CHAPTER 2

Structural Estimation in Urban Economics

Thomas J. Holmes[*], Holger Sieg[†]
[*]University of Minnesota and Federal Reserve Bank of Minneapolis, Minneapolis, MN, USA
[†]University of Pennsylvania, Philadelphia, PA, USA

Contents

Handbook of Regional and Urban Economics, Volume 5A
ISSN 1574-0080, http://dx.doi.org/10.1016/B978-0-444-59517-1.00002-7

Abstract

Structural estimation is a methodological approach in empirical economics explicitly based on economic theory, in which economic modeling, estimation, and empirical analysis are required to be internally consistent. This chapter illustrates the structural approach with three applications in urban economics: (1) discrete location choice, (2) fiscal competition and local public good provision, and (3) regional specialization. For each application, we first discuss broad methodological principles of model selection and development. Next we treat issues of identification and estimation. The final step of each discussion is how estimated structural models can be used for policy analysis.

Keywords

Structural estimation, Fiscal competition, Public good provision, Regional specialization

JEL Classification Codes

R10, R23, R51

2.1. AN INTRODUCTION TO STRUCTURAL ESTIMATION

Structural estimation is a methodological approach in empirical economics explicitly based on economic theory. A requirement of structural estimation is that economic modeling, estimation, and empirical analysis be internally consistent. Structural estimation can also be defined as theory-based estimation: the objective of the exercise is to estimate an explicitly specified economic model that is broadly consistent with observed data. Structural estimation, therefore, differs from other estimation approaches that are either based on purely statistical models or based only implicitly on economic theory.[1] A structural estimation exercise typically consists of the following three steps: (1) model selection and development, (2) identification and estimation, and (3) policy analysis. We discuss each step in detail and then provide some applications to illustrate the key methodological issues that are encountered in the analysis.

[1] For example, the most prominent approach in program evaluation is based on work by Neyman (1923) and Fisher (1935), who suggested evaluating the impact of a program by using potential outcomes that reflect differences in treatment status. The objective of the exercise, then, is typically to estimate average treatment effects. This is a purely statistical model, which is sufficiently flexible such that it has broad applications in many sciences.

2.1.1 Model selection and development

The first step in a structural estimation exercise is the development or selection of an economic model. These models can be simple static decision models under perfect information or complicated nonstationary dynamic equilibrium models with asymmetric information.

It is important to recognize that a model that is suitable for structural estimation needs to satisfy requirements that are not necessarily the same requirements that a theorist would typically find desirable. Most theorists will be satisfied if an economic model captures the key ideas that need to be formalized. In structural estimation, we search for models that help us understand the real world and are consistent with observed outcomes. As a consequence, we need models that are not rigid, but are sufficiently flexible to fit the observed data. Flexibility is not necessarily a desirable property for a theorist, especially if the objective is to analytically characterize the properties of a model.

Theorists are typically reluctant to work with parameterized versions of their model, since they aim for generality. An existence proof is, for example, considered to be of limited usefulness by most theorists if it crucially depends on functional form assumptions. Flexible economic models often have the property that equilibria can only be computed numerically—that is, there are no analytical solutions. Numerical computations of equilibria require a fully parameterized and numerically specified model. The parametric approach is, therefore, natural to structural modeling in microeconomics as well as to much of modern quantitative macroeconomics. Key questions, then, are how to determine the parameter values and whether the model is broadly consistent with observed outcomes. Structural estimation provides the most compelling approach to determine plausible parameter values for a large class of models and to evaluate the fit of the model.

2.1.2 Identification and estimation

Structural estimation also requires that we incorporate a proper error structure into the economic model. Since theory and estimation must be internally consistent, the model under consideration needs to generate a well-specified statistical model.[2] Any economic model is, by definition, an abstraction of the real world. As a consequence, it cannot be an exact representation of the "true" data-generating process. This criticism is not specific to structural estimation, since it also applies to any purely statistical modeling and estimation approach. We are interested in finding economic models that, in the best-case scenario, cannot be rejected by the data using conventional statistical hypothesis or specification tests. Of course, models that are rejected by the data can also be very helpful and improve our knowledge. These models can provide us with guidance on how to improve our modeling approach, generating a better understanding of the research questions that we investigate.

[2] Notice that this is another requirement that is irrelevant from a theorist's perspective.

A standard approach for estimating structural models requires the researcher to compute the optimal decision rules or the equilibrium of a model to evaluate the relevant objective function of an extremum estimator. It is a full-solution approach, since the entire model is completely specified on the computer. In many applications, it is not possible to use canned statistical routines to do this. Rather, the standard approach involves programming an economic model, though various procedures and routines can be pulled off the shelf to use in solving the model.[3] The step of obtaining a solution of an economic model for a given set of parameters is called the "inner loop" and often involves a fixed point calculation (i.e., taking as given a vector of endogenous variables, agents in the model make choices that result in the same vector of endogenous variables, satisfying the equilibrium conditions). There is also an "outer loop" step in which the parameter vector is varied and a maximization problem is solved to obtain the parameter vector that best fits the data according to a given criterion. The outer/inner loop approach is often called a "nested fixed point" algorithm.

Whenever we use nested fixed point algorithms, the existence and uniqueness of equilibrium are potentially important aspects of the analysis. Uniqueness of equilibrium is not a general property of most economic models, especially those that are sufficiently flexible to be suitable for structural estimation. Moreover, proving uniqueness of equilibrium can be rather challenging.[4] Nonuniqueness of equilibrium can cause a number of well-known problems during estimation and counterfactual comparative static analysis. Sometimes we may want to condition on certain observed features of the equilibrium and only impose a subset of the equilibrium conditions. By conditioning on observed outcomes, we often circumvent a potential multiplicity of equilibria problems.

Another potential drawback of the full-solution estimation approach is that it is computationally intensive. We are likely to hit the feasibility constraints quickly because of the well-known curses of dimensionality that are encountered, for example, in dynamic programming.[5]

It is, therefore, often desirable to derive estimation approaches that do not rely on full-solution approaches. Often we can identify and estimate the parameters of a model using necessary conditions of equilibrium, which can take the form of first-order conditions, inequality constraints, or boundary indifference conditions. We call these "partial solution" approaches.[6] These approaches are often more elegant than brute force

[3] A useful reference for algorithms to solve economic models is Judd (1998). Another standard reference for numerical recipes in C programming is Press et al. (1988).

[4] For example, the only general uniqueness proofs that we have for the Arrow–Debreu model rely on high-level assumptions about the properties of the excess demand function.

[5] See Rust (1994) for a discussion of computational complexity within the context of dynamic discrete choice models.

[6] Some of the most compelling early applications of partial solution methods in structural estimation are those of Heckman and MaCurdy (1980) and Hansen and Singleton (1982). See Holmes (2011) for a recent example of an application of an inequality constraint approach used to estimate economies of density.

approaches, but they are more difficult to derive, since they typically exploit specific idiosyncratic features of the model. Finding these approaches requires a fair bit of creativity.

A parametric approach is not necessary for identification or estimation. It can be useful to ask the question whether our model can be identified under weak functional form assumptions. Those approaches, then, typically lead us to consider nonparametric or semiparametric approaches for identification or estimation. Notice that identification and estimation largely depend on the available data—that is, the information set of the econometrician. Thus, identification and estimation are closely linked to the data collection decisions made by the researchers.

Once we have derived and implemented an estimation procedure, we need to determine whether our model fits the data. Goodness of fit can be evaluated on the basis of moments used in estimation or moments that are not used in estimation. We would also like to validate our model—that is, we would like to use some formal testing procedures to determine whether our model is consistent with the data and not seriously misspecified. A number of approaches have been proposed in the literature. First, we can use specification tests that are typically based on overidentifying conditions. Second, we can evaluate our model on the basis of out-of-sample predictions. The key idea is to determine whether our model can predict the observed outcomes in a holdout sample. Finally, we sometimes have access to experimental data that may allow us to identify certain treatment or causal effects. We can then study whether our theoretical model generates treatment effects that are of similar magnitude.[7]

2.1.3 Policy analysis

The third and final step of a structural estimation exercise consists of policy analysis. Here, the objective is to answer the policy questions that motivated the empirical analysis. We can conduct retrospective or prospective policy analysis.

Retrospective analysis evaluates an intervention that happened in the past and is observed in the sample period. One key objective is to estimate treatment effects that are associated with the observed policy intervention. Not surprisingly, structural approaches compete with nonstructural approaches. As pointed out by Lucas (1976), there are some compelling reasons for evaluating a policy change within an internally consistent framework. The structural approach is particularly helpful if we are interested in nonmarginal or general equilibrium effects of policies.

Prospective analysis focuses on new policies that have not been enacted. Again, evaluating the likely impact of alternative policies within a well-defined and internally consistent theoretical framework has some obvious advantages. Given that large-scale

[7] Different strategies for model validation are discussed in detail in Keane and Wolpin (1997) and Todd and Wolpin (2006).

experimental evaluations of alternative policies are typically expensive or not feasible in urban economics, the structural approach is the most compelling one in which to conduct prospective policy analysis.

2.1.4 Applications

Having provided an overview of the structural approach, we now turn to the issue of applying these methods in urban and regional economics. We focus on three examples that we use to illustrate broad methodological principles. Given our focus on methodology, we acknowledge that we are not able to provide a comprehensive review of various articles in the field that take a structural estimation approach.[8] Our first application is location choice. This is a classic issue, one that was addressed in early applications of McFadden's Nobel Prize–winning work on discrete choice (McFadden, 1978). As noted earlier, structural estimation projects typically require researchers to write original code. The literature on discrete choice is well developed, practitioner's guides are published, and reliable computer code is available on the Web.

Our second application considers the literature on fiscal competition and local public good provision. One of the key functions of cities and municipalities is to provide important public goods and services such as primary and secondary education, protection from crime, and infrastructure. Households are mobile and make locational decisions based, at least in part, on differences in public goods, services, and local amenities. This analysis combines the demand side of household location choice with the supply side of what governments offer. Since the focus is on positive analysis, political economy models are used to model the behavior of local governments. In this literature, one generally does not find much in the way of canned software, but we provide an overview of the basic steps for working in this area.

The third application considers recent articles related to the allocation of economic activity across space, including the Ahlfeldt et al. (2014) analysis of the internal structure of the city of Berlin and the Holmes and Stevens (2014) analysis of specialization by industry of regions in the United States. We use the discussion to highlight (1) the development of the models, (2) identification and the basic procedure for estimation, and (3) how the models can be used for policy analysis.

2.2. REVEALED PREFERENCE MODELS OF RESIDENTIAL CHOICE

A natural starting point for a discussion of structural estimation in urban and regional economics is the pioneering work by Daniel McFadden on estimation of discrete choice

[8] For example, we do not discuss a number of articles that are squarely in the structural tradition, such as those of Holmes (2005), Gould (2007), Baum-Snow and Pavan (2012), Kennan and Walker (2011), or Combes et al. (2012).

models. One of the main applications that motivated the development of these methods was residential or locational choice. In this section, we briefly review the now classic results from McFadden and discuss why urban economists are still struggling with some of the same problems that McFadden studied in the early 1970s.

The decision-theoretical framework that underlies modern discrete choice models is fairly straightforward. We consider a household i that needs to choose among different neighborhoods that are indexed by j. Within each neighborhood there are a finite number of different housing types indexed by k. A basic random utility model assumes that the indirect utility of household i for community j and house k is given by

$$u_{ijk} = x_j'\beta + z_k'\gamma + \alpha(y_i - p_{jk}) + \epsilon_{ijk}, \tag{2.1}$$

where x_j is a vector of observed characteristics of community j, z_k is a vector of observed housing characteristics, y_i is household income, and p_{jk} is the price of housing type k in community j. Each household chooses the neighborhood-housing pair that maximizes utility. One key implication of the behavioral model is that households make deterministic choices—that is, for each household there exists a unique neighborhood-house combination that maximizes utility.

McFadden (1974) showed how to generate a well-defined econometric model that is internally consistent with the economy theory described above. Two assumptions are particularly noteworthy. First, we need to assume that there is a difference in information sets between households and econometricians. Although households observe all key variables, including the error terms (ϵ_{ijk}), econometricians observe only x_j, z_k, y_i, and p_{jk}, and a set of indicators, denoted by d_{ijk}, where $d_{ijk} = 1$ if household i chooses neighborhood j and house type k and $d_{ijk} = 0$ otherwise. Integrating out the unobserved error terms then gives rise to well-behaved conditional choice probabilities that provide the key ingredient for a maximum likelihood estimator of the parameters of the model.

Second, if the error terms are independent and identically distributed across i, j, and k and follow a type I extreme value distribution, we obtain the well-known conditional logit choice probabilities:

$$\Pr\{d_{ijk} = 1 | x, z, p, y_i\} = \frac{\exp\{x_j'\beta + z_k'\gamma + \alpha(y_i - p_{jk})\}}{\sum_{n=1}^{J}\sum_{m=1}^{K}\exp\{x_n'\beta + z_m'\gamma + \alpha(y_i - p_{nm})\}}. \tag{2.2}$$

A key advantage of the simple logit model is that conditional choice probabilities have a closed-form solution. The only problem encountered in estimation is that the likelihood function is nonlinear in its parameters. The estimates must be computed numerically. All standard software packages will allow researchers to do that. Standard errors can be computed using the standard formula for maximum likelihood estimators.

One unattractive property of the logit model is the independence of irrelevant alternatives property. It basically says that the ratio of conditional choice probabilities of two products depends only on the relative utility of those two products. Another (related) unattractive property of the simple logit model is that it generates fairly implausible substitution patterns for the aggregate demand. Own and cross-price elasticities are primarily functions of a single parameter (α) and are largely driven by the market shares and not by the proximity of two products in the characteristic space.

