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

November 10, 2011

Abstract

Cities exist because of the productivity gains arising 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 data and city-level panel data to estimate our model's parameters by the Generalized Method of Moments. The estimates imply that local agglomeration has an economically and statistically significant impact on the growth rate of per capita consumption, raising it by about 10 percent. Lending credibility to this finding, we demonstrate that our model reproduces important features of the cross-section of cities.

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, Sevi Rodriguez-Mora, José Víctor Ríos Rull, Marcelo Veracierto, Mark Wright, numerous seminar participants, referees and editor for helpful comments. The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

1 Introduction

Cities emerge because of productivity gains that accompany the clustering of production and workers. Also known as agglomeration effects, these gains arise from superior 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. We are the first to answer this question. To do so, we build and estimate 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. Production and housing are location specific; local infrastructure is a produced durable input into finished land that is used in production and housing; and there are local agglomeration effects that offset congestion from crowding more workers onto the same finished land, as proposed by Ciccone and Hall (1996). We model an agglomeration externality by allowing total factor productivity (TFP) at a location to increase with the location's output density—output per acre of finished land in production. However, agents do not take this externality into account when making their decisions. We study the model's competitive equilibrium and show that along the balanced growth path, per capita consumption growth is related to the endogenously determined rate of increase in finished land prices, as well as to a parameter governing the effect of agglomeration on productivity net of congestion.

If land prices have no trend, then agglomeration has no impact on growth. 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, the curtailment of land use due to rising land prices causes output density and therefore TFP to grow faster than they would otherwise. We estimate that land prices indeed have a positive secular trend, growing 1.0 percent a year between 1978 and 2009. Combined with our estimate of the net 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 percent.

Our strategy for identifying the size of agglomeration effects builds on Lucas (2001), who identifies these effects with the use of variation in land rents within a city. Similarly, we use variation in housing rents across cities. Essentially our estimator finds the degree of agglomeration that generates a time-series and cross-sectional pattern of housing rents, output prices and labor quality that most closely matches the pattern of high-skill workers' wages. To estimate the size of agglomeration effects, we use annual panel data for 22 cities from 1978 to 2009. This estimate, along with estimates of several other model parameters, are inputs into our estimate of the increase in consumption growth due to agglomeration. We use aggregate time series data to estimate these other parameters. Our Generalized Method of Moments (GMM) estimation of the impact of local agglomeration on aggregate growth accounts for the sampling uncertainty in both the micro and macro data.

Our model is most closely related to the one in Rossi-Hansberg and Wright (2007). Their focus is on integrating a traditional microeconomic model of cities into an aggregate growth framework. A key difference between our two models is the role played by land. In their model land is not a factor of production. However, workers inelastically demand one unit of land for housing, and land is differentiated by its distance from the center of the city, with more distant land entailing higher commuting costs. A city expands by freely adding land at the perimeter. In contrast, in our model all land in a city is identical, and it is an elastic input to both production and housing. A city expands by adding costly infrastructure to raw land to create additional finished land. Infrastructure accumulation leads to aggregate growth by adding finished land to existing cities. We view our model as capturing the idea that cities expand by adding jobs and housing on the periphery.

Our focus on agglomeration also differs from that of Rossi-Hansberg and Wright (2007), who instead study a human capital externality. While we do not consider human capital externalities explicitly, we do account for variation in human capital in our empirical work. Following Ciccone and Peri (2006), in our estimation we model the labor input as a combination of low- and high-skilled workers and use this representation of labor to control for the composition of the workforce. Doing so is important because of the clear connection between agglomeration and human capital. For example, it may be that more dense urban areas attract higher human capital workers. Not considering such behavior would lead to biases in our estimates.

We organize the paper 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 calibrating and studying our model. Section 7 concludes.

2 The Model

Our model extends the neo-classical growth model to include location-specific production and housing; local infrastructure as a produced, durable input into finished land that is used in production and housing; and local agglomeration effects as proposed by Ciccone and Hall (1996). We begin by describing the Ciccone-Hall model of agglomeration. We then outline the decentralized competitive environment of the growth model, describe the planning problem that yields the equilibrium outcomes, and characterize the model's balanced growth path.

2.1 Modeling Agglomeration

We work with the model of agglomeration in Ciccone and Hall (1996). In their review of the extensive theoretical literature on agglomeration, they point out that these models stem from the insight that when local markets are more active, it is profitable to produce a larger number of differentiated intermediate inputs. This increased variety of inputs leads to greater productivity in the production of final goods. Ciccone and Hall (1996)'s contribution is to formulate a simple reduced form model of agglomeration that is suitable for empirical analysis and that captures the essential idea that local productivity can be increasing in the density of economic activity. Since we work with their model, we first review its microeconomic foundations.

In the Ciccone-Hall model production per acre of land at a location is a constant-returns function of labor and a nontransportable composite service input. This second input is a constant-elasticity-of-substitution function of an endogenous number of service varieties. Though not essential, Ciccone and Hall (1996) assume monopolistic competition by the service providers. Fixed costs and free entry yield a zero-profit condition that determines the number of varieties. Denser acres of land have greater variety because more intermediate service producers can break even. This in turn yields a reduced form for the production of composite services in which labor productivity is increasing in the number of varieties. With density leading to variety, and variety leading to productivity, the model ends up with a reduced form relationship between density and productivity.

We make two key changes to this model. In reality, firms compete with housing for land, thus complicating the measurement of density in the model and the data. Ciccone and Hall (1996), who do not explicitly consider housing, assume that all land at a location is used to evaluate density. This assumption is justified if producers utilize a fixed fraction of land in all locations. Instead, we assume that the model's measure of density includes only land used by firms. Since firms compete with housing for land in the model, housing rents can be used to measure the cost of land. This mapping plays an important role in our estimation.

The second change involves the fact that cities grow, and therefore productive land also grows. Ciccone and Hall (1996) deal with this issue by identifying locations with fixed-size counties. We address it by assuming that production can occur on land only after having infrastructure put in place. Specifically, finished land is produced using a constant-return-to-scale function of raw land and infrastructure. Cities grow via the accumulation of infrastructure, and diminishing returns to infrastructure captures the idea that the best quality land is developed first. We justify these assumptions by the obvious importance of roads, sewers, electricity and other factors that are tied to a location and necessary for housing and production.[1]

[1] A related model is in Weill and Van Nieuwerburgh (2010), who study house price dispersion with a multi-city model featuring land accumulation. In their model preferences are linear, there is no aggregate growth, land is an input into housing, but not production, and agglomeration is absent.

