problem. Each of the 3109 sites could either have or not have another stable long-run equilibrium. This means that there are 2^{3109} configurations, or approximately 9×10^{935} cases. Randomization-based solvers such as simulated annealing would have a tough go with so many permutations to sift through. Fortunately, mixed-integer linear programs are well suited to this sort of problem, and we are able to solve these problems in minutes rather than in ages of the universe.

13

4.6 Implementing the linear program

Conditioned on observing employment, wages, and land areas, and the calibrated parameter values, the equilibrium conditions (equations 1–4.2 can be written as a system of linear equations in a vector of endogenous variables

$$\mathbf{x_i} \equiv [\gamma_i, \delta_{1,i}, \delta_{2,i}, \log \phi_i, \log \psi_i, \log \sigma, \log Y_i, \log K_i, \log p_i, \partial \log V_i / \partial \log X_i]. \tag{7}$$

For example, taking logs of equation 1 and rearranging yields

$$\log \phi + \log \sigma - \log Y_i - \omega \log K_i = -(1 - \beta - \omega) \left[\log(1 - \alpha) + \log H_i - \beta \log X_i \right], \tag{8}$$

where the right-hand-side of equation 8 consists only of calibrated parameters or data. Note that each equation 1–4.2 can be rewritten in a similar way—where the right hand side is composed only of calibrated parameters and data, and the left hand side is a linear equation in the endogenous variables.

Thus, the equilibrium conditions at each site can be summarized by the linear problem $\mathbf{A_i x_i} = \mathbf{b_i}$, in which

$$\mathbf{A_i} \equiv \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & -1 & \omega & 0 & 0 \\ 0 & \log X_i & (\log X_i)^2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ -X_i & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ -X_i & \frac{1-\theta}{1-\omega} & 2\frac{1-\theta}{1-\omega} \log X_i & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

and

$$\mathbf{b_i} \equiv \begin{bmatrix} -(1 - \beta - \omega) \left[\log(1 - \alpha) + \log H_i - \beta \log X_i \right] \\ 0 \\ -(1 - \theta)] \left[\log w_i + \log(1 - \theta) \right] - \theta(\log \alpha + \log H_i - \log X_i) \\ \log r - \log \omega \\ \log X_i - \log \beta + \log w_i \\ -\log(1 - \beta - \omega) + \log(1 - \alpha) + \log H_i \\ -(\beta(1 - \theta) - 1 + \omega)/(1 - \omega) \end{bmatrix}.$$

Note that the seventh row defines $V'(X_i)$, and rows 1–6 correspond to equations 1–6. In addition, we require that the data be in a stable equilibrium—so $V'(X_i) < 0$ is the final site-specific equilibrium condition, appended to $\mathbf{A_i}$ and $\mathbf{b_i}$.

The global equilibrium is defined by $8 \times I$ (i.e., the number of rows of **A** times the number of locations) constraints and $10 \times I$ (the number of columns of **A** times the number of locations) endogenous variables.

5   Bounds on multiplicity (χ)

The model above is consistent with both a very broad and a very limited scope for multiplicity in long-run equilibria. In this section, we solve for bounds on multiplicity using the model above. Specifically, what is the minimum (maximum) number of sites where alternative equilibrium allocations of population are feasible?

After we take as given both the calibrated parameters above and the observations on X, w, and H, there are two sets of remaining parameters. First, the ϕ and ψ are productivity and amenity shifters that are allowed to be site-specific. Second, the γ, δ_1, and δ_2 are parameters that influence, roughly speaking, the returns to scale as a site. We dub these "stretcher" parameters because they affect the slope of the V curve. The shifters are neutral in scale, with a rise in ϕ or ψ increasing V at all levels of X. In contrast, the stretchers are complements with X, at least when evaluated over a large enough span of X, because if γ, δ_1, or δ_2 goes up, the V curve rises more at very large X than at very small X. We treat the stretcher parameters in three distinct ways: fixed at a default value (label "fixed" below), constrained to the the same across sites ("global"), or site-specific ("local"). The default values for these parameter allow for congestion, but not agglomeration, but we relax these assumptions below.

