Drs. Muth and Mills meet Dr. Tiebout: Integrating Location-Specific Amenities into Multi-Community Equilibrium Models[*]

Dennis Epple

Tepper School of Business, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213. Email: epple@andrew.cmu.edu

Brett Gordon

Columbia Business School, 3022 Broadway, New York, NY 10027. Email: brg2114@columbia.edu

Holger Sieg

Tepper School of Business, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213. Email: holgers@andrew.cmu.edu

Abstract. We consider the problem of integrating spatial amenities into locational equilibrium models with multiple jurisdictions. We provide sufficient conditions under which models that assume a single housing price in each community continue to apply in the presence of location-specific amenities that vary both within and across communities. If these conditions are satisfied, the models, estimation methods, and results in Epple and Sieg (1999) are valid in the presence of (potentially unobserved) location-specific amenities. We also show how to construct sufficient statistics that capture location specific spatial heterogeneity. We apply these techniques using data from the Pittsburgh metropolitan area. We find that these amenity measures capture proximity to important local employment centers as well as heterogeneity in school quality within a given school district. (JEL Classification: C51, H31, R12)

1

1 Introduction

Much research in urban economics and regional science has focused on improving our understanding of the relationship between housing markets and the urban economy. One important focal point of research has been to study the impact of locational amenities and spatial heterogeneity on equilibrium sorting patterns of households.[1] During the past decades, researchers have developed equilibrium models in which households (or firms) sort themselves within an urban area taking into consideration differences in spatial amenities (such as proximity to central business districts) and local housing market conditions. Variation in amenities across locations impacts the demand for locations. This in turn gives rise to variation in land and housing prices. The importance of such spatial patterns may be magnified through endogenous variation in neighborhoods, variation which may reflect local social interactions, heterogeneity in household demographic characteristics and preferences, or spatial agglomeration effects on firm productivities.[2]

Parallel to the development of spatial equilibrium models, there has been much progress in theoretical research that analyzes local public good provision in a system of local jurisdictions.[3] These models take as their starting point the idea that households are at least potentially mobile. Communities differ according to their levels of public good provision, tax rates, and local housing market conditions. Each household takes these factors into account in choosing a community. If local public good provision is decentralized via local majority rule then the level of public goods will depend on characteristics of the community's residents. Households will sort among communities according to tastes and income, so that households with similar preferences for local public goods will tend to live in the same community. Since the population of each community is endogenous, the set of households who live in the community and the decisive voters in the community are jointly determined

---

[1]This literature starts with the classic papers by Alonso (1964), Mills (1967), and Muth (1969).

[2]Firm location decisions in the presence of agglomeration effects are studied in Lucas and Rossi-Hansberg (2002).

[3]This literature was inspired by Tiebout (1956). See, for example, Epple, Filimon, and Romer (1984), Goodspeed (1989), Epple and Romer (1991), Nechyba (1997a, 1997b) and Benabou (1996a, 1996b), Fernandez and Rogerson (1996, 1998), Nechyba (2000, 2003), Benabou (2002), Fernandez and Rogerson (2003), and Rothstein (2006).

in equilibrium. When voting and thereby collectively determining the level of public good provision within a community, voters take into consideration the interaction among housing market equilibrium, mobility, and public good provision.

While these two sets of models attempt to explain the same observed sorting patterns, they rely on different mechanisms. In Tiebout models, household preferences for location are primarily driven by differences in publicly provided goods and tax rates across jurisdictions in the metropolitan area. In Muth-Mills models, spatial variation in amenities is the key driving force behind sorting patterns. Tiebout models typically ignore spatial heterogeneity while most spatial models ignore the importance of competition among jurisdictions.[4] These abstractions are naturally of concern. One purpose of this paper is to provide conditions under which models that assume a single housing price in each community continue to apply in the presence of location-specific amenities that vary both within and across communities.

Such an aggregation result is not only interesting from a purely theoretical perspective, but also has important implications for empirical work. There has been much interest lately in studying identification and estimation of locational equilibrium models.[5] A theory based estimation approach differs from a traditional regression based empirical approach in several important ways. The estimation procedure is directly derived from a well specified underlying locational equilibrium model. A key purpose of this approach is thus to identify and estimate the primitives of a theoretical model (such as preferences, technologies, and distributions of endowments) given the available data sources. This approach then provide a direct test of the goodness of fit of the underlying locational equilibrium model. Moreover, once we have estimated all parameters of an underlying model, we can consider out-of-sample predictions to validate the model. Alternatively, we can perform counter-factual policy experiments to determine the likely impact of policy interventions such as tax or

---

[4]Exceptions are deBartolome and Ross (2003) and Hanushek and Yilmaz (2007).

[5]See, among others, Epple and Sieg (1999), Epple, Romer, and Sieg (2001), Wu and Cho (2003), Sieg, Smith, Banzhaf, and Walsh (2004), Calabrese, Epple, Romer, and Sieg (2006), Ferreyra (2007), Walsh (2007), and Epple, Peress, and Sieg (2009). Bergstrom, Rubinfeld, and Shapiro (1982), Rubinfeld, Shapiro, and Roberts (1987), Nesheim (2001), Bajari and Kahn (2004), Bayer, McMillan, and Reuben (2004), Schmidheiny (2006), Bayer, Ferreira, and McMillan (2007), and Ferreira (2005) are examples of related empirical approaches which are based on more traditional discrete choice models or hedonic frameworks.

expenditure reforms. Finally, we can estimate welfare measures that are broadly consistent with partial or general equilibrium effects. A theory based approach thus allows researchers to address a variety of important research questions that are outside the scope of traditional regression based empirical analyses.[6]

It is, therefore, natural to explore the implications of spatial heterogeneity using estimation strategies that are based on locational equilibrium models. This paper shows that the basic ideas that underlie many of the recent empirical approaches remain valid. In particular, under the conditions we develop here, the models, estimation methods, and results in Epple and Sieg (1999) and Epple, Romer, and Sieg (2001) are valid in the presence of such location-specific amenities. This result is a direct consequence of the aggregation property since the representation of preferences that imposes the existence of a single housing price for each community is a valid characterization of preferences among communities. As a consequence we can estimate household preferences by focusing on sorting *among* communities, ignoring sorting *within* communities. Another consequence of this result is that estimation approaches that entail solving for equilibrium as part of estimation algorithm as discussed in Calabrese, Epple, Romer, and Sieg (2006) are also valid in the presence of spatial variation of amenities within communities.