2.2 Economic Environment

The Ciccone-Hall model of agglomeration is partial equilibrium. To embed it in a dynamic stochastic general equilibrium model, we must take a stand on market structure and on 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 relocation fee. Here we consider a natural benchmark: the case of complete markets in state-contingent consumption claims, perfect competition, and costless labor mobility.

There is an infinitely lived representative household with a unit measure of homogeneous members who each period supply a unit of labor inelastically. The household maximizes the expected present value of its members' utility. Preferences depend on locally provided housing services, h, and a consumption good that is freely traded across a fixed measure of locations called cities, C. Each period the household freely allocates across cities its workers, n, and the existing stocks of business capital and residential structures capital, K_b and K_s. It also chooses for the next period infrastructure in each city, k_f, and the total quantities of business capital and residential structures. We assume these decisions are made after the household observes each city's output density (output per unit of finished land) and exogenous productivity shock, z. The household takes all prices and the distribution of productivity and density as given. That is, it behaves competitively and does not take into account the effect of its actions on the density of production in each location.

Each city also 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 then rent to local goods producers and housing service providers. Housing service providers rent residential structures, k_s, and finished land, l_h, to produce housing services, which they then sell to workers. Goods producers rent business capital, k_b, and labor from the household, and finished land, l_b, to produce the city-specific intermediate good, y, with a technology that depends on the exogenous city-specific productivity shocks and the city's output density, $\bar{y}/\bar{l}_b$. There are also final good producers who combine city-specific intermediate goods to produce a composite good that can be transformed 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 an optimization problem with side conditions. While we focus on the competitive equilibrium in our empirical work, we state the planning problem here to describe the details of the model's specification succinctly.

Let $q_t(z^t)$ denote the time t distribution of cities across productivity histories z^t. Idiosyncratic technology evolves as a stationary discrete Markov chain. Taking into account the fact that in the competitive equilbrium households perfectly insure their members against consumption risk, the planner's problem is:

$$\max_{\substack{\{C_t, K_{bt+1}, K_{st+1}, y(z^t), l_b(z^t), l_h(z^t), n(z^t), \\ k_b(z^t), k_s(z^t), k_{ft+1}(z^t), h(z^t)\}_{t=0}^{\infty}}} \left[\sum_{t=0}^{\infty} \beta^t \ln C_t + \psi \sum_{t=0}^{\infty} \beta^t \sum_{z^t} q_t(z^t) n(z^t) \ln \frac{h(z^t)}{n(z^t)} \right]$$

$$\begin{aligned}
\text{subject to} \qquad & C_t + P_{bt}\left[K_{bt+1} - (1-\kappa_b)K_{bt}\right] \\
& \quad + P_{st}\left[K_{st+1} - (1-\kappa_s)K_{st}\right] \\
& \quad + P_{ft}\sum_{z^t} q_t(z^t)\left[k_{ft+1}(z^t) - (1-\kappa_f)k_{ft}(z^{t-1})\right] \\
& \quad \leq \left[\sum_{z^t} q_t(z^t) y(z^t)^{\eta}\right]^{\frac{1}{\eta}} && (1) \\
y(z^t) &\leq [A_t z_t]^{(1-\alpha)\phi} \left[\frac{\bar{y}(z^t)}{\bar{l}_b(z^t)}\right]^{\frac{\lambda-1}{\lambda}} l_b(z^t)^{1-\phi} k_b(z^t)^{\alpha\phi} n(z^t)^{(1-\alpha)\phi}, \ \forall z^t && (2) \\
h(z^t) &\leq l_h(z^t)^{1-\omega} k_s(z^t)^{\omega}, \ \forall z^t && (3) \\
l_h(z^t) + l_b(z^t) &\leq k_{ft}(z^{t-1})^{\zeta}, \ \forall z^t && (4) \\
\sum_{z^t} q_t(z^t) k_b(z^t) &\leq K_{bt} && (5) \\
\sum_{z^t} q_t(z^t) k_s(z^t) &\leq K_{st} && (6) \\
\sum_{z^t} q_t(z^t) n(z^t) &\leq 1 && (7)
\end{aligned}$$

and $K_{b0}, K_{s0}, k_f(z_0), \{A_t, P_{bt}, P_{st}, P_{ft}, z^t\}_{t=0}^{\infty}, \bar{y}(z^t)$ and $\bar{l}_b(z^t)$ given. The variable A_t denotes the aggregate level of Hicks-neutral technology, and the variables P_{bt}, P_{st} and P_{ft} denote the rates at which consumption goods can be transformed into the different kinds of investment. These technology variables are a source of growth in the model, but for now we set them to one in every period. The parameters α, ϕ, ω, and ζ all lie in the interval $[0, 1]$, $\eta \leq 1$, $\psi > 0$ and $\lambda \geq 1$. Equation (1) is the aggregate resource constraint for final goods. Equation (2) defines the production function for intermediate goods, which is identical to the one in Ciccone and Hall (1996) except that the land input is endogenous and not necessarily equal to the endowment of raw land in the city. The parameter λ governs the impact of output density on local TFP, with $\lambda = 1$ corresponding to no impact. Equations (3) and (4) are the local resource constraints for housing services and finished land. We normalize the quantity of raw land in a city to one. Lastly, equations (5)–(7) are the aggregate resource constraints for business capital, residential structures and employment.

We obtain competitive equilibrium allocations as a solution to this optimization problem such that $y(z^t) = \bar{y}(z^t)$ and $l_b(z^t) = \bar{l}_b(z^t)$. Prices corresponding to an equilibrium are easy to obtain from the constraint's Lagrange multipliers. The details of this mapping along with the first order conditions of the planning problem are described in Davis, Fisher, and Whited (2011) (DFW). There we also show how the first order conditions for the planning problem correspond to the first order conditions of the model's agents 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. Therefore the competitive equilibrium exists and is unique and the allocations are Pareto-efficient. Based on an argument in Kehoe, Levine, and Romer (1992), we suspect that for $\lambda > 1$ and sufficiently close to one, there also exists a unique equilibrium, because the model without density effects has a unique equilibrium. Nevertheless, equilibrium existence and uniqueness with $\lambda > 1$ remain open questions. However, in Section 6 we solve the model with several values of λ, including our estimates. In each case we find an equilibrium that appears to be unique. Since the planner does not take into account the density externality when solving its problem, competitive equilibria in the $\lambda > 1$ case are inefficient.

Several features of our model are worth emphasizing. First, for most parameter values and technology processes, the model exhibits non-trivial cross-sectional variation in housing rents, wages and output prices. This feature is important because we exploit cross-sectional variation in these variables in our empirical work. Because infrastructure capital is predetermined in every city, as indicated by (1), idiosyncratic technology shocks induce variation in housing rents. Variation in wages occurs because preferences imply that housing services' consumption value is constant. Consequently, in a competitive equilibrium workers in high-wage cities enjoy relatively low housing services and pay high housing rents. Workers stay in low-wage cities because they enjoy relatively high housing services. The price of city-specific output varies because of the specification of the final good aggregator in (1), as long as $\eta < 1$.