One way to solve this problem is to relax the assumption that idiosyncratic tastes are independent across locations and houses. McFadden (1978) suggested modeling the distribution of the error terms as a generalized extreme value distribution, which then gives rise to the nested logit model. In our application, we may want to assume that idiosyncratic shocks of houses within a given neighborhood are correlated owing to some unobserved joint neighborhood characteristics. A main advantage of the nested logit model is that conditional choice probabilities still have closed-form solutions, and estimation can proceed within a standard parametric maximum likelihood framework. Again, most major software packages will have a routine for nested logit models. Hence, few technical problems are involved in implementing this estimator and computing standard errors. The main drawback of the nested logit is that the researcher has to choose the nesting structure before estimation. As a consequence, we need to have strong beliefs about which pairs of neighborhood-house choices are most likely to be close substitutes. We, therefore, need to have detailed knowledge of the neighborhood structure within the city that we study in a given application.

An alternative approach, one that avoids the need to impose a substitution structure prior to estimation and can still generate realistic substitution patterns, is based on random coefficients.[9] Assume now that the utility function is given by

$${}_{ijk} = x_j'\beta_i + z_k'\gamma_i + \alpha_i(y_i - p_{jk}) + \epsilon_{ijk}, \tag{2.3}$$

where γ_i, β_i, and α_i are random coefficients. A popular approach is based on the assumption that these random coefficients are normally distributed. It is fairly straightforward to show that substitutability in the random coefficient logit model is driven by observed housing and neighborhood characteristics. Households that share similar values of random coefficients will substitute between neighborhood-housing pairs that have similar observed characteristics.

A key drawback of the random coefficient model is that the conditional choice probabilities no longer have closed-form solutions and must be computed numerically. This process can be particularly difficult if there are many observed characteristics, and hence high-dimensional integrals need to be evaluated. These challenges partially led to the development of simulation-based estimators (see Newey and McFadden, 1994 for some basic results on consistency and asymptotic normality of simulated maximum likelihood estimators). As discussed, for example, in Judd (1998), a variety of numerical algorithms have been developed that allow researchers to solve these integration

[9] For a detailed discussion, see, for example, Train (2003).

problems. A notable application of these methods is that of Hastings et al. (2006), who study sorting of households among schools within the Mecklenburg Charlotte school district. They evaluate the impact of open enrollment policies under a particular parent choice mechanism.[10]

Demand estimation has also focused on the role of unobserved product characteristics (Berry, 1994). In the context of our application, unobserved characteristics may arise at the neighborhood level or the housing level. Consider the case of an unobserved neighborhood characteristic. The econometrician probably does not know which neighborhoods are popular. More substantially, our measures of neighborhood or housing quality (or both) may be rather poor or incomplete. Let ξ_j denote an unobserved characteristic that captures aspects of neighborhood quality that are not well measured by the researcher. Utility can now be represented by the following equation:

$$u_{ijk} = x_j'\beta_i + z_k'\gamma_i + \alpha_i(y_i - p_{jk}) + \xi_j + \epsilon_{ijk}. \tag{2.4}$$

This locational choice model is then almost identical in mathematical structure to the demand model estimated in Berry et al. (1995). The key insight of that article is that the unobserved product characteristics can be recovered by matching the observed market shares of each product. The remaining parameters of the model can be estimated by using a generalized method of moments estimator that uses instrumental variables to deal with the correlation between housing prices and unobserved neighborhood characteristics. Notice that the Berry–Levinsohn–Pakes estimator is a nested fixed point estimator. The inner loop inverts the market share equations to compute the unobserved product characteristics. The outer loop evaluates the relevant moment conditions and searches over the parameter space.

Estimating this class of models initially required some serious investment in programming, since standard software packages did not contain modules for this class of models. Now, however, both a useful practitioner's guide (Nevo, 2000) and a variety of programs are available and openly shared. This change illustrates an important aspect of structural estimation. Although structural estimation may require some serious initial methodological innovations, subsequent users of these techniques often find it much easier to modify and implement these techniques.[11] Notable articles that introduced this empirical approach to urban economics are those of Bayer (2001), Bayer et al. (2004), and Bayer et al. (2007), who estimate models of household sorting in the Bay Area.

[10] Bayesian estimators can also be particularly well suited for estimating discrete choice models with random coefficients. Bajari and Kahn (2005) adopt these methods to study racial sorting and peer effects within a similar framework.

[11] Computation of standard errors is also nontrivial, as discussed in Berry et al. (2004). Most applied researchers prefer to bootstrap standard errors in these models.

Extending these models to deal with the endogenous neighborhood characteristics or peer effects is not trivial. For example, part of the attractiveness of a neighborhood may be driven by the characteristics of neighbors. Households may value living, for example, in neighborhoods with a large fraction of higher-income households because of the positive externalities that these families may provide. Three additional challenges arise in these models. First, peer effects need to be consistent with the conditional choice probabilities and the implied equilibrium sorting. Second, endogenous peer effects may give rise to multiplicity of equilibria, which creates additional problems in computation and estimation. Finally, the standard Berry–Levinsohn–Pakes instrumentation strategy, which uses exogenous characteristics of similar house-neighborhood pairs, is not necessarily feasible anymore, since we are dealing with endogenous neighborhood characteristics that are likely to be correlated with the unobserved characteristics.[12] Finding compelling instruments can be rather challenging. Some promising examples are given by Ferreira (2009), who exploits the impact of property tax limitations (Proposition 13) in California on household sorting. Galliani et al. (2012) exploit random assignment to vouchers to construct instruments in their study of the effectiveness of the Moving to Opportunity housing assistance experiment.

Researchers have also started to incorporate dynamic aspects into the model specification. Locational choices and housing investments are inherently dynamic decisions that affect multiple time periods. As a consequence, adopting a dynamic framework involves some inherent gains. In principle, we can follow Rust (1987), but adopting a dynamic version of the logit model within the context of locational choice is rather challenging. Consider the recent article by Murphy (2013), who estimates a dynamic discrete choice model of land conversion using data from the Bay Area. One key problem is measuring prices for land (and housing). In a dynamic model, households must also forecast the evolution of future land and housing prices to determine whether developing a piece of land is optimal. That creates two additional problems. First, we need to characterize price expectations based on simple time series models. Second, we need one pricing equation for each location (assuming land or housing (or both) within a neighborhood is homogeneous), which potentially blows up the dimensionality of state space associated with the dynamic programming problem.[13] Some user guides are available for estimating dynamic discrete choice models, most notably the chapter by Rust (1994). Estimation and inference is fairly straightforward as long as one stays within the parametric maximum likelihood framework.

[12] Bayer and Timmins (2005) and Bayer et al. (2007) provide a detailed discussion of these issues in the context of the random utility model above. See also the survey articles on peer effects and sorting in this handbook. Epple et al. (2014) estimate a game of managing school district capacity, in which school quality is largely defined by peer effects.

[13] Other promising examples of dynamic empirical approaches are those of Bishop (2011), who adopts a Hotz–Miller conditional choice probabilities estimator, and Bayer et al. (2012). Yoon (2012) studies locational sorting in regional labor markets, adopting a dynamic nonstationary model.

Thanks to the requirement to disclose estimation codes by a variety of journals, some software programs are also available that can be used to understand the basic structure of the estimation algorithms. However, each estimation exercise requires some coding.

Finally, researchers have worked on estimating discrete choice models when there is rationing in housing markets. Geyer and Sieg (2013) develop and estimate a discrete choice model that captures excess demand in the market for public housing. The key issue is that simple discrete choice models give rise to biased estimators if households are subject to rationing and, thus, do not have full access to all elements in the choice set. The idea of that article is to use a fully specified equilibrium model of supply and demand to capture the rationing mechanism and characterize the endogenous (potentially latent) choice set of households. Again, we have to use a nested fixed point algorithm to estimate these types of models. The key finding of this chapter is that accounting for rationing implies much higher welfare benefits associated with public housing communities than simple discrete choice estimators that ignore rationing.

2.3. FISCAL COMPETITION AND PUBLIC GOOD PROVISION

We next turn to the literature on fiscal competition and local public good provision. As noted above, one key function of cities and municipalities is to provide important public goods and services. Households are mobile and make locational decisions based on differences in public goods, services, and local amenities. The models developed in the literature combine the demand side of household location choice, which are similar to the ones studied in the previous section, with political economy models that are used to model the behavior of local governments.

We start Section 2.3.1 by outlining a generic model of fiscal competition that provides the basic framework for much of the empirical work in the literature. We develop the key parts of the model and define equilibrium. We also discuss existence and uniqueness of equilibrium and discuss key properties of these models. We finish by discussing how to numerically compute equilibria for more complicated specifications of the model, and we discuss useful extensions.

In Section 2.3.2, we turn to an empirical issue. We start by broadly characterizing the key predictions of this class of models and then develop a multistep approach that can be used to identify and estimate the parameters of the model. We finish this section by discussing alternative estimators that rely less on functional form assumptions.

In Section 2.3.3, we turn to policy analysis. We consider two examples. The first example considers the problem of estimating the willingness to pay for improving air quality in Los Angeles. We discuss how to construct partial and general equilibrium measures that are consistent with the basic model developed above. Our second application considers the potential benefits of decentralization and compares decentralized with centralized outcomes within a general equilibrium model.

2.3.1 Theory

The starting point of any structural estimation exercise is a theoretical model that allows us to address key research questions. In this application, we consider fiscal competition and public good provision within a system of local jurisdictions.[14] This literature blends the literature on demand for public goods and residential choice with the literature on political economy models of local governments that characterize the supply of public goods and services.

2.3.1.1 Preferences and heterogeneity

We consider an urban or metropolitan area that consists of J communities, each of which has fixed boundaries. Each community has a local housing market, provides a (congestable) public good g, and charges property taxes, t. There is a continuum of households that differ by income, y. Households also differ by tastes for public goods, denoted by α. Note that unobserved heterogeneity in preferences is a key ingredient in any empirical model that must be consistent with observed household choices, since households that have the same observed characteristics typically do not make the same decisions.

Households behave as price takers and have preferences defined over a local public good, housing services, h, and a composite private good, b. Households maximize utility with respect to their budget constraint:

$$\begin{aligned} &\max_{(h,b)} \; U(\alpha, g, h, b) \\ &\text{s.t.} \;\; (1+t)\, p^h \, h \;=\; y \;-\; b, \end{aligned} \tag{2.5}$$

which yields housing demand functions $h(p,\ y;\ \alpha,\ g)$. The corresponding indirect utility function is given by

$$V(\alpha, g, p, y) \;=\; U(\alpha, g, h(p, y, \alpha), y - ph(p, y, \alpha, g)), \tag{2.6}$$

where $p = (1 + t)p^h$. Consider the slope of an indirect indifference curve in the $(g,\ p)$-plane:

$$M(\alpha, g, p, y) = -\,\frac{\partial V(\alpha, g, p, y)/\partial g}{\partial V(\alpha, g, p, y)/\partial p}. \tag{2.7}$$

If $M(\cdot)$ is monotonic in y for given α, then indifference curves in the $(g,\ p)$-plane satisfy the single-crossing property. Likewise, monotonicity of $M(\cdot)$ in α provides a single crossing for given y. As we will see below, the single-crossing properties are key to characterizing both the sorting and the voting behavior of households. One challenge encountered in structural

[14] Our theoretical model builds on previous work by Ellickson (1973), Westhoff (1977), Epple et al. (1984), Goodspeed (1989), Epple and Romer (1991), Nechyba (1997), Fernandez and Rogerson (1996), Benabou (1996a,b), Durlauf (1996), Fernandez and Rogerson (1998), Epple and Platt (1998), Glomm and Lagunoff (1999), Henderson and Thisse (2001), Benabou (2002), Rothstein (2006), and Ortalo-Magne and Rady (2006).

estimation is to find a flexible parameterization of the model that is not overly restrictive.[15] A promising parameterization of the indirect utility function is given below:

$$V(g,p,y,\alpha) = \left\{ \alpha g^{\rho} + \left(e^{\frac{y^{1-\nu}-1}{1-\nu}} \, e^{-\frac{Bp^{\eta+1}-1}{1+\eta}} \right)^{\rho} \right\}^{1/\rho}, \tag{2.8}$$

where α is the relative weight that a household assigns to the public goods. Roy's identity implies that the housing demand function is given by

$$h = B\, p^{\eta}\, y^{\nu}. \tag{2.9}$$

Note that η is the price elasticity of housing and ν is the income elasticity. This demand function is a useful characterization of the demand, since it does not impose unitary income or price elasticities.[16] Note that this utility function satisfies the single-crossing property if $\rho < 0$.

2.3.1.2 Household sorting

One objective of the model is to explain household sorting among the set of communities. There are no mobility costs, and hence households choose j to maximize

$$\max_{j} \; V(\alpha, g_j, p_j, y). \tag{2.10}$$

Define the set C_j to be the set of households living in community j:

$$C_j = \{(\alpha, y) | V(\alpha, g_j, p_j, y) \geq \max_{i \neq j} \; V(\alpha, g_i, p_i, y)\}. \tag{2.11}$$

Figure 2.1 illustrates the resulting sorting in the (p, g)-space. It considers the case of three communities denoted by $j-1$, j, and $j+1$. It plots the indifference curve of a household that is indifferent between $j-1$ and j, denoted by $y_{j-1}(\alpha)$. Similarly, it plots the indifference curve of a household that is indifferent between j and $j+1$, denoted by $y_j(\alpha)$. Note that for a given level of α, the household that is indifferent between j and $j+1$ must have higher income than the household that is indifferent between $j-1$ and j, and as a consequence, we have $y_j(\alpha) > y_{j-1}(\alpha)$. Single crossing then implies that the household with higher income levels must have steeper indifference curves than the household with lower income levels. Finally, Figure 2.1 also plots the indifference curve of a household with income given by $y_j(\alpha) > y > y_{j-1}(\alpha)$. This household will strictly prefer to live in community j.