While it is reassuring to know that spatial heterogeneity within communities can be ignored under certain conditions, there are clearly many useful applications in which it is desirable to have reliable measures of location specific spatial amenities. Unfortunately, spatial amenities are often observed with error or sometimes hard to measure at all. It is, therefore, useful to explore indirect approaches for measuring spatial amenities. Another contribution of this paper is that we show how to construct simple sufficient statistics that capture location specific spatial heterogeneity if the assumptions needed for aggregation are satisfied. These measures are based on the observed value of housing per unit of land and exploit recent advances in the estimation of housing production functions developed in

---

[6]See also the survey paper by Thomas Holmes in this edition of the journal that provides a general discussion of theory based estimation procedures and their applications to regional sciences.

4

Epple, Gordon, and Sieg (2009).[7] We illustrate this new technique for measuring amenities using an application which analyzes housing construction and household sorting in the Pittsburgh Metropolitan area. We use a parametric approach to estimate the production function of housing and construct amenity measures. We find that these amenity measures capture proximity to important local employment centers as well as heterogeneity in school quality within a given school district.

The rest of the paper is organized as follows. Section 2 discusses how to extend locational equilibrium models to account for spatial heterogeneity. Section 3 provides a set of sufficient conditions guaranteeing the existence of a well-defined (aggregate) locational equilibrium model that is consistent with spatial variation in amenities. Section 4 provides a new approach for measuring spatial amenities. Section 5 provides some empirical results to illustrate the new technique for measuring amenity values. Section 6 offers some conclusions and discusses important areas of future research.

2   A Model

In this section we outline how to construct a locational equilibrium model that accounts for spatial amenities. Consider a metropolitan area divided into J jurisdictions with fixed, historically-determined boundaries. Each jurisdiction provides a local public good, g, and finances expenditure on the good with an ad valorem tax, t, on the value of housing.

A key departure from the models considered in the previous literature is that we assume there are I_j geographic areas or locations, indexed by i, within municipality j.[8] Each location has land area that is continuously divisible. We denote by p^l the price per unit

---

[7]Using the same techniques, we can also show how to estimate the price of housing services in each community. This research is related to Sieg, Smith, Banzhaf, and Walsh (2002) who consider a model in which housing services are provided from multiple inputs and derives the associated indirect utility function, demonstrating that a single price index captures the role of housing inputs when the production function for housing services exhibits constant returns to scale.

[8]The larger jurisdictions that comprise the metropolitan area are referred to as "communities." The smaller areas within each community are referred to as 'locations." We suppress the location specific index (i, j) where this can be done without causing confusion.

5

land and notice that this price is location specific. We depart from the classic monocentric urban model by assuming that the metropolitan plain need not be featureless. Instead, locations in the metropolitan area may differ in a vector of amenities. These amenities, taken by private market participants as exogenously given at each location, may include, inter alia, air quality, aesthetic appeal of the natural setting, distances to places of employment, commerce, education, recreation, and noisy streets. We denote by A the amenity level at a location. We assume that the amenity vector is constant within location but varies across locations.

In each location there is a firm, which behaves competitively and produces a housing good q from two factors, land l and structures m, using a constant returns to scale technology, $q = q(m, l)$. Land can be purchased at a price p^l and structures at price p^m. The price for structures is constant within the metro area. The price of q is denoted by p^q. Solving the firm's optimization problem and assuming that land is in fixed supply, we obtain an aggregate housing supply function denoted by $Q^s(p^q)$ that is monotonically increasing in the price of housing p^q.

The metropolitan area is inhabited by a continuum of households varying in preference parameter α and endowed income y. The metropolitan population is normalized to one and has density $f(\alpha, y)$. Households obtain utility from amenities, A, from housing consumption, q, from a locally provided public good, g, and from the numeraire bundle, b. Households have the utility function $U(\alpha, g, q, A, b)$. Households face the budget constraint $y = (1+t)p^q q + b$, where, as explained above, the price of housing, p^q, may vary with location, and the price of the numeraire good, is the same at all locations and normalized to be 1. We assume that all households are price takers.

It is useful to model the household choice problem as occurring in two stages. In the first stage, the household determines the optimal housing and consumption plans for each location given the prices of housing and the prize for the composite consumption good. In the second stage, the household chooses the location at which it obtains the highest utility. Suppressing the location superscript, we can write the first stage of the household's problem

6

at a location with housing price p^q and amenities A as:

$$(1) \qquad \max_q \quad U(\alpha, g, q, A, b)$$

$$s.t. \quad y = (1+t)\, p^q\, q\, +\, b$$

We denote the implied indirect utility function of household (y, α) as $V(\alpha, g, p^q(1+t), A, y)$, and the associated demand function for housing services as $q_d(p^q(1+t), A; y, \alpha)$.

Based on our characterization of household preferences, we can then characterize the sorting patterns of households. Let C_{ij} denote the set of households residing in location (i, j):

$$(2) \qquad C_{ij} = \Big\{ (\alpha, y) \ \Big| \ V(\alpha, y, g_j, p_{ij}^q(1+t_j), A_{ij}, y)$$

$$\geq \max_{km \neq ij} = V(\alpha, y, g_k, p_{mk}^q(1+t_k), A_{mk}, y) \Big\}$$

The housing market clears in location (i, j) if the following holds:

$$(3) \qquad \int_{C_{ij}} q_d(g_j, p_{ij}^q(1+t_j), A_{ij}, y, \alpha)\, f(\alpha, y)\, dy\, d\alpha = Q_{ij}^s(p_{ij}^q)$$

Aggregating over the set of of locations that form a community, we can compute total tax revenues for each community. Budget balance in jurisdiction j requires:

$$(4) \qquad t_j \sum_{i \in I_j} p_{ij}^q \int_{C_{ij}} q_d(p_{ij}^q, A_{ij}; y, \alpha) f(\alpha, y)\, dy\, d\alpha = c(g_j)$$

where $c(g_j)$ denotes the costs of providing g_j.