Second, as (5) indicates, business capital can be reallocated across cities each period without any cost, and city-specific investment is not subject to a non-negativity constraint. These assumptions are important for our empirical work because they allow us to eliminate hard-to-measure capital from our estimation. Their plausibility can be gauged by evaluating the incidence of disinvestment in calibrated versions of the model. In the calibrated model considered in Section 6 the incidence of disinvestment is negligible.

Lastly, labor mobility is costless. This does not pose a problem for our empirical work. Our empirical strategy is built from the first order conditions of the intermediate goods producers, landlords and housing service providers along with the household's first order conditions for capital accumulation. For many plausible specifications, these first order conditions are invariant to the nature of the mobility technology.

2.4 Balanced Growth

To study growth we assume

$$\begin{aligned} P_{xt} &= \gamma_x^{-t},\ x = s, b, f; \\ A_t &= \gamma_a^t; \\ N_t &= \gamma_n^t. \end{aligned}$$

The γ_x's describe how fast the technology used to produce x-type investment goods rises relative to the technology used to produce consumption. So, for example, if $\gamma_s < 1$ the structures investment technology grows at a slower rate than the consumption technology. In a competitive equilibrium the P_x's equal the price in consumption units of x-type investment goods. Therefore with $\gamma_x \neq 1$ investment prices exhibit trends. We consider trends in investment prices since there are such trends empirically and, similar to Greenwood et al. (1997), these trends influence our calculations. We also assume trends in Hicks-neutral technology, A_t, and in population, N_t, given by γ_a and γ_n. 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 Rossi-Hansberg and Wright (2007).[2]

We now derive the balanced growth rate of per capita consumption as a function of the underlying exogenous growth rates of technology and population. Along a balanced growth path the investment prices grow at the inverse rates of relative investment technical change:

$$g_{p_x} = \gamma_x^{-1}, \;\; x = s, b, f, \tag{8}$$

where g_{p_x} denotes the balanced growth rate of the x-type investment price. The aggregate resource constraint (1) implies that per capita consumption growth equals the per capita growth in consumption units of each of the three aggregate capital stocks. Equivalently,

$$g_c = g_{p_x} g_{k_x}, \;\; x = s, b, f, \tag{9}$$

where g_c denotes the growth rate of consumption and g_{k_x} denotes the growth rate of x-type capital. The city-specific land constraints (4) imply that per capita finished land grows as

$$g_l = \gamma_n^{\zeta-1} g_{k_f}^{\zeta}. \tag{10}$$

By solving the city resource constraints (2) for city-specific output, we show that

$$g_c = \gamma_a^{(1-\alpha)\delta} {g_l}^{1-\delta} {g_{k_b}}^{\alpha\delta}. \tag{11}$$

[2]See DFW for how to transform the model with growth into the stationary planning problem described in the previous section.

The parameter $\delta \equiv \phi\lambda$ measures the net effect of agglomeration on productivity. If the agglomeration externality, λ, is large enough to offset the diminishing returns due to congestion, ϕ, then $\delta > 1$. Replacing g_l and g_{k_b} in (11), using (8)–(10), and solving for g_c yields

$$g_c = \gamma_a^{\frac{(1-\alpha)\delta}{1-\delta\alpha+(\delta-1)\zeta}} \gamma_b^{\frac{\alpha\delta}{1-\delta\alpha+(\delta-1)\zeta}} \gamma_n^{\frac{(1-\zeta)(\delta-1)}{1-\delta\alpha+(\delta-1)\zeta}} \gamma_f^{\zeta\frac{1-\delta}{1-\delta\alpha+(\delta-1)\zeta}}. \tag{12}$$

Equation (12) relates per capita consumption growth to exogenous technology and population growth. With $0 < \zeta < 1$ per capita consumption growth is increasing in neutral and business investment technologies. If $\delta > 1$, so that agglomeration effects exceed the effects of congestion, then per capita consumption growth is increasing in population growth and decreasing in the growth rate of the infrastructure technology. The positive effect of population growth is due to its positive impact on density as long as $\zeta < 1$. Empirically, technical change in infrastructure production is slower than for consumption goods, $\gamma_f < 1$. In this case, if $\delta > 1$, then relatively slow infrastructure technical change limits the expansion of finished land, thereby raising density and consumption growth.

A more intuitive expression for g_c comes from the first order condition for finished land in intermediate good production. Because of Cobb-Douglas technology, this equation states that expenditures on finished land by intermediate good producers in a city equal a constant fraction of that city's output in consumption units. Therefore, per capita consumption growth equals the product of the growth rates of the per capita stock of finished land and the growth in the average price of land:

$$g_c = g_l g_{p_l}. \tag{13}$$

Replacing g_l in (11) using (13) yields

$$g_c = \gamma_a \gamma_b^{\frac{\alpha}{1-\alpha}} g_{p_l}^{\frac{\delta-1}{\delta(1-\alpha)}}. \tag{14}$$

Equation (14) relates per capita consumption growth to neutral and business investment technological change and to growth in the price of finished land. The dependence on neutral and investment technologies is familiar from models of investment-specific technical change. The additional term shows that the impact of agglomeration on growth depends on the presence of a trend in the price of finished land. If land prices are growing, $g_{p_l} > 1$, then land growth does not keep up with output growth, and the density of economic activity grows. If agglomeration effects outweigh congestion effects, $\delta > 1$, then this mechanism provides a source of per capita consumption growth. Without the effect of density on productivity, $\lambda = 1$ and $\delta = \phi$, equation (14) implies

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

Assuming $g_{p_l} > 1$ and $0 < \phi < 1$, per capita consumption growth is lower than predicted by technical change alone, because of decreasing returns to land in production. If agglomeration and congestion effects cancel, $\delta = 1$, then per capita consumption growth is the same as in the neoclassical growth model with neutral and investment-specific technical change.

