Macroeconomic Implications of Agglomeration*

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

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

Toni M. Whited
University of Rochester
toni.whited@simon.rochester.edu

March 21, 2013

Abstract

Cities exist because of the productivity gains that arise from clustering production and workers, a process called agglomeration. How important is agglomeration for aggregate growth? This paper constructs a dynamic stochastic general equilibrium model of cities and uses it to estimate the effect of local agglomeration on aggregate growth. We combine aggregate time-series and city-level panel data to estimate the model's parameters via generalized method of moments. The estimates imply a statistically and economically significant impact of local agglomeration on the growth rate of per capita consumption, raising it by about 10%.

JEL Classification Numbers: E0, O4, R0
Keywords: economic growth, productivity, increasing returns, agglomeration, density

*We thank Rosella Argenziano, Marco Bassetto, V.V. Chari, Jonathan Heathcote, François Ortalo-Magné, Erwan Quintin, Sevi Rodriguez-Mora, José Víctor Ríos Rull, Marcelo Veracierto, Mark Wright, numerous seminar participants, the referees, and the editor for helpful comments. Yunjeen Kim provided outstanding research assistance. The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

1 Introduction

Cities emerge because of productivity gains that accompany the clustering of production and workers. Also known as agglomeration effects, these gains arise from better matching of workers and jobs, knowledge spillovers that accelerate the adoption of new technologies, expanded opportunities for specialization, scale economies in the provision of common intermediate inputs, and lower transportation costs. Because agglomeration effects give rise to cities, most economic activity, and therefore growth, occurs in cities. A question that naturally arises is the extent to which local agglomeration contributes to aggregate growth. Our paper answers this question by estimating a dynamic stochastic general equilibrium model of cities and growth in which local agglomeration affects per capita consumption growth.

Our model extends the neoclassical growth model along three dimensions. First, production and housing are location-specific; second, local infrastructure is a produced durable input into finished land, which is in turn used in production and housing; and third, local agglomeration effects offset the congestion that stems from crowding more workers onto the same finished land, as proposed by Ciccone and Hall (1996). Agglomeration is modeled as an externality in which total factor productivity (TFP) at a location is increasing in the location's output density – total output per acre of finished land in production. We study the model's competitive equilibrium in which agents do not take into account the density externality.[1] Along the model's balanced growth path, per capita consumption growth depends on the endogenously determined rate of increase in finished land prices and a parameter governing the effect of agglomeration on productivity.

In this setting, agglomeration affects growth only if land prices exhibit a trend. With a positive trend in land prices, agglomeration affects growth via two channels. First, firms economize on land when it becomes more costly, thereby increasing congestion and lowering growth. This effect is analogous, though opposite in direction, to the effect on capital of falling equipment prices in Greenwood, Hercowitz, and Krusell (1997). Second, when rising land prices curtail land use, density and therefore TFP grow faster than they would otherwise. We estimate that land prices indeed have a positive secular trend, growing 1.0% per year between 1978 and 2009. Combined

[1] We do not prove that this equilibrium is unique – a point we discuss further below.

with our estimate of the effect of agglomeration on local productivity, this estimate of the growth rate of land prices implies that agglomeration, by itself, raises per capita consumption growth by an economically and statistically significant 10.2%.

Our strategy for identifying the size of the density externality builds on Lucas (2001), whose identification uses variation in land rents within a city. Similarly, we use variation in land rents both across cities and over time to estimate our model's parameters jointly via generalized method of moments (GMM). We can identify the density externality in this estimation because it induces co-variation between TFP growth and an instrumental-variable prediction of growth in land rents. Predicted increases in land rents lead firms to economize on land, thereby raising density and TFP growth. In our model, TFP growth is positively correlated contemporaneously with *realized* land rents regardless of whether the density externality is operative. Nevertheless, our instrumental variables strategy identifies the magnitude of the externality because predicted land rents are orthogonal to the TFP shock that drives the contemporaneous correlation between realized TFP and land rents.

Our estimation combines evidence from aggregate time series and a panel of 22 U.S. cities for the years 1978 to 2009. We estimate the local effects of agglomeration using both data sets. The panel data contain information on land rents and the necessary inputs to a conventional measure of city-specific TFP, in which we control for heterogeneity in the work force following Ciccone and Peri (2006). We use aggregate time-series data to estimate other model parameters. Some of these parameters enter into our measurement of TFP, and some we use, along with our estimate of the size of agglomeration effects, to measure the impact of agglomeration on growth. Our estimation accounts for the sampling uncertainty in both the micro and macro data.

While our paper is novel for using city-level data to quantify the impact of city-level agglomeration on aggregate growth within a fully-specified model, it follows a substantial theoretical literature that rationalizes aggregate growth models with evidence on cities. Eaton and Eckstein (1997), Black and Henderson (1999), and Rossi-Hansberg and Wright (2007) study endogenous growth driven by human capital externalities of the kind introduced by Lucas (1988); Duranton (2007) considers the Grossman and Helpman (1991) quality ladder model of endogenous growth; and Córdoba (2008) studies balanced growth in a neoclassical growth model with exogenous technical change. Only Black and Henderson (1999) and Córdoba (2008) include a connection between agglomeration and aggregate growth, but they do not attempt to quantify the contribution of agglomeration to aggregate growth.

The key distinguishing characteristic of our framework is the role played by land. Duranton (2007) and Córdoba (2008) have no explicit role for land. In the other models, land is differentiated by its distance from the city center, with more distant land entailing either higher commuting costs, as in Black and Henderson (1999) and Rossi-Hansberg and Wright (2007), or lower productivity, as in Eaton and Eckstein (1997). In these models, cities expand by adding land to the perimeter at constant marginal cost. In our model, finished land is a homogeneous input to both housing and production. Cities expand by adding housing and production and by adding costly infrastructure to raw land. Infrastructure accumulation thus contributes to aggregate growth by adding increasingly costly finished land to existing cities.[2]

The paper is organized as follows. Section 2 describes our structural model. Sections 3 and 4 then describe the empirical strategy and the data. Section 5 discusses and provides some validation of the estimates. In Section 6 we provide additional validation of the estimates by solving and simulating our model. Section 7 concludes.

2 The Model

Because it is central to our empirical work, we begin by describing the Ciccone-Hall model of agglomeration. We then outline the decentralized competitive growth model that incorporates their framework, describe the planning problem that yields the equilibrium outcomes, and characterize the model's balanced growth path.

2.1 Modeling Agglomeration

Duranton and Puga (2004) review the extensive theoretical literature on the micro-foundations of urban agglomeration, emphasizing gains from specialization and variety, improved labor market matching, and learning. Ciccone and Hall (1996) argue

[2] Another related paper is Weill and Van Nieuwerburgh (2010), who study house price dispersion within a multi-city model that features land accumulation, but no agglomeration effects or growth.

that many of these ideas lead to the same implication: productivity is rising in the density of economic activity. We therefore work with their simple reduced form. In addition, this framework is convenient because it is also easily embedded in a standard growth framework and is particularly amenable to empirical analysis.

In the Ciccone-Hall model, production per acre of land at a location is a constant-returns function of labor and a non-transportable composite service input. This second input is a constant-elasticity-of-substitution function of an endogenous number of service varieties. Fixed costs of providing services and free entry yield a zero-profit condition that determines the number of varieties. Denser, higher-productivity acres of land have greater variety because more intermediate service producers can break even. This connection between density and variety in turn yields an expression for the production of composite services in which labor productivity increases with the number of varieties. With density leading to variety, and variety leading to productivity, the model yields a reduced-form relationship between density and productivity.

Land is exogenously fixed in the Ciccone-Hall model. In reality, land occupied by cities changes over time. Furthermore, we show that assuming a fixed stock of productive land biases upwards our estimates of the affect of agglomeration on growth. We therefore depart slightly from the Ciccone-Hall model by assuming that land is productive only after having infrastructure put in place. The land occupied by a city grows via the accumulation of infrastructure, with diminishing returns to infrastructure capturing the idea that the best quality raw land gets developed first. This approach is easily justified by the obvious importance of roads, sewers, electricity, and other factors that are tied to a location and necessary for housing and production.

2.2 Economic Environment

To embed the Ciccone-Hall model of agglomeration in a dynamic stochastic general equilibrium model, we must take a stand on market structure and the nature of insurance arrangements among the model's agents. In addition, we need to describe how workers move from one city to another, for example by paying a re-location fee. Here we consider a natural benchmark: complete markets in state-contingent consumption claims, perfect competition, and costless labor mobility.

There is an infinitely lived representative household with a measure N of homogeneous members who inelastically supply a unit of labor each period, taken to be a year. The household maximizes the expected present value of its members' utility, which depends on locally provided housing services, h, and a consumption good that is freely traded across a fixed measure of locations called cities. Because of complete markets, per capita consumption in each city is identical and equal to aggregate per capita consumption, C.[3] Each period the household freely allocates across cities its workers, business capital, and residential structures: n, k_b and k_h, with k_b, and k_h chosen from the predetermined aggregate stocks, K_b and K_h. The household also chooses for the next period these two aggregate stocks and infrastructure in each city, k_f. These decisions are made after the household observes each city's output density, x, defined as total production of the local intermediate good per unit of finished land, and an exogenous productivity shock, z, which evolves as a stationary discrete Markov process. The household takes all prices and the distribution of productivity and density as given, and in particular does not take into account the effect of its actions on the density of production in each location.

In addition to workers, each city contains landlords, providers of housing services, and intermediate good producers. Landlords rent local infrastructure capital from the household and combine it with raw land to produce finished land, which they rent to local goods producers and housing service providers. Housing service providers rent residential structures and finished land, l_h, to produce housing services, which they sell to workers. Intermediate goods producers rent business capital and labor from the household, and finished land, l_b, to produce the city-specific intermediate good, y. Output in a city with TFP a is given by:

$$y \leq a l_b^{1-\phi} k_b^{\alpha\phi} n^{(1-\alpha)\phi}. \tag{1}$$

TFP is specified as:

$$a = [\Upsilon z]^{(1-\alpha)\phi} \, x^{\frac{\lambda-1}{\lambda}}, \tag{2}$$

where Υ denotes the aggregate level of Hicks-neutral technology. Factor share parameters α and ϕ lie in the interval $[0, 1]$. Congestion affects production according to ϕ; if $\phi = 1$, then congestion has no impact. The effect of output density x on local TFP

[3] Lower case letters denote city-level allocations, and upper case letters denote economy-wide aggregates.

is denoted by $\lambda \geq 1$, with $\lambda = 1$ corresponding to no impact. Equations (1) and (2) combined are equivalent to the specification in Ciccone and Hall (1996) except that the finished land input does not equal the endowment of raw land in the city.