5.1   Computation of $V(X_i)$ over a grid of counterfactual city sizes

Identifying when alternative feasible solutions exist is not directly possible with a linear program because $\log V(X)$ is nonlinear in counterfactual employment levels. Our solution is to compute $\log V(X_{ij})$ over a fine grid of J discrete values of X_{ij}, for all sites i. (Here, $j = 1, ..., J$ indexes counterfactual employment levels for each site i. For the computations below, $J = 100$.) Because equation 4 is linear in the endogenous and unobserved variables **x**, we can append a set of gridded computation constraints to the linear program.

The next step is to define an auxiliary binary variable $\chi_i \in \{0, 1\}$ that indicates for each location whether, given parameters and data, population

15

is uniquely determined ($\chi_i = 0$) or not ($\chi_i = 1$). This makes the problem a MIP rather than a pure linear program.

5.2 Minimize χ

We first consider the minimum number of sites where multiplicity is possible, i.e., the maximum number of sites where population is uniquely determined by fundamentals.

The following mixed-integer linear program attempts to find the ϕ_i, ψ_i, $\delta_{1,i}$, $\delta_{2,i}$ and γ_i such that the number of locations with multiplicity is minimized.

$$\min \sum_{i=1}^{I} \chi_i$$

subject to

$$
\begin{aligned}
\mathbf{A} \cdot \mathbf{x} &= \mathbf{b} \\
\mathbf{G_1} \cdot \mathbf{x} - \mathbf{V} &= \mathbf{G_2} \\
tol \times \text{sign}(X_{ij} - X_i) \times V(X_{ij}) &\leq \chi_i \\
&\forall i, j
\end{aligned}
$$

$$\chi_i \text{ binary}$$

where $\mathbf{V}$ is a vector of computed V_{ij} for each location i and discretized population grid point X_j and $\mathbf{G_1}$ and $\mathbf{G_2}$ are matrices of coefficients from equation 4, that defines $\log V_{ij}$ as a linear function of endogenous variables, parameters, and counterfactual employment levels X_{ij}.

The third set of constraints, $tol \times \text{sign}(X_{ij} - X_i) \times V(X_{ij}) \leq \chi_i$, $\forall i.j$, define χ_i for each site. tol is a very small number, i.e. 1×10^{-3}. $V(X_{ij})$ is the computed indirect utility for each grid point X_j. $\text{sign}(X_{ij} - X_i)$ is the sign of the difference between grid point X_{ij} and observed population X_i. Thus, the sign of the left hand side of the inequality is $+1$ for $X_{ij} > X_i$ and -1 for $X_{ij} < X_i$.

Consider the case where population is uniquely determined for location i, as in Figure 3, leftmost panel. In other words, indirect utility $V(X_{ij})$ monotonically declines with population X_{ij}. In this case, the $\text{sign}(X_{ij} - X_i)$ transforms the left hand side into a function that is weakly negative over the entire range of X_{ij}, i.e., $tol \times \text{sign}(X_{ij} - X_i) \times V(X_{ij}) \leq 0$.

As the problem tries to minimize $\sum \chi_i$, this implies that optimal $\chi_i^* = 0$ for sites with no multiplicity. In contrast, consider the case where population is not uniquely determined for location i, as in Figure 3, right panel. Then, $tol \times \text{sign}(X_{ij} - X_i) \times V(X_{ij}) > 0$ for some X_{ij}, so $0 < \chi_i^*$, which translates to $\chi_i = 1$ on account of χ being binary. In this setup, the mixed-integer

linear program will attempt to choose parameters to avoid multiplicity and guarantee uniqueness at each site, thus allowing for $\chi_i = 0$ at as many sites are possible.

The model is compatible with little to no multiplicity, as shown in Table 2. Columns 1-3 indicate the class of restrictions on the stretcher parameters $(\gamma, \delta_1, \delta_2)$, while column 4 reports the results from the χ-minimization program. For row A.1, we choose the default parameter values that shut down agglomeration, and, as expected, the MIP solver finds that χ can be zero for all sites. When we allow for agglomeration parameters to be positive, the routine still indicates a lower bound on multiplicity at exactly zero sites. This holds up even if we allow the stretcher parameters to be site-specific ("local").