[15] We will discuss nonparametric or semiparametric identification below.

[16] To avoid stochastic singularities, we can easily extend the framework and assume that the housing demand or expenditures are subject to an idiosyncratic error that is revealed to households after they have chosen the neighborhood. This error term thus enters the housing demand, but does not affect the neighborhood choice. Alternatively, we can assume in estimation that observed housing demand is subject to measurement error. We follow that approach in our application.

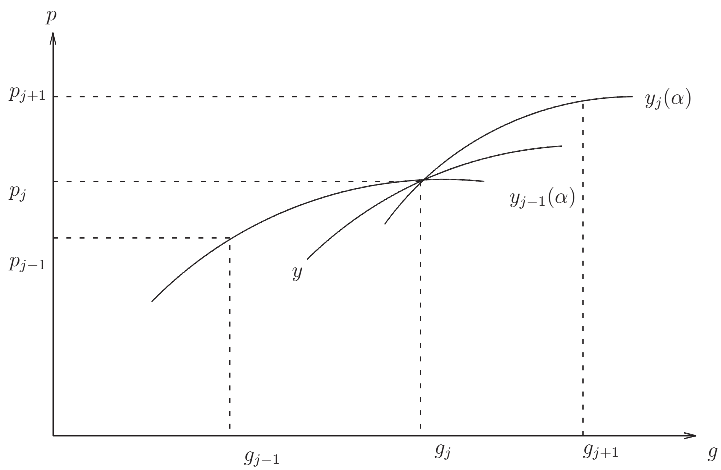

Figure 2.1 Sorting in the (p, g)-space.

Alternatively, we can characterize household sorting by deriving the boundary indifference loci $\alpha_j(y)$, which are defined as

$$V(\alpha_j(y), g_j, p_j, y) = V(\alpha_j(y), g_{j+1}, p_{j+1}, y), \tag{2.12}$$

and are hence the inverse of $y_j(\alpha)$. Given our parameterization, these boundary indifference conditions can be written as

$$\ln\alpha - \rho\left(\frac{y^{1-\nu}-1}{1-\nu}\right) = \ln\left(\frac{Q_{j+1}-Q_j}{g_j^{\rho}-g_{j+1}^{\rho}}\right) \equiv K_j, \tag{2.13}$$

where

$$Q_j = e^{-\frac{\rho}{1+\eta}\left(Bp_j^{\eta+1}-1\right)}. \tag{2.14}$$

Figure 2.2 illustrates the resulting sorting of households across communities in equilibrium in the $(\ln y, \ln\alpha)$-space. The loci passing through the K-intercepts characterize the boundary indifference conditions. The loci passing through the L-intercepts characterize the set of decisive voters within each community (as explained in detail below).

2.3.1.3 Community size, housing markets, and budgets

A measure of the size (or market share) of community j is given by

$$n_j = P(C_j) = \int_{C_j} f(\alpha, y)\, dy\, d\alpha. \tag{2.15}$$

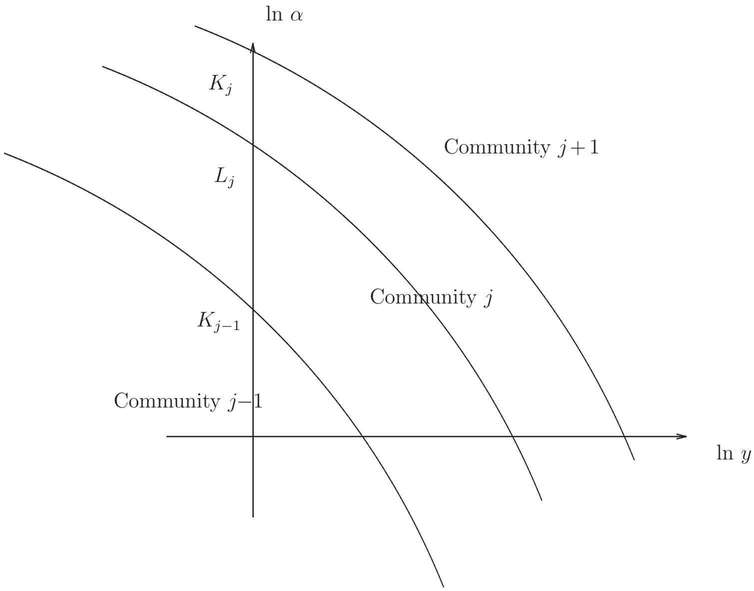

Figure 2.2 The distribution of households across and within communities.

Aggregate housing demand is defined as

$$H_j^d = \int_{C_j} h(p_j, \alpha, \gamma)\ f(\alpha, \gamma)\ d\gamma\ d\alpha. \tag{2.16}$$

Housing is owned by absentee landlords, and the aggregate housing supply in community j depends on the net-of-tax price of housing p_j^h and a measure of the land area of community j denoted by l_j. Hence, we have that

$$H_j^s = H(l_j, p_j^h). \tag{2.17}$$

A commonly used housing supply function is given by $H_j^s = l_j\ [p^h]^\tau$. Note that τ is the price elasticity and l_j is a measure of the availability of land. Housing markets need to clear in equilibrium for each community.

The budget of community j must be balanced. This implies that

$$t_j\ p_j^h \int_{C_j} h(p_j, \alpha, \gamma)\ f(\alpha, \gamma)\ d\gamma\ d\alpha\ /\ P(C_j)\ =\ c(g_j), \tag{2.18}$$

where $c(g)$ is the cost per household of providing g.[17]

Next we endogenize the provision of local public goods, assuming that residents vote on fiscal and tax policies in each community. Fernandez and Rogerson (1996) suggest the following timing assumptions:

1. Households choose a community of residence having perfect foresight of equilibrium prices, taxes, and spending in all communities.

[17] A linear cost function is commonly used in quantitative work—that is, $c(g) = c_0 + c_1 g$.

2. The housing markets clear in all communities.
3. Households vote on feasible tax rates and levels of public goods in each community. Hence, the composition of each community, the net-of-tax price of housing, and the aggregate housing consumption are determined prior to voting. Voters treat the population boundaries of each community and the housing market outcomes as fixed when voting. This timing assumption then implies that the set of feasible policies at the voting stage is given by the following equation:

$$p_j(g) = p_j^h + \frac{c(g_j)}{H_j/P(C_j)}. \tag{2.19}$$

This set is also sometimes called the government-services possibility frontier (GPF) in the literature.

Consider a point (g^*, p^*) on the GPF. We say that (g^*, p^*) is a majority rule equilibrium if there is no other point on the GPF $(\hat{g}, \hat{p})$ that would beat (g^*, p^*) in a pairwise vote.[18]

A voter's preferred level of g is then obtained by maximizing the indirect utility function $V(\alpha, g_j, p_j, y)$ subject to the feasibility constraint derived above. Single crossing implies that for any level of income y, the single-crossing properties imply that households with higher (lower) values of α will have higher (lower) demands for local public goods. As a consequence, there exists a function $\tilde{\alpha}_j(y)$ that characterizes the set of pivotal voters. This function is implicitly defined by the following condition:

$$\int_0^\infty \int_{\alpha_{j-1}(y)}^{\tilde{\alpha}_j(y)} f(\alpha, y)\, d\alpha\, dy = \frac{1}{2} P(C_j). \tag{2.20}$$

Given our parameterization, the locus of decisive voters is given by

$$\ln\alpha - \rho\left(\frac{y^{1-\nu}-1}{1-\nu}\right) = L_j = \ln\left(\frac{B\, e^{-\rho \frac{Bp_j^{\eta+1}-1}{1+\eta}}\, p_j^{\eta}\, p_j'(g)}{g_j^{\rho-1}}\right). \tag{2.21}$$

See Figure 2.2 for an illustration of this locus.

2.3.1.4 *Equilibrium*

Definition 2.1

An *intercommunity equilibrium* consists of a set of communities, $\{1, \ldots, J\}$; a continuum of households, C; a distribution, P, of household characteristics α and y; and a partition of C across communities $\{C_1, \ldots, C_J\}$, such that every community has a positive population—that is, $0 < n_j^* < 1$; a vector of prices and taxes, $(p_1^*, t_1^*, \ldots, p_J^*, t_J^*)$; an

[18] Note that in this model, sincere voting is a dominant strategy.

allocation of public good expenditures, $(g_1^*, \ldots, g_J^*)$; and an allocation, (h^*, b^*), for every household (α, y), such that the following hold:

1. Every household, (α, y), living in community j maximizes its utility subject to the budget constraint[19]

$$(h^*, b^*) = \arg\max_{(h,b)} U(\alpha, g_j^*, h, b)$$
$$\text{s.t. } p_j^* h = y - b.$$

2. Each household lives in one community and no household wants to move to a different community—that is, for a household living in community j, the following holds:

$$V(\alpha, g_j^*, p_j^*, y) \geq \max_{i \neq j} V(\alpha, g_i^*, p_i^*, y). \tag{2.22}$$

3. The housing market clears in every community:

$$\int_{C_j} h^*(p_j^*, y, \alpha) f(\alpha, y) \, dy \, d\alpha = H_j^s\left(\frac{p_j^*}{1 + t_j^*}\right). \tag{2.23}$$

4. The population of each community, j, is given by

$$n_j^* = P(C_j^*) = \int_{C_j} f(\alpha, y) \, dy \, d\alpha. \tag{2.24}$$

5. The budget of every community is balanced:

$$\frac{t_j^*}{1 + t_j^*} p_j^* \int_{C_j} h^*(p_j^*, y, \alpha) f(\alpha, y) \, dy \, d\alpha \, / \, n_j = c(g_j^*). \tag{2.25}$$

6. There is a voting equilibrium in each community: Over all levels of (g_j, t_j) that are perceived to be feasible allocations by the voters in community j, at least half of the voters prefer (g_j^*, t_j^*) over any other feasible (g_j, t_j).

Existence of equilibrium can be shown under a number of regularity conditions discussed in Epple et al. (1984, 1993). In general, there are no uniqueness proofs, and there is some scope for nonuniqueness in these types of models. Multiple equilibria can arise, since it is possible that different endogenous levels of public good provision are consistent with optimal household decisions and market clearing conditions. As a consequence, these equilibria will have different endogenous housing prices and sorting patterns across communities. However, Calabrese et al. (2006) prove that there can only be one equilibrium that is consistent with a given distribution of community sizes and community ranking; that is, different equilibria will result in different size distributions and (p, g) orderings.

[19] Strictly speaking, all statements only have to hold for almost every household; deviations of behavior of sets of households with measure zero are possible.

2.3.1.5 Properties of equilibrium

Given that we have defined an equilibrium for our model, it is desirable to characterize the properties of equilibria. From the perspective of structural estimation, these properties are interesting, since they provide (a) some predictions that can potentially be tested and (b) necessary conditions that can be exploited to form orthogonality conditions for an estimator.[20]

Epple and Platt (1998) show that for an allocation to be a locational equilibrium, there must be an ordering of community pairs, $\{(g_1, p_1), \ldots, (g_J, p_J)\}$, such that we have the following:

1. Boundary indifference. The set of border individuals are indifferent between the two communities: $I_j = \left\{(\alpha, y) \mid V(\alpha, g_j, p_j, y) = V(\alpha, g_{j+1}, p_{j+1}, y)\right\}$.
2. Stratification. Let $y_j(\alpha)$ be the implicit function defined by the equation above. Then, for each α, the residents of community j consist of those with income, y, given by $y_{j-1}(\alpha) < y < y_j(\alpha)$.
3. Increasing bundles. Consider two communities i and j such that $p_i > p_j$. Then, $g_i > g_j$ if and only if $y_i(\alpha) > y_j(\alpha)$.
4. Majority voting equilibrium exists for each community and is unique.
5. The equilibrium is the preferred choice of households (y, α) on the downward-sloping locus $\tilde{y}_j(\alpha)$ satisfying $\int_{\underline{\alpha}}^{\overline{\alpha}} \int_{\underline{y}}^{\tilde{y}_j(\alpha)} f_j(y, \alpha)\, dy\, d\alpha = 0.5 P(C_j)$.
6. Households living in community j with (y, α) to the northeast (southwest) of the $\tilde{y}_j(\alpha)$ locus in the (α, y)-plane prefer a tax that is higher (lower) than the equilibrium.

We will show below how to exploit these properties to estimate the parameters of the model.

2.3.1.6 Computation of equilibrium

Since equilibria can only be computed numerically, we need an algorithm to do so. Note that an equilibrium is characterized by a vector $(t_j, p_j, g_j)_{j=1}^{J}$. To compute an equilibrium, we need to solve a system of $J \times 3$ nonlinear equations: budget constraints, housing market equilibria, and voting conditions. We also need to check second order conditions once we have found a solution to the system of equations.

Computing equilibria is essential to conducting counterfactual policy analysis, especially if we have strong reasons to believe that policy changes can have substantial general equilibrium effects. It is also important if we want to use a nested fixed point approach to estimation. We will discuss these issues in the next sections in detail.

2.3.1.7 Extensions

Peer effects and private schools

Calabrese et al. (2006) develop an extended model with peer effects. The quality of local public good provision, denoted by q, depends on expenditures per household, g, and a measure of peer quality, denoted by $\overline{y}$:

[20] We will show in Section 2.3.2 how to use spatial indifference loci and voting loci to construct an estimator for key parameters of the model.

$$q_j = g_j \left(\frac{\overline{y}_j}{\overline{y}}\right)^{\phi}, \tag{2.26}$$

where peer quality can be measured by the mean income in a community,

$$\overline{y}_j = \int_{C_j} y f(\alpha, y)\, dy\, d\alpha \,/\, n_j. \tag{2.27}$$

Ferreyra (2007) also introduced peer effects as well as private school competition within a model with a fixed housing stock to study the effectiveness of different school voucher programs.