Final good producers combine the city-specific intermediate goods to produce a composite final good that can be transformed linearly into consumption and capital goods. Landlords, housing service providers, intermediate goods producers, and final goods producers maximize profits taking all prices, productivity shocks and density as given. There is no aggregate uncertainty.

2.3 Competitive Equilibrium

The competitive equilibrium for this economy can be found as the solution to a planning problem with side conditions. While we exploit characteristics of the competitive equilibrium to describe our estimation strategy, the planning problem is useful to describe the details of the model's specification succinctly. Let $\mu_t(z^t)$ denote the time t distribution of cities across productivity histories. The planning problem is:

$$\max_{\substack{\{C_t, K_{bt+1}, K_{ht+1}, y(z^t), l_b(z^t), l_h(z^t), n(z^t), \\ k_b(z^t), k_h(z^t), k_{ft+1}(z^t), h(z^t)\}_{t=0}^{\infty}}} \sum_{t=0}^{\infty} \beta^t \ln C_t + \psi \sum_{t=0}^{\infty} \beta^t \sum_{z^t} \mu_t(z^t) n(z^t) \ln \frac{h(z^t)}{n(z^t)}$$

$$\begin{aligned}
\text{subject to:} \quad & C_t + \Gamma_{bt}\left[K_{bt+1} - (1-\kappa_b)K_{bt}\right] + \Gamma_{ht}\left[K_{ht+1} - (1-\kappa_h)K_{ht}\right] \\
& + \Gamma_{ft} \sum_{z^t} \mu_t(z^t)\left[k_{ft+1}(z^t) - (1-\kappa_f)k_{ft}(z^{t-1})\right] \\
& \leq \left[\sum_{z^t} \mu_t(z^t) y(z^t)^{\eta}\right]^{1/\eta} \qquad (3)
\end{aligned}$$

$$y(z^t) \leq \left[\Upsilon_t z_t\right]^{(1-\alpha)\phi} x(z^t)^{\frac{\lambda-1}{\lambda}} l_b(z^t)^{1-\phi} k_b(z^t)^{\alpha\phi} n(z^t)^{(1-\alpha)\phi}, \ \forall z^t \qquad (4)$$

$$h(z^t) \leq l_h(z^t)^{1-\omega} k_h(z^t)^{\omega}, \ \forall z^t \qquad (5)$$

$$l_h(z^t) + l_b(z^t) \leq k_{ft}(z^{t-1})^{\zeta}, \ \forall z^t \qquad (6)$$

$$\sum_{z^t} \mu_t(z^t) k_b(z^t) \leq K_{bt} \qquad (7)$$

$$\sum_{z^t} \mu_t(z^t) k_h(z^t) \leq K_{ht} \qquad (8)$$

$$\sum_{z^t} \mu_t(z^t) n(z^t) \leq N_t \qquad (9)$$

with $K_{b0}, K_{h0}, k_f(z_0), x(z^t)$, and $\{\Upsilon_t, \Gamma_{bt}, \Gamma_{ht}, \Gamma_{ft}, N_t, z^t\}_{t=0}^{\infty}$ given. The variables Γ_{bt}, Γ_{ht}, and Γ_{ft} denote the rates at which consumption goods can be transformed into the different kinds of investment. These variables are a source of growth in the model, but for now we set them to one in every period. We also assume for now that population is constant, $N_t = 1$. The parameters ω, ζ, η, and ψ denote, respectively, the share of structures in housing services, the returns to augmenting raw land with infrastructure, the elasticity of substitution of the city goods in the production of the final good, and the weight on housing services in preferences. We restrict $\omega \in [0, 1]$, $\zeta \in [0, 1]$, $\eta \leq 1$, and $\psi \geq 0$. Equation (3) is the aggregate constraint on final goods. Equation (4) combines (1) and (2) for each productivity history. Equations (5) and (6) are the local resource constraints for housing services and finished land. In the latter we normalize the quantity of raw land in a city to one. Lastly, equations (7)–(9) are the aggregate resource constraints for business capital, residential structures, and employment.[4]

Following Kehoe, Levine, and Romer (1992), we find competitive equilibrium allocations as the solution to this optimization problem such that $x(z^t) = y(z^t)/l_b(z^t)$. Equilibrium prices then come from the Lagrange multipliers on the constraints. The details of this mapping along with the first order conditions of the planning problem are described in Davis, Fisher, and Whited (2013) (DFW). There we also show how the first order conditions for the planning problem correspond to the agents' first order conditions in the competitive equilibrium. We refer to the agents' first order conditions below when describing our empirical strategy.

We have limited results on the existence and uniqueness of the competitive equilibrium. When $\lambda = 1$, the planning problem has a concave objective and convex constraint set so a unique competitive equilibrium exists, and the allocations are Pareto-efficient. When $\lambda > 1$, the allocations are not efficient, and we have no general existence and uniqueness results.[5] The issue is that the externality induces

[4]The underlying micro-foundations of the density externality involve local intermediate goods, which do not appear anywhere in the statement of the model because it is written in terms of final goods. Labor and capital include the factors of production used to produce local intermediate goods. See Ciccone and Hall (1996) for more details.

[5]The growth and cities papers cited in the introduction focus on efficient allocations derived from planning problems that take into account any externalities. The exception is Black and Henderson (1999), who examine competitive allocations that are inefficient. They argue that the information and technological requirements necessary to sustain efficient allocations are implausible.

increasing returns to labor, which in turn implies that some cities could be unoccupied. However, in DFW we show that for the special case without housing, $\psi = 0$, if $\eta\lambda\phi < 1$, there is a unique equilibrium in which all cities are occupied. The restriction on parameters guarantees that the complementarity of city-specific goods and congestion from diminishing returns to finished land provide sufficient curvature to offset the increasing returns. Including housing provides additional curvature, so we expect that existence and uniqueness extend to the general case. Below we discuss numerical solutions to the model with housing and $\lambda > 1$. Nevertheless, equilibrium uniqueness remains an open question in this case.

We now emphasize several important features of our model, starting with the role of finished land in determining the effect of agglomeration on aggregate growth. For agglomeration to influence growth, output density must grow, and density growth only happens if the supply of finished land grows more slowly than output. Finished land growth in turn depends on the degree of diminishing returns to infrastructure in production, ζ. This parameter is therefore key to the ultimate effect of agglomeration on growth. Land also plays an essential role in our model because it has competing uses in production and housing. This model feature is important for our empirical analysis because it implies that we can price finished land using housing rents.

Our model exhibits non-trivial cross-section and time-series variation in housing rents, wages, and output prices, all of which are important in our empirical work. Because local infrastructure is predetermined, idiosyncratic productivity shocks induce variation in housing rents. Variation in wages occurs because the household balances sending workers to high productivity cities with their preferences for housing. The price of city-specific output varies if $\eta < 1$ in the production of the final goods from intermediate goods. Notice that because infrastructure is predetermined, wages, rents, and output prices are predictable. This predictability justifies our instrumental variables strategy to identify the key parameter λ described below.

The model includes several assumptions that we have made for convenience, but only some affect our empirical work. First, non-infrastructure capital is costlessly allocated across cities contemporaneously with the TFP shock, and we do not impose a non-negativity constraint on city-level investment. These assumptions imply a single economy-wide rental price for non-infrastructure capital and thus allow us to avoid measuring local prices for capital services in our empirical work. While the non-negativity assumption is essential for this result, we can relax the capital allocation assumption. In particular, if the TFP shock occurs one period after the allocation of non-infrastructure capital, then rental rates are equal across cities up to an expectation error. Our instrumental variables estimation is robust to having this kind of error term. Second, infrastructure can be put in place within a single year, and labor mobility is costless. These assumptions are inconsequential because they can be relaxed without affecting the equations we use in our estimation. Third, we assume that the density externality immediately impacts TFP and that firms make factor input decisions contemporaneously with the TFP shock. Relaxing these assumptions can affect our estimation, and we examine the consequences when we discuss our empirical results. Finally, all of these assumptions can be relaxed without affecting the balanced growth path discussed next.

2.4 Balanced Growth

To study growth, we assume:

$$\begin{aligned}
\Gamma_{jt} &= \gamma_j^{-t},\ j = b, f, h; \\
\Upsilon_t &= \gamma_a^t; \\
N_t &= \gamma_n^t.
\end{aligned}$$

The parameter $\gamma_j > 0$ is the growth rate of the technology for producing the j-type investment good relative to that for consumption. If $\gamma_j \neq 1$, the corresponding investment technology exhibits a trend. We consider trends in the investment technologies because they exist empirically and, similar to Greenwood et al. (1997), they influence our estimates. We also allow for trends in Hicks-neutral technology and in population, given by $\gamma_a > 0$ and $\gamma_n > 0$. Along the balanced growth path, the average growth of city-specific output equals the constant rate of aggregate output growth. The city-specific variables in the model's competitive equilibrium fluctuate around this path. This concept of balanced growth is equivalent to the one considered by Eaton and Eckstein (1997), Rossi-Hansberg and Wright (2007), and Córdoba (2008).[6]

[6] The transformation of the model with growth into the stationary planning problem considered in the previous section is standard. See DFW for the details.