Moreover, we can close the model and show that a voting equilibrium exits. By exploiting the aggregation results developed in this paper, the existence of voting equilibrium can be established using the approach developed in prior research.[9] Since our focus here is on developing the aggregation results, we do not detail the analysis of the voting problem.

---

[9]For a detailed discussion, see, for example, Epple, Romer, and Sieg (2001).

An equilibrium for this model is then defined as follows:

Definition 1 *An intercommunity equilibrium consists of a set of communities; a continuum of households; a distribution of household characteristics; and a partition of households across communities and locations such that every location has a positive population; a vector of prices, a vector of tax rates; an allocation of public goods; and an allocation of private goods for every household, such that:*

1. *Every household, living in a given location maximizes its utility subject to the budget constraint.*

2. *Each household lives in one location and no household wants to move to a different community.*

3. *The housing market clears in every location.*

4. *The budget of every community is balanced.*

5. *There is a voting equilibrium in each community: Over all levels of taxes and expenditures that are perceived to be feasible allocations by the voters in each community, at least half of the voters prefer the equilibrium levels of taxes and expenditures over any other feasible combination.*

In the above definition, we have avoided further notational complexity by assuming all urban locations are occupied. The model can be extended in an obvious way to the case where the amenity levels at some locations are sufficiently low that the locations are unoccupied in equilibrium.

3   Aggregation

In this section, we provide sufficient conditions under which the model discussed above allows a representation such that a single housing price in each community continues to

apply in the presence of location-specific amenities. We show this to be the case when the following additional assumptions apply.

Assumption 1 *The bundle of amenities can be mapped into a uni-dimensional index, $a(A)$, common across households, with utility increasing in the value of the index: $U_a(\alpha, g, q, a, b) > 0$.*

Assumption 1 is strong and rules out, for example, models with horizontal taste differentiation in which the uni-dimensional aggregation index is type specific. Thus households do not use the same aggregation rule.[10] It also rules out multiple index models in which a single index that captures the impact of all amenities on the utility of households does not exist

Employing Assumption 1, we suppress A until we return later to our empirical application. Note that Assumption 1 implies that a change in the metropolitan equilibrium associated, for example, with a change in the population distribution $f(\alpha, y)$ will not change the ordering of housing prices within any community. That is, for every j, the ordering of p_{ij}^q across locations $i \in I_j$ is the same in any equilibrium.

Assumption 2 *The utility function can be written $U(\alpha, g, q, a, b) = U(\alpha, g, h(q, a), b)$.*

Assumption 2 states that there is a housing index, h, that incorporates the effects of land and non-land inputs, and amenities.

Assumption 3 *The index of housing services is multiplicatively separable: $h = q \cdot a(A)$.*

Using assumptions 1 through 3, the choice problem of a consumer at a location with amenity a and price per unit housing services p^q is then:

$$(5) \qquad \max_q U(\alpha, g_j, q \cdot a, y - p^q(1 + t_j)q),$$

---

[10]As a consequence, there are interesting applications for which Assumption 1 might be restrictive.

Substituting $h = q \cdot a$, this may alternatively be written:

$$(6) \qquad V\left(\alpha, g_j, \frac{p^q(1+t_j)}{a}, y\right) \; = \; \max_h U\left(\alpha, g, h, y - \frac{p^q(1+t_j)}{a}h\right)$$

Within a community, the consumer chooses among locations with differing amenity levels to maximize $V(\cdot)$. It follows from the above indirect utility function that any consumer (α, y) given a choice of locations $i \in I_j$ with differing amenity levels will prefer the location with lowest $\frac{p_{ij}^q}{a_{ij}}$. Thus, in equilibrium, all locations $i \in I_j$ must have a constant price per unit of h, $\frac{p_{ij}^q}{a_{ij}} = p_j^h$. Thus, the indirect utility function for a consumer located anywhere in the community can be written $V(\alpha, g_j, p_j, y)$, where $p_j = p_j^h(1 + t_j)$. We have thus shown the following result:

Proposition 1 *Under assumptions 1 through 3, the consumer choice problem can be expressed in terms of a single price, p_j^h, per community that is invariant to location specific amenities within the community.*

The following example, illustrates the main result in Proposition 1. For simplicity we assume in this example that there is only one community $(J = 1)$ and set the tax rate equal to zero $t = 0$. Consider a household that has preferences defined over a local public good g, a continuously varying local amenity r, consumption of housing services $q(r)$, and a composite private good b:

$$(7) \qquad U \; = \; \left\{\alpha\, g^\rho \; + \; B_0\left(r^{-\delta}\, q^B\, b^{(1-B)}\right)^\rho\right\}^{1/\rho}$$

where B_0 is conveniently defined. The budget constraint of a household is given by:

$$(8) \qquad y \; = \; p^q\, q \; + \; b$$

Solving this optimization problem yields:

$$(9) \qquad B\, y \; = \; p^q\, q$$

$$(1 - B)\, y \;=\; b$$

Substituting these optimality conditions into the utility function implies that the indirect utility function can be written as:

$$(10) \qquad V(r_i) \;=\; \left\{ \alpha\, g^\rho \;+\; \left[(r_i^{\frac{\delta}{B}} \, p_i^q)^{-B} \, y \right]^\rho \right\}^{1/\rho}$$

Define the amenity-adjusted housing services h as:

$$(11) \qquad h \;=\; r^{-\delta/B} \, q$$

Solving for q and substituting into the budget constraint, we see that the price of h must satisfy:

$$(12) \qquad p_i^h \;=\; r_i^{\frac{\delta}{B}} \, p_i^q$$

Households choose among locations to maximize V. Since there can be no arbitrage opportunities in equilibrium, it must be the case for any pair of locations i and j that allocations must satisfy the following equal utility condition:

$$(13) \qquad V(r_j) \;=\; V(r_i)$$

This equal utility condition implies that the price of amenity adjusted housing services p_i^h is constant and does not depend on i:

$$(14) \qquad p_i^h \;=\; p^h \quad , i = 1, .., I$$

Hence the indirect utility function does not depend on r and can be written as:

$$(15) \qquad V \;=\; \left\{ \alpha\, g^\rho \;+\; \left[p_h^{-B} \, y \right]^\rho \right\}^{1/\rho}$$

We now adopt the usual single-crossing conditions for multi-community equilibrium models. Let $S(\alpha, y, g, p) = \frac{dp}{dg}|_{V=\bar{V}}$

Assumption 4 $S(\alpha, y, g, p)$ *is increasing in* α *for all* y *and increasing in* y *for all* α.