Amenities and heterogeneity

One key drawback of the model above is that it assumes that households only sort on the basis of local public good provisions. It is possible to account for exogenous variation in amenities without having to change the structure of the model, as discussed in Epple et al. (2010a). Allowing for more than one endogenous public good is difficult, however, because it is hard to establish the existence of voting equilibrium when voting over multidimensional policies. As a consequence, the empirical literature in fiscal competition has primarily considered the model discussed above.

Dynamics

Benabou (1996b), Benabou (2002), and Fernandez and Rogerson (1998) reinterpret the model above using an overlapping generations approach to study fiscal competition. In their models, young individuals do not make any decisions. Hence, individuals make decisions only at one point in time. Epple et al. (2012) then extend the approach and develop an overlapping generations model in which individuals make decisions at different points during the life cycle. This model captures the differences in preferred policies over the life cycle and can be used to study the intergenerational conflict over the provision of public education. This conflict arises because the incentives of older households without children to support the provision of high-quality educational services in a community are weaker than the incentives of younger households with school-age children.

Epple et al. show that the observed inequality in educational policies across communities not only is the outcome of stratification by income, but also is determined by the stratification by age and a political process that is dominated by older voters in many urban communities with low-quality educational services. The mobility of older households creates a positive fiscal externality, since it creates a larger tax base per student. This positive tax externality can dominate the negative effects that arise because older households tend to vote for lower educational expenditures. As a consequence, sorting by age can reduce the inequality in educational outcomes that is driven by income sorting.[21]

[21] Only a few studies have analyzed voting in a dynamic model. Coate (2011) models forward-looking behavior in local elections that determine zoning policies. He is able to use a more general approach to voting by adopting an otherwise simpler structure in which there is limited housing choice and heterogeneity and housing prices are determined by construction costs.

2.3.2 Identification and estimation

The second step involved in structural estimation is to devise an estimation strategy for the parameters of the model. At this stage, a helpful approach is to check whether the model that we have written down is broadly consistent with the key stylized facts that we are trying to explain. In the context of this application, we know that community boundaries rarely change (Epple and Romer, 1989). As a consequence, we do not have to deal with the entry or exit of communities. We also know that there is a large amount of variation in housing prices, mean income, expenditures, and property taxes among communities within most US metropolitan areas. Our model seems to be well suited for dealing with those sources of heterogeneity. At the household level, we observe a significant amount of income and housing expenditure heterogeneity both within and across communities. Again, our model is broadly consistent with these stylized facts.

2.3.2.1 The information set of the econometrician

Before we develop an estimation strategy, an essential step is to characterize the information set of the econometrician. Note that this characterization largely depends on the available data sources. If we restrict our attention to publicly available aggregate data, then we can summarize the information set of the econometrician for this application as follows. For all communities in a single metropolitan area, we observe tax rates and expenditures; the marginal distribution of income and community sizes; and a vector of locational amenities, denoted by x. Housing prices are strictly speaking not observed, but can be estimated as discussed in Sieg et al. (2002). Alternatively, they need to be treated as latent.[22]

2.3.2.2 Predictions of the model

Next, it is useful to summarize the key predictions of the model:

1. The model predicts that households will sort by income among the set of communities.
2. The model predicts that household sorting is driven by differences in observed tax and expenditure policies, which are, at least, partially capitalized in housing prices.
3. The model predicts that observed tax and expenditure policies must be consistent with the preferences of the decisive voter in each community.

We need to develop a strategy to test the predictions of the model in an internally consistent way.

[22] Microdata that contain locational identifiers at the local level are available only through census data centers.

2.3.2.3 Household sorting by income

More formally, the model predicts the *distribution of households by income* among the set of communities. Intuitively speaking, we can test this prediction of the model by matching the predicted marginal distribution of income in each community, $f_j(y)$, to the distribution reported in the US census.

To formalize these ideas, recall that the size of community j is given by

$$P(C_j) = \int_{-\infty}^{\infty} \int_{K_{j-1}+\rho\frac{y^{1-\nu}-1}{1-\nu}}^{K_j+\rho\frac{y^{1-\nu}-1}{1-\nu}} f(\ln\alpha, \ln y)\ d\ln\alpha\ d\ln y. \tag{2.28}$$

One key insight that facilitates estimation is that we can (recursively) express the community-specific intercepts, $(K_0, \ldots, K_J)$, as functions of the community sizes, $(P(C_1), \ldots, P(C_J))$, and the parameters of the model:

$$\begin{aligned} &K_0 = -\infty, \\ &K_j = K_j(K_{j-1}, P(C_j) \mid \rho, \mu_y, \sigma_y, \mu_\alpha, \sigma_\alpha, \lambda, \nu), \quad j = 1, \ldots, J-1, \\ &K_J = \infty. \end{aligned} \tag{2.29}$$

The intuition for this result is simple.[23] By definition, $K_0 = -\infty$, which establishes the lower boundary for community 1. As we increase the value of K_1, we push the boundary locus that characterizes the indifference between communities 1 and 2 to the northwest in Figure 2.2. We keep increasing the value of K_1 until the predicted size of the population of community 1 corresponds to the observed population size. This step of the algorithm then determines K_1. To determine K_2, we push the boundary locus that characterizes the indifference between communities 2 and 3 to the northwest by increasing the value of K_2. We continue in this way until all values of K_j have been determined.[24] Finally, note that one could also start with the richest community and work down.

Let q be any given number in the interval (0, 1), and let $\zeta_j(q)$ denote the qth quantile of the income distribution—that is, $\zeta_j(q)$ is defined by $F_j[\zeta_j(q)] = q$. We observe the empirical income distribution for each community. An estimator of $\zeta_j(q)$ is given by

$$\zeta_j^N(q) = F_{j,N}^{-1}(q), \tag{2.30}$$

where $F_{j,N}^{-1}(\cdot)$ is the inverse of the empirical distribution function.

The qth quantile of community j's income distribution predicted by the model is defined by the following equation:

[23] For a formal proof, see Epple and Sieg (1999).

[24] Note that this algorithm is similar to the share inversion algorithm proposed in Berry (1994) for random utility models.

$$\int_{-\infty}^{\ln(\zeta_j(q))} \int_{K_{j-1}+\rho\frac{y^{1-\nu}-1}{1-\nu}}^{K_j+\rho\frac{y^{1-\nu}-1}{1-\nu}} f(\ln\alpha, \ln y)\ d\ln\alpha\, d\ln y = q\, P(C_j). \tag{2.31}$$

Given the parameterization of the model, the income distributions of the J communities are completely specified by the parameters of the distribution function, $(\mu_y, \mu_\alpha, \lambda, \sigma_y, \sigma_\alpha)$, the slope coefficient, ρ, the curvature parameter, ν, and the community-specific intercepts, $(K_0, \ldots, K_J)$.

Epple and Sieg (1999) use estimates of the 25% quantile, the median, and the 75% quantiles. For notational simplicity, we combine the $3 \times J$ restrictions into one vector:

$$e_N(\theta_1) = \left\{ \begin{array}{l} \ln(\zeta_1(0.25,\ \theta_1)) \ - \ \ln(\zeta_1^N(0.25)) \\ \ln(\zeta_1(0.50,\ \theta_1)) \ - \ \ln(\zeta_1^N(0.50)) \\ \ln(\zeta_1(0.75,\ \theta_1)) \ - \ \ln(\zeta_1^N(0.75)) \\ \cdots \\ \ln(\zeta_J(0.25,\ \theta_1)) \ - \ \ln(\zeta_J^N(0.25)) \\ \ln(\zeta_J(0.50,\ \theta_1)) \ - \ \ln(\zeta_J^N(0.50)) \\ \ln(\zeta_J(0.75,\ \theta_1)) \ - \ \ln(\zeta_J^N(0.75)) \end{array} \right\}, \tag{2.32}$$

where θ_1 is the vector of parameters identified at this stage. Epple and Sieg (1999) show that we can identify and estimate only the following parameters at this stage: $\mu_{\ln y}$, $\sigma_{\ln y}$, λ, $\rho/\sigma_{\ln\alpha}$, and ν.

If the model is correctly specified, the difference between the observed and the predicted quantiles will vanish as the number of households in the sample goes to infinity. The estimation is simplified, since the quantiles of the income distribution of community j depend on (p_j, g_j) only through K_j, which can be computed recursively using the observed community sizes. We can, therefore, estimate a subset of the underlying structural parameters of the model using the following minimum distance estimator:

$$\begin{aligned} \theta_1^N \ &= \arg\min_{\theta_1 \in \Theta_1} \ \{e_N(\theta_1)'\ A_N\ e_N(\theta_1)\} \\ \text{s.t. } K_j \ &= \ K_j(K_{j-1}, P(C_j) \mid \theta_1), \ \ j = 1, \ldots, J-1, \end{aligned}$$

where θ_1 is the unknown parameter vector, and A_N is the weighting matrix. This is a standard nonlinear parametric estimator. Standard errors can be computed using the standard formula described in Newey and McFadden (1994). Note that we need the number of households and not necessarily the number of communities to go to infinity in order to compute asymptotic standard errors.

Epple and Sieg (1999) find that the estimates have plausible values and high precision. The overall fit of the income quantiles is quite remarkable, especially given the fact that the model relies on only a small number of parameters. The model specification is rejected using conventional levels of significance. Rejection occurs largely because we cannot match the lower quantiles for the poor communities very well.

Epple et al. (2010c) show that it is possible to nonparametrically identify and estimate the joint distribution of income and tastes for public goods.[25] More important, the analysis in Epple et al. (2010c) shows that the rejection of the model reported in Epple and Sieg (1999) is primarily driven by the parametric log-normality assumptions. If one relaxes this assumption while maintaining all other parametric assumptions made above, one cannot reject the model above solely on the basis of data that characterize community sizes and local income distributions. By construction of the semiparametric estimator developed in Epple et al. (2010c), we obtain a perfect fit of the observed income distribution for each community. We, therefore, conclude that the type of model considered above is fully consistent with the observed income distributions at the community level.

2.3.2.4 Public good provision

The first stage of the estimation yields a set of community-specific intercepts, K_j. Given these intercepts, the levels of public good provision that are consistent with observed sorting by income are given by the following recursive representation:

$$g_j = \left\{ g_1^{\rho} - \sum_{i=2}^{j} (Q_i - Q_{i-1}) \exp(-K_i) \right\}^{1/\rho}. \tag{2.33}$$

To obtain a well-defined econometric model, we need to differentiate between observed and unobserved public good provision. A natural starting point would be to assume that observed public good provision, measured by expenditures per capita, is a noisy measure of the true public good provision.

A slightly more general model specification assumes that the level of public good provision can be expressed as an index that consists of observed characteristics of community j denoted x_j and an unobserved characteristic denoted ϵ_j:

$$g_j = x_j' \gamma + \epsilon_j, \tag{2.34}$$

where γ is a parameter vector to be estimated. The first component of the index $x_j' \gamma$ is local government expenditures with a coefficient normalized to be equal to 1. The characteristic ϵ_j is observed by the households, but is unobserved by the econometrician. We assume that $E(\epsilon_j | z_j) = 0$, where z_j is a vector of instruments. Define

$$m_j(\theta) = g_j - x_j' \gamma. \tag{2.35}$$

[25] Technically speaking, the marginal distribution of income is identified. In addition, one can identify only a finite number of points on the distribution of tastes conditional on income. These points correspond to the points on the boundary between adjacent neighborhoods. For points that are not on the boundary loci, we can provide only lower and upper bounds for the distribution. These bounds become tighter as the number of differentiated neighborhoods in the application increases.

We can estimate the parameters of the model using a generalized method of moments estimator, which is defined as follows:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left\{ \frac{1}{J} \sum_{j=1}^{J} z_j \, m_j(\theta) \right\}' V^{-1} \left\{ \frac{1}{J} \sum_{j=1}^{J} z_j \, m_j(\theta) \right\}, \tag{2.36}$$

where z_j is a set of instruments. Epple and Sieg (1999) suggest using the functions of the rank of the community as instruments. Hence, we can identify and estimate the following additional parameters: γ, $\mu_{\ln\alpha}$, $\sigma_{\ln\alpha}$, ρ, and η. Epple and Sieg (1999) find that the estimates are reasonable and that the fit of the model is good. Standard errors can be approximated using the standard formula described in Newey and McFadden (1994). Note that we need the number of communities to go to infinity to compute asymptotic standard errors.

2.3.2.5 Voting

The model determines tax rates, expenditures on education, and mean housing expenditures for each community in the metropolitan area. We need to determine whether these levels are consistent with optimal household sorting and voting in equilibrium. Again, we can take a partial-solution approach and use necessary conditions that voting imposes on observed tax and expenditure policies. This approach was taken in Epple et al. (2001). They find that the simple voting model discussed above does not fit the data. More sophisticated voting models perform better.

Alternatively, we can take a full-solution approach and estimate the remaining parameters of the model using a nested fixed point algorithm. The latter approach is taken in Calabrese et al. (2006). They modify the equilibrium algorithm discussed in Section 2.3.1.7 and compute equilibrium allocations that satisfy (a) optimal household sorting, (b) budget balance, and (c) majority rule equilibrium, and that are consistent with the observed community sizes. These allocations are an equilibrium in the sense that a housing supply function exists for each community that generates a housing market equilibrium. We can then match the equilibrium values for expenditures, tax rates, and average housing consumption to the observed ones using a simulated maximum likelihood estimator. That article confirms the results in Epple et al. (2001) that the simple model does not fit the data. However, an extended model, in which the quality of public goods depends not only on expenditures, but also on local peer effects, significantly improves the fit of the model.