We now use the model's resource constraints to derive the balanced growth rate of per capita consumption as a function of the exogenous growth rates of technology and population. The aggregate resource constraint (3) implies that per capita consumption and output growth both equal per capita growth in consumption units of each of the three aggregate capital stocks. In particular,

$$g_c = \gamma_j^{-1} g_{k_j}, \quad j = b, f, h, \tag{10}$$

where g_c denotes the growth rate of consumption and g_{k_j} denotes the growth rate of j-type capital. The city-specific land constraints (6) imply that the growth rate of finished land per capita, g_l, satisfies:

$$g_l = \gamma_n^{\zeta-1} g_{k_f}^{\zeta}. \tag{11}$$

Imposing the equilibrium condition $x = y/l_b$ in the city-specific output resource constraint (4) and solving for output, we obtain the final ingredient of the balanced growth formula:

$$g_c = \gamma_a^{(1-\alpha)\delta} {g_l}^{1-\delta} {g_{k_b}}^{\alpha\delta}, \tag{12}$$

where the parameter

$$\delta \equiv \phi\lambda$$

measures the effect of agglomeration on productivity net of diminishing returns from congestion. Replacing g_l and g_{k_b} in (12) using (10) and (11) and solving for g_c yields:

$$g_c = \gamma_a^{\frac{(1-\alpha)\delta}{1-\delta\alpha+(\delta-1)\zeta}} \gamma_b^{\frac{\alpha\delta}{1-\delta\alpha+(\delta-1)\zeta}} \gamma_f^{\frac{\zeta(1-\delta)}{1-\delta\alpha+(\delta-1)\zeta}} \gamma_n^{\frac{(1-\zeta)(\delta-1)}{1-\delta\alpha+(\delta-1)\zeta}}. \tag{13}$$

Equation (13) relates per capita consumption growth to the growth in technology and population. This relationship depends on the size of agglomeration effects, δ, and the returns to infrastructure in producing finished land, ζ. Examining the first two exponents in (13) shows that agglomeration amplifies the impact of neutral and business investment technological growth, but only with diminishing returns to infrastructure, $\zeta < 1$. Intuitively, when $\zeta = 1$, finished land grows too fast and limits the growth of density. For example, we see from (10) and (11) that if $\zeta = \gamma_f = 1$, then land growth keeps pace with output growth and density does not grow at all. Next, if $\delta > 1$, consumption growth increases with population growth when $\zeta < 1$, and consumption growth decreases with infrastructure technology growth when $\zeta > 0$.[7] Empirically, infrastructure technology grows more slowly than consumption, so $\gamma_f < 1$. In this case, if $\delta > 1$, then relatively slow infrastructure technical change limits the expansion of finished land, thereby raising density and consumption growth. In sum, diminishing returns to infrastructure are critical if agglomeration is to increase consumption growth via growth in population, neutral technology, or business investment technology. However, even if $\zeta = 1$, agglomeration can raise consumption growth because of the relatively slow growth of finished land that arises from $\gamma_f < 1$.

A more intuitive expression for g_c can be derived from the first order condition for finished land in intermediate goods production. This condition states that expenditures on finished land by intermediate goods producers in a city equal a constant fraction of that city's output in consumption units. Therefore, consumption growth equals the product of the growth rates of the stock of finished land and the average rent on land, g_{p_l}:

$$g_c = g_l g_{p_l}. \tag{14}$$

Replacing g_l in (12) using (14) yields:

$$g_c = \gamma_a \gamma_b^{\frac{\alpha}{1-\alpha}} g_{p_l}^{\frac{\delta-1}{\delta(1-\alpha)}}. \tag{15}$$

Equation (15) relates per capita consumption growth to neutral and business investment technological change and to growth in the rental price of land. If agglomeration and congestion effects cancel, $\delta = 1$, then consumption growth is the same as in the neoclassical growth model with neutral and investment-specific technical change and without land. The additional term shows that the impact of agglomeration on growth depends on the presence of a trend in land rents. If rents are growing, $g_{p_l} > 1$, then equation (14) implies land growth does not keep pace with output growth, and the density of economic activity grows. If agglomeration dominates congestion, $\delta > 1$, then this mechanism is an additional source of growth. Without the effect of density on productivity, $\lambda = 1$ and $\delta = \phi$, equation (15) implies:

$$g_c = \gamma_a \gamma_b^{\frac{\alpha}{1-\alpha}} g_{p_l}^{\frac{\phi-1}{\phi(1-\alpha)}}.$$

[7] The dependence of per capita economic growth on population growth is a feature shared by models of endogenous technical change. Although some empirical tests of this dependence (e.g. Levine and Renelt 1992), are inconclusive, Kremer (1993) provides supportive evidence.

If $g_{p_l} > 1$ and $0 < \phi < 1$, consumption growth is lower than predicted by technical change alone, because of decreasing returns to land in production.

3 Empirical Strategy

We now discuss how to estimate of the effect of local agglomeration on aggregate consumption growth by using equation (13), which relates consumption growth to underlying technology and population growth. In a competitive equilibrium investment good (and capital) prices equal the inverse of the corresponding investment technology. Therefore we can estimate γ_a from (13) as:

$$\tilde{\gamma}_a = \tilde{g}_c^{\frac{(1-\alpha)\delta+(\zeta-1)(\delta-1)}{(1-\alpha)\delta}} \tilde{\gamma}_n^{\frac{(1-\zeta)(1-\delta)}{(1-\alpha)\delta}} \tilde{g}_{p_b}^{\frac{\alpha}{(1-\alpha)}} \tilde{g}_{p_f}^{\zeta\frac{1-\delta}{(1-\alpha)\delta}}, \tag{16}$$

where $\tilde{x}$ denotes the point estimate of x and g_{p_j} denotes the growth rate of the price of investment good j. Next, we construct g_c^*, the counterfactual growth rate of consumption without agglomeration, $\delta = \phi$, from (13) and substitute for $\tilde{\gamma}_a$ using (16):

$$g_c^* = \tilde{g}_c^{\frac{1-\alpha\delta-\zeta(1-\delta)}{\lambda[1-\alpha\phi-\zeta(1-\phi)]}} \tilde{\gamma}_n^{\frac{(1-\lambda)(1-\zeta)}{\lambda[1-\alpha\phi-\zeta(1-\phi)]}} \tilde{g}_{p_f}^{\frac{\zeta(1-\lambda)}{\lambda[1-\alpha\phi-\zeta(1-\phi)]}}, \tag{17}$$

where δ corresponds to its empirical value. We use

$$\Lambda = \frac{\tilde{g}_c - g_c^*}{g_c^* - 1} \tag{18}$$

to measure the increase in per capita consumption growth due to agglomeration.

Estimating Λ is the objective of our empirical work. To this end, we need estimates of the growth rates g_c, g_{p_b}, g_{p_f}, and γ_n, as well as the parameters λ, α, ϕ, and ζ. Our strategy is to use relationships predicted by the model to form a sufficient number of moment conditions to identify all the parameters and growth rates necessary for obtaining Λ. We identify λ using panel data for 22 cities in the United States from 1978 to 2009. For the remaining parameters and growth rates, we use aggregate time series on prices and quantities. Our GMM estimation integrates the micro and macro data so that estimates of the standard error of Λ includes sampling uncertainty in both. The discussion below is relatively brief; see DFW for more detail.

3.1 Identification with a Panel of US Cities

We identify λ using the model's predicted relationship between measured TFP and the rental price of land in units of the locally produced good. Imposing the equilibrium conditions that $x = y/l_b$ in each city and applying a log transformation to the definition of TFP in (2) yields TFP for city i at date t:

$$\ln a_{it} = \frac{\lambda - 1}{\lambda}\left[\ln y_{it} - \ln l_{bit}\right] + (1-\alpha)\phi\left[\ln z_{it} + \ln \Upsilon_t\right]. \tag{19}$$

We do not have data on city-level output or land to measure the density externality directly. However, the density externality can be expressed in terms of observable factor prices by exploiting the intermediate goods producers' and housing service providers' first order conditions for land use. Before making these substitutions, for any generic variable, x_{it}, we subtract the cross-sectional time t mean of $\ln x_{it}$ and define:

$$\hat{x}_{it} \equiv \ln\left(x_{it}\right) - \frac{1}{N}\sum_{j=1}^{N}\ln\left(x_{jt}\right).$$

By construction, $\hat{x}_{it}$ is mean zero in every year. This transformation eliminates variation in the aggregate technology. Substituting factor prices into (19) then yields:

$$\hat{a}_{it} = \frac{\lambda - 1}{\lambda}\hat{r}_{it} + (1-\alpha)\phi\hat{z}_{it} \tag{20}$$

with local-good-denominated land rents, $\hat{r}_{it}$, measured as:

$$\hat{r}_{it} = \frac{1}{1-\omega}\hat{r}_{hit} - \hat{p}_{yit}, \tag{21}$$

where r_h denotes the rental price of housing services and p_y denotes the output price.

Our estimation of λ is built around (20) with conventional measurement of TFP and data on housing rents and output prices. Estimation must address the model's prediction that the idiosyncratic technology term, $\hat{z}_{it}$, is correlated with local-output-denominated land rents. We address this endogeneity by adopting a natural stochastic process for $\hat{z}_{it}$ that is motivated by two key empirical observations about cities "Gibrat's" law" and "Zipf's law," (Gabaix 1999). Gibrat's law for cities states that their populations follow similar growth processes with a common mean and variance. Gabaix (1999) shows that if cities satisfy Gibrat's law, then the probability that the normalized size of a city is greater than some level S is given by $\Pr(\text{Size} > S) = a/S^b$, where $b = 1/(1 - S_{\min}/\bar{S})$, $S_{\min}$ is the lower barrier of a reflected random walk in $\ln S$, and $\bar{S}$ is mean city size. In the case of city population, $S_{\min}$ is presumably small relative to $\bar{S}$ so that $b \simeq 1$. Zipf's law is that $b = 1$ for cities.