5.3 Maximize χ

The second problem we consider is the *maximum* number of locations where multiplicity is possible, i.e., the minimum number of sites where population is uniquely determined by fundamentals.

The following MILP attempts to find ψ_i, ψ_i, $\delta_{1,i}$, $\delta_{2,i}$ and γ_i such that the number of locations with multiplicity is maximized.

$$\max \sum_{i=1}^{I} \chi_i$$

subject to

$$\mathbf{Ax} = \mathbf{b}$$
$$\mathbf{G_1} \cdot \mathbf{x} - \mathbf{V} = \mathbf{G_2}$$

$$\mu_i - \text{sign}(X_{ij} - X_i) \times V(X_{ij}) - Ms_{ij} \leq 0 \quad \forall i,j$$
$$\sum_j s_{ij} = J - 1 \quad \forall i$$

$$\chi_i - M * z_i \leq 0 \quad \forall i$$
$$-\mu_i + M * z_i \leq M \quad \forall i$$

$$\chi_i - My_i \leq 0 \quad \forall i$$
$$-V(X_{i,1}) + My_i \leq M \quad \forall i$$

$$V(X_{i,J}) < 0 \quad \forall i$$
$$\chi_i, s_{ij}, y_i, z_i \text{ binary}$$

17

The first two sets of constraints are identical to the minimization problem.

The next two lines specify a set of constraints that attempt to find the maximum value μ_i of the sign-transformed indirect utility curve, i.e., $\max\left[\text{sign}(X_{ij} - X_i) \times V(X_{ij})\right]$. Note that s_{ij} is a binary variable and M is a large number, i.e., 1×10^3. The summing-up constraint, $\sum s_{ij} = J - 1$, ensures that $s_{ij} = 1$ for all but one grid point X_{ij} per site. Note that if $s_{ij} = 1$, then the previous inequality is dominated by the "big M," and is not relevant for the maximization problem.[10] If however, for some grid point X_{ij}, $s_{ij} = 0$, then the constraints require that μ_i be no less than than the maximum value of $V(X_{ij})$.

The subsequent two lines specify a set of constraints that link χ_i to μ_i. These constraints implement the logical statement: If $\mu_i \leq 0$, then $\chi_i \leq 0$, else $\chi_i \leq M$. In words, if the maximum of the sign-transformed indirect utility curve is less than equal to zero (i.e., population is uniquely determined by fundamentals), then the value of χ_i for that site is restricted to be no more than zero. Since the problem is trying to maximize $\sum \chi_i$, it will attempt to find a feasible solution to avoid this constraint.

The next two lines specify a set of technical constraints. These constraints implement the logical statement: If $V(X_{i,1}) \leq 0$, then $\chi_i \leq 0$. In words, if indirect utility at the minimum $X_i j$ is negative, then set χ_i to be no greater than 0. We can show numerically that if computed indirect utility at the minimum $X_i j$ is negative, then any alternative equilibria (if they exist) are unstable.

Finally, the last line specifies a set of "no-black-hole" constraints. We require computed indirect utilities at the maximum X_{ij} to be strictly negative, i.e. below equilibrium utility across locations. This ensures that in equilibrium, no site might be so attractive as to absorb the entire population.

As the problem is attempting to maximize $\sum \chi_i$, it will attempt to find parameters so that $z_i = 1$ (according to the fifth constraint). Note that if $z_i = 1$, then $\mu_i \geq 0$, according to the sixth constraint. Thus, the problem will attempt to find parameters so that μ_i is positive.