2.3.2.6 Identifying and estimating housing supply functions

Finally, we briefly discuss how to estimate the housing supply function. If one treats the prices of land and structures as known, few methodological problems arise. However, the key problem encountered in estimating the supply function of housing is that the quantity of housing services per dwelling and the price per unit of housing services are not observed by the econometrician. Instead, we observe the value (or rental expenditures) of a housing unit, which is the product of the price per unit of housing services and the quantity of housing services per dwelling.[26]

Epple et al. (2010b) provide a new flexible approach for estimating the housing production function that treats housing quantities and prices as latent variables. Their approach to identification and estimation is based on duality theory. Assuming that the housing production function satisfies constant returns to scale, one can normalize output in terms of land use. Although we do not observe the price or quantity of housing, we often observe the value of housing per unit of land. The key insight of that article is that the price of housing is a monotonically increasing function of the value of housing per unit of land. Since the price of housing is unobserved, the attention thus focuses on the value of housing per unit of land instead. Constant returns to scale and free entry also imply that profits of land developers must be zero in equilibrium. One can exploit the zero profit condition and derive an alternative representation of the indirect profit function as a function of the price of land and value of housing per unit of land. Differentiating the alternative representation of the indirect profit function with respect to the (unobserved) price of housing gives rise to a differential equation that implicitly characterizes the supply function per unit of land. Most important, this differential equation depends only on functions that can be consistently estimated by the econometrician. Using a comprehensive database of recently built properties in Allegheny County, Pennsylvania, they found that this new method provides reasonable estimates for the underlying production function of housing and the implied housing supply function.

2.3.3 Policy analysis

Once we have found a model that fits the data well and passes the standard specification tests, we can use the model to perform counterfactual policy analysis. Here, we consider two applications. The first one estimates welfare measures for air quality improvements. The second application focuses on the benefits of decentralization.

2.3.3.1 Evaluating regulatory programs: the Clean Air Act

An important need is to evaluate the efficiency of public regulatory programs such as the Clean Air Act. Most methods commonly used in cost–benefit analyses are designed to consider relatively small projects that can be evaluated within a partial equilibrium framework. Sieg et al. (2004) show how to use the methods discussed above to develop an approach for evaluating the impact of large changes in spatially delineated public goods

[26] This problem is similar to the omitted price problem that is encountered in the estimation of production functions. That problem arises because researchers typically observe only revenues and not prices and quantities. If there is a large local or regional variation in product prices, revenues are not a good proxy for quantity.

or amenities on economic outcomes. They study Los Angeles, which has been the city in the United States with the worst air quality. As a consequence, we have access to high-quality data because southern California has a good system of air quality monitors. Between 1990 and 1995, southern California experienced significant air quality improvements. Ozone concentrations were reduced by 18.9% for the study area as a whole. Ozone changes across communities ranged from a 2.7% increase to a 33% decline. In Los Angeles County, the number of days that exceeded the federal 1 h ozone standard dropped by 27% from 120 to 88 days. We want to estimate welfare measures for these improvements in air quality.

One important distinction is to differentiate between partial and general equilibrium welfare measures. As pointed out by Scotchmer (1986, pp. 61–62), "an improvement to amenities will induce both a change in property values and a change in the population of the improved area. Short-run benefits of an improvement are those which accrue before the housing stock, or distribution of population, adjusts. Long-run benefits include the benefits which accrue when the housing stock and distribution of population change. The literature has not dwelled on the distinction between benefits in the short run and long run, probably because the value of *marginal* improvements is the same in both cases." Consider the case in which we exogenously change the level of public good provision in each community from g_j to $\overline{g}_j$. In our application, the change in public good provision arises from improvements in air quality that are due to federal and state air pollution policies. The conventional partial equilibrium Hicksian willingness to pay, $\mathrm{WTP}_{\mathrm{PE}}$, for a change in public goods is defined as follows:

$$V(\alpha, y - \mathrm{WTP}_{\mathrm{PE}}, \overline{g}_j, p_j) = V(\alpha, y, g_j, p_j). \tag{2.37}$$

Households will adjust their community locations in response to these changes. Such an analysis implies that housing prices can change as well. An evaluation of the policy change should reflect the price adjustments stemming from any changes in community-specific public goods. We can define the general equilibrium willingness to pay as follows:

$$V(\alpha, y - \mathrm{WTP}_{\mathrm{GE}}, \overline{g}_k, \overline{p}_k) = V(\alpha, y, g_j, p_j), \tag{2.38}$$

where $k(j)$ indexes the community chosen in the new (old) equilibrium. Since households may adjust their location, the subscripts for $(\overline{g}_k, \overline{p}_k)$ need not match (g_j, p_j).

Using data from Los Angeles in 1990, Sieg et al. (2004) estimate the parameters of a sorting model that is similar to the one discussed in the previous sections. They find that willingness to pay ranges from 1% to 3% of income. The model predicts significant price increases in communities with large improvements in air quality and price decreases in communities with small air quality improvements. Partial equilibrium gains are thus often offset by price increases. At the school district level, the ratio of general to partial equilibrium measures ranges from 0.28 to 8.81, with an average discrepancy of nearly 50%. Moreover, there are large differences between the distributions of gains in partial versus general equilibrium.

Sieg et al. (2004) use the projected changes in ozone concentrations for 2000 and 2010, together with the estimates for household preferences for housing, education, and air quality, to conduct a prospective analysis of policy changes proposed by the Environmental Protection Agency. They measure general equilibrium willingness to pay for the policy scenarios developed for the prospective study as they relate to households in the Los Angeles area. Estimated general equilibrium gains from the policy range from $33 to $2400 annually at the household level (in 1990 dollars).[27]

2.3.3.2 Decentralization versus centralization

One of the key questions raised in the seminal article of Tiebout (1956) is whether decentralized provision of local public goods, together with sorting of households among jurisdictions, can result in an efficient allocation of resources. It is not difficult to construct some simple examples in which allocations are not efficient in Tiebout models (Bewley, 1981). However, this question is more difficult to answer once we consider more realistic models. Moreover, we would like to have some idea about the quantitative magnitude of potential inefficiencies.

Calabrese et al. (2012) attempt to answer both sets of questions. First, they derive the optimality conditions for a model that is similar to the one developed in Section 2.3.1. They show that an efficient differentiated allocation must satisfy a number of fairly intuitive conditions. First, the social planner relies on lump-sum taxes and sets property taxes equal to zero. The planner does not rely on distortionary taxes. Second, the level of public good provision in each community satisfies the Samuelson condition. Finally, each household is assigned to a community that maximizes the utility of the household. The last condition is not obvious because of the fiscal externalities that households provide.

The second step of the analysis, then, is to try to quantify the potential efficiency losses that arise in equilibria. They calibrated the model and compared welfare in property tax equilibria, both decentralized and centralized, with the efficient allocation. Inefficiencies with decentralization and property taxation are large, dissipating most if not all of the potential welfare gains that efficient decentralization could achieve. In property tax equilibria, centralization is frequently more efficient! An externality in community choice underlies the failure to achieve efficiency with decentralization and property taxes: poorer households crowd richer communities and free ride by consuming relatively little housing, thereby avoiding taxes. They find that the household average compensating variation for adopting the multijurisdictional equilibrium is $478. The per household

[27] Tra (2010) estimates a random utility model using a similar data set for Los Angeles. His findings are comparable to the ones reported in Sieg et al. (2004). Wu and Cho (2003) also study the role of environmental amenities in household sorting. Walsh (2007) estimates a model that differentiates between publicly and privately provided open space to study policies aimed at preventing urban sprawl in North Carolina.

compensating variation for land owners is −\$162. Hence, the decentralized Tiebout equilibrium implies a welfare loss equal to \$316 per household. This equals 1.3% of 1980 per household income.

2.4. THE ALLOCATION OF ECONOMIC ACTIVITY ACROSS SPACE

Understanding how economic activity is allocated across space is a core subject in urban and regional economics. This section considers two applications related to the topic: the regional specialization of industry and the internal structure of cities. We begin by developing models used in the two applications and discuss identification and estimation. Finally, we address various issues that need to be confronted when using the estimated models to evaluate the effects of counterfactual policies.

Although the focus is on methodology, we want to emphasize the interesting questions that can be addressed with structural models along the lines that we discuss. The first application is a model in which locations specialize in industries. With a successful quantitative model, we can evaluate questions such as how investments in transportation infrastructure affect the pattern of regional specialization. The second application is a model of where people live and work in a city, and it takes into account economies of density from concentrating workers and residents in particular locations. If we succeed in developing a computer-generated quantitative model of the city, we can evaluate how regulations, subsidies, or investments in infrastructure affect where people live and work, and how these policies affect levels of productivity and welfare. Note that, befitting its importance for the field, other chapters in this handbook delve into various aspects of the allocation of economic activity across space. In particular, Chapter 5, by Combes and Gobillon, reviews empirical findings in the literature on agglomeration, including results from structural approaches.[28] And Chapter 8, by Duranton and Puga, reviews the theoretical and empirical literature on urban land use. Although the other chapters focus primarily on results, again, the focus here is on methodology.

2.4.1 Specialization of regions

The first application is based on articles that apply the Eaton and Kortum (2002) model of trade to a regional context, with regions the analog of countries. Note that in our second application on the internal structure of cities that follows, we will assume that workers are mobile across different locations in a city. In contrast, here in our first application, there is no factor mobility across locations; only goods flow. Donaldson (forthcoming) applies the framework to evaluate the regional impact of investments in transportation infrastructure. Holmes and Stevens (2014) apply the framework to evaluate the effects of increased imports from China on the regional distribution of manufacturing within the United States. In the exposition, we focus on the Holmes and Stevens (2014) version.

[28] See also Combes et al. (2011) and Rosenthal and Strange (2004).

2.4.1.1 Model development

Suppose there is a continuum of different goods in an industry, with each good indexed by $\omega \in [0, 1]$. There are J different locations indexed by j. For expositional simplicity, assume for now there is a single firm at location j that is capable of producing good ω. Let $z_{\omega,j}$ be the firm's productivity, defined as output per unit input, and let w_j be the cost of one input unit at location j. Let $\mathbf{z}_\omega \equiv \left(z_{\omega,1}, z_{\omega,2}, \ldots, z_{\omega,J}\right)$ denote the vector of productivity draws across all firms, and let $F(\mathbf{z}_\omega)$ be the joint distribution. There is a transportation cost to ship goods from one location to another. As is common in the literature, we assume iceberg transportation costs. Specifically, to deliver one unit from j to k, $d_j^k \geq 1$ units must be delivered. Assume $d_j^j = 1$ and $d_j^k > 1$, $k \neq j$—that is, there is no transportation cost for same-location shipments, but there are strictly positive costs for shipments across locations. The cost for firm j to deliver one unit to k is then

$$c_{\omega,j}^k = \frac{w_j d_j^k}{z_{\omega,j}}. \tag{2.39}$$

The minimum cost of serving k over all J source locations is

$$\underline{c}_\omega^k = \min_j c_{\omega,j}^k, \tag{2.40}$$

and let $\underline{j}^k$ be the firm solving (2.40), the firm with the lowest cost to sell to k. If the joint distribution $F(\mathbf{z}_\omega)$ is continuous, the lowest-cost firm $\underline{j}^k$ is unique except for a set of measure zero. If firms compete on prices in a Bertrand fashion in each market k, the most efficient firm for k, firm $\underline{j}^k$, gets the sale. For a given product ω, the likelihood the firm at j is the most efficient for k depends on the joint distribution of productivity draws, transportation costs d_j^k, and input costs $(w_1, w_2, \ldots, w_J)$.

Eaton and Kortum (2002) make a particular assumption on the joint distribution $F(\mathbf{z}_\omega)$ that yields an extremely tractable framework. Specifically, productivity draws of individual firms are assumed to come from the Fréchet distribution. The draws across firms are independent, and the cumulative distribution function (c.d.f.) for a firm at location j is given by

$$F_j(z) = \mathrm{e}^{-T_j z^{-\theta}}. \tag{2.41}$$

The *shape* parameter θ governs the curvature of the distribution and is constant across locations; the lower θ, the greater the variation in productivity draws across firms. The *scale* parameter T_j allows locations to differ in mean productivity; the higher T_j, the higher the average productivity drawn by a firm at location j. Let $G_j^k(c)$ be the c.d.f. of the cost c_j^k of firm j to ship goods to k. This can be derived by plugging (2.39) into (2.41). It is convenient to write the equation in terms of the complement of the c.d.f. (the probability of drawing above c_j^k):

$$1 - G_j^k\left(c_j^k\right) = e^{-T_j\left(w_j d_j^k\right)^{-\theta}\left(c_j^k\right)^{\theta}}. \tag{2.42}$$

This equation has the same functional form as (2.41), only now the scale parameter takes wages and transportation costs into account. Consider the c.d.f. $\underline{G}^k(\underline{c}^k)$ of $(\underline{c}^k)$, the lowest cost across all sources. Writing the equation in terms of its complement, we calculate the probability that the cost is higher than $\underline{c}^k$ at all locations—that is,

$$\begin{aligned} 1-\underline{G}^k(\underline{c}^k) &= \prod_{j=1}^{J}\left[1-G_j^k(\underline{c}^k)\right] \\ &= e^{-\sum_{j=1}^{J} T_j\left(w_j d_j^k\right)^{-\theta}\left(\underline{c}^k\right)^{\theta}}. \end{aligned} \tag{2.43}$$

Note that the shape of the functional form of (2.43) is the same as (2.42), only now the scale factor is the sum of the scale factors of the cost distributions across the different locations. This is a convenient property of the Fréchet. Moreover, straightforward calculations yield the following expression for the probability that the firm at j is the lowest-cost source for serving location k:

$$\pi_j^k = \frac{T_j\left(w_j d_j^k\right)^{-\theta}}{\sum_{s=1}^{J} T_s\left(w_s d_s^k\right)^{-\theta}}. \tag{2.44}$$

This formula is intuitive. The numerator is an index of firm j's efficiency to sell at k, varying proportionately with the productivity parameter T_j, and inversely with input costs and transportation costs to get from j to k. The formula takes firm j's efficiency relative to the sum of the efficiency indices across all source locations. In Eaton and Kortum (2002), firms price competitively. Bernard et al. (2003) extend the framework to an oligopoly setting. Under the assumption that demand has constant elasticity, both treatments show that the share of sales at location k, sourced from location j, is given by formula (2.44). Hence, if X^k denotes total industry expenditure at location k, and Y_j^k the sales of firms at j to k, and if Y_j equals total sales at j to all destinations, then

$$Y_j = \sum_{k=1}^{S} Y_j^k = \sum_{k=1}^{S} \frac{T_j\left(w_j d_j^k\right)^{-\theta}}{\sum_{s=1}^{J} T_s\left(w_s d_s^k\right)^{-\theta}} X^k. \tag{2.45}$$

This is a useful equation that links expenditures and sales at each location with the location-level productivity parameters, input prices, and transportation costs. From the formula, we can see that an industry will tend to concentrate at a particular location j if its productivity is high, if input costs are low, and if the costs of transportation to locations with high expenditures are low.[29] The second application below uses the same

[29] Anderson and van Wincoop (2003) derive a similar equation in an alternative formulation.