Although population need not inherit the distribution of productivity, we verify in Section 6 that if productivity is (approximately) exponentially distributed in our model, then population, which is endogenous, is as well. Furthermore, we discuss evidence that city-specific technology is indeed exponentially distributed in its upper tail. These considerations motivate the following process for idiosyncratic technology:

$$\begin{aligned}
\ln z_{it+1} &= \max\left\{\ln z_{it} + \varepsilon_{it+1}, \ln z_{\min,t}\right\};\\
\ln z_{\min,t} &= \ln z_{\min} + \ln \bar{z}_t;\\
\bar{z}_t &= E_t z_{it};\\
\varepsilon_{it+1} &\sim N(0, \sigma_\varepsilon^2).
\end{aligned}$$

This process characterizes city-specific technology as a random walk with a reflecting barrier. Gabaix (1999) shows that there exists an invariant distribution for the differences $\ln z_t - \ln z_{\min,t}$ and that the invariant distribution has an exponential upper tail. In our empirical work, we sample 22 cities from the upper tail of the city distribution and assume that $\ln z_{it} > \ln z_{\min,t}$ for these cities over the sample period.

With this stochastic process $\Delta\hat{z}_{it} = \varepsilon_{it}$, where Δ is the first-difference operator. It then follows from (20) that:

$$\Delta\hat{a}_{it} = \frac{\lambda - 1}{\lambda}\Delta\hat{r}_{it} + (1-\alpha)\phi\varepsilon_{it}. \tag{22}$$

Although ε_{it} is correlated with the other variables in (22), we can easily address this endogeneity with instrumental variables. Because the error term in (22) is independently distributed across time and space, *any* variable dated $t-1$ and earlier is orthogonal to ε_{it}, and hence is a valid instrument. Consequently, we estimate λ using the moment conditions:

$$E\left\{\left[\Delta\hat{a}_{it} - \frac{\lambda-1}{\lambda}\Delta\hat{r}_{it}\right] \times \hat{v}_{it-j} \mid \Omega\right\} = 0 \tag{23}$$

for valid instruments $\hat{v}_{it-j}$, $j \geq 1$ and for conditioning information Ω equal to a date t, a city i, or the null set. We discuss our selection of instruments below.

It is apparent from equation (23) that our estimation identifies λ through both the time series and the cross-section covariation of TFP growth with the instrumental variable forecast of growth in local-good-denominated land rents. A positive correlation between TFP growth and predicted land rents indicates that $\lambda > 1$ and the externality is operative. Predicted increases in land rents lead firms to economize on land, thereby raising density and TFP growth. In our model, TFP growth is positively correlated contemporaneously with *realized* land rents regardless of the density externality. Nevertheless, our instrumental variables strategy identifies λ because predicted rents are orthogonal to the TFP shock driving the contemporaneous correlation. In practice, other shocks, such as to product demand, property taxes, and amenities, contribute to variation in land rents. The presence of such shocks is helpful because it reduces the standard error of our estimate of λ. It is straightforward to introduce these extra sources of variation into the model, but doing so does not change (23) because the relevant first order conditions are unchanged.

To implement the estimation, we measure log TFP as the difference between the log of output and the factor-share weighted sum of the log of factor inputs. From (1),

$$\ln a_{it} = \ln y_{it} - (1-\phi)\ln l_{bit} - \alpha\phi \ln k_{it} - (1-\alpha)\phi \ln n_{it}. \tag{24}$$

We use the first order conditions for intermediate good producers' and housing service providers' factor input choices to replace the quantities that we cannot measure with observable housing rents, wages, or local good prices. Combining this substitution with the equality of rental rates of non-infrastructure capital across cities, we obtain the following measurement equation for TFP:

$$\hat{a}_{it} = \frac{1-\phi}{1-\omega}\hat{r}_{hit} + (1-\alpha)\phi\hat{w}_{it} - \hat{p}_{yit}, \tag{25}$$

where w_{it} denotes the local wage rate.

So far we have assumed that workers are homogeneous. Ignoring cross-sectional variation in labor quality biases this measure of TFP because denser urban areas attract higher skilled workers. To address the bias, we follow Ciccone and Peri (2006) and consider "skilled" and "unskilled" workers as imperfect substitutes in producing total labor services. We now derive a version of (25) that incorporates heterogeneous workers in a way that does not affect the balanced growth path of the model.

Suppose that the effective labor input in (1), n, is a constant elasticity of substitution composite of unskilled, n_u, and skilled labor, n_s:

$$n = \left[\sigma n_u^{\xi} + (1-\sigma) n_s^{\xi}\right]^{1/\xi},$$

where $0 < \sigma < 1$ and $\xi \leq 1$. Effective labor satisfies the intermediate good producers' first order condition for labor as before. Let w_u and w_s denote the wage rates of unskilled and skilled workers. The first order conditions for intermediate goods producers' choices of unskilled and skilled labor can be used to express the wages of skilled workers as:

$$w_s = (1-\sigma)\sigma^{1/\xi - 1}(1-\alpha)\phi w s^{1/\xi - 1} m^{\xi - 1},$$

where w denotes the implicit wage for the composite labor input, $s \equiv (w_u n_u + w_s n_s)/(w_u n_u)$, and $m \equiv n_s/n_u$. Substituting for composite wages, w, in (25), we arrive at the following measure of TFP used in our estimation:

$$\hat{a}_{it} = \frac{1-\phi}{1-\omega}\hat{r}_{hit} + (1-\alpha)\phi\left[\hat{w}_{sit} - \frac{1-\xi}{\xi}\hat{s}_{it} - (\xi - 1)\,\hat{m}_{it}\right] - \hat{p}_{yit}. \tag{26}$$

Equation (26) reduces to equation (25) if $\xi = 1$, that is, if unskilled and skilled labor are perfect substitutes. We estimate the substitutability parameter ξ with λ using the same set of moment conditions in (23).

3.2 Identification with Aggregate US Data

Identifying λ and ξ using (23) requires that we also estimate ω, α, and ϕ. To complete the measurement of Λ, we need to estimate g_c, g_{p_f}, γ_n, and ζ. We identify the growth rates using the moment conditions:

$$E\left\{(\ln X_t - \ln(g)t)\cdot t\right\} \quad = \quad 0, \;\; X = C, P_f, N \text{ and } g = g_c, g_{p_f}, \gamma_n, \tag{27}$$

where E denotes the unconditional expectations operator and P_f denotes the price of infrastructure investment goods. Equation (27) assumes that $\ln X_t$ is de-meaned prior to the analysis. (Our standard errors incorporate the sampling uncertainty in these means.)

We identify ω, the share of structures in housing services, using the housing service providers' first order conditions, which equate two ratios: the ratio of residential structures income to finished land income and the ratio of the respective share parameters in the production function for housing services. We cannot measure these income flows, so instead we measure the values of the underlying assets and relate these values to the income flows using an arbitrage condition for the provision of finished land.

Finished land is raw land combined with infrastructure. It follows that in each city the value of a unit of finished land satisfies the following arbitrage condition:

$$p_{lit} = E_{t|i} \left\{ \frac{1}{R} \left[r_{lit+1} + (1-\kappa_f)^{\zeta} p_{lit+1} \right] \right\}, \tag{28}$$

where $E_{t|i}$ is the expectation at time t for city i; r_{li} is the rental rate of finished land in city i; p_{li} is the price of finished land in city i; and $R \equiv 1/\beta$ is the interest rate. Equation (28) states that the value of an additional unit of finished land equals the expected discounted rent from that land plus the value next period of the finished land with the un-depreciated infrastructure. Along a balanced growth path:

$$E_t p_{lit+1} = g_{p_l} E_t p_{lit}, \tag{29}$$

where E_t is the expectation over all cities at time t. Since $E_t E_{t|i} x_{it} = E_t x_{it}$ for any random variable x_{it}, it follows from (28) and (29) that:

$$E_t p_{lit} = \frac{g_{p_l}}{R - (1-\kappa_f)^{\zeta} g_{p_l}} E_t r_{lit}. \tag{30}$$

That is, the average value of finished land is proportional to the average rent on that land.

The moment condition we use to identify ω is then:

$$E \left\{ \frac{\sum p_{lit} l_{hit}}{\sum (P_{ht} k_{hit} + p_{lit} l_{hit})} \left[\frac{\omega}{1-\omega} \frac{R/g_{p_l} - (1-\kappa_f)^{\zeta}}{R/g_{p_h} + \kappa_h - 1} + 1 \right] - 1 \right\} = 0, \tag{31}$$

where the summations are over all cities and P_h is the price of residential structures. We estimate the depreciation rates in this expression as their sample average values:

$$E \left\{ \kappa_j - \frac{P_{jt} D_{jt}}{P_{jt} K_{jt}} \right\} = 0, \;\; j = f, h \tag{32}$$

where D_{jt} is real depreciation of j-type capital. The growth rate of the price of residential structures, g_{p_h}, is identified using the analogue of (27). To estimate g_{p_l}, we use another implication of the housing service providers' first order conditions:

$$E \left\{ \left(\ln E_t r_{hit} - \left[(1-\omega) \ln(g_{p_l}) + \omega \ln(g_{p_h}) \right] t \right) \cdot t \right\} = 0. \tag{33}$$

Equation (33) states that the growth rate of the (de-meaned) log of average housing services rent is a weighted average of the log growth rates of finished land and residential structures prices. This relationship relies on structures and land rents inheriting the trends of their respective asset prices, which in turn follows from the household's first order conditions for capital accumulation.

Next we consider ζ, α, and ϕ. To identify ζ, we exploit connections between value and income analogous to (30), which allow us to derive the identifying moment condition:

$$E\left\{\frac{R/g_{p_l}-(1-\kappa_f)^{\zeta}}{R/g_{p_f}-(1-\kappa_f)}\zeta-\frac{\sum P_{ft}k_{fit}}{\sum\left(p_{lit}l_{bit}+p_{lit}l_{hit}\right)}\right\}=0.$$

We identify α and ϕ using labor's share of total income and finished land's share of non-labor income:

$$E\left\{\frac{\sum w_{it}n_{it}}{\sum\left[w_{it}n_{it}+r_{lit}l_{bit}+r_{bt}k_{bit}\right]}-\phi(1-\alpha)\right\}=0;$$

$$E\left\{\frac{\sum p_{lit}l_{bit}}{\sum\left[P_{bt}k_{bit}+p_{lit}l_{bit}\right]}\left[\frac{\alpha\phi}{1-\phi}\frac{R/g_{p_l}-(1-\kappa_f)^{\zeta}}{R/g_{p_b}+\kappa_b-1}+1\right]-1\right\}=0,$$

where r_b and P_b are the rental rate and price of business capital. The second condition again exploits connections between value and income analogous to (30) and adds g_{p_b} and κ_b as parameters to estimate. We identify these additional parameters using the analogues of (27) and (32).