There is significant scope for multiplicity, although not for all sets of parameter restrictions. These results are also seen in Table 2, but in columns 5-9. As above, we first consider a default set of stretcher parameters that allows for congestion but not for agglomeration. Specifically, $\gamma = 1.5e - 06$, $\delta_1 = 0$, and $\delta_2 = 0$. The upper bound on the number of sites with

---

[10]Such "big M" strategies are commonly used in mixed-integer-linear programming to find approximate solutions to nonlinear equations, for example.

multiplicity is zero, as shown in row A.1. This should come as little surprise because we have shut down agglomeration. Indeed, for no $\delta_2 = 0$ is max $\chi > 0$, as seen in A.1, B.1-2, C.1-2, and D.3. This is also consistent with the analysis above: δ_2 must be greater than zero to obtain V curve with three intersections (Figure 3, rightmost panel). Relaxing this $\delta_2 = 0$ assumption is not enough alone, however; we see in C.2 that no single δ_2 can be found that allows for multiplicity, if we fix the other stretcher parameters at their default levels.

Nevertheless allowing for more flexible parameters options yields scope for multiple equilibria. For example, row C.3 displays the output from a program in which the δ_2 are allowed to be site-specific ("local") in which we find that as much as 86% of sites could be characterized by multiplicity. In Panel D, we allow for local heterogeneity in one of the other stretcher parameters, and we see that as much as all (γ local, row D.2) or all but one (δ_1 local, row D.1) sites could have the possibility of another long-run equilibrium. Panel E further relaxes the assumption in Panel D by letting parameters that were fixed at their defaults be globally flexible (i.e., not fixed *ex ante* at a specific value, but constrained to be common values across sites). Allowing all three to be local (site-specific) yields an upper bound of all sites possessing an additional stable, long-run equilibrium.

In Table 3, we explore the scope for multiplicity on a grid for the stretcher parameters. For Panel A, we allow for site-specific γ but fix the δ at specific values for each χ-maximization exercise over a grid of plausible δ. The first cell (in which max $\sum \chi_i / J = 0$) is equivalent to Table 2, Panel B.1, column 5. Moving up in $\{\delta_1, \delta_2\}$ space (right and/or down on the grid), we see this unique-equilibrium result is robust in the neighborhood around $\delta_1 = \delta_2 = 0$. It is also robust throughout the first column for all $\delta_2 = 0$, which concords with the results in Table 2, various panels. Nevertheless, as both δ get sufficiently far from zero, scope for multiplicity emerges. The upper bound on χ peaks in the vicinity of $\delta_1 = .07$ and $\delta_2 = .0185$. Blank cells in the grid indicate that no feasible solution was possible. This reflects the fact that, if the agglomeration forces are so large, the model cannot find a stable equilibrium over the (amply wide) grid of city sizes.

In results not shown, we also created a version of Panel A for the $\chi-$minimization exercise. This table would have the same upper-triangular form, but with zeros throughout the non-blank cells. This is true for a site-specific γ or for a γ that is constrained to be the same across all sites. How is this possible even in the presence of strong agglomeration forces? While it is true that the δ might generate a curve with just the right wiggles, it is relatively easy to simply perturb those curves up or down with the shifters,

Fraction of sites with multiple equilibria possible ($\chi = \in \{0,1\}$) for various parameter restrictions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				min $\Sigma \chi$	max $\Sigma \chi$		max $\Sigma \chi$		
	Restrictions on parameters						Solved parameter values		
	γ	$\delta 1$	$\delta 2$	mean χ	mean χ	Time to solve	$\gamma \times$1e6	$\delta 1$	$\delta 2$
A. 1	Fixed	Fixed	Fixed	0	0	44s	1.5	.00	.000
B. 1	Global	Fixed	Fixed	0	0	37s	0	.00	.000
2	Fixed	Global	Fixed	0	0	40s	1.5	.46	.000
3	Fixed	Fixed	Global	0	0	52s	1.5	.00	.025
C. 1	Local	Fixed	Fixed	0	0	42s	0	.00	.000
2	Fixed	Local	Fixed	0	0	54s	1.5	.53	.000
3	Fixed	Fixed	Local	0	.8601	24s	1.5	.00	.027
D. 1	Fixed	Local	Local	0	.9997	5118s	1.5	.24	.014
2	Local	Fixed	Local	0	1	48s	3.0	.00	.027
3	Local	Local	Fixed	0	0	874s	8.6	1.25	.000
E. 1	Global	Local	Local	0	1	4667s	0.4	.21	.014
2	Local	Global	Local	0	1	433s	1.7	.39	.004
3	Local	Local	Global	0	1	454s	0.8	.36	.005
F. 1	Local	Local	Local	0	1	39s	4.9	.11	.022