Fréchet magic to derive tractable expressions of equilibrium commuting flows between different locations in the same city.

2.4.1.2 Estimation and identification

We now turn to the issue of estimation and identification. To impose more structure on transportation costs, let m_j^k be the distance in miles between locations j and k, and assume the iceberg transportation cost depends only on distance—that is, $d_j^k = f(m_j^k)$, where $f(0) = 1$, and $f'(m) > 0$. Next, define a function $h(m)$ by

$$h(m_j^k) \equiv \left(d_j^k\right)^{-\theta} = f(m_j^k)^{-\theta}. \tag{2.46}$$

We can think of this as a *distance discount*. It equals 1 when the distance is zero and strictly declines as the distance increases, depending on the rate at which the iceberg transportation cost increases, as well as the shape parameter θ of the productivity distribution. Next, define $\gamma_j \equiv T_j w_j^{-\theta}$, a composite of the technology measure T_j, the wage at j, and the shape parameter θ. In a partial equilibrium context, where the wage w_j is fixed and the technology level T_j is exogenous, the composite parameter γ_j can be treated in a structural way now. We discuss alternatives in the discussion of policy below.

Using our definitions of $h(m_j^k)$ and γ_j, we can then rewrite (2.45) as

$$Y_j = \sum_{k=1}^{S} \frac{\gamma_j h(m_j^k)}{\sum_{s=1}^{J} \gamma_s h(m_s^k)} X^k, \quad j = 1, \ldots, J. \tag{2.47}$$

Suppose for the sake of discussion that the distance discount function $h(\cdot)$ is known for the particular industry under consideration. Suppose we have data $\left\{ Y_j, X^k, m_j^k, \text{all } j \text{ and } k \right\}$—that is, the value of production at each location, absorption at each location, and distance information. The vector of cost efficiencies $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_J)$ is identified from the set of equations given by (2.47). The identification is subject to a rescaling by a positive multiplicative constant, so a normalization is required, e.g., $\gamma_1 = 1$, if $Y_1 > 0$. See Proposition A.1 in the appendix of Ahlfeldt et al. (2014) for a proof that a unique $\boldsymbol{\gamma}$ exists that solves (2.47), again subject to a normalization. The appendix in Holmes and Stevens (2014) describes an iterative procedure to obtain a solution as a fixed point. Think of the γ_j as a location-level fixed effect that is solved for to exactly fit the data. Redding and Sturm (2008) and Behrens et al. (2013) perform similar calculations.

The above consideration takes as given the distance discount $h(m)$. Suppose the discount is unknown *a priori*. In this case, data on the distances that shipments travel are useful. A long tradition in the trade literature examines how trade flows vary with distance; one example is the gravity model considered in Anderson and van Wincoop (2003). Here, we focus on the approach taken in Holmes and Stevens (2014). In the census data used in the study, total shipments originating across all plants at a given location j are observed (this is Y_j). In addition, an estimate of absorption at each destination (i.e., X^k) is also obtained. In addition to these aggregate quantities, the article employs data from a random sample of individual transactions, for which the origin and destination are provided. Let the distance discount function be parameterized by a vector $\boldsymbol{\eta}$—that is, we write $h(m, \boldsymbol{\eta})$. The article jointly estimates $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_J)$ and $\boldsymbol{\eta}$ by choosing $(\boldsymbol{\gamma}, \boldsymbol{\eta})$ to maximize the likelihood of the shipment sample, subject to $(\boldsymbol{\gamma}, \boldsymbol{\eta})$, satisfying (2.47) for the given values of Y_j and X^k. If shipments in the data tend to go short distances, the estimated distant discount $h(m, \hat{\boldsymbol{\eta}})$ will tend to drop sharply with distance (examples in the data include industries like ready-mix cement and ice). In cases in which shipments travel long distances, the estimated distance discount will be relatively flat at 1 (an example is medical equipment).

2.4.2 Internal structure of cities

Our discussion is based on the work of Ahlfeldt et al. (2014), who estimate a structural model of the city of Berlin. (See also Duranton and Puga (2015) in this volume for a discussion of the work of Ahlfeldt et al. (2014) that complements ours.) Theories of the internal structure of cities focus on flows of commuters from their place of residence to their place of work, and the spillover benefits from economies of density. The city of Berlin provides a fascinating context because of the way the Berlin Wall blocked such flows. The paper uses data for periods before, during, and after the existence of the Berlin Wall to estimate a rich model that simultaneously takes into account both commuter and spillover flows.

The paper builds on a long tradition in urban economics research on the internal structure of cities, dating back to the literature on the monocentric model of the city. This classic early model is useful for illustrating theoretical points, such as how a change in commuting costs affects land prices. Yet this abstraction, in which land is used for residence and not for production, and where all residents commute to work at a single point, does not correspond to what actual cities look like. Lucas and Rossi-Hansberg (2002) provided an important generalization in which land is used for both residence and production. Yet again, this structure aims at theoretical points, and one abstraction is that a city is a perfect circle with uniform rings. Furthermore, there is no worker heterogeneity, with the implication that all workers living in a given part of the city would commute to the same place for work. Ahlfeldt et al. (2014) estimate a structural model of an actual city, and its approach departs from these various simplifications. Their model explicitly takes into account that land features are not uniform over space and that cities are not circles. It takes into account that individuals are heterogeneous and may vary in their match quality with particular employers, and in match quality with particular places to live. Finally, the model allows for spillovers to arise on the consumption side as well as on the production side.

2.4.2.1 Model development

We provide a brief overview of the modeling setup. Individuals are freely mobile and choose whether or not to live in the city, and if so, where to live and where to work, from a choice of J discrete locations. Firms are also freely mobile about where to produce, and a given parcel of land can be used for production or residence. Productivity varies across locations, because of the exogenous features of land, as well as endogenously, through the levels of neighboring employment and the resulting spillovers. Specifically, the productivity index A_j at location j is given by

$$A_j = \Upsilon_j^{\lambda} a_j, \tag{2.48}$$

where a_j is the exogenous location quality, and Υ_j is aggregated spillovers received by j from all other city locations, defined by

$$\Upsilon_j = \sum_{k=1}^{J} e^{-\delta m_j^k} \widetilde{Y}_k, \quad \lambda \geq 0, \delta \geq 0. \tag{2.49}$$

In this expression, $\widetilde{Y}_k$ is employment at location k, and m_j^k is the distance between locations i and j. The parameter δ governs how rapidly spillovers decline with distance. The parameter λ determines how the aggregated spillovers convert into productivity gains. Analogously, there is an exogenous consumption amenity level b_j at location j and an endogenous spillover component from neighboring residents, with the same functional form as for the production side, but with different parameters. The last pieces of the model relate to individual choice. Individuals who choose to live in the city obtain match quality draws for every possible combination of where they might live and where they might work. Commuting costs create tension between these two considerations. Besides commuting costs and match quality, individuals need to take into account how wages vary by location in their decision of where to work. In the decision of where to live, they need to take into account housing rents and consumption amenities.

Note that the model is very flexible and general in the way that exogenous productivity a_j is free to vary across locations. Analogously, the exogenous consumption amenity b_j is free to vary. Allowing for this generality is important because if this variation exists and we ignore it, we might mistakenly attribute all the observed concentration of employment or residence to spillovers, when exogenous variations in land quality also play a role.

For technical convenience, analogous to the first application, Ahlfeldt et al. (2014) make use of the Fréchet structure of Eaton and Kortum (2002), regarding the distribution of workplace/residence match qualities. The assumption yields a tractable approach.

2.4.2.2 Estimation and identification

In our first application, the logic behind the identification of location-specific productivities and distance discounting (the parameters given by ($\boldsymbol{\gamma}$, $\boldsymbol{\eta}$)) is straightforward. The issues are more complex in the Ahlfeldt et al. (2014) model of residential and worker location within a city. We highlight two challenges in particular. First, separating out natural advantage (given by the exogenous productivity component a_j at each location j) from knowledge spillovers (the elasticity λ listed above) is intrinsically difficult. Suppose we see in the data that at locations with a high density of workers, land rents are high. Is this because locations with high exogenous productivity a_j are attracting a large number of workers and this bids up rents? Or does causation go the other way, such that locations with a high concentration of workers are more productive, which in turn bids up rents? Or does the answer lie somewhere in between?

The second issue is that when there are knowledge spillovers, there is a potential for multiple equilibria to exist at given values of the model's structural parameters. For example, workers might cluster at point A just because everyone else is clustering there (i.e., the cluster is self-fulfilling). Perhaps an alternative equilibrium also exists where workers cluster at some different point B. The possibility of multiplicity has potential implications for estimation and identification as well as for policy analysis.

Ahlfeldt et al. (2014) confront these issues by exploiting the historical context of the Berlin Wall going up and coming down. They treat these events as quasi-experimental variation that can be used to identify the structural parameters of the model. Data were collected at a fine geographic level, 16,000 city blocks, and include the number of residents X_t^j in block j at time t, the number of workers $Y_{j,t}$ employed at j at time t, and the rental price of land $r_{j,t}$ at time t for block j. The wage at location j plays the same role in the Ahlfeldt et al. (2014) model as the productivity variable T_j plays in the industry specialization application, and there is a formula in Ahlfeldt et al. (2014) that is analogous to (2.45). Location-level wages are unobserved and are inferred in a way that is analogous to the way that unobserved location-level productivities were inferred in the regional specialization application.

Let $\boldsymbol{\beta}$ be a vector that collects all of the various parameters of the model, such as the knowledge spillover elasticity λ and the spatial discount parameter δ that appear in the productivity specification (2.48). Let $a_{j,t}$ and $b_{j,t}$ be the natural advantage parameters for production and consumption at location j at time t, which we write in vector form as $\mathbf{a}_t$ and $\mathbf{b}_t$, with elements for each of the J locations. Let $(\mathbf{X}_t,\mathbf{Y}_t,\mathbf{r}_t)$ be the vector of data that contains the number of residents, number of workers, and the rental rate for each block. Although there may be multiple equilibria, a key result of the paper is that for a fixed parameter vector $\boldsymbol{\beta}$ and a given data realization $(\mathbf{X}_t,\mathbf{Y}_t,\mathbf{r}_t)$, there exists unique values of $(\mathbf{a}_t,\mathbf{b}_t)$ consistent with equilibrium.[30] For intuition, recall the earlier discussion that if in the data we see high concentration and high rents, we can account for these findings by giving all the credit to natural advantage and none to spillovers, or all of the credit to spillovers and none to natural advantage, or something in between. But in the present discussion, when we take the parameter vector $\boldsymbol{\beta}$ as given, as well as the data, we are *fixing* the credit given to spillovers, and the resulting values $(\mathbf{a}_t,\mathbf{b}_t)$ can be thought of as the residual credit that must be given to natural advantage, in order

[30] This is uniqueness, subject to some normalizations.

for the equilibrium conditions to hold. So in terms of estimation, the second issue noted above, about the potential multiplicity of equilibrium, ends up not being a concern.

We now turn to the first challenge, disentangling spillovers and natural advantage. Following the above discussion, for a given set of model parameters and the observed data, the article infers the implied values of natural advantage in production a_j and consumption amenity b_j for each location j. The key identifying assumption is that any changes in these natural advantage variables over time are unrelated to the distance of a location from the Berlin Wall. The article estimates significant levels of spillovers for both production and consumption. Remarkably, the estimates based on what happened between 1936 and 1986, when the Berlin Wall went up, are very similar to the estimates based on 1986 and 2006, when the Berlin Wall went down. The key feature of the data that drives estimates of spillovers is that after the Berlin Wall was erected, land prices collapsed near it. The pattern reversed when the Berlin Wall was taken down. To understand how this works in the model, suppose we shut down knowledge spillovers. The sharp drops in land prices near the Berlin Wall imply that natural advantage must have systematically declined near the Berlin Wall. This is inconsistent with the identifying assumption.

2.4.3 Policy analysis

As emphasized in Section 2.1, a key benefit of the structural approach to empirical work is that prospective policy analysis can be conducted with the estimated model. At the beginning of this section, we mentioned a variety of interesting policy issues that can be addressed with the class of models discussed here. Now we focus on a particular case that is useful for illustrating methodological points. In the model of industry specialization, we evaluate how opening up the domestic industry to foreign competition affects the regional distribution of production. Holmes and Stevens (2014) conduct such an exercise by evaluating the regional impact of imports from China, and here we consider a simpler version of the experiment.

Following our discussion above of the regional specialization model, we begin with our estimates of the vector $\boldsymbol{\gamma}$ of cost efficiency indices across locations and the parameters $\boldsymbol{\eta}$ governing distance discounts $h(m, \boldsymbol{\eta})$. Suppose imports are initially banned. The specific policy change we consider is to allow imports, subject to a quota. Suppose the world market is such that imports will flow in, up to the quota. Suppose the quota is set in such a way that the value of imports will equal 5% of the total domestic market. Assume for simplicity that all imports must go through the same port, which is at some new location $J+1$, and the distance discount from here to other locations follows the same distance discount estimated in the first stage. Assume that the industry under consideration is relatively small, such that imports do not affect wages. Finally, make Cobb-Douglas assumptions about consumer utility so that relative spending shares on the industry X^k/X^j between any pair of locations k and j do not change.