In sum, these moment conditions are sufficient to obtain the six proximate inputs into Λ: g_c, g_{p_f}, λ, α, ϕ, and ζ. Estimating these inputs in turn requires moment conditions to identify all the depreciation rates and investment price growth rates in the model, plus ω. Only two parameters in our model are not used: the weight on housing services in preferences, ψ, and the parameter governing the elasticity of substitution of the city goods, η (β is implicit in R). In total, there are 13 moment conditions involving aggregate variables.

3.3 Estimation

We estimate Λ in three steps. First, we collect the 13 moment conditions involving aggregate variables into a vector-valued function $\Psi\left(X_t,\theta\right)$, so that:

$$E\Psi\left(X_t,\theta\right)=0. \tag{34}$$

Here, X_t is a vector of the aggregate variables included in these moment conditions, and θ is a parameter vector given by:

$$\theta \equiv \left[\kappa_b, \kappa_s, \kappa_f, g_{p_l}, g_{p_b}, g_{p_h}, g_{p_f}, \gamma_n, g_c, \alpha, \phi, \omega, \zeta\right]'.$$

Because the dimensions of Ψ and θ are equal, this system of moment conditions is exactly identified. We estimate equation (34) by GMM, in which we use a Newey-West weight matrix with a lag length of 2. Next, we estimate λ and ξ using the moment conditions in (23) and the measure of TFP given in (26). This estimation requires that we plug in the estimates of ω, α, and ϕ from the first step into (23). To account for the sampling variation associated with these three plug-in parameters, we adjust the weight matrix using their respective influence functions as in Newey and McFadden (1994). We also cluster the weight matrix at the city level. Third, we substitute the point estimates for g_c, g_{p_f}, $\delta = \lambda\phi$, α, ϕ, and ζ into (18) to obtain Λ. To calculate the sampling variance of Λ, we need the joint covariance matrix of these six parameters, which we calculate by stacking the influence functions of the parameters as shown by Erickson and Whited (2002). After this calculation, a standard application of the delta method gives the variance of Λ. See DFW for details.

We now describe our choice of instruments, v_{it-j}, and conditioning set, Ω, needed for estimation of (23). Recall that any variable dated $t-1$ or earlier is orthogonal to ε_{it} and hence is a valid instrument. To allow for the possibility that our data contain *i.i.d.* measurement error, we use variables dated $t-2$ and earlier because first-difference measurement errors are correlated with instruments dated at $t-1$. Next, two kinds of conditioning sets have been proposed for traditional dynamic panel estimation, with fixed sample length T and large N. For any instrument x, the moment condition:

$$E\{x_{it-2}\varepsilon_{it} \mid \Omega\} = 0$$

holds for each date, $\Omega = t$, for each city, $\Omega = i$, and therefore also unconditionally. In one approach, (e.g. Holtz-Eakin, Newey, and Rosen 1988) expectations of each date t cross-section are used so that the empirical counterpart to the moment condition is evaluated across cities at each date t:

$$\sum_{i=1}^{N} x_{it-2}\varepsilon_{it} = 0. \tag{35}$$

This approach involves $T-2$ moment conditions for each instrumental variable x. In the other, (e.g. Anderson and Hsiao 1982) the unconditional expectation is used and the sample moment condition is evaluated over cities and time:

$$\sum_{t=3}^{T}\sum_{i=1}^{N} x_{it-2}\varepsilon_{it} = 0. \tag{36}$$

This second approach involves one moment condition per instrumental variable.

The first approach has the advantage that it exploits more information and therefore is more efficient. It is usually applied in cases with large N and small T because the cross-sectional average is likely a good measure of the cross-sectional expectation. In this case, levels are usually used as instruments for growth rates to conserve observations. In contrast, in our case with relatively large T and small N, the second approach might be warranted because with small N the cross-sectional average could be a poor estimate of the cross-sectional expected value.

The choice between the two approaches depends on the quality of the instruments in each approach. We cannot use the weak-instrument tests surveyed in Stock, Wright, and Yogo (2002) because they do not apply to nonlinear GMM estimators. Instead, we make our instrument choice using the test in Wright (2003) for under-identification of a nonlinear GMM model. The null hypothesis of this test is that the gradient of the moment conditions with respect to the parameters is not of full rank. It is therefore literally a test of under-identification rather than of weak identification. Although we cannot therefore test for weak instruments, the Wright test is conservative in the sense that its size is less than its significance value.

The variables in our instrument set are the measured variables in (26) plus log house prices and log per capita income. We consider four possibilities: twice-lagged instruments in levels and twice-lagged instruments in differences, where in each case we consider moment conditions based on both (35) and (36). For the moment conditions from (35), we never reject the null of an under-identified model. For the moment conditions from (36), we strongly reject the null of an under-identified model (p-value of 0.006) for the instruments in differences, but we cannot reject the null of an under-identified model for the instruments in levels (p-value of 0.051). We therefore use the moment conditions based on (36) with twice lagged differences as instruments.

To assuage concerns about the validity of asymptotic theory for a small panel such as ours, we conduct a Monte Carlo experiment in which we estimate our model on simulated data that approximates our actual data closely in terms of means, variances, covariances, and autocorrelations. Details are in DFW. We find that our procedure produces unbiased estimates and that the t-tests on our parameters of interest tend to under-reject slightly in samples of the size we consider; that is, they are conservative.

4 Data

This section describes our annual data, which span the period 1978-2009. We begin with the panel data for the moment conditions (23) and then discuss the aggregate data for the moment conditions (34). Details concerning the data are in DFW.

4.1 Panel Data

A location in our model is a land mass with at least some infrastructure where households live and work. We use Metropolitan Statistical Areas (MSAs) to represent these locations. The geography of an MSA is a natural empirical counterpart to our locations because the criteria for defining them is explicitly economic, based on commuting-to-work patterns.[8] MSAs are defined as contiguous combinations of counties, but the included counties differ over time and by data source. However, each MSA comprises a core set of counties that is common across data sources and time. This core includes a substantial majority of the population corresponding to any particular MSA definition.

MSA-level panel data on output prices, the labor market, and housing rents come from several sources. We create MSA-specific price indices for output, p_{yit}, with data from the Bureau of Economic Analysis (BEA). These indices are weighted averages of industry-specific price indices, with weights equal to the share of total earnings paid to employees in each industry. The mix of industries varies across MSAs and over time within each MSA, and prices vary by industry, so output prices vary by MSA.

[8] Ciccone and Hall (1996) define a location to be a county. Rozenfeld, Rybski, Gabaix, and Makse (2011) propose an alternate definition of a location based satellite imagery that emphasizes geographically dense areas of integrated economic activity.

We construct MSA-level labor market variables with data from the March Current Population Survey. We define skilled workers to have at least four years of college, and unskilled workers to have fewer than four years of college. The average hourly wage of skilled workers, w_{sit}, is total wages paid to skilled workers divided by their total hours worked; the ratio of skilled to unskilled labor, m_{it}, is the ratio of total hours worked for these two groups; and the inverse share of wages going to unskilled workers, s_{it}, is the total wage bill divided by wages paid to unskilled workers.

We construct MSA-level data on housing rents, r_{hit}, by merging data from the 1990 Decennial Census of Housing (DCH) with housing rental price indices from the Bureau of Labor Statistics (BLS). We estimate the level of housing rents by MSA in 1990 with the DCH data. We then use the MSA-specific BLS price indices for shelter to extrapolate rents backwards from 1990 to 1978 and forwards to 2009. Annual average housing rents are measured by averaging across the MSAs.

When estimating (23), we use two additional MSA-level variables as instruments: per-capita personal income and house prices. Personal income is from the BEA's Local Area Personal Income Tables. House prices are measured using the repeat-sales price indices for existing homes produced by the Federal Housing Finance Agency.

After merging all non-missing data on output prices, housing rents, and labor market variables, the sample includes 22 MSAs covering the 1978-2009 period. This sample of cities accounts for 37% of the U.S. population. With the exception of average housing rents in equation (33), the moment conditions involving these variables are based on within-year deviations from averages, and therefore in these cases we do not adjust nominal variables for overall price inflation. Average housing rents are deflated by the consumption price index described in the next sub-section.

4.2 Aggregate Data

Data on aggregate capital stocks, their depreciation rates, and their price indices are from the Fixed Asset Tables, produced by the BEA. Residential structures, K_{st}, are the BEA measure of residential fixed assets. Business capital, K_{bt}, is defined as all fixed assets (private and government) plus consumer durable goods, less residential structures and infrastructure capital. The aggregate stock of infrastructure capital,

$K_{ft} \equiv \sum k_{fit}$, is defined as all government-owned highways and streets, transportation structures, power structures, and sewer and water systems, plus all privately owned power and communication structures, transportation structures, and structures related to water supply, sewage and waste disposal, public safety, highways and streets, and conservation and development. We think of infrastructure as any structure that makes the underlying land more productive for any end user, business or household. Our measurement of infrastructure capital is largely consistent with the definition proposed by Cisneros (2010) as "basic systems that bridge distance and bring productive inputs together." Our data for depreciation, D_{st}, D_{bt} and D_{ft}, and investment price indices are consistent with our definitions of the capital stocks. The real investment prices, P_{st}, P_{bt}, and P_{ft}, are the investment price indices divided by our consumption deflator, described below.