Notes: Solved parameters are (a) not necessarily unique and (b) reported as the mean across sites if the parameter is site-specific ("Local"). If the parameters are local, they are allowed to be site specific. If the parameters are global, they are flexible, but constrained to be the same across sites. If the parameters are fixed, they are set as in row A.1, columns 7-9.

Table 2: Fraction of sites with multiple equilibria possible ($\chi \in \{0,1\}$) for various parameter restrictions

Maximum fraction of sites with multiple equilibria possible ($\chi=1$) for various combinations of agglomeration parameters (δ)

Panel A: γ is site-specific

								$\delta 2$							
$\delta 1$	0	.003	.007	.010	.013	.017	.020	.023	.027	.030	.033	.037	.040	.043
.00	0	0	0	0	0	.002	.68	.89	.79	.79	.58			
.03	0	0	0	0	0	.09	.93	.84	.77	.75				
.07	0	0	0	0	0	.49	.89	.76	.79	.54				
.10	0	0	0	0	.02	.92	.81	.76	.65					
.13	0	0	0	0	.24	.88	.74	.83						
.17	0	0	0	0	.88	.77	.77							
.20	0	0	0	.06	.86	.72	.73							
.23	0	0	0	.76	.71	.84								
.27	0	0	0	.84	.74									
.30	0	0	.35	.66	.80									
.33	0	0	.77	.90										
.37	0	0	.73											
.40	0	.58												
.43	0													
.47														
.50														

Panel B: γ and $\delta 1$ are site-specific

							$\delta 2$							
0	.003	.007	.010	.013	.017	.020	.023	.027	.030	.033	.037	.040	.043
0	1	1	1	1	1	1	1	1	.89	.61			

Notes: Blank cells indicate no feasible solution.

Table 3: Maximum fraction of sites with multiple equilibria possible ($\chi=1$) for various combinations of agglomeration parameters (δ)

21

and thereby leave the curve with only one intersection with $\overline{V}$.

Finally, in Panel B, we allow both γ and δ_1 to be site-specific and attempt to maximize χ over a grid of δ_2. Here we see only small values of δ_2 are sufficient to allow for indeterminacy of equilibria across all sites. As above, however, very large values of δ_2 reduce the scope for multiplicity and then render the model infeasible.

6   Bounds on gains (or losses) from reallocation

Beyond the scientific interest in the determinants of city size, the theme of multiplicity raises the specter of misallocation. In the present context, can we find a set of shifter and stretcher parameters that are consistent with the data and calibrated parameters, but would also, given the model, permit an alternate (equilibrium) allocation of employment across sites? If so, would any of these alternative allocations produce higher (or lower) utility? What is the minimum (maximum) level of utility $\overline{V}$ that can be achieved by reallocating population across sites? The corresponding MIP can be written as follows:

$$\max \overline{V}$$

subject to

$$
\begin{array}{cccc}
\mathbf{Ax} & = & \mathbf{b} & \\
\mathbf{G_1} \cdot \mathbf{x} - \mathbf{V} & = & \mathbf{G_2} & \\
V(X_{i,J}) & < & 0 & \forall i \\
& & & \\
V_{i,j+1} - \overline{V} & \leq & +M \times (1 - Z_{ij}) & \forall i,j \\
-V_{i,j} + \overline{V} & \leq & +M \times (1 - Z_{ij}) & \forall i,j \\
\sum_{j=1}^{J} Z_{ij} & = & 1 & \forall i \\
\end{array}
$$

$$
\begin{array}{ccc}
Z_{ij} \text{ binary} & & \\
\sum_i \sum_j Z_{ij} \left(X_{i,j} + X_{i,j+1} \right) / 2 & = & X_{total}
\end{array}
$$