3 Empirical Strategy

We now propose how to measure the effect of local agglomeration on aggregate consumption growth using the expression relating consumption growth to underlying technology and population growth, (12). Since growth in investment prices identifies investment technologies, (8), we can estimate γ_a from (12) as

$$\tilde{\gamma}_a = \tilde{g}_c^{\frac{(1-\alpha)\delta+(\zeta-1)(\delta-1)}{(1-\alpha)\delta}} \tilde{\gamma}_n^{\frac{(1-\zeta)(1-\delta)}{(1-\alpha)\delta}} \tilde{g}_{p_b}^{\frac{\alpha}{(1-\alpha)}} \tilde{g}_{p_f}^{\zeta\frac{1-\delta}{(1-\alpha)\delta}} \tag{15}$$

where $\tilde{x}$ denotes the point estimate of x. Next, construct g_c^*, the counterfactual growth rate of consumption without agglomeration, $\delta = \phi$, from (12) and substitute for $\tilde{\gamma}_a$ using (15):

$$g_c^* = \tilde{g}_c^{\frac{\phi[1-\alpha\delta-\zeta(1-\delta)]}{\delta[1-\alpha\phi-\zeta(1-\phi)]}} \tilde{\gamma}_n^{\frac{(\phi-\delta)(1-\zeta)}{\delta[1-\alpha\phi-\zeta(1-\phi)]}} \tilde{g}_{p_f}^{\frac{\zeta(\phi-\delta)}{\delta[1-\alpha\phi-\zeta(1-\phi)]}}. \tag{16}$$

We use

$$\Lambda = \frac{\tilde{g}_c - g_c^*}{g_c^* - 1} \tag{17}$$

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

Estimating Λ is the objective of our empirical work. To obtain Λ we need estimates of the growth rates g_c, g_{p_f} and γ_n and the parameters δ, α, ϕ and ζ. Our identification 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 use our model's first order conditions for intermediate goods firms' and housing service providers' choices of factor inputs combined with a panel of wages, housing rents and output prices for 22 US cities to identify δ. For the remaining parameters and growth rates, we use moment conditions involving aggregate data derived from agents' first order conditions and constraints evaluated along the balanced growth path. Our GMM estimation integrates the micro and macro data so that estimates of standard errors account for sampling uncertainty in both. The discussion below is relatively brief. See DFW for more detailed derivations.

3.1 Identification with Panel of US Cities

Consider any two cities from our model, indexed i and j, in some arbitrary period. Solving for output in (2) and dividing output in city i by that in city j yields

$$\frac{y_i}{y_j} = \left[\frac{z_i}{z_j}\right]^{\delta(1-\alpha)} \left[\frac{l_{bi}}{l_{bj}}\right]^{1-\delta} \left[\frac{k_{bi}}{k_{bj}}\right]^{\delta\alpha} \left[\frac{n_i}{n_j}\right]^{\delta(1-\alpha)} .$$

Next, we use the intermediate good producers' first order condition for renting capital to eliminate the capital stock term in this last equation. We obtain

$$\frac{y_i/l_{bi}}{y_j/l_{bj}} = \left[\frac{z_i}{z_j}\right]^{\frac{\delta(1-\alpha)}{1-\delta\alpha}} \left[\frac{n_i/l_{bi}}{n_j/l_{bj}}\right]^{\frac{\delta(1-\alpha)}{1-\delta\alpha}} \left[\frac{p_{yi}}{p_{yj}}\right]^{\frac{\delta\alpha}{1-\delta\alpha}} , \tag{18}$$

where p_{yi} denotes the price of goods from city i. Equation (18) is closely related to equation (19) in Ciccone and Hall (1996), which is the basis for their estimation equation. It differs in two key respects. First, because Ciccone and Hall (1996) assume intermediate goods are perfect substitutes in producing final goods, $\eta = 1$, their equation does not contain any output prices. So if $\eta < 1$ their estimation is subject to an omitted variable bias. Second, they assume all land in a given location is used in production, while we have competing uses for land. Since the allocation of land between residential and non-residential uses in U.S. cities is unavailable, we cannot use (18) as a basis for estimation.

Instead, we use the intermediate good producers' first order conditions for finished land and labor to replace the quantity of land in (18) with its price. We also use the housing service providers' first order condition for finished land to link the price of land to housing rents. After making these substitutions, taking logs, summing over all cities j for each i, and re-introducing time subscripts we have

$$\hat{w}_{it} = \frac{1}{1-\omega}\frac{\delta-1}{\delta(1-\alpha)}\hat{r}_{hit} + \frac{1}{\delta(1-\alpha)}\hat{p}_{yit} + \hat{z}_{it}, \tag{19}$$

where w denotes wages and r_h denotes the rental price of housing services. Equation (19) uses the definition

$$\hat{x}_i \equiv \ln(x_i) - \frac{1}{N}\sum_{j=1}^{N} \ln(x_j)$$

for any variable x. By construction $\hat{x}_i$ is mean zero in every year. This transformation eliminates aggregate fluctuations by pulling out a "time effect."

So far we have assumed that workers are homogeneous, which is questionable empirically. Not accounting for cross-sectional variation in labor quality could lead to biased estimates of density effects. To address this issue, we follow Ciccone and Peri (2006) who consider "skilled" and "unskilled" workers as imperfect substitutes in producing total labor services. We now derive a version of (19) that incorporates heterogeneous workers in a way that does not affect the balanced growth path of our model.

Suppose that the effective labor input in (2), n, is a constant elasticity of substitution composite of unskilled, n_u, and skilled labor, n_e:

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

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_e denote the wages 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

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

where w denotes the implicit wage for the composite labor input, n, $s \equiv (w_u n_u + w_e n_e)/(w_u n_u)$, and $m \equiv n_e/n_u$. Substituting for composite wages, w, using (19), it is straightforward to derive

$$\hat{w}_{eit} = \frac{1}{1-\omega}\frac{\delta-1}{\delta(1-\alpha)}\hat{r}_{hit} + \frac{1}{\delta(1-\alpha)}\hat{p}_{yit} + \frac{1-\xi}{\xi}\hat{s}_{it} + (\xi-1)\,\hat{m}_{it-1} + \hat{z}_{it}. \tag{20}$$

Equation (20) reduces to equation (19) if $\xi = 1$, that is, if unskilled and skilled labor are perfect substitutes.

To build an estimation strategy around (20) we must address the fact that the model predicts that the idiosyncratic technology term $\hat{z}_{it}$ is correlated with the other right-hand-side variables. We address this endogeneity by exploiting the implications of a particular stochastic process for $\hat{z}_{it}$. Our specification of this process is motivated by the model's structure and evidence on the cross-section of cities.

Two key empirical observations about cities are "Gibrats law" and "Zipf's law," *c.f.* Gabaix (1999). Any model of the cross-section of cities should be consistent with the phenomena described by these laws. Gibrat's law for cities states that their populations follow similar growth processes. In particular they share a common mean equal to the mean city population growth rate and a common 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 S is given by $\Pr(\text{Size} > S) = a/S^b$ with $b \simeq 1$. More precisely, $b = 1/(1 - S_{\min}/\bar{S})$, where $S_{\min}$ 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.