We construct value shares as follows. Data for measuring the share of housing value attributable to finished land, $\sum p_{lit} l_{hit}$ and $\sum (P_{st} k_{sit} + p_{lit} l_{hit})$, are from an update to the data described in Davis and Heathcote (2007). To measure the share of land in the sum of the values of business capital and land, we need $\sum p_{lit} l_{bit}$ and $\sum (P_{bt} k_{bit} + p_{lit} l_{bit})$, which we obtain from the BEA's Flow of Funds Accounts and Fixed Assets Tables. To measure the share of finished land value attributable to infrastructure, we need measures of the value of infrastructure capital, $\sum P_{ft} k_{fit}$, and the value of all finished land, $\sum (p_{lit} l_{bit} + p_{lit} l_{hit})$. The former is just the product of the price and quantity of infrastructure described above, $P_{ft} K_{ft}$. The business portion of the latter is from the Flow of Funds Accounts, and the housing portion is from Davis and Heathcote (2007). Labor income, $\sum w_{it} n_{it}$, is from the National Income and Product Accounts (NIPA). Total income, $\sum (w_{it} n_{it} + r_{lit} l_{bit} + r_{bt} k_{bit})$, is defined as Gross Domestic Income, as reported by the BEA, plus an estimate of the service flow from the stock of durable goods and less housing services.

Finally, we construct real per capita aggregate consumption as nominal consumption divided by its price index, divided again by the population. Nominal consumption is total consumption from the NIPA, less consumption of housing services and expenditures on durable consumption goods, plus government consumption expenditures and our estimate of the service flow from the stock of consumer durable goods. The price index for consumption is consistent with this quantity measure. The annual population data are from the BLS and correspond to the non-institutional population over the age of 16.

5 Empirical Findings

If $\lambda = 1$, there is no agglomeration, and it cannot affect aggregate growth. Therefore, we begin by confirming that agglomeration is operative; that is, $\lambda > 1$. We then discuss our baseline estimates of the effect of agglomeration on growth, Λ, as well as the remaining parameter estimates. Finally, we discuss several model specification tests and consider the effects of some of our assumptions on the baseline estimates.

5.1 Confirmation of $\lambda > 1$

The top row of Table 1 displays the baseline GMM estimate of λ. This estimation produces point estimates of all of the parameters and accounts for sampling uncertainty in the measurement of TFP and the rental price of land.[9] We estimate λ to be 1.069 and with a standard error of 0.03. This point estimate is significantly larger than 1, and it implies that a 10% increase in output per unit of land increases local TFP by 0.62%. We thus conclude that the agglomeration externality is operating in our sample of cities.

To build confidence in this conclusion, and to illustrate the source of identification of λ, we consider several informal estimation strategies that in principle should yield similar results. Suppose we measure land prices and TFP using (21) and (26), with the parameters taken from the same GMM estimation that produces our estimate of λ, and then use these estimated variables as data in (22).[10] In this case, we can use linear methods to estimate λ from the coefficient on land prices, which is a function of λ. We consider both OLS and 2SLS estimation, where in each case we use the data in three ways: we pool the entire panel, we consider each cross-section separately, and we consider each time series separately. In the second two cases, we report the average estimate and calculate standard errors, as in Fama and MacBeth (1973). For 2SLS, we use the same instrument set as we do in our GMM estimation.

[9] We assume an interest rate of $R = 1.05$. Our estimates are largely insensitive to this choice, and our standard errors are nearly invariant to including the sampling variation from estimating R.

[10] We report the complete set of GMM parameter estimates in Table 2.

The rest of Table 1 reports these results. The OLS and 2SLS panel estimates are close to those based on cross-section or time-series data only. We thus confirm the model's prediction that identification comes from both dimensions, so that the samples should yield similar information. All but the 2SLS cross-section estimates are statistically significant. Lending credibility to the instruments underlying the baseline estimates, the F-statistic from the first stage of 2SLS with the pooled sample (not shown) is 40.8, far exceeding the criterion for instrument quality in Stock et al. (2002).

Figure 1 plots the pooled data on measured TFP and land rents, along with the associated OLS and 2SLS regression lines, with the former based on contemporaneous land rents and the latter based on land rents predicted at $t-2$ by the instruments. These regression lines have slopes equal to $(\tilde{\lambda}-1)/\tilde{\lambda}$, where $\tilde{\lambda}$ refers to the corresponding point estimate of λ in Table 1. Notice the clustering of observations along the 2SLS regression line. To confirm this impression, we re-estimate the slope of the regression line, trimming the data if in absolute value measured TFP exceeds 0.1 and predicted land rents exceed 0.05. The resulting slope of 0.086 is nearly identical to its unconditional value of 0.090; outliers are not driving our results.

5.2 Baseline Estimates of Λ and its Underlying Parameters

Having confirmed that agglomeration is operative in our sample of cities, we now address the question of its contribution to aggregate growth. Table 2 reports the parameter estimates from our baseline GMM estimation, along with asymptotic standard errors. We estimate that local agglomeration raises aggregate per capita consumption growth by $\Lambda = 10.2\%$ (standard error 5%), which we find despite our relatively small estimate of the net impact of density on the level of productivity, $\delta = \phi\lambda = 1.041$ (0.016). We conclude that the aggregate effect of statistically significant local agglomeration is itself significant.

Are these agglomeration effects economically significant? The estimate in Table 2 of per capita consumption growth of 1.7 percent per year would be 1.51% in the absence of local agglomeration. From our model, it is straightforward to show that per capita consumption and housing would have to be 3% larger each year in perpetuity to compensate individuals for the absence of the aggregate growth that stems from local agglomeration.[11] This calculation is equivalent to \$360 billion in 2009 alone, and the present value of this compensation is over \$7.5 trillion. Thus, seemingly small local effects of agglomeration have large aggregate consequences.

Our estimate of the impact of local agglomeration on *aggregate growth* is related to a large literature that measures the effect of local agglomeration on the *local level* of firm productivity and wages. These studies estimate the percentage increase in productivity or wages due to agglomeration, holding fixed all factors of production. Our estimate of 4.1% for the net effect of agglomeration on local productivity, δ, is in the middle of the range of 2% to 6%, as estimated recently by Combes, Duranton, and Gobillon (2008) and surveyed by Rosenthal and Strange (2004).[12]

We now discuss the remaining parameters in Table 2, focusing on the most noteworthy. We estimate the share of infrastructure in finished land, ζ, to be 0.545 (0.053). We are unaware of any previous estimates of this parameter. Given our estimate of finished land's share of production, $1-\phi$, infrastructure's share of production is 1.4%, and raw land's share of production is 1.2%. The latter share is close to the value reported by Ciccone (2002). Our estimate of ζ, combined with estimates of g_{p_f}, g_c, g_{p_l} and γ_n, also has implications for the growth rate of per capita finished land, g_l. In particular, g_l can be expressed as $(g_c/g_{p_f})^{\zeta}\gamma_n^{\zeta-1}$ or, equivalently, g_c/g_{p_l}. We do not impose the over-identifying restriction implied by these two expressions in estimation, but we verify that it is not rejected below. Furthermore, we cannot reject the null that both expressions equal 1 (p-value 0.99). That is, our estimates imply that along the balanced growth path, finished land per person is constant, and the density of economic activity is growing at the same rate as per capita consumption, 1.7% a year.

We estimate that the price of business capital declines at about 0.5% per year (g_{p_b}), and the prices of infrastructure and residential structures rise at 0.6% and 0.8% per year (g_{p_f} and g_{p_s}). To understand the role of these trends, we re-estimate our model imposing zero trends, $g_{p_b} = g_{p_f} = g_{p_s} = 1$. The data resoundingly reject this

[11] This calculation uses the value of $\psi = 0.2$ discussed in Section 6. See DFW for the details.

[12] As the latter authors describe, researchers focus on what are called "urbanization" and "localization" effects. Urbanization describes the impact of local output or employment density on the wages and productivity of all industries in a location. Our approach fits into this framework. Localization describes the impact of the size of an industry in a location on wages and productivity in that industry and location. We do not consider localization, although evidence described by Henderson (2003) and Rosenthal and Strange (2003) suggests it is also an important determinant of local productivity.

restriction, but conditional on it, the point estimate for the share of infrastructure in finished land, ζ, is approximately 1. When $\zeta = 1$ and $g_{p_f} = 1$, equation (17) in the model implies that agglomeration has no impact on consumption growth. Thus, these trends in capital prices are key to our finding of an economically and statistically significant impact of agglomeration on growth.

Our estimate of $\xi = 0.545$ (0.035) implies an elasticity of substitution between unskilled and skilled labor of 2.2. This estimate is somewhat larger than the range of 1.3 to 1.7 reported by Autor, Katz, and Krueger (1998).[13] However, as Autor et al. (1998) emphasize, substantial uncertainty exists concerning the magnitude of this elasticity. For example, the results in Katz and Murphy (1992) imply that our point estimate lies within a 95% confidence interval.

5.3 Tests of Model Specification

Table 3 displays results from five tests of the validity of our specification. The first two rows correspond to tests for serial correlation of the residuals in (23). If the model is correctly specified, the residuals should exhibit autocorrelation only up to order one with classical *i.i.d.* measurement error. We test the null hypotheses of no second- or third-order serial correlation with a nonlinear version of the test in Arellano and Bond (1991), finding no rejections. The third test is the Hansen (1982) and Sargan (1958) J-test of the over-identifying restrictions implicit when estimating (23). The p-value for this test indicates that the over-identifying restrictions are not rejected. The fourth row reports our test of the balanced-growth-path restriction $g_{p_l} = g_c^{1-\zeta}\gamma_n^{1-\zeta}g_{p_f}^{\zeta}$, which we do not impose in estimation. Table 3 indicates we do not reject the hypothesis that this condition is satisfied by our estimates. In the last row we test a condition concerning various growth rates that would hold if finished land was in constant supply (as opposed to finished land per person, which we find is not growing.) Because finished land does, in fact, grow, our model would be rejected if this condition held. However, we easily reject the condition.

[13] See also Heckman, Lochner, and Taber (1998) and Krusell, Ohanian, Ríos-Rull, and Violante (2000).

5.4 Effects of Model Assumptions on Estimates

We now investigate how some of our assumptions influence the results, beginning with two departures from Ciccone and Hall (1996) embedded in the model: less than perfect substitutability of city-specific goods and growth of finished land. Assuming perfect substitutability, $\eta = 1$, involves estimating (23) without using output prices in the measures of TFP and land rents.[14] Assuming no finished land growth involves setting $g_{p_f} = g_{p_l} = g_c\gamma_n$ and dropping the moment conditions used to estimate g_{p_f} and g_{p_l} from (34). We consider this case even though we reject it (see Table 3) because it shows how allowing for finished land growth affects our estimates.