The following necessary conditions for equilibrium then apply.

Proposition 2 *Consider an equilibrium allocation in which no two communities have the same housing prices. For such an allocation to be a locational equilibrium – no-one wishes to move – there must be an ordering of community pairs,* $\{(g_1, p_1), ..., (g_J, p_J)\}$, *such that:*

1. *Boundary Indifference: The set of "border" individuals between any two adjacent communities are indifferent between the two communities. This set is characterized* $\{(\alpha, y)|V(\alpha, g_j, p_j, y) = V(\alpha, g_{j+1}, p_{j+1}, y)\}$ $j = 1, ..., J-1$

2. *Stratification: Let* $\alpha(y)$ *be the implicit function defined by equation (10). Then, for each level of income* y, *the residents of community* j *consist of those with tastes,* α , *given by:* $\alpha_{j-1}(y) < \alpha < \alpha_j(y)$

3. *Increasing Bundles: Consider two communities* k *and* j *such that* $p_k > p_j$. *Then* $g_i > g_j$ *if and only if* $\alpha_k(y) > \alpha_j(y)$.

Propositions 1 through 2 establish that Assumptions 1 through 3 are sufficient to justify estimation of multi-community equilibrium models while ignoring intra-community variation in amenities. If we adopt the same functional form for the indirect utility function employed in previous empirical work such as Epple and Sieg (1999), then the implied housing demand functions, and house values are valid in the presence of location-specific amenities. If we know community populations, we can then estimate utility function parameters while ignoring amenity variation, as we have done in Calabrese, Epple, Romer and Sieg (2006). Proposition 1, coupled with the assumption that community populations are known, thus permits solution of the model without the need either to characterize location-specific amenities or to solve for the price per unit of housing services at all locations. This is

an exceedingly valuable decomposition of the problem from an empirical perspective. Housing price per unit in one community can be normalized to one. Amenity-adjusted housing prices in remaining communities are determined by usual the boundary-indifference conditions, where boundaries here are in the $\{\alpha, y\}$ space. Since individuals within a community are indifferent among locations within the community, results are invariant to which types within a community are located at a physical boundary.

4   Measuring Spatial Amenities

We now turn to the development of a measure of the spatial amenities a_{ij}. Note that the key problem encountered in the analysis is that neither p_q nor p_h are observed by the econometrician. However, we observe the value of $v = V/L$. As shown in Epple et al. (2009), there exists a monotonic relationship between v and p_q if the housing production function of housing satisfies constant returns to scale. To illustrate this result consider the value of housing per unit land which can be written as:

$$(16) \qquad\qquad v \;=\; p_q\, q_s(p_q) \;=\; w(p_q)$$

where $q_s(p_q)$ is the supply of housing per unit of land which is monotonically increasing in p_q. As a consequence $w(p_q)$ is monotonically increasing in p_q and hence we have $p_q = w^{-1}(v)$.

Under the assumptions in Section 3, equilibrium requires that $p_q = p_h\, a$. Hence:

$$(17) \qquad\qquad a = \frac{w^{-1}(v)}{p_h}$$

Thus, given a functional form for the housing production function, we can infer the amenity level from the observed value of a house per unit land up to a community specific factor of proportionally.

To gain some additional insights, let us consider the special case where $q(L, M)$ is of the

CES variety.[11] In that case, the production function per unit of land is given by

$$(18) \qquad q(L, M)/L = q(m) = [\alpha m^\rho + (1 - \alpha)]^{1/\rho}$$

Consider the producer's profit function per unit of land which is given by

$$(19) \qquad \pi = p_q [\alpha m^\rho + (1 - \alpha)]^{1/\rho} - p_m m - p_l$$

Since m can be measured in arbitrary units and the price of m does not depend on the location, we can normalize p_m to one. The first-order condition for m implies that

$$(20) \qquad m = \frac{1}{1 - \alpha} \left[\left(\frac{1}{\alpha p_q} \right)^{\rho/(1-\rho)} - \alpha \right]^{-1/\rho}$$

After substituting this into the CES production function, we obtain the following supply function per unit of land:

$$(21) \qquad q(p_q) = \left\{ \alpha \left[\frac{1}{1 - \alpha} \left[\left(\frac{1}{\alpha p_q} \right)^{\rho/(1-\rho)} - \alpha \right]^{-1/\rho} \right]^\rho + (1 - \alpha) \right\}^{1/\rho}$$

and $w(p_q) = p_q q(p_q)$. Given parameter values for α and ρ, we can be numerically invert $w(p_q)$ to obtain a.

To make this procedure operational, we need to estimate α and ρ. Referring back to the firm's problem, the first-order condition for m implies:

$$(22) \qquad m^{\rho-1} = \frac{(1 - \alpha)}{\alpha p_l}$$

Next, note that $m = v - p_l$. Substituting this into the above and taking logs yields

$$(23) \qquad \ln(v - p_l) = \frac{1}{\rho - 1} \ln \left(\frac{1 - \alpha}{\alpha} \right) + \frac{1}{1 - \rho} \ln(p_l)$$

---

[11]Epple et al. (2009) provide a flexible approach that does not require functional form assumptions.

For simplicity, we assume that $\ln(v - p_l)$ is measured with classical error. We then obtain a well-defined regression model that can be used to estimate α and ρ.[12]

5   Empirical Results

For our empirical application, we examine data from Allegheny County in Pennsylvania, which contains the greater metropolitan area of Pittsburgh. Our main data source is from real estate assessment data collected from the Allegheny County web site, based on assessments by the Office of Property Assessments.[13] The web site provides access to a database with detailed information about all properties, both residential and commercial, in the entire county. The database is updated on a yearly basis, and contains a wide array of information about the property, such as details about the owner, usage of the property, sales and tax information, and building characteristics.