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

$$\ln z_{it+1} = \max\left\{\ln z_{it} + \varepsilon_{it+1}, \ln z_{\min,t}\right\}; \tag{21}$$

$$\ln z_{\min,t} = \ln z_{\min} + \ln \bar{z}_t; \tag{22}$$

$$\bar{z}_t = E_t z_{it}; \tag{23}$$

$$\varepsilon_{it+1} \sim N(0, \sigma_\varepsilon^2). \tag{24}$$

With this process city-specific technology is characterized by 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 all these cities over the sample period.

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

$$\Delta \hat{w}_{eit} = \frac{1}{1-\omega}\frac{\delta-1}{\delta(1-\alpha)}\Delta \hat{r}_{hit} + \frac{1}{\delta(1-\alpha)}\Delta \hat{p}_{it} + \frac{1-\xi}{\xi}\Delta \hat{s}_{it} + (\xi-1)\Delta \hat{m}_{it} + \varepsilon_{it}. \quad (25)$$

The innovation term ε_{it} is correlated with the other right-hand side variables, so to use (25) to estimate the structural parameters, we need to find instrumental variables. This task is simple relative to using (20) because our model predicts that *any* variable dated $t-1$ and earlier is orthogonal to ε_{it} and hence is a valid instrument. Consequently, we identify δ and ξ using the moment conditions

$$E\left\{\left[\begin{array}{c} \Delta \hat{w}_{eit} - \frac{1}{1-\omega}\frac{\delta-1}{\delta(1-\alpha)}\Delta \hat{r}_{hit} \\ -\frac{1}{\delta(1-\alpha)}\Delta \hat{p}_{yit} - \frac{1-\xi}{\xi}\Delta \hat{s}_{it} - (\xi-1)\Delta \hat{m}_{it} \end{array}\right] \times \hat{v}_{it-j}\right\} = 0 \quad (26)$$

for instruments $\hat{v}_{it-j}$, $j \geq 1$. We discuss our choice of instruments below.

Essentially, our estimator finds the δ that generates a time-series and cross-sectional pattern of housing rents, output prices and labor quality that most closely matches the pattern of high-skill workers' wages.[3] For instance, if high-skill workers' wage growth is positively correlated with growth in housing rents, holding fixed the growth rates of technology, output prices and labor quality, then $\delta > 1$. This situation arises in our model because firms economize on land in high rent cities, which reduces wages because of congestion, but increases wages through the density externality. With $\delta > 1$, the density effects outweigh the congestion effects, and wages increase.

The theory says equation (25) holds exactly in the data, in the sense that the city-specific productivity shock (unobserved to the econometrician) is the error term.[4] While city-specific productivity shocks are the only source of cross-sectional variation in our model, our panel estimation is nonetheless robust to including additional shocks in the model. These extra sources of variation would not change (25) because this equation is derived from agents' first order conditions, which continue to hold with

[3] Of course, because we are using instrumental variables, our identifying information is from the forecastable component of these variables.

[4] In our estimation we allow for the possibility that the error term includes measurement error.

additional shocks. For example, suppose there are shocks to the demand for city-specific goods. Such shocks would affect wages, rents and output prices, while leaving exogenous productivity unaffected. Equation (25) would hold exactly, but the demand shocks would contribute to the variation of wages, rents and prices. Therefore, we do not expect the right-hand-side variables in equation (25) to account for all the variation in wages (in an OLS sense), but they remain useful for identifying δ.[5]

3.2 Identification with Aggregate US Data

Identifying δ using (26) requires that we also estimate ω and α, and to complete the measurement of Λ we need to estimate g_c, g_{p_f}, γ_n, ϕ and ζ. We identify the growth rates using the moment conditions

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

where E denotes the unconditional expectations operator. Equation (27) assumes that $\ln X_t$ is de-meaned prior to the analysis. (Our standard errors address the sampling uncertainty in these means.)

To identify ω, the share of structures in housing services, we use the housing service providers' first order conditions, which equate two ratios: 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 in a specific city 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 $E_{t|i}$ denotes expectation at t conditional on i, r_{li}, is the rental price of finished land in city i, p_{li} is the capital price of finished land in city i, and $R \equiv 1/\beta$ is the

[5] As additional exogenous shocks are added to the model, the R^2 of an OLS regression of high-skill wages on the right-hand-side variables in (25) is driven toward zero. With the data described in Section 4 the R^2 of this regression is 0.37.

interest rate. Equation (28) states that the value of an additional unit of finished land at date t in city i equals the expected discounted rent from that land plus next period's value 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 time t conditional expectations operator. Since $E_t E_{t|i} x_t = E_t x_t$ for random variable x_t, 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

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

where the summations are over all cities. We estimate the depreciation rates in this expression as their sample average values:

$$E\left\{\kappa_x - \frac{P_{xt} D_{xt}}{P_{xt} K_{xt}}\right\} = 0,\ x = s, f, \tag{32}$$

where D_{xt} is real depreciation of x-type capital. The growth rate g_{p_s} is identified using the analogue of (27). To estimate g_{p_l} we use another implication of the housing service providers' first order conditions:

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

Equation (33) says 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 follows from the household's first order conditions for capital accumulation.

It remains to identify α, ϕ and ζ. We identify the first two parameters using labor's share of total income and finished land's share of non-labor income:

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

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

The second condition 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). The parameter ζ is the share of infrastructure in finished land production. By exploiting value to income mappings analogous to (30), we derive the moment condition used to identify ζ:

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

Now we have described a sufficient number of moment conditions to obtain Λ. In order to identify Λ's six proximate inputs, g_c, g_{p_f}, δ, α, ϕ and ζ, we have to include moment conditions to identify all the model's depreciation rates and investment price growth rates, plus ω. Only two parameters in our model are not used: ψ and 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.[6] To begin, 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:

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

Because this system of moment conditions is exactly identified, the dimensions of Ψ and θ are equal. The first step is to estimate equation (34) by GMM, in which we use a Newey-West weight matrix with a lag length of 2. In the second step we estimate δ and ξ using the moment conditions in (26). This estimation requires that we plug in the estimates of ω and α from the first step into (26). To account for the sampling variation associated with these two 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. The third step is to substitute the point

[6] See DFW for additional details

estimates for g_c, g_{p_f}, δ, α, ϕ, and ζ into (17) 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 parameters' influence functions as shown by Erickson and Whited (2002). After this calculation, a standard application of the delta method gives the variance of Λ.