Table 4 displays estimates of Λ for these two perturbations, along with the direct inputs into its calculation and the corresponding baseline estimates reproduced from Table 2. Omitting output prices raises the point estimate of Λ by almost 30%. This result is entirely due to the larger estimate of δ, which in turn is indicative of positive correlations in our sample between output prices and both skilled wages and housing rents. While the estimate of Λ is larger, it is also less precisely estimated. We therefore cannot reject the hypothesis that $\Lambda = 0$, although we still reject $\delta = \phi$. Interestingly, the estimate of δ in this case is very close to estimates in Ciccone and Hall (1996). Assuming a fixed supply of finished land also raises the point estimate of Λ by about 30%, even though it lowers the estimate of δ. The increase in Λ occurs primarily because in this counterfactual, g_{p_f} is larger than its baseline estimate.

Now we consider relaxing our assumptions that the density externality immediately impacts TFP and that allocation decisions are contemporaneous with the productivity shock. The details of this analysis are in DFW. Briefly, if we assume the density externality takes up to two years to affect TFP, our point estimates of δ and Λ rise by several tenths of a percent, with roughly proportional increases in sampling uncertainty. If allocations are chosen one period in advance of the productivity shock, our estimation is unaffected, but for each additional period we must add one more lag to the instruments for them to remain valid. We consider decisions made two to three years in advance and find that our point estimates of δ and Λ rise substantially, in some cases more than doubling. Not surprisingly, adding lags to the instruments

[14] With $\eta = 1$ and $\delta > 1$, the model might not have a unique solution. Nevertheless, it is interesting to consider this case to assess the effect of including output prices.

reduces their predictability so that sampling uncertainty rises in these cases.

6 Additional Validation of the Empirical Findings

In this section we use model-generated data to study the model's implications for the distribution of city sizes and the viability of our strategy for estimating the key agglomeration parameter λ. We verify that the model is consistent with Gibrat's and Zipf's laws and show that our empirical strategy accurately identifies λ.

6.1 Calibration and Solution

For simplicity, we abstract from growth, assume that $\omega = 0$ so that housing consists only of land, and assume that labor is homogeneous, $\xi = 1$. We set the consumption share of housing, ψ, to 0.2, which is the average consumption share of shelter in our sample from NIPA. For the other parameters in Table 2, we use the corresponding baseline estimates. Below we discuss the parameter determining the degree of substitutability in final goods production, η.

We assume a productivity process that is isomorphic to the one underlying the estimation:

$$\ln z_t = \max \left\{ \gamma_z + \ln z_{t-1} + \varepsilon_t, \ln z_{\min} \right\}, \tag{37}$$

where $\gamma_z < 0$ and ε_t is *i.i.d.* normally distributed with mean 0 and variance σ_ε^2. We normalize the lower bound $z_{\min}$ so that the mean of z is 1. This process induces an invariant distribution for z_t that is approximately exponential away from its lower tail.

Is (37) plausible? To answer this question, we need a measure of the *level* of technology, but it cannot be measured with our data because of unidentified fixed effects that arise from measuring output prices with an index. However, we can use the results from Davis, Fisher, and Veracierto (2011), who construct a measure of the level of city-specific technology using a similar model. They find that technology is exponentially distributed for the 200 largest U.S. cities. Indeed, plots of the log rank of technology versus the log level of technology are remarkably similar to the analogous plots for population, with the exception of different slopes. Population has a slope near -1 (Zipf's law) and technology has a slope of about -5. We refer to these slope values as Zipf coefficients. The parameter γ_z controls the Zipf coefficient corresponding to (37).

Rather than calibrate γ_z to match a Zipf coefficient for technology of -5 we take advantage of results in Luttmer (2007) to simplify our computations. Luttmer (2007) considers a competitive equilibrium model, in which the distribution of firm size is driven by a reflected random walk productivity process. He shows that such a process induces a firm size distribution with a Zipf coefficient that is proportional to the Zipf coefficient corresponding to the productivity process. We have verified numerically that the induced distribution of population in our model shares this property. Therefore, without loss of generality, we choose γ_z so that the Zipf coefficient for productivity equals -20 and set $\sigma_\varepsilon = 0.001$. These settings simplify our computations by limiting the domain of the productivity process.[15] With this driving process, the volatility of employment growth is about one-third its empirical counterpart, which is appropriate because productivity shocks are the only source of variation in our model, but one of many in our data.

We solve the model as follows. For given values of consumption, aggregate investment, and a worker's shadow value, the first order conditions produce the cross-sectional allocation of capital and workers. The solution is obtained by finding a fixed point in the economy-wide variables that satisfies four side conditions: consumption and investment exhaust aggregate production, the allocation of workers sums to one, city-level investment in infrastructure sums to aggregate investment, and the allocation of non-infrastructure capital exhausts its stock.[16]

6.2 Gibrat's and Zipf's Laws

We now consider the model's implications for the size distribution of cities. To assess whether the model is consistent with Gibrat's law, we examine its implication that city growth is independent of city size. To this end, we estimate an OLS regression of log population growth on log population level using a large sample of simulated

[15] Low absolute-value Zipf coefficients imply very wide domains that require many grid points.

[16] In the cases we have considered, the algorithm is well-behaved and yields a unique solution from multiple sets of starting values for consumption, aggregate investment, and a worker's shadow value.

data. The regression coefficient is positive and significantly different from zero, but it is extremely small and the R^2 of the regression is .0001. We conclude that the model is consistent with Gibrat's law.

We assess the model's implications for Zipf's law by examining whether it can match the ratio of the Zipf coefficient for technology to that for population, which in the data is about 5 to 1. The parameter governing the degree of substitutability among city-specific goods, η, can be chosen to match this ratio. Figure 2 shows log rank versus log level for technology and population from a sample drawn from the steady state distribution with $\eta = 0.927$.[17] Population is approximately exponential with a Zipf coefficient (the slope of the OLS regression line) that is five times smaller than for technology. These plots of population and technology strongly resemble those displayed for the 200 largest cities by Davis et al. (2011).

The striking flattening of the population distribution arises because factors are reallocated from low to high productivity cities, thereby making population smaller than otherwise in low productivity cities and larger in high productivity cities. This reallocation increases the cross-sectional variance of population relative to technology but leaves rank unchanged. Agglomeration amplifies this effect; when we shut down the density externality by setting $\lambda = 1$, the ratio of Zipf coefficients falls to 4. The extent of the flattening is positively related to η; with goods more easily substitutable, the incentive to reallocate resources rises. Note that with $\eta = 0.927$, our condition for all cities being occupied, $\eta\lambda\phi < 1$, is easily satisfied at our point estimates. Thus, the model is consistent with Zipf's law.

6.3 Estimation with Model-Generated Data

Table 5 shows results from estimating (22) on simulated data with OLS and (23) with GMM for two values of λ. We consider both a large sample (222 cities for 499 years) and a sample the size of our data set. For this exercise we use (25) to measure TFP, assuming correct values for the parameters.[18] Variables are only predictable

[17]The discrete approximation to (37) is relatively poor in the upper extremes and so we exclude the highest 5% of the population distribution. We also exclude the lowest 5% of the population distribution because Zipf's law does not apply to the lower tail.

[18]The Cobb-Douglas production function implies that estimating the parameters in (25) would yield the values without any sampling uncertainty.

one period ahead in our model, so for GMM we use the first lag of the instruments, which include log growth rates of wages, land rents, and output prices. Because of the model's large contemporaneous correlation between TFP and land rents, the OLS estimates are severely upward biased and thus completely uninformative about the value of λ. Our GMM implementation of instrumental variables estimation, on the other hand, does a very good job of finding the correct value of λ, independent of the sample size, although the standard errors with the small sample are relatively large.[19] Including additional sources of variation in land rents, as is the case in our data, would lower the standard errors.

7 Conclusion

Our work suggests many areas for future research, and we conclude by summarizing a few. One area needing attention is a proof of equilibrium existence and uniqueness for the general version of our model. Other limitations of our analysis include its assumptions of convenience, which may affect our conclusions. While we have addressed some of these issues in our empirical work, others remain. For example, the impact of assuming perfect mobility of non-infrastructure capital remains to be determined. Despite these shortcomings, it would be interesting to apply our methodology to quantify the role of other mechanisms in growth. Another interesting avenue for further study is additional predictions of the model for the cross-section of cities. Davis et al. (2011) have begun this work. Finally, our model can be useful for understanding how the economy responds to aggregate shocks. For example, the co-location of housing and production in our model endogenizes complementarities that Fisher (2007) shows helps to reconcile theory with aggregate evidence on investment.

References

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*(1), 47–82.

[19] We conjecture that our actual OLS estimates in Table 1 are close to our GMM and 2SLS estimates because OLS allows for attenuation bias induced by measurement error, while our GMM and 2SLS estimation corrects for this bias.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Autor, D. H., L. F. Katz, and A. B. Krueger (1998). Computing inequality: Have computers changed the labor market? *The Quarterly Journal of Economics 113*(4), 1169–1213.

Black, D. and V. Henderson (1999). A theory of urban growth. *Journal of Political Economy 107*(2), 252–284.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review 46*(2), 213–227.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *American Economic Review 86*(1), 54–70.

Ciccone, A. and G. Peri (2006). Identifying human-capital externalities: Theory with applications. *Review of Economic Studies 73*(2), 381–412.

Cisneros, H. (2010). America's essential infrastructure: A key to competitiveness. In M. Underhill (Ed.), *The Handbook of Infrastructure Investing*. John Wiley and Sons.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics 63*(2), 723–742.

Córdoba, J.-C. (2008). On the distribution of city sizes. *Journal of Urban Economics 63*(1), 177–197.

Davis, M., J. Fisher, and M. Veracierto (2011). The role of housing in labor reallocation. Revision to Chicago Fed working paper no. 2010-18.