The Office of Property Assessments provides two types of assessment: one from the county and one based on the 'full market' value. The assessments under each type are decomposed into estimates for the land value, the building value, and the total value. As noted above, we also observe the most recent sale price and date for property. Though using market transactions would be preferable to market assessments, we chose the latter for three reasons. First, many of these sales took place number of years ago; nearly half of all the transactions took place prior to 1995. Second, of the 371,251 properties for which we have sales information, the sales price of 83,039 of them, or 22%, are listed as \$1. It is likely that these properties were sold either family members or close friends. Finally, in July of 2004, the Pennsylvania State Tax Equalization Board performed a study of all properties that had been sold to determine how close the assessed values were to the sale prices. They found that on average the assessments were within 2.5% of the sale price. For these reasons, we chose the assessed market values since they appear to be accurate and

---

[12]Epple et al. (2009) discuss how to identify and estimate the underlying production function under much weaker conditions.

[13]http://www2.county.allegheny.pa.us/RealEstate/

15

give us a consistent measure of housing value.

The complete database lists 561,174 properties. After eliminating all non-residential properties and those that are listed as condemned or abandoned, we are left with 423,556 observations. We successfully geocoded – matched to longitude and latitude coordinates – 370,178 of these properties. Properties that were not geocoded were usually those that lacked sufficient address information to be precisely located, and frequently were outside the city limits. We used the coordinates to assign each property to the proper Census block, which allowed us to match income data from the Census SF-3 long-form at the block group level to each address.[14] We also construct a second subsample which focuses on recently built units. This sample only contains houses that were built after 1995.[15]

The distance to the city center is a commonly used amenity. However, due to Pittsburgh's unusual geography, a simple 'straight line' distance measure does not adequately capture the actual time it takes to travel from one point to another. For this reason, we use travel time data from the Southwestern Pennsylvania Commission (SPC) for Allegheny County. The SPC divided the county into 995 traffic zones of varying size, roughly distributed according to traffic and population density. The city of Pittsburgh is covered by 465 zones. The SPC provided us with estimated travel times from each zone to another, under both congested and uncongested conditions. We matched each address in the database to its corresponding traffic zone, and retrieved the travel times to the designated city center traffic zone.

While the city itself is a single, large school district, it is divided into a set of eight high school attendance zones. To match each city address in our data set to the correct attendance zone, the Pittsburgh Public School District provided us with the address feeder table for the attendance zones. This table consists of a large set of street names and number ranges along with the corresponding attendance zone. Unfortunately, the table is not an exhaustive list of all the possible city addresses. We were able to match 85,231 of the 90,676

---

[14]The block group is the most disaggregated level at which the Census makes income data available.

[15]The theory of production technically speaking only applies to new housing construction, and not existing units that were build in the past using potentially different technologies.

city addresses with a direct match of the street name and a street number that fell in the appropriate range. For the remaining 5,445 addresses, we assigned each to the attendance zone of the closest matched residence. Table 1 provides some summary descriptive statistics of our data.

We estimated equation (23) using a large subset of our original data containing 359,272 observations. This subset was obtained after eliminating properties which did not have positive lot area sizes and market values listed, as well as those properties for which we were unable to match with travel time data. An endogeneity problem arises in estimation if the errors capture productivity shocks. The use of instruments in estimation largely deals with this endogeneity problem. We use the travel time to the city center as instrument for the price of land. Our identifying assumption is then that there are no spatial patterns in productivity shocks within the metropolitan area. For example, productivity shocks must be uncorrelated with proximity to the city center. Results for both least squares and IV regression are presented in Table 2, along with the values of α and ρ implied by the coefficient estimates. All the parameter estimates are significant at the 99% level. Standard errors for α and ρ calculated using simulation techniques. Table 3 reports the results for the sub-sample of recently constructed houses since the theory applies strictly speaking only for new construction. We find similar results.

Substituting the estimates for α and ρ from the IV regression into equation (21), we numerically inverted it to obtain the amenity estimates a_{ij} for each residence within each community up to the constant p_h.[16] Within the city limits of Pittsburgh, we have data on 91,767 residences. For those residences, we have defined the communities as consisting of the eight high school attendance zones.

Figure 1 shows amenity plots for the entire region of interest with amenity values color-coded into one of five quantiles. (dark values = higher amenities, white = lowest amenity). The downtown of Pittsburgh is located at the junction of three rivers. Figure 1 shows

---

[16]All of the following estimates were repeated with the parameter values from the least squares case, with qualitatively similar conclusions.

that there is a high-value amenity region to the east of the city center. This area contains Carnegie Mellon University, the University of Pittsburgh hospital system, and Squirrel Hill and Shadyside. These are generally viewed as being among the most desirable neighborhoods in the city.

We can decompose the amenity into endowed and publicly determined components and check whether these values are consistent with our expectations. For our endowed amenity, we used the driving time from that location to the city center under congested road conditions.[17] For our endogenous amenity, we used high school attendance zones to capture differences in school quality. This approach is motivated by the work of Black (1999) who investigated differences in housing values near boundaries of school attendance zones. Table 4 reports results of logged amenity values regressed on travel time to the city center and attendance zone dummy variables. As expected, the coefficient on travel time is negative and highly significant. The estimates for the high school attendance are reasonable for the most part. The main outlier is the point estimate for Peabody high school. We find a significant positive amenity effect despite the fact that test scores for that school are generally lower than average.[18]

While Pittsburgh has a distinct downtown area, major local employment centers can be found in universities, hospitals, and regional business districts. To understand the effect that such centers have on local amenities, we estimate the radius of amenity influence that Carnegie Mellon University has on the surrounding housing values. A dummy variable was created to indicate whether a property was within a particular radius from Carnegie Mellon University, and we added this variable to the regressions from Table 4. The results for various sizes of radii are presented in Table 5. We find that the proximity to Carnegie Mellon University is only valued with a 0.5 mile radius. This non-linearity is consistent with a "walking distance" proximity amenity.

---

[17]Results were similar when using uncongested travel times.

[18]Ranking the high schools by the 2005 average PSSA math scores, we obtain Allderdice, Brashear, Schenley, Langley, Carrick, Oliver, Westinghouse, and Peabody. The rankings are similar for reading and writing scores.