Putting all of these assumptions together, we see that the policy is equivalent to creating a new location $J + 1$, with its own efficiency index γ_{J+1} and no consumption—that is, $X_{J+1} = 0$—holding fixed the cost efficiency indices of the other locations γ_j, $j \leq J$, and the distance discounts $h(m, \boldsymbol{\eta})$. For any given value of γ_{J+1}, we can use Equation (2.47), now extended to sum up to $J + 1$, to solve for the sales of each location Y_j^{new}, where "new" means after the policy change. The higher γ_{J+1}, the greater are imports Y_{J+1}^{new} and the lower domestic production at each location Y_j^{new}, $j \leq J$. We pick γ_{J+1} such that the value of imports Y_{J+1}^{new} is 5% of the domestic market. We then compare Y_j^{new} with Y_j^{old} to examine the regional impact of trade. In general, the effects vary across locations, depending on the role of transportation costs (domestic producers near the port will be hurt more than others), a location's productivity, and the productivity of a location's neighbors.

We now have in place an example structural model, for which we laid out the issues of estimation and identification, and have presented an illustrative policy experiment. Next we use the example to address various issues.

First, notice that we were able to conduct this particular experiment without having to unpack the estimated distance function $h(m, \eta)$ into underlying parts. Remember this is a composite of other parameters. We are able to do this because the underlying policy change being considered leaves distance discounting alone. Of course, there are other policy changes, such as infrastructure investment to reduce transportation costs, for which we would need estimates of these deeper structural parameters to conduct policy analysis. Donaldson (forthcoming) needs these deeper structural parameters in his analysis of the productivity effects of the introduction of the railroad network in India. A key step in his analysis is his use of data on how price varies across space to directly infer transportation costs and how these costs changed after the railroad network was introduced.[31]

Second, we left wages unchanged. If the industry being considered accounts for a significant share of a particular location's employment, then the policy experiment will lead to local wage changes. That is, the cost efficiency parameter $\gamma_j = T_j w_j^{-\theta}$ being held fixed in the exercise now varies. If this is a concern, the analysis must be extended to incorporate a structural model of regional wages. In addition, the shape parameter θ of the productivity distribution needs to be estimated.

Third, we left the productivity parameter T_j unchanged. This is appropriate if productivity reflects natural advantage, but is a concern if knowledge spillovers are potentially important. Suppose, in particular, that the location productivity scaling parameter takes the following form, analogous to that in Ahlfeldt et al. (2014):

$$T_j = a_j N_j^{\lambda}, \tag{2.50}$$

[31] For a related analysis, see also Duranton et al. (2014).

where a_j is natural advantage, N_j is industry employment at j, and λ is the knowledge spillover elasticity. So far we have implicitly assumed that $\lambda = 0$, so $T_j = a_j$, but now we consider $\lambda > 0$. In Eaton and Kortum (2002), equilibrium expenditure on inputs at location j is a fraction $\frac{\theta}{1+\theta}$ of revenue, or $w_j N_j = \frac{\theta}{1+\theta} Y_j$. Solving for N_j and substituting (2.50), we can write cost efficiency at j as

$$\gamma_j = T_j w_j^{-\theta} = a_j \left(\frac{\frac{\theta}{1+\theta} Y_j}{w_j} \right)^{\lambda} w_j^{-\theta}. \tag{2.51}$$

Now suppose we also have data on wages at j. If we take θ and λ as known, following our discussion above, we can solve (2.47) for a unique solution vector $\mathbf{a} = (a_1, a_2, \ldots, a_J)$, subject to a normalization. With this setup in place, the analysis can proceed in two ways. The ideal procedure, if feasible, is to go back to the estimation stage to develop a strategy for estimating θ and λ. For example, as in Ahlfeldt et al. (2014), it may be possible to obtain instruments that can be used to construct orthogonality conditions that are satisfied by the vector $\mathbf{a}$ of natural advantages. If estimation of θ and λ is not feasible, then researchers can take a second approach that takes the form of robustness analysis. The estimates under the identifying assumption that $\lambda = 0$ provide the baseline case, and the policy experiment under this assumption is discussed first. Next is a discussion of how results would change if knowledge spillovers are introduced. A variety of estimates of λ can be found in the literature, as discussed in this volume. A value of $\lambda = 0.10$ is generally considered on the high end. Turning to the θ parameter, note that $\frac{\theta}{1+\theta}$ is the variable cost share of revenues. Thus a broad range of θ from 3 to 9 is equivalent to variable cost shares that range from 0.75 to 0.90. This broad range nests values that have been obtained in various applications in the literature (e.g., $\theta = 8.28$ in Eaton and Kortum, 2002). Now consider re-estimating the model over a grid of θ and λ satisfying $\theta \in [3, 9]$ and $\lambda \in [0, 0.10]$ and resimulating the policy experiment for each case. This provides a range of estimates for the policy effects, with $\lambda = 0$ corresponding to the benchmark case. (In that limit, the choice of θ is irrelevant for the policy experiment.) It may very well be that the baseline results are relatively robust to these alternative assumptions. Transportation cost may be the primary force determining the relative impact of imports across regions (i.e., where those locations closest to ports are affected the most), and knowledge spillovers might be a secondary consideration. If so, the proposed robustness analysis will make this clear. In any case, this discussion highlights how the structural empirical approach yields models that can be built upon and enriched. Rather than speculate about how allowing for agglomeration economies can change an answer, the model can be extended and the answer to the question simulated.

We conclude this discussion of policy experiments by coming back to the issue of multiple equilibria. In the baseline version with $\lambda = 0$, equilibrium is unique. As is well understood in the literature, multiple equilibria may be possible when $\lambda > 0$. In this case, there is positive feedback, where adding more production lowers costs, increasing the incentive to have still more production, and there are potentially multiple places where an industry might agglomerate. Suppose there is a policy intervention and there are multiple equilibria given the model estimates. Which equilibrium is the relevant one? This issue can be a difficult one, but we can make some observations. First, although multiplicity is possible when $\lambda > 0$, there might be enough curvature (e.g., transportation costs or congestion costs) such that there is an unique equilibrium. If researchers verify uniqueness, this addresses the issue. Second, equilibrium might be unique locally in the vicinity of the baseline case. If the policy intervention is small, a sensible approach may be to focus on the comparative statics of the local equilibrium. Third, it may be possible to estimate the selection process for equilibria, as in Bajari et al. (2010a).

2.4.4 Relation to entry models in the industrial organization literature

When spillovers exist in the models discussed above, interactions are created between decision makers. The study of interactions between decision makers is a general problem in economics. Recently, extensive work has been done on this class of models in the industrial organization literature, focusing on developing partial-solution approaches to study entry by firms into markets, and in particular incorporating dynamics. Here, we connect the discussion above to this literature.

In environments considered in the industrial organization literature, there are often relatively few decision makers, in which case taking into account that entry is discrete may be important. Urban and regional applications often abstract from discreteness in the underlying economic environment, as in the examples above, and this abstraction can be useful when a relatively large number of decision makers are interacting. As research in urban and regional applications takes advantage of new data sets at high levels of geographic resolution, it permits the study of interactions at narrow levels, where there may be relatively few decision makers. In such cases, taking discreteness into account may be useful, and the discussion here illustrates the discrete case. In any case, the partial-solution approaches discussed below can also be scaled up to include cases of large numbers of interacting agents.[32] As a starting point for the discussion, a useful step is to review the classic discrete choice model of social interactions in Brock and Durlauf (2001). We can think of this as the approximate state of the literature at the time of publication of the previous handbook (see Durlauf, 2004). In the model, an agent is making a decision where the agent's payoff depends on the decisions of the other agents. Labeling variables to represent the context of a model of industry agglomeration, suppose that at a given location j, there are I potential entrants indexed by i. Let a_j be a measure of the natural

[32] See, for example, Weintraub et al. (2008).

advantage of location j. Let N_j be the total number of firms that enter at location j. Define U_{ij}^E and U_{ij}^N to be firm i's profit from entering or not entering market j, and suppose profits take the following form:

$$U_{ij}^E = \beta^E + \beta^a a_j + \beta^N N_j + \varepsilon_{ij}^E, \tag{2.52}$$

$$U_{ij}^N = \varepsilon_{ij}^N. \tag{2.53}$$

In this specification, β^a is the weight on natural advantage, and β^N is the weight on firm interactions. The shocks ε_{ij}^E and ε_{ij}^N are independent and identically distributed and are private information observed only by potential entrant i. In a Nash equilibrium, firms will take as given the strategies of the other firms, which specify how their entry decisions will depend on their private shocks. Taking as given these entry strategies by the other firms, let EN_j be the expected count of firm entry perceived by a given firm, conditional on the given firm itself entering. Note $EN_j \geq 1$, because the count includes the firm itself. Substituting expected entry EN_j into the payoff U_{ij}^E, firm i enters if

$$\beta^E + \beta^a a_j + \beta^N EN_j + \varepsilon_{ij}^E \geq \varepsilon_{ij}^N, \tag{2.54}$$

which can be written as a cutoff rule in terms of the difference in shocks,

$$\varepsilon_{ij}^E - \varepsilon_{ij}^N \geq f_{ij}(EN_j) \equiv -\left(\beta^E + \beta^a a_j + \beta^N EN_j\right). \tag{2.55}$$

Thus, starting out with a perceived value of expected entry EN_j, we derive the entry rule (2.55), from which we can calculate expected entry. An equilibrium is a fixed point where EN_j maps to itself. As highlighted in Brock and Durlauf (2001), if β^N is positive and large, there can be multiple equilibria. If expected entry is high, then with $\beta^N > 0$, entry is more attractive and high entry is self-fulfilling. If the coefficient on natural advantage β^a is positive, entry will tend to be higher in locations with higher natural advantage.[33]

In terms of estimation, Brock and Durlauf (2001) note that if the private shocks are extreme values and if EN_j is observed, then the parameters β^E, β^a, and β^N can be estimated as a standard logit model. Although EN_j may be increasing in a_j, it does so in a nonlinear fashion (through the discrete entry). Since a_j and EN_j are not perfectly collinear, β^a and β^N are separately identified. This is in contrast to the earlier *linear-in-means* formulation in Manski (1993), where it was noted that the analog of EN_j in the model was linear in the analog of a_j, implying that the analogs of β^a and β^N were not separately identified. Researchers are often uncomfortable about relying heavily on functional form assumptions to obtain identification. There is great value in coming up with exclusion restrictions based on the economics of the problem. For example, suppose potential

[33] Note that this monotonicity claim regarding natural advantage a_j ignores complications that may arise with comparative statics when multiple equilibria exist.

entrants vary in productivity ω_i, and suppose the profitability of entry U_{ij}^E above is modified to include an additional term $\beta^\omega \omega_i$—that is,

$$U_{ij}^E = \beta^E + \beta^\omega \omega_i + \beta^a a_j + \beta^N N_j + \varepsilon_{ij}^E. \tag{2.56}$$

Assume that firm productivities are common knowledge. With $\beta^\omega > 0$, and everything else the same, the higher ω_i, the more likely firm i is to enter. This sets up an exclusion restriction, where a higher value of productivity $\omega_{i'}$ for some other firm i' has no direct effect on firm i's profitability and affects profitability only indirectly by affecting the likelihood of entry by firm i'.

We now connect the discussion to recent developments in the industrial organization literature. This literature has long been interested in analysis of games with payoff structures such as (2.52), though typically the focus has been on environments in which the interaction parameter β^N is negative—that is, agents are worse off when others enter. For example, if the market is the drugstore market, a firm will be worse off if it has to share the market with more competitors, and in addition the added competition will put downward pressure on prices (Bresnahan and Reiss, 1991). The recent literature has focused on dynamics.[34] Going back to the problem as described above, we find dynamics add two elements. First, agents who decide to enter consider not only current profits but also future profits and how future entry will evolve. Second, when agents make entry decisions, in general there may already be incumbent firms in the industry. Although the literature is typically motivated by cases in which $\beta^N < 0$, the technical developments also apply for $\beta^N > 0$.

Let y_{ijt} be an indicator variable that firm i is an incumbent in location j at time i (i.e., entered previously), and let $\mathbf{y}_t = (y_{1jt}, y_{2jt}, \ldots, y_{Ijt})$ be the vector listing incumbent status. Analogously, let $\boldsymbol{\omega}$ be the vector of firm productivities. The state of the industry at the beginning of time t at j is $\mathbf{s}_{jt} = (a_j, \boldsymbol{\omega}, \mathbf{y}_t)$—that is, location natural advantages, firm productivities, and a list of firms that have entered. Let a firm's current period payoff when it participates in market j in period t be given by (2.56). It is straightforward to see how the nested fixed point works here: for a given set of parameters, solve for equilibrium and then vary the parameters to best fit the data according to some metric. However, for computational tractability, the recent literature has focused on two-step approaches, following techniques developed by Hotz and Miller (1993), for discrete choice in labor market applications. The idea is to estimate behavioral relationships in a first stage and then in a second stage back out the parameters that rationalize the behavior.