We now describe our choice of instruments. Recall that our model implies that ε_{it} in (25) is correlated with the other right-hand side variables, so our estimation must account for this endogeneity. Our model also facilitates this task by implying that any variable dated $t-1$ or earlier is orthogonal to ε_{it} and hence is a valid instrument. We want to account for the possibility of classical *i.i.d.* measurement error so we consider variables dated $t-2$ and earlier.[7]

In traditional dynamic panel estimation, with fixed sample length T and large N, two kinds of instruments have been proposed. For any instrument x, the moment condition

$$Ex_{it-2}\varepsilon_{it} = 0$$

holds at each date t and for each city i. In one approach, *e.g.* Holtz-Eakin, Newey, and Rosen (1988), 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 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 advantage of the first approach is that it exploits more information and so is more efficient. It is usually applied in cases with large N and small T because the cross-sectional average can be expected to be a good measure of the cross-sectional expectation. In this case levels are usually used as instruments for growth rates to

[7] In general, if the first date of the instruments is $t-j$, then our estimation is robust to errors that follow a moving average process of order no greater than $j-1$.

conserve observations. On the other hand, with relatively large T and small N, as is in our case, the second approach may be warranted because with small N the cross-sectional average may be a poor estimate of the cross-sectional expected value.

In our small panel the choice between the two approaches depends less on asymptotic properties and more on the strength of the instrument sets for each approach. Therefore, to choose between the two approaches, we use the test in Wright (2003) for underidentification 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 underidentification rather than of weak identification. Because there is no exact finite sampling theory for nonlinear GMM estimators, the weak-instrument tests surveyed in Stock, Wright, and Yogo (2002) cannot be extended to our case. Although we cannot therefore literally 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 all of the measured variables in (25) plus house prices and 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 (35) and on (36). For the moment conditions based on (35), we never reject the null of an underidentified model. For the moment conditions based on (36), we strongly reject the null of an underidentified model (p-value of 0.006) for the instruments in differences, but we cannot reject the null of an underidentified 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 conducted a Monte Carlo experiment in which we estimate our model on simulated data that is designed to approximate our actual data closely in terms of means, variances, covariances, and autocorrelations. 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. Details are in DFW.

4 Data

In this section we briefly describe our data, which is annual and spans the period 1978-2009. We begin with the panel data for the moment conditions (26) and then discuss the aggregate data for the moment conditions (34). For a detailed description of our data see DFW.

4.1 Panel Data

A location in our model is a developed area where households live and work. To represent these locations, we use Metropolitan Statistical Areas (MSAs), which are fixed and contiguous combinations of counties. They contain land that is initially minimally developed, for example farm land, and that eventually becomes more intensively developed for urban purposes. Therefore, the finished land accumulation process in the model is present in our data. The geography of an MSA is a natural empirical counterpart to our locations because the criteria for defining an MSA are explicitly economic, based on commuting-to-work patterns. Furthermore, the data we need are available for MSAs.[8]

We construct MSA-level panel data on output prices, the labor market, and housing rents from several sources. We create MSA-specific price indexes for output, p_{yit}, from data from the Bureau of Economic Analysis (BEA). These indexes are weighted averages of industry-specific price indexes, with weights equal to the share of total earnings paid to employees by 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.

We construct MSA-level labor market variables with data from the March Current Population Survey. High-skill workers are those with at least four years of college, and low-skill workers are those with less than four years of college. The average hourly wage of high-skilled workers, w_{eit}, is total wages paid to high-skill workers divided

[8] Ciccone and Hall (1996) define a location to be a county. Rozenfeld, Rybski, Gabaix, and Makse (2011) propose an alternative geography to MSAs and counties based on satellite imagery that emphasizes geographically dense areas of integrated economic activity. The data we need for our empirical work are not available for their geographical areas nor for counties.

by their total hours worked; the ratio of high-skill to low-skill labor, m_{it}, is the ratio of total hours worked for these two groups; and s_{it}, the inverse share of wages going to low-skill workers, is measured as the ratio of the total wage bill to wages paid to low-skill 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 indexes 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 indexes for shelter to extrapolate rents backwards from 1990 to 1978 and forwards to 2009. Annual average housing rents are measured by averaging across our sample of MSAs.

We use two additional MSA-level variables as instruments when estimating (25), 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 indexes 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 time period. This sample of cities accounts for 37 percent of the US 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 capital stocks, their depreciation rates, and their price indexes are from the Fixed Asset Tables, produced by the BEA. Residential structures, K_{st}, is 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 less infrastructure capital. Infrastructure capital, K_{ft}, is defined to be 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 indexes are consistent with our definitions of the capital stocks. The real investment prices, P_{st}, P_{bt} and P_{ft}, are the investment price indexes divided by our consumption deflator, described below.

Measuring the value shares requires data from several sources. Data for evaluating 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 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.

Real per capita consumption is measured as follows. Nominal consumption is defined as total consumption as reported in 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. Real per-capita aggregate consumption is then nominal consumption divided by its price index, divided again by the population.

5 Empirical Findings

Table 1 reports our baseline parameter estimates along with asymptotic standard errors.[9] We estimate local agglomeration raises aggregate per capita consumption growth by $\Lambda = 10.2$ percent (standard error 5 percent), which we find despite our relatively small estimate of the net impact of density on productivity, $\delta = 1.041$ (0.016). The standard error on Λ masks the fact that our estimates speak clearly on the narrower question of whether the effect of local agglomeration on aggregate growth is statistically significant. It is straightforward to see from (16) that agglomeration has a positive impact on growth if and only if $\delta > \phi$, equivalently, $\lambda > 1$. The point estimate $\phi = 0.974$ (0.002) and the relatively small sampling uncertainty in estimating ϕ and δ is such that we easily reject the hypothesis that $\delta = \phi$. We conclude that the local and aggregate effects of agglomeration are statistically significant.

Are these effects of agglomeration economically significant? The estimated growth rate of per capita consumption of 1.7 percent per year (shown in the table) would be 1.51 percent in the absence of local agglomeration. It is straightforward to establish that per capita consumption and housing would have to be 3 percent larger each period in perpetuity to compensate individuals for the absence of the aggregate growth due to local agglomeration.[10] This calculation is equivalent to 360 billion dollars in 2009 alone. The present value of this compensation is over 7.5 trillion dollars. Seemingly small local effects of agglomeration have very large aggregate consequences.

While we are the first to estimate the impact of local agglomeration on *aggregate growth*, there is a large literature that measures the effect of local agglomeration on the *local level* of firm productivity and wages. These studies attempt to measure the percent increase in productivity or wages due to agglomeration at a point in time, holding fixed all factors of production. Our estimate of 4.1 percent for the net effect of agglomeration on local productivity is in the middle of the range of 2 to 6 percent, as estimated recently by Combes, Duranton, and Gobillon (2008) and surveyed by Rosenthal and Strange (2004).[11]

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

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

[11] As the latter authors describe, researchers focus on what are called "urbanization" effects and "localization" effects. Urbanization describes the impact of local output or employment density on

We now discuss the remaining parameters in Table 1, focusing on the most noteworthy. Infrastructure's share of finished land, ζ, is estimated 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 percent, and raw land's is 1.2 percent. 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 did 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 hypothesis that both expressions are equal to 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 percent a year.