6 Conclusions and Future Research

As all readers of this paper know, modeling entails adopting assumptions that abstract to a greater or lesser degree from the complexity encountered in practice, assumptions chosen to facilitate tractability while preserving features that are essential to the problem being studied. In this paper, we have presented assumptions that permit analysis of household choice of community, housing markets, and collective choice within communities while abstracting from amenity variation within communities. From an econometric perspective, these assumptions permit estimation of structural econometric models using data aggregated to the community level (Epple and Sieg, 1999; Epple, Romer, and Sieg, 2001; Calabrese, Epple, Romer, and Sieg, 2006; Sieg, Smith, Banzhaf, and Walsh, 2004). These models can be solved for multi-community equilibrium and can, in turn, be used to study a variety of policy issues (Smith, Sieg, Banzhaf, and Walsh, 2004; Calabrese, Epple, and Romano, 2009).

In applying the framework of this paper, the researcher treats as communities a set of geographic subareas within a metropolitan area. Thus, to apply the framework, the researcher must choose a definition of a community. For example, for some applications, a community might be defined to be a municipality. For other applications, a researcher might define school districts to be communities, or, opting for a more fine-grained definition, the researcher might define school attendance zones to be communities. The more fine grained the definition of community, the more comfortable the researcher will be with Assumptions 1, 2, and 3 as a characterization within-community spatial variation of amenities. Of course, a more fine grained the definition creates more substantial demands in estimation and computation of equilibrium. However, as Epple and Sieg (1999) and Calabrese, Epple, Romer, and Sieg (2006) show, computation of equilibrium and estimation is readily undertaken for a model with 92 communities. There are no apparent barriers to applying the model to a much larger number of communities.

We recognize, of course, that there are alternative frameworks. Our goal here has been twofold. One is to reassure readers who might have concluded that our framework cannot

accommodate spatial variation in amenities. More importantly, by formalizing conditions for incorporating spatial amenities, we hope to broaden the scope of application of the framework.

We conclude with some comments with respect to the broader agenda that lies ahead. Empirical research in local public economics and urban economics that is based on rigorous, theoretical equilibrium models has significantly improved our understanding of the role that local public goods, neighborhood peer effects and spatial amenities play in explaining observed household sorting patterns. As a consequence, researchers now have a quantitative framework for analyzing the impact of local public policies on housing prices and household sorting patterns. Previous empirical studies have analyzed a variety of different policies including local educational and school finance reform measures, improvements in air quality, policies aimed to prevent urban sprawl and increase access to open space, and local redistribution. Empirical analysis based on rigorous economic theory has allowed researchers to address a variety of important research questions such as the estimation of welfare measures which can take in consideration household adjustments in response to large policy changes. While there has been much progress over the past decade, it is fair to say that there is plenty of room for future research. There are many open research questions at the intersection of urban economics, regional science, and local public finance. Some of these can be answered using the current generation of urban sorting models. Others are more difficult and will require new models and empirical frameworks. In the remainder of the paper, we discuss some of the most pressing challenges and opportunities.

Most current models also do not account for peer effects in residential sorting or use only crude measures — typically the mean of some distribution of household characteristics.[19] This approximation does not account for some salient features of human interactions. Moreover, we need better models that account for the interaction of neighborhood and peer effects with other amenities or local public goods. For example, good local schools are par-

---

[19]The empirical literature incorporating peer and neighborhood effects in local public good provision includes early research by Henderson, Mieszkowski, and Sauvageau (1978), and Summers and Wolfe (1977), and the more recent resurgence reflected, for example, in work of Cutler, Glaeser, and Vigdor (1999), Hoxby (2000), and Rothstein (2006).

tially a function of the quality of student body, the quality of parents that volunteer inside the school and monitor teachers and administrators. Clearly, there are trade-offs. Spending more resources on schools should improve the quality of the teachers and thus reduces the need for monitoring by the parents. We are just beginning to understand these issues.

Research also needs to incorporate more realistic models of the housing market. Models that treat housing as a perfectly divisible homogeneous good are convenient, but often fail to capture important heterogeneity among housing. Standard differentiated product models that have been used in the industrial organization literature offer an alternative. However, estimating housing production functions or supply functions of housing treating housing as a differentiated product is challenging.

In addition empirical research needs to account for institutional features and urban housing market policies such as public housing and Section 8 vouchers.[20] Locational choices and mobility of low income households are often constrained by access and availability to affordable housing. There is very little research that the studies the impact of urban housing and renewal polices within a well-defined quantitative equilibrium model. Moreover, there is some empirical evidence that zoning laws can be important in restricting housing supply in desirable neighborhoods.[21]

Future research also needs to explore alternative models of endogeneizing public good provision and test the assumptions that underlie these models. There is some research on the validity of the myopic median voter model. But voters might be more sophisticated. Alternatively, there are many policies which may not reflect the preferences of median voters. Collective decisions may not be made by voters, but bureaucrats or elected officials that have their own agendas. Thus alternative explanations may be needed. There is much need for quantitative research investigating the incentives and constraints faced by local politicians, and decision makers.

---

[20]Epple, Geyer, and Sieg (2009) study entry and exit in urban housing programs. They estimate an equilibrium search model which accounts for rationing due to excess demand in equilibrium.

[21]Glaeser and Gyourko (2005), Glaeser, Gyourko, and Saks (2005), and Brueckner and Rosenthal (2005) have forcefully argued that understanding housing supply is key to understanding the growth and decline of urban areas.

Finally, most research has been conducted within a static framework. Researchers need to introduce dynamics in household locational choices and land development. These models will allow us to get a better understanding of mobility patterns, based on more realistic assumptions capturing the costs and benefits of mobility.[22]

---

[22]Some early dynamic work include Benabou (1996a, 1996b), Durlauf (1996), Fernandez and Rogerson (1998), and Glomm and Lagunoff (1999). More recently Ortalo-Magne and Rady (2006) have worked on the role of uncertainty in housing prices on residential sorting. Epple, Romano, and Sieg (2009) study the impact of limited mobility on residential within an overlapping generations model. Bayer, McMillan, Murphy, and Timmins (2007) estimate a dynamic housing demand model. Murphy (2007) considers the dynamics of land conversion and housing development.

References

Alonso, W. (1964). *Location and Land Use: Towards a General Theory of Land Rent.* Cambridge University Press. Cambridge.