To explain this, suppose first that the state $\mathbf{s}_{jt} = (a_j, \boldsymbol{\omega}, \mathbf{y}_t)$ is common knowledge for industry participants and is also observed by the econometrician studying the problem (we come back to this below). Moreover, in cases in which there are multiple equilibria, assume the same equilibrium is played conditional on the state $\mathbf{s}_{jt}$ across all the sample

[34] See Aguirregabiria and Mira (2010) for a survey.

locations in the data. Given $\mathbf{s}_{jt}$, entry decisions will depend on the realizations of the shocks ε_{ij}^{E} and ε_{ij}^{N} for each i and j, and will induce a probability of entry $p_{ij}(\mathbf{s}_{jt})$ for each firm i at j, given $\mathbf{s}_{jt}$. This is a conditional choice probability. Since $\mathbf{s}_{jt}$ is observed by the econometrician, we can obtain an estimate of $\hat{p}_{ij}(\mathbf{s}_{jt})$ from the sample averages. The estimated values $\hat{p}_{ij}(\mathbf{s}_{jt})$ from the first stage summarize an agent's choice behavior. In the second stage, various approaches can recover the structural parameters from the first stage estimates of choice behavior. For the sake of brevity, we consider a simple special case: entry is static (lasts for one period), in which case payoffs look exactly like (2.52). Let $\widehat{E_iN_j}(\mathbf{s}_{jt})$ be an estimate of the expected count of entering firms from the perspective of firm i, given that it enters and given the state. This is constructed as

$$\widehat{E_iN_j}(\mathbf{s}_{jt}) = 1 + \sum_{k \neq i} \hat{p}_{kj}(\mathbf{s}_{jt}). \tag{2.57}$$

If firm i enters, it counts itself in addition to the expected value of all other potential entrants. Now substitute $\widehat{E_iN_j}(\mathbf{s}_{jt})$ for EN_j into (2.56), and the structural parameter vector $\beta = (\beta^E, \beta^\omega, \beta^a, \beta^N)$ can be estimated as a standard logit model.[35] The simplicity of the approach is the way in which it takes a potentially complicated model with game-theoretical interactions and boils it down to the estimation of a much more tractable decision-theoretical model. Notice that in the estimation procedure just described, it was not necessary even once to solve for the equilibrium.

Having sketched the approach, we now connect it to our earlier discussion of the work of Ahlfeldt et al. (2014), beginning with the issue of how the potential for multiplicity of equilibria factors into the analysis. In Ahlfeldt et al. (2014), no assumptions about equilibrium selection are made, whereas in the two-step approach, it is necessary to assume that the same equilibrium is played conditional on $\mathbf{s}_{jt}$. Ahlfeldt et al. (2014) provide a full-solution approach. In contrast, the two-step approach is a partial-solution method, and the technical simplicities that it delivers are purchased at the cost of an additional assumption.

Next, recall that Ahlfeldt et al. (2014) are very flexible about allowing for unobserved natural advantage. But ultimately, the paper is able to do this because of the information obtained from the quasi-experimental variation of the Berlin Wall going up and coming back down. The two-step method assumes that the econometrician sees $\mathbf{s}_{jt}$, which is everything except for the private temporary firm-specific shocks ε_{ijt}^{E} and ε_{ijt}^{N}. This limitation is a serious one, because the natural expectation is that industry participants have information about locations that an econometrician would not see. Recent work has generalized the two-step approaches to allow for an unobserved, persistent, location-specific quality shock (see Aguirregabiria and Mira, 2007; Arcidiacono and Miller,

[35] Bajari et al. (2010b) provide a useful treatment of nonparametric approaches to estimating static models of interactions.

2011; and the discussion in Aguirregabiria and Nevo, 2013). The approach can be viewed as a random effects formulation as opposed to a fixed effect formulation. In particular, permanent location-specific unobserved shocks themselves are not identified, but rather the distribution of the shock is identified. For example, if the pattern in the data is that some locations tend to have persistently low entry levels while other locations have persistently high entry levels, holding fixed the same observable state $\mathbf{s}_{jt}$, this would be rationalized by some dispersion in the random effect.

Two-step approaches have been applied to some topics in urban and regional economics, albeit in only a limited number of cases so far. One example is the work of Suzuki (2013), which uses the approach to examine how land use regulations affect entry and exit in the hotel industry. Another is the work of Bayer et al. (2012), which uses this kind of approach to estimate a model of the demand for housing. In the model, homeowners have preferences over the characteristics of their neighbors and so have to forecast how a neighborhood will evolve. This approach is analogous to a firm making an entry decision in a market and forecasting whether subsequent entry will take place.

An interesting aspect of the two-step approach is the way it provides a bridge between structural estimation and descriptive work. The essence of the first stage is the description of behavior. Yet from this approach, the description of behavior has an interpretation in terms of an equilibrium relationship in a formal model.

2.5. CONCLUSIONS

Structural estimation requires creativity and tenancy; good economic modeling skills; a deep understanding of econometric methods; computational, programming, and data management skills; and an interest in and understanding of public policy. We hope that this survey article will inspire other researchers who are not afraid to work on hard and challenging problems to explore structural estimation approaches in urban economics.

Moving forward, it is not too hard to predict that computer-aided decision making will play a much larger role in the future. Computational capacities, in terms of both software and hardware, will continue to improve. This capacity will provide researchers with the opportunity to develop more powerful algorithms designed to solve complex and challenging problems. By combining the computational power and accuracy of machines with human ingenuity and creativity, we will able to solve problems that seem completely intractable at this point.

Structural estimation can be viewed as one compelling method for providing quantitative models and algorithms that can be used within a broader framework of decision support systems. In other areas of economics, such as asset pricing and portfolio management, consumer demand analysis, or monetary policy, structurally estimated models are already commonly used to help households, firms, and government agencies make more informed decisions. The challenge is to develop quantitative models in urban and regional economics that are equally successful. The next generations of urban economists will need to rise to this challenge.

ACKNOWLEDGMENTS

We thank Nate Baum-Snow, Gilles Duranton, Dennis Epple, Vernon Henderson, Andy Postlewaite, and Will Strange for helpful discussions and detailed comments. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis, the Federal Reserve Board, or the Federal Reserve System.

REFERENCES

Aguirregabiria, V., Mira, P., 2007. Sequential estimation of dynamic discrete games. Econometrica 75, 1–53.
Aguirregabiria, V., Mira, P., 2010. Dynamic discrete choice structural models: a survey. J. Econom. 156, 38–67.
Aguirregabiria, V., Nevo, A., 2013. Recent developments in empirical IO: dynamic demand and dynamic games. In: Acemoglu, D., Arellano, M., Deckel, E. (Eds.), Advances in Economics and Econometrics. In: Tenth World Congress, vol. 3. Cambridge University Press, Cambridge, pp. 53–122.
Ahlfeldt, G., Redding, S., Sturm, D., Wolf, N., 2014. The economics of density: evidence from the Berlin Wall. NBER Working paper 20354, July 2014.
Anderson, J., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. Am. Econ. Rev. 93, 170–192.
Arcidiacono, P., Miller, R., 2011. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. Econometrica 79, 1823–1867.
Bajari, P., Kahn, M.E., 2005. Estimating housing demand with an application to explaining racial segregation in cities. J. Bus. Econ. Stat. 23, 20–33.
Bajari, P., Hong, H., Krainer, J., Nekipelov, D., 2010a. Estimating static models of strategic interactions. J. Bus. Econ. Stat. 28, 469–482.
Bajari, P., Hong, H., Ryan, S., 2010b. Identification and estimation of a discrete game of complete information. Econometrica 78, 1529–1568.
Baum-Snow, N., Pavan, R., 2012. Understanding the city size wage premium. Rev. Econ. Stud. 79, 88–127.
Bayer, P., 2001. Exploring differences in the demand for school quality: an empirical analysis of school choice in California, Working paper.
Bayer, P., Timmins, C., 2005. On the equilibrium properties of locational sorting models. J. Urban Econ. 57, 462–477.
Bayer, P., McMillan, R., Rueben, K., 2004. The causes and consequences of residential segregation: an equilibrium analysis of neighborhood sorting, Working paper.
Bayer, P., Ferreira, F., McMillan, R., 2007. A unified framework for measuring preferences for schools and neighborhoods. J. Polit. Econ. 115, 588–638.
Bayer, P., McMillan, R., Murphy, A., Timmins, C., 2012. A dynamic model of demand for houses and neighborhoods, Working paper.
Behrens, K., Mion, G., Murata, Y., Sudekum, J., 2013. Spatial frictions. IZA DP Working paper No. 7175.
Benabou, R., 1996a. Equity and efficiency in human capital investments: the local connection. Rev. Econ. Stud. 63, 237–264.
Benabou, R., 1996b. Heterogeneity, stratification and growth: macroeconomic effects of community structure and school finance. Am. Econ. Rev. 86, 584–609.
Benabou, R., 2002. Tax and education policy in a heterogeneous-agent economy: maximize growth and efficiency? Econometrica 70, 481–517.

Bernard, A., Eaton, J., Jensen, J.B., Kortum, S., 2003. Plants and productivity in international trade. Am. Econ. Rev. 93, 1268–1290.
Berry, S., 1994. Estimating discrete-choice models of product differentiation. Rand J. Econ. 25, 242–262.
Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.
Berry, S., Linton, O., Pakes, A., 2004. Limit theorems for estimating parameters of differentiated product demand systems. Rev. Econ. Stud. 71, 613–654.
Bewley, T.F., 1981. A critique of Tiebout's theory of local public expenditures. Econometrica 49, 713–740.
Bishop, K., 2011. A dynamic model of location choice and hedonic valuation, Working paper.
Bresnahan, T.F., Reiss, P.C., 1991. Entry and competition in concentrated markets. J. Polit. Econ. 99, 977–1009.
Brock, W., Durlauf, S., 2001. Discrete choice with social interactions. Rev. Econ. Stud. 68, 235–260.
Calabrese, S., Epple, D., Romer, T., Sieg, H., 2006. Local public good provision: voting, peer effects, and mobility. J. Public Econ. 90, 959–981.
Calabrese, S., Epple, D., Romano, R., 2012. Inefficiencies from metropolitan political and fiscal decentralization: failures of Tiebout competition. Rev. Econ. Stud. 79, 1081–1111.
Coate, S., 2011. Property taxation, zoning, and efficiency: a dynamic analysis. NBER Working paper 17145.
Combes, P., Duranton, G., Gobillon, L., 2011. The identification of agglomeration economies. J. Econ. Geogr. 11, 253–266.
Combes, P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2012. The productivity advantages of large cities: distinguishing agglomeration from firm selection. Econometrica 80, 2543–2594.
Donaldson, D., forthcoming. Railroads of the Raj: Estimating the impact of transportation infrastructure. Am. Econ. Rev.
Duranton, G., Puga, D., 2015. Urban land use. In: Duranton, G., Henderson, J.V., Strange, W. (Eds.), Handbook of Regional and Urban Economics, vol. 5. Elsevier, Amsterdam, pp. 467–560.
Duranton, G., Morrow, P., Turner, M., 2014. Roads and trade: evidence from the US. Rev. Econ. Stud. 81 (2), 681–724.
Durlauf, S., 1996. A theory of persistent income inequality. J. Econ. Growth 1, 75–93.
Durlauf, S., 2004. Neighborhood effects. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam, pp. 2173–2242.
Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70, 1741–1779.
Ellickson, B., 1973. A generalization of the pure theory of public goods. Am. Econ. Rev. 63, 417–432.
Epple, D., Platt, G., 1998. Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes. J. Urban Econ. 43, 23–51.
Epple, D., Romer, T., 1989. On the flexibility of municipal boundaries. J. Urban Econ. 26, 307–319.
Epple, D., Romer, T., 1991. Mobility and redistribution. J. Polit. Econ. 99, 828–858.
Epple, D., Sieg, H., 1999. Estimating equilibrium models of local jurisdictions. J. Polit. Econ. 107, 645–681.
Epple, D., Filimon, R., Romer, T., 1984. Equilibrium among local jurisdictions: toward an integrated treatment of voting and residential choice. J. Public Econ. 24, 281–308.
Epple, D., Filimon, R., Romer, T., 1993. Existence of voting and housing equilibrium in a system of communities with property taxes. Reg. Sci. Urban Econ. 23, 585–610.
Epple, D., Romer, T., Sieg, H., 2001. Interjurisdictional sorting and majority rule: an empirical analysis. Econometrica 69, 1437–1465.
Epple, D., Gordon, B., Sieg, H., 2010a. Drs. Muth and Mills meet Dr. Tiebout: integrating location-specific amenities into multi-community equilibrium models. J. Reg. Sci. 50, 381–400.
Epple, D., Gordon, B., Sieg, H., 2010b. A new approach to estimating the production function for housing. Am. Econ. Rev. 100, 905–924.
Epple, D., Peress, M., Sieg, H., 2010c. Identification and semiparametric estimation of equilibrium models of local jurisdictions. Am. Econ. J. Microecon. 2, 195–220.
Epple, D., Romano, R., Sieg, H., 2012. The life cycle dynamics within metropolitan communities. J. Public Econ. 96, 255–268.
Epple, D., Jha, A., Sieg, H., 2014. Estimating a game of managing school district capacity as parents vote with their feet, Working paper.

Fernandez, R., Rogerson, R., 1996. Income distribution, communities, and the quality of public education. Q. J. Econ. 111, 135–164.
Fernandez, R., Rogerson, R., 1998. Public education and income distribution: a dynamic quantitative evaluation of education-finance reform. Am. Econ. Rev. 88, 813–833.
Ferreira, F., 2009. 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. Am. Econ. Rev. 97, 789–817.
Fisher, R., 1935. Design of Experiments. Hafner, New York.
Galliani, S., Murphy, A., Pantano, J., 2012. Estimating neighborhood choice models: lessons from a housing assistance experiment, Working paper.
Geyer, J., Sieg, H., 2013. Estimating an model of excess demand for public housing. Quant. Econ. 4, 483–513.
Glomm, G., Lagunoff, R., 1999. A dynamic Tiebout theory of voluntary vs involuntary provision of public goods. Rev. Econ. Stud. 66, 659–677.
Goodspeed, T., 1989. A reexamination of the use of ability-to-pay taxes by local governments. J. Public Econ. 38, 319–342.
Gould, E., 2