The first three sets of constraints are identical to the min/max χ problem. The next three lines specify a set of constraints that—in effect–find the alternative equilibrium at each site for a given value of $\overline{V}$. The two inequalities insure that such that $V_{i,j+1} \leq \overline{V} \leq V_{i,j}$ for some j. The summing-up constraint $\sum Z_{ij} = 1$, combined with Z being binary, specifies that this relation

hold at exactly one grid point X_{ij} for each site i. Thus, for $Z_{ij} = 1$, the bounding constraints hold, but for the rest of the grid points X_{ij}, $Z_{ij} = 0$ and the bounding constraints are effectively ignored. (Note that the bounding constraints $V_{i,j+1} \leq \overline{V} \leq V_{i,j}$ implicitly require that any new equilibrium allocation be stable, because $V_{i,j+1} \leq V \leq V_{i,j}$.) The final equation assures that the new allocation of employment across sites sums to the total. Population is pinned midway between grid points because a closer approximation would not be a linear function of parameters. (Future work will refine this approximation with a 'big M' strategy for the interpolation and/or with a finer X_{ij} grid.)

In Figure 4, we show sample results from such an exercise. In this simulation, a 15% increase in utility is possible, if we assume that all three of the stretcher parameters are site-specific. (This is an example and not the maximum $\bar{V}$ nor is it a unique reallocation for $\log \bar{V} = .15$.) Panel A plots log employment in the data (x axis) versus in the simulation (the "counterfactual" on the y axis). Panels B-D shows the $\tilde{V}$ curves for several examples sites and illustrate the types of transitions that would occur. (We label these sites with numbers, not county names, to protect the innocent.) The dashed lines at zero denote the original (current) level of log indirect utility while the dashed lines at .15 denote the utility level ($\log \bar{V}$) after reallocation. Equilibrium (both observed and counterfactual) are seen at the starred downward sloping intersections of the $\tilde{V}$ curve with the respective dashed lines. Panel B shows a site that is in the low equilibrium but that might be bumped up to a high one. Such sites would grow tremendously in this counterfactual world and must absorb population from elsewhere. Such sites are the circles above the diagonal in Panel A.

The remaining sites would shrink, but in two distinct ways. Panel C shows two sites whose equilibria are uniquely defined. Reallocating population away from these counties reduces congestion and land dilution there. Panel D shows instead the example of a site that shrinks down to a low equilibrium, but this new low equilibrium is well below the original low equilibrium that was consistent with a $\log \bar{V} = 0$.