We estimate that the price of business capital declines at about 0.5 percent per year (g_{p_b}), and the prices of infrastructure and residential structures rise at 0.6 percent and 0.8 percent per year (g_{p_f} and g_{p_s}). To understand the role of these trends in capital prices, we re-estimate our model imposing zero trends, $g_{p_b} = g_{p_f} = g_{p_s} = 1$. The data resoundingly reject this restriction, but conditional on it, the point estimate for the share of infrastructure in finished land, ζ, is approximately 1. When $\zeta = 1$ and $g_{p_f} = 1$, our model implies agglomeration has no impact on consumption growth (equation 16). 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).[12] However, as Autor et al. (1998) emphasize, substantial uncertainty exists concerning this elasticity's

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.

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

magnitude. For example, the results of Katz and Murphy (1992) imply that our point estimate lies within a 95 percent confidence interval.

The plausibility of the parameter estimates in Table 1 depends on whether our model is correctly specified. Table 2 displays results from five statistical tests of the validity of our specification. The first two rows correspond to tests for serial correlation of the residuals in (26). Allowing for classical *i.i.d.* measurement error, if the model is correctly specified, the residuals should exhibit autocorrelation only up to order one. 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 (26). The p-value corresponding to 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 did not impose in estimation. Table 2 indicates we do not reject the hypothesis that this condition is satisfied by our estimates. In the last row we test a condition that would hold if finished land were in constant supply (as opposed to finished land per person, which we find is not growing.) Because we observe the quantity of finished land growing, our model would be rejected if this condition held. However, we easily reject it.

Finally, we investigate how two departures from Ciccone and Hall (1996) affect our estimates: substitutability of city-specific goods and growth of finished land. Table 3 displays results corresponding to our baseline specification (reproduced from Table 1) and two variations of it. The specification with perfect substitutability, $\eta = 1$, involves estimating (26) without the output price term. The no-finished-land-growth specification 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 latter case even though we reject the restriction (see Table 2) because it illustrates the role in our estimation played by growth in finished land. The table does not report all the parameter estimates, only Λ and the direct inputs into its calculation that could change.

Table 3 shows that omitting output prices raises the point estimate of Λ by almost 30 percent. This result is entirely due to the larger estimate of δ, which in turn is indicative of a positive correlation between output prices and high skill wages in our sample – the omitted variable bias is positive. Notice that although the point estimate of Λ is larger, it is also less precisely estimated so that in this case we cannot reject the hypothesis that $\Lambda = 0$, although we still reject $\delta = \phi$. Interestingly, the point estimate of $\delta = 1.057$ in this case is very close to estimates by Ciccone and Hall (1996), who assume perfect substitutability. As indicated by Table 3, assuming a fixed supply of finished land, another assumption made by Ciccone and Hall (1996), also raises the point estimate of Λ by about 30 percent, even though it lowers the estimate of δ. Inspection of (16) reveals that the counterfactual rate of per capita consumption growth absent agglomeration is decreasing in g_{p_f} at our point estimates. Consequently, the increase in the point estimate of Λ primarily is driven by g_{p_f} being larger than its baseline estimate in this counterfactual.

6 Additional Validation of the Empirical Findings

Our estimate of Λ depends on the structural model we have specified. The model involves some seemingly strong assumptions, so we need to assess their impact on our findings. Furthermore, if the model has predictions that are consistent with the data beyond those we have already considered, it would provide further validation of the empirical findings. This section considers calibrated solutions to our model to address these issues. We have three main results. First, when there is a density externality, $\lambda > 1$, the model appears to have a unique solution for parameter configurations in the neighborhood of our estimates. Second, our simplifying assumption of reversible business investment may not be quantitatively important. Third, the calibrated model has desirable empirical predictions, and obvious generalizations that are consistent with our estimation strategy have the potential to account for the data even better. Our discussion here is brief; for more details see DFW.

6.1 Calibration

For simplicity and without loss of generality, we assume $\omega = 0$ and $\xi = 1$ so that housing consists only of land and labor is homogeneous. We set $\psi = .2$, which is the average consumption share of shelter in our sample, and we set $\eta = .82$. This value, which controls the elasticity of substitution between city-specific goods, is from Alvarez and Shimer (2011). We then solve the model with the remaining parameters set to their baseline values in Table 1.

The idiosyncratic technology process is

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

where ε_t is *i.i.d.* normally distributed with mean zero and variance σ_ε^2. This process is isomorphic to the one underlying our estimation summarized by equations (21)-(24). With $\gamma_z < 0$ for fixed $z_{\min}$ this process has an invariant distribution in z_t, which is required to solve our model. The upper tail of this distribution is exponential.

Is the specification (37) plausible? To address this question we need a measure of the *level* of technology, which is not possible in our framework because of unidentified fixed effects arising from technological initial conditions and the way we measure output prices. Davis, Fisher, and Veracierto (2011) construct a measure of the level of city-specific technology in a model similar to ours under the assumptions of perfect capital and labor mobility and homogeneous labor. While their measure is not exactly consistent with our framework, it still should be informative. Davis et al. (2011) find that technology is exponentially distributed for the largest 200 US cities. Indeed plots of log rank of technology versus log level of technology are remarkably similar to the analogous plots for population, except that the slopes are different. While for population the slope is near -1 (Zipf's law), for technology it is near -2.5. We refer to these slope values as Zipf coefficients.

We are able to choose γ_z to match a Zipf coefficient equal to -2.5. The variance of the innovation, σ_ε, is calibrated to match our estimate of the cross-sectional variance of employment growth described in DFW. We approximate (37) with a discrete Markov chain that yields a good approximation.

6.2 Findings

Our algorithm for solving for the model's steady state is described in DFW. For $\lambda > 1$ we do not have a theorem establishing existence and uniqueness of the steady state. However, in practice, we have had little trouble solving for it with values of λ up to

1.15, and we have obtained a solution for our calibrated parameter configuration. For the cases we have considered, our algorithm yields a unique solution from multiple sets of starting values.