Bajari, P. and Kahn, M. (2004). Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities. *Journal of Business and Economic Statistics, 23(1)*, 20–33.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy, 115 (4)*, 588–638.

Bayer, P., McMillan, D., Murphy, A., and Timmins, C. (2007). A Dynanic Model of the Housing Market. Working Paper.

Bayer, P., McMillan, R., and Reuben, K. (2004). The Causes and Consequences of Residential Segregation: An Equilibrium Analysis of Neighborhood Sorting. Working Paper.

Benabou, R. (1996a). Equity and Efficiency in Human Capital Investments: The Local Connection. *Review of Economic Studies, 63(2)*, 237–64.

Benabou, R. (1996b). Heterogeneity, Stratification and Growth: Macroeconomic Effects of Community Structure. *American Economic Review, 86*, 584–609.

Benabou, R. (2002). Tax and Education Policy in a Heterogeneous-Agent Economy: Maximize Growth and Efficiency?. *Econometrica, 70(2)*, 481–517.

Bergstrom, T., Rubinfeld, D., and Shapiro, P. (1982). Micro-Based Estimates of Demand Functions for Local School Expenditures. *Econometrica, 50*, 1183–1205.

Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary School Education. *Quarterly Journal of Economics, 114(2)*, 577–600.

Brueckner, J. and Rosenthal, S. (2005). Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns Be Rich?. Working Paper.

Calabrese, S., Epple, D., and Romano, R. (2009). Inefficiencies from Metropolitan Political and Fiscal Decentralization: Failures of Tiebout Competition. Working Paper.

Calabrese, S., Epple, D., Romer, T., and Sieg, H. (2006). Local Public Good Provision: Voting, Peer Effects, and Mobility. *Journal of Public Economics, 90 (6-7)*, 959–81.

Cutler, D., Glaeser, E., and Vigdor, J. (1999). The Rise and Decline of the American Ghetto. *Journal of Political Economy, 107*, 455–506.

deBartolome, C. and Ross, S. (2003). Equilibria with Local Governments and Commuting: Income Sorting vs. Income Mixing. *Journal of Urban Economics, 54(1)*, 1–20.

Durlauf, S. (1996). A Theory of Persistent Income Inequality. *Journal of Economic Growth, 1*, 75–93.

Epple, D., Filimon, R., and Romer, T. (1984). Equilibrium Among Local Jurisdictions: Toward an Integrated Treatment of Voting and Residential Choice. *Journal of Public Economics*, *24*, 281–304.

Epple, D., Geyer, J., and Sieg, H. (2009). Entry and Exit in Urban Housing Programs: Estimating an Equilibrium Search Model with Rationing. Working Paper.

Epple, D., Gordon, B., and Sieg, H. (2009). A New Approach to Estimating the Production Function for Housing. *American Economic Review*, *forthcoming*.

Epple, D., Peress, M., and Sieg, H. (2009). Household Sorting and Neighborhood Formation. Working Paper.

Epple, D., Romano, R., and Sieg, H. (2009). The Dynamics of Metropolitan Communities. Working Paper.

Epple, D. and Romer, T. (1991). Mobility and Redistribution. *Journal of Political Economy*, *99(4)*, 828–858.

Epple, D., Romer, T., and Sieg, H. (2001). Interjurisdictional Sorting and Majority Rule: An Empirical Analysis. *Econometrica*, *69*, 1437–1465.

Epple, D. and Sieg, H. (1999). Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy*, *107 (4)*, 645–681.

Fernandez, R. and Rogerson, R. (1996). Income Distribution, Communities, and the Quality of Public Education. *Quarterly Journal of Economics*, *111 (1)*, 135–164.

Fernandez, R. and Rogerson, R. (1998). Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education-Finance Reform. *American Economic Review*, *88(4)*, 813–833.

Fernandez, R. and Rogerson, R. (2003). Equity and Efficiency: An Analysis of Education Finance System. *Journal of Political Economy*, *111*, 858–97.

Ferreira, F. (2005). You Can Take It with You: Proposition 13 Tax Benefits, Residential Mobility, and Willingness to Pay for Housing Amenities. Working Paper.

Ferreyra, M. (2007). Estimating the Effects of Private School Vouchers in Multi-District Economies. *American Economic Review*, *97*, 789–817.

Glaeser, E. and Gyourko, J. (2005). Urban Decline and Durable Housing. *Journal of Political Economy*, *113*, 345–75.

Glaeser, E., Gyourko, J., and Saks, R. (2005). Urban Growth and Housing Supply. Working Paper.

Glomm, G. and Lagunoff, R. (1999). A Dynamic Tiebout Theory of Voluntary vs Involuntary Provision of Public Goods. *Review of Economic Studies*, *66*, 659–678.

Goodspeed, T. (1989). A Reexamination of the Use of Ability-to-Pay Taxes by Local Governments. *Journal of Public Economics*, *38 (3)*, 319–342.

Hanushek, E. and Yilmaz, K. (2007). The Complementarity of Tiebout and Alonso. *Journal of Housing Economics*, *16(2)*, 243–61.

Henderson, V., Mieszkowski, P., and Sauvageau, Y. (1978). Peer Group Effects and Educational Production Functions. *Journal of Public Economics, 10,* 97–106.

Hoxby, C. (2000). The Effects of Class size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics, 115,* 1239–1285.

Lucas, R. and Rossi-Hansberg, E. (2002). On the Internal Structure of Cities. *Econometrica, 70 (4),* 1445–76.

Mills, E. (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review, 57,* 197–210.

Murphy, A. (2007). A Dynanic Model of Housing Supply. Working Paper.

Muth, R. (1969). *Cities and Housing.* University of Chicago Press. Chicago.

Nechyba, T. (1997a). Existence of Equilibrium and Stratification in Local and Hierarchical Tiebout Economies with Property Taxes and Voting. *Economic Theory, 10(2),* 277–304.

Nechyba, T. (1997b). Local Property and State Income Taxes: the Role of Interjurisdictional Competition and Collusion. *Journal of Political Economy, 105 (2),* 351–84.

Nechyba, T. (2000). Mobility, Targeting and Private School Vouchers. *American Economic Review, 90,* 130–146.

Nechyba, T. (2003). Centralization, Fiscal Federalism, and Private School Attendance. *International Economic Review, 44,* 179–204.