Note that the utility gains come from the reduced dilution/congestion in all three Panels. By construction, two equilibria that could be obtained at a given site and a given $\bar{V}$ would generate the same utility level locally. This is important for thinking about understanding how the gains from reallocation would take place in such a model. Simply flipping one site from high to low and another site from low to high, in a way that satisfies the total population constraint, would not produce gains, again by virtue of what it means for there to be two equilibria at each of these two sites.
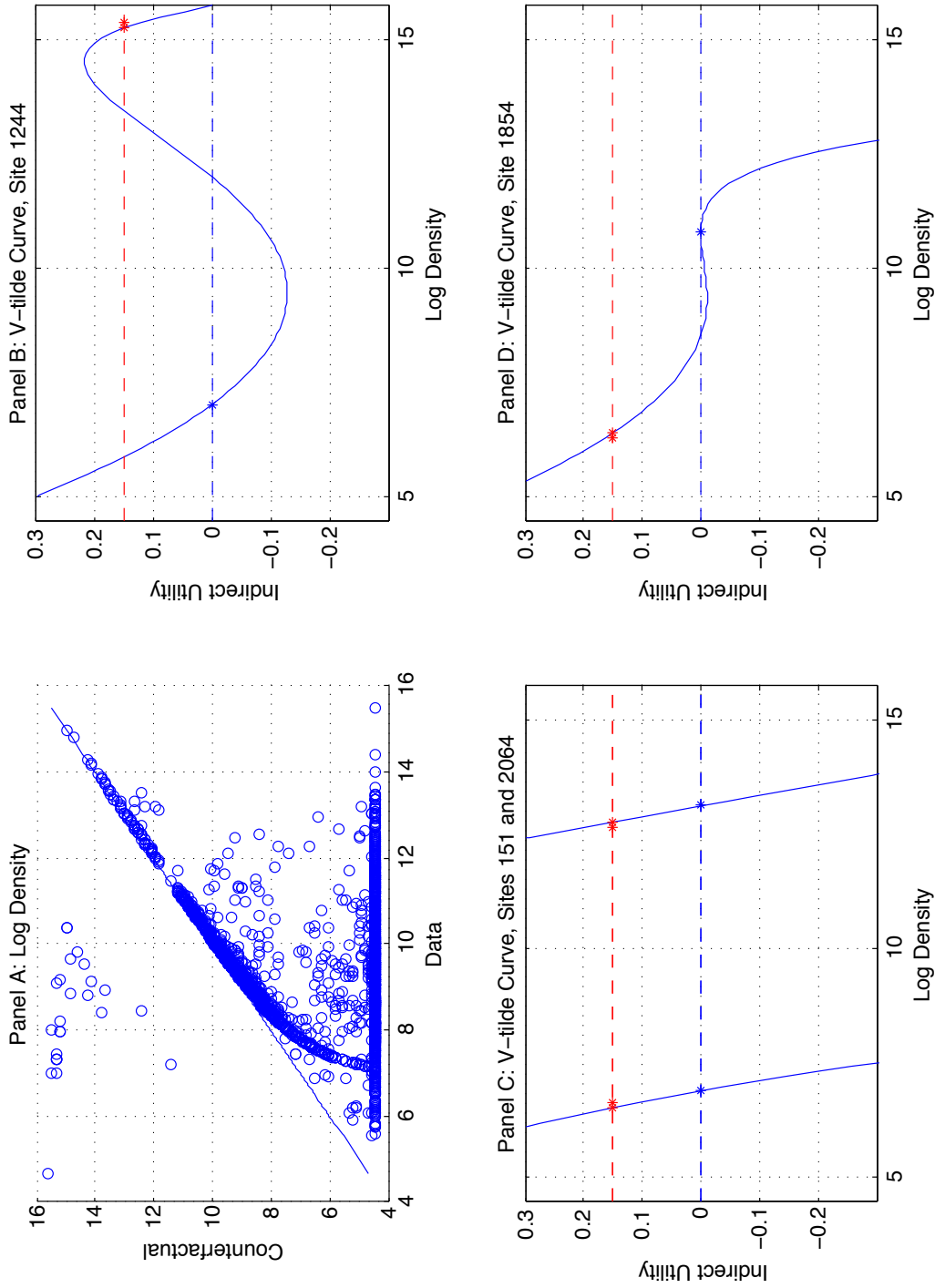
Figure 4: Reallocation and utility gains

This is distinct from Paul David's (1985) story about the QWERTY keyboard, in which he claimed they were potentially large benefits from having chosen a different keyboard layout. The key difference between David's example and this one is whether the equilibrium arises through constraints or marginal conditions. In the keyboard example, we are at a corner solution: essentially everyone has specialized into a particular layout (QWERTY), but the next person to arrive on the scene would still strictly prefer to learn the QWERTY layout. Thus the keyboard layouts are analogous to the 'black hole' locations that we sought to rule out by assumption above. In our model, there are no black-hole cities but rather equilibria are all determined by the marginal worker being indifferent between one site and another.

Why does this QWERTY comparison matter for our intuition? Consider the usual informal theorizing in which someone laments that it would be nice if accidents of history had brought us to have the current distribution of population in places with milder winters or nicer beaches or more reliable water supply or some such. Of course, this lament is nonsensical if all of the equilibria are uniquely determined. But note that it is also hard to understand this comment even in our model in the case where multiplicity is possible. The amenities mentioned above should, in equilibrium, already be priced.

7   Conclusions

How can an equilibrium model of city sizes accommodate both the historical evidence of path dependence and the empirical size distribution of cities? One solution is to allow for multiple equilbria in city sizes and locational fundamentals that are non-neutral with respect to city size. Of course, allowing for such flexibility comes at a cost in terms of identification. We leave an investigation of such a model for future work.