Davis, M., J. Fisher, and T. Whited (2013). Supplementary materials for "Macroeconomic Implications of Agglomeration". Unpublished Manuscript.

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

Duranton, G. (2007). Urban evolutions: The fast, the slow, and the still. *American Economic Review 97*(1), 197–221.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In V. Henderson and J. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4 of *Handbooks in Economics*, pp. 2063–2117. North-Holland: Elsevier Science, North-Holland.

Eaton, J. and Z. Eckstein (1997). Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics 27*(4–5), 443–474.

Erickson, T. and T. Whited (2002). Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory 18*(3), 776–799.

Fama, E. F. and J. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Fisher, J. D. (2007). Why does household investment lead business investment over the business cycle? *Journal of Political Economy 115*(1), 141–168.

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

Greenwood, J., Z. Hercowitz, and P. Krusell (1997). Long run implications of investment-specific technological change. *American Economic Review 87*(3), 342–362.

Grossman, G. M. and E. Helpman (1991). Quality ladders in the theory of growth. *Review of Economic Studies 58*(1), 43–61.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

Heckman, J. J., L. Lochner, and C. Taber (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics 1*(1), 1–58.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics 53*(1), 1–28.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*(6), 1371–1395.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kehoe, T. J., D. K. Levine, and P. M. Romer (1992). On characterizing equilibria of economies with externalities and taxes as solutions to optimization problems. *Economic Theory 2*(1), 43–68.

Kremer, M. (1993). Population growth and technological change: One million b.c. to 1990. *Quarterly Journal of Economics 103*(3), 681–716.

Krusell, P., L. Ohanian, J. V. Ríos-Rull, and G. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1054.

Levine, R. and D. Renelt (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review 82*(4), 942–963.

Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*(1), 3–42.

Lucas, R. E. J. (2001). Externalities and cities. *Review of Economic Dynamics 4*(2), 245–274.

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

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Vol. 4*, pp. 2111–2245. North-Holland.

Rosenthal, S. S. and W. C. Strange (2003). Geography, industrial organization, and agglomeration. *The Review of Economics and Statistics 85*(2), 377–393.

Rosenthal, S. S. and W. C. Strange (2004). Evidence on the nature and sources of agglomeration economies. In V. Henderson and J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, pp. 2119–2171. North Holland.

Rossi-Hansberg, E. and M. L. J. Wright (2007). Urban structure and growth. *Review of Economic Studies 74*(2), 597–624.

Rozenfeld, H., D. Rybski, X. Gabaix, and H. Makse (2011). The area and populatin of cities: New insights from a different perspective on cities. *American Economic Review 101*(5), 2205–2225.

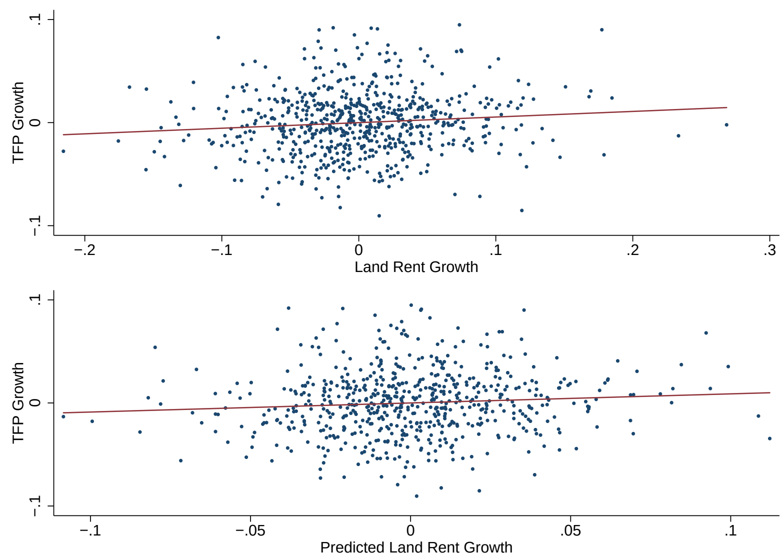
Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*(3), 393–415.

Stock, J., J. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*(4), 518–527.

Weill, P.-O. and S. Van Nieuwerburgh (2010). Why has house price dispersion gone up? *Review of Economic Studies 77*(4), 1567–1601.

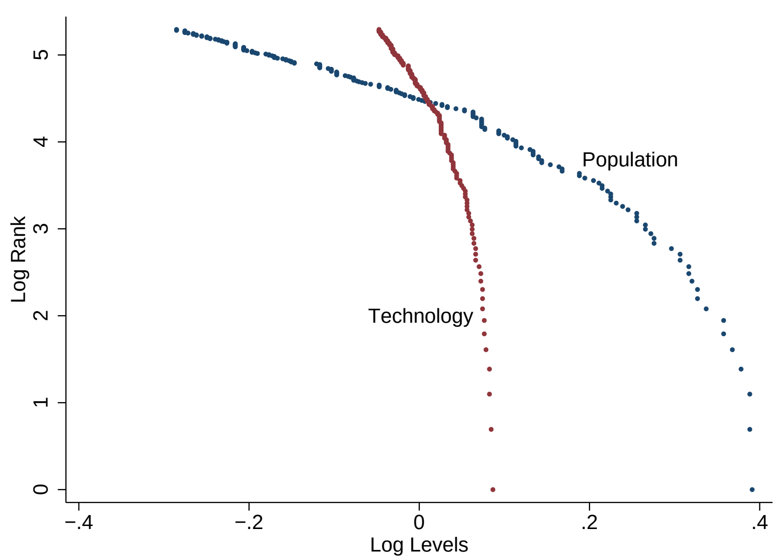
Wright, J. (2003). Detecting lack of identification in GMM. *Econometric Theory 19*(2), 322–330.

Figure 1: OLS and 2SLS with Panel Data



Note: Data are from 22 cities over 29 years. Six TFP outliers are not in the plots to conserve on white space. The regression lines are based on all the data.

Figure 2: Equilibrium Distribution of Population and Technology



Source: Simulations from model.

Table 1: Estimates of λ

Methodology	Estimate	Std. Error
GMM with estimated TFP & Land Price	1.069	0.030
OLS with TFP & Land Price as data		
Panel Data	1.054	0.016
Mean of 22 Time Series	1.074	0.019
Mean of 29 Cross-Sections	1.088	0.032
2SLS with TFP & Land Price as data		
Panel Data	1.099	0.029
Mean of 22 Time Series	1.084	0.036
Mean of 29 Cross-Sections	1.065	0.045

Note: GMM estimates and standard errors are from estimating equations (23) and (34) with $R = 1.05$. Standard errors for GMM include the sampling uncertainty arising from estimating the parameters in the TFP and land price measurement equations (26). OLS and 2SLS estimates are from estimating (22) taking TFP and land prices as data.

Table 2: Baseline Parameter Estimates

Parameter	Description	Estimate	Std. Error
Λ	Effect of agglomeration on growth	0.102	0.050
δ	Net effect of density on productivity	1.041	0.018
ζ	Infrastructure share in finished land	0.545	0.053
ϕ	Non-land income share	0.974	0.002
α	Capital's share of non-land income	0.299	0.002
ω	Structure's share of housing	0.668	0.024
ξ	Skilled-unskilled labor substitutability	0.545	0.035
γ_n	Population growth	1.012	0.000
g_{p_f}	Growth of infrastructure prices	1.006	0.002
g_{p_b}	Growth of business capital prices	0.995	0.001
g_{p_s}	Growth of residential structures price	1.008	0.001
g_{p_l}	Growth of finished land prices	1.010	0.003
g_c	Per capita consumption growth	1.017	0.001
κ_f	Infrastructure depreciation rate	0.021	0.000
κ_b	Business capital depreciation rate	0.107	0.001
κ_s	Structures depreciation rate	0.016	0.000

Note: Estimates and standard errors are from estimating equations (18), (23), and (34). Estimates are based on $R = 1.05$.

Table 3: Specification Tests

Test	Statistic	p-value
$AR(2)$	-1.022	0.307
$AR(3)$	0.264	0.792
J	9.535	0.090
$H_0 : g_{p_l} = g_c^{1-\zeta}\gamma_n^{1-\zeta}g_{p_f}^{\zeta}$	-1.135	0.128
$H_0 : g_{p_f} = g_{p_l} = g_c\gamma_n$	13.128	0.001

Note: The first two rows correspond to Arellano and Bond (1991) tests of the null hypotheses of no second-order and no third-order residual serial correlation in equation (23). The third rows corresponds to the Hansen (1982) and Sargan (1958) J-test of the over-identifying restrictions in (23). The last rows correspond to Wald tests of the indicated null hypotheses. The tests in rows 1, 2, and 4 are t-tests. The J-test in row 3 has five degrees of freedom and the Wald test in row 5 has two degrees of freedom.

Table 4: Effects of Model Assumptions on Baseline Parameter Estimates

Model	Λ	δ	ζ	ϕ	α
Baseline	0.102	1.041	0.545	0.974	0.299
	(0.050)	(0.018)	(0.053)	(0.002)	(0.002)
$\eta = 1$	0.128	1.057	0.545	0.974	0.299
	(0.069)	(0.020)	(0.053)	(0.002)	(0.002)
$g_{p_f} = g_{p_l} = g_c\gamma_n$	0.133	1.032	0.571	0.983	0.306
	(0.052)	(0.010)	(0.016)	(0.001)	(0.002)

Notes: Standard errors are in parentheses. In both perturbations to the baseline, estimation is based on the same set of instruments as in the baseline.

Table 5: Estimates of λ with Simulated Data

	$\lambda = 1.069$		$\lambda = 1$	
	Estimate	Std. Error	Estimate	Std. Error
OLS				
Large Sample	1.366	0.000	1.368	0.000
Small Sample	1.370	0.004	1.373	0.004
GMM				
Large Sample	1.0728	0.0030	1.0052	0.0041
Small Sample	1.0780	0.0382	1.0217	0.0512

Notes: The large samples include 222 cities for 499 years and the small sample is the same size as our empirical sample, 22 cities for 29 years.