Our strategy for estimating Λ relies on business investment being reversible. A more appealing assumption would be that this investment is at least partially irreversible. However, if business investment never turns negative in equilibrium, then we lose nothing by proceeding as if it is in fact reversible – the irreversibility constraints may not bind in practice. It turns out that among the largest cities comprising 37% of the population (corresponding to our empirical sample) business investment never turns negative. Moreover, while it is not crucial for our empirical strategy, it turns out that infrastructure investment is effectively irreversible as well; less than .01 percent of infrastructure capital leaves the largest cities in any given year.

We now describe how our model compares to the cross-section of MSAs, beginning with Zipf's law. The population distribution generated by the exogenous technology process is approximately exponential with a Zipf coefficient of -1.0, just as in the data. That the population distribution inherits the exponential distribution may not be surprising, but the flattening of the population distribution relative to that for technology is striking. This flattening happens because the equilibrium support of the population distribution is wider than that for the technology distribution. If factors were allocated in proportion to technology, then the population distribution would be identical to that for technology. However, in the model 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. Agglomeration tends to amplify these effects; when we shut down the density externality by setting $\lambda = 1$ the Zipf coefficient for population falls to -1.2.

Table 4 reports standard deviations of wages, rents and output prices relative to employment, and correlations of these variables with employment, in the model and data. (See DFW for how we estimate the statistics.) Two cases are reported. One corresponds to the estimated value of the agglomeration parameter, $\lambda = 1.08$, and the other to the case of $\lambda = 1$, in which the remaining parameters are left at their calibrated values. We study log growth rates in the model. The empirical counterpart to the log growth rate of variable x_{it} is $\Delta\hat{x}_{it}$.

The model is a qualitative success with respect to all the statistics in Table 4. Wages, rents and prices are all less volatile than employment, as in the data, and the signs of the correlations are the same in the model as in the data. The correlations are driven by the key feature of the model that workers are re-allocated to cities with the best productivity. With land constrained in the short run, rents rise with employment. Output prices fall because they are diminishing in the quantity of output. The positive correlation between employment and wages reflects that while reallocation tends to mitigate the effect of productivity changes on wages, it does not dominate it. The main impact of agglomeration is to amplify fluctuations. It has little impact on co-movement.[13]

We suspect that including additional shocks and frictions would help to improve the model's quantitative predictions. A positive shock to city-specific output demand would raise employment and output prices, increasing the relative volatility of prices and moving the correlation between prices and employment toward zero. Shocks to property taxes would move land rents oppositely to employment driving the correlation of rents and employment toward zero.[14] Finally, including labor mobility costs would break the tight connection between wages and productivity, thus improving both the relative volatility of wages and their co-movement with employment. While these additions to the model would improve its empirical plausibility, they would impact neither our estimation strategy nor the results we obtain.

7 Conclusion

We study the effect of local agglomeration on aggregate growth, providing three main contributions. First, we extend the neo-classical growth model to include location-specific production and housing; local infrastructure as a produced, durable input into finished land that is used in production and housing; and local agglomeration effects as proposed by Ciccone and Hall (1996). We derive the balanced growth path

[13]The small impact of agglomeration on co-movement relates to *unconditional* moments. This does not mean that agglomeration is unimportant for *conditional* moments, such as those we use in our estimation.

[14]Barlevy and Fisher (2011) find that changes in property taxes are highly correlated with changes in house prices even after accounting for other variables thought to influence house prices.

of this model and show how per capita consumption growth depends on exogenous productivity growth, endogenous growth in finished land prices, and a parameter measuring the net impact of agglomeration on local productivity. Second, we show how to use our model with a combination of panel and aggregate data to estimate the impact of local agglomeration on aggregate growth and address sampling uncertainty. Third, we apply this methodology to U.S. data and find that agglomeration has an economically and statistically significant positive impact on per capita consumption growth, raising it by about 10 percent. To our knowledge this is the first such estimate. Furthermore, we present substantial evidence in support of our model's specification, validating our findings.

Many areas of future research are suggested by this paper and we conclude by summarizing a few. First, our finding of a significant contribution of agglomeration to per capita consumption growth needs further consideration both empirically and theoretically. Second, given the important role creation and destruction of cities plays in Rossi-Hansberg and Wright (2007), the robustness of our findings to making endogenous the number of cities should be examined. Third, our methodology for assessing the contribution of agglomeration to growth can be applied to other models to quantify the role of various mechanisms in growth. Fourth, our model has many predictions for the cross section of cities we have not considered but are worth doing so. For example, it has predictions for patterns of migration.

Finally, our model should be useful for understanding how the economy responds to aggregate shocks. For example, a key feature of the business cycle is that residential investment leads non-residential investment. Fisher (2007) shows how a reduced form model of complementarities between housing and production reconciles this observation with business cycle theory. Our model endogenizes these complementarities through the co-location of housing and production.

References

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

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

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.

Barlevy, G. and J. Fisher (2011). Mortgage choices and housing speculation. Revision to Chicago Fed working paper no. 2010-12.

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.

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

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 (2011). Supplementary materials for "macroeconomic implications of agglomeration". Unpublished Manuscript.

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

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.

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.

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.

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

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

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*. 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*, 1567–1601.

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

Table 1: Baseline Parameter Estimates

Parameter	Description	Estimate	Std. Error
Λ	Agglomeration's effect 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 (17), (26) and (34). Estimates are based on $R = 1.05$.

Table 2: 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 (26). The third rows corresponds to the Hansen (1982) and Sargan (1958) J-test of the over-identifying restrictions in (26). The last rows correspond to Wald tests of the indicated null hypotheses. The tests in lines 1,2 and 4 are t-tests. The J-test has five degrees of freedom and the Wald test in line 5 has two degrees of freedom.

Table 3: 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 4: Employment and Prices in the Cross-section of Cities

Statistic	Data	Model, $\lambda > 1$	Model, $\lambda = 1$
$\text{sd}(\Delta\hat{n})$	1.01	1.01	0.86
$\text{sd}(\Delta\hat{w})/\text{sd}(\Delta\hat{n})$	0.71	0.49	0.47
$\text{sd}(\Delta\hat{r_h})/\text{sd}(\Delta\hat{n})$	0.32	0.89	0.86
$\text{sd}(\Delta\hat{p}_y)/\text{sd}(\Delta\hat{n})$	0.75	0.26	0.26
$\text{corr}(\Delta\hat{n}, \Delta\hat{w})$	0.49	0.91	0.91
$\text{corr}(\Delta\hat{n}, \Delta\hat{r_h})$	0.32	0.75	0.74
$\text{corr}(\Delta\hat{n}, \Delta\hat{p}_y)$	−0.12	−0.99	−0.99

Note: The notation in the first column is: $\text{sd}(x)$ corresponds to "standard deviation of x" and $\text{corr}(x, y)$ corresponds to "correlation between x and y." See DFW for how we obtain both the empirical and theoretical statistics.