References

Albouy, David, and Bryan Stuart. 2014. "Urban Population and Amenities." NBER working paper 19919.

Bleakley, Hoyt, and Jeffrey Lin. 2012*a*. "Portage and Path Dependence." *Quarterly Journal of Economics*, 127(2): 587–644.

Bleakley, Hoyt, and Jeffrey Lin. 2012*b*. "Thick Markets and Churning in the Labor Market: Evidence from U.S. Cities." *Journal of Urban Economics*, 72(2–3): 87–103.

Brinkman, Jeffrey. 2013. "Congestion, Agglomeration, and the Structure of Cities." Federal Reserve Bank of Philadelphia working paper 13-25.

Brooks, Leah, and Byron Lutz. 2014. "Vestiges of Transit: Urban Persistence at a Micro Scale." Working paper.

Chatterjee, Satyajit. 2006. "A Quantitative Assessment of the Role of Agglomeration Economies in the Spatial Concentration of U.S. Employment." Federal Reserve Bank of Philadelphia working paper 06-20.

David, Paul. 1985. "Clio and the Economics of QWERTY." *American Economic Review*, 75: 332–337.

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

Desmet, Klaus, and Esteban Rossi-Hansberg. 2013. "Urban Accounting and Welfare." *American Economic Review*, 103(6): 2296–2327.

Diamond, Peter. 1982. "Aggregate Demand Management in Search Equilibrium." *Journal of Political Economy*, 90(5): 881–894.

Diamond, Rebecca. 2013. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." Working paper.

Haines, Michael R., and the Inter-university Consortium for Political and Social Research. 2004. *Historical, Demographic, Economic, and Social Data: The United States, 1790–2000.* Colgate University and Inter-university Consortium for Political and Social Research.

Hanlon, Walker. 2014. "Temporary Shocks and Persistent Effects in the Urban System: Evidence from British Cities after the U.S. Civil War." Working paper.

Harari, Mariaflavia. 2014. "Cities in Bad Shape: Urban Geometry in India." Working paper.

Haughwout, Andrew F., and Robert P. Inman. 2001. "Fiscal policies in open cities with firms and households." *Regional Science and Urban Economics*, 31: 147–180.

Helpman, Elhanan. 1998. "The Size of Regions." In *Topics in Public Economics: Theoretical and Applied Analysis.* , ed. David Pines, Efraim Sadka and Itzhak Zilcha, 33–54. Cambridge University Press.

Henderson, J.V. 1974. "The Sizes and Types of Cities." *American Economic Review*, 64(4): 640–656.

Jedwab, Remi, and Alexander Moradi. 2014. "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa." Working paper.

Jefferson, Thomas. 1781. *Notes on the State of Virginia.*

Krugman, Paul. 1991. "History Versus Expectations." *Quarterly Journal of Economics*, 106(2): 651–667.

Lee, Sanghoon, and Jeffrey Lin. 2013. "Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income." Federal Reserve Bank of Philadelphia working paper 13-48.

Lee, Sanghoon, and Qiang Li. 2013. "Uneven Landscapes and City Size Distributions." *Journal of Urban Economics*, 78: 19–29.

Michaels, Guy, and Ferdinand Rauch. 2013. "Resetting the Urban Network: 117-2012." Centre for Economic Policy Research Discussion Paper 9760.

Murphy, Kevin M. 1986. "Specialization and Human Capital." PhD diss. University of Chicago.

Rappaport, Jordan. 2008*a*. "Consumption Amenities and City Population Density." *Regional Science and Urban Economics*, 38: 533–552.

Rappaport, Jordan. 2008*b*. "A Productivity Model of City Crowdedness." *Journal of Urban Economics*, 65: 715–722.

Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf. 2011. "History and Industry Location: Evidence from German Airports." *Review of Economics and Statistics*, 93(3): 814–931.

Villareal, Carlos. 2012. "The Persisting Influence of Historical Marshes on the Rental Price of Housing in New York City 1830-1940." Working paper.