Nesheim, L. (2001). Equilibrium Sorting of Heterogeneous Consumers Across Locations: Theory and Empirical Implications. Dissertation, University of Chicago.

Ortalo-Magne, F. and Rady, S. (2006). Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints. *Review of Economics Studies, 73,* 459–85.

Rothstein, J. (2006). Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition Among Jurisdictions. *American Economic Review, 96,* 1333–1350.

Rubinfeld, D., Shapiro, P., and Roberts, J. (1987). Tiebout Bias and the Demand for Local Public Schooling. *Review of Economics and Statistics, 69,* 426–437.

Schmidheiny, K. (2006). Income segregation and local progressive taxation: Empirical evidence from Switzerland. *Journal of Public Economics, 90,* 429–458.

Sieg, H., Smith, V. K., Banzhaf, S., and Walsh, R. (2002). Interjurisdictional Housing Prices in Locational Equilibrium. *Journal of Urban Economics, 52,* 131–153.

Sieg, H., Smith, V. K., Banzhaf, S., and Walsh, R. (2004). Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods. *International Economic Review, 45 (4),* 1047–1077.

Smith, V. K., Sieg, H., Banzhaf, S., and Walsh, R. (2004). General Equilibrium Benefits of Environmental Improvements: Projected Ozone Reductions for the LA Air Basin. *Journal of Environmental Economics and Management*, *47*, 559–584.

Summers, A. and Wolfe, B. (1977). Do Schools Make a Difference?. *American Economic Review*, *67*, 639–652.

Tiebout, C. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, *64 (5)*, 416–424.

Walsh, R. (2007). Endogenous open space amenities in a locational equilibrium. *Journal of Urban Economics*, *61 (2)*, 319–344.

Wu, J. and Cho, S. (2003). Estimating households' preferences for environmental amenities using equilibrium models of local jurisdictions. *Scottish Journal of Political Economy*, *50*, 189–206.

Tables and Figures

Table 1: Descriptive Statistics

Full Sample of Residential Real Estate ($N = 359{,}272$)					
Variable	Mean	Median	Stdev	Min	Max
value per unit of land, v	14.30	11.47	13.70	0.004	698.10
price of land, p_l	2.86	2.92	2.68	0.001	100.60
lot area (sq. ft.)	13665	6600	57560	165	6141960
travel time (min.)	23.27	22	10.23	1	62
Post-1995 Sample of Residential Real Estate ($N = 6{,}362$)					
Variable	Mean	Median	Stdev	Min	Max
value per unit of land, v	21.44	14.29	26.91	0.15	366.62
price of land, p_l	3.32	2.28	3.86	0.05	41.75
lot area (sq. ft.)	26756	15507	52197	540	1207483
travel time (min.)	29.12	30	9.47	1	59

The size of the full residential sample is 359,272, the size of the post-1995 sample is 6,362.

Table 2: Production Function Estimates: Full Sample

	OLS	Std Err	IV	Std Err
Constant	1.40	(0.0019)	1.04	(0.0043)
Slope	0.92	(0.0018)	1.36	(0.0054)
R^2	0.64		0.49	
α	0.8208	(0.0023)	0.6821	(0.0022)
ρ	-0.0869	(0.0006)	0.2675	(0.0008)

Table 3: Production Function Estimates: Post 1995

	OLS	Std Err	IV	Std Err
Constant	1.70	(0.0087)	1.37	(0.0300)
Slope	0.95	(0.0073)	1.36	(0.0382)
R^2	0.82		0.66	
α	0.8581	(0.1029)	0.7333	(0.0357)
ρ	-0.0556	(0.0026)	0.2636	(0.0087)

Table 4: Amenity Regressions

	Full Sample		Post-1995	
Amenity	Coefficient	Std Err	Coefficient	Std Err
Travel Time	-0.01603	0.00019	-0.00411	0.00037
Brashear	0.09606	0.00188	0.23545	0.01998
Oliver	0.01325	0.00209	0.11045	0.02768
Schenley	0.07471	0.00243	0.15881	0.04155
Langley	0.01300	0.00243	-0.09123	0.04572
Peabody	0.21198	0.00216	0.29649	0.03027
Allderdice	0.22537	0.00204	0.15307	0.04061
Westinghouse	0.06711	0.00291	0.02414	0.04061
Constant	-0.00846	0.00265	-0.25800	0.01163
Adj-R^2	0.2123		0.1023	
Obs.	91,767		6,362	

Carrick HS is the omitted school.

Table 5: Local Employment Center Regressions

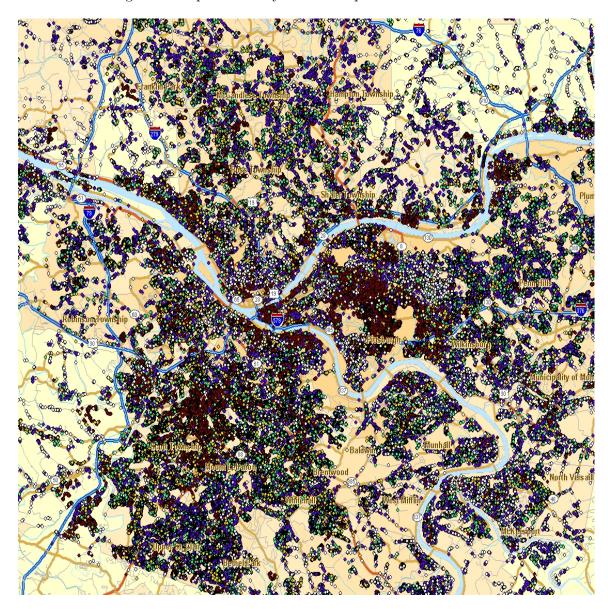| Radius	Dummy Value	Std Err	Adj-R^2
0.1	0.09639	0.00524	0.2152
0.2	0.07841	0.00252	0.2206
0.3	0.03230	0.00202	0.2145
0.4	0.03519	0.00198	0.2150
0.5	0.04106	0.00189	0.2164
0.75	-0.00168	0.00181	0.2123
1.0	-0.01504	0.00216	0.2127

Figure 1: Map of Amenity Values Properties in the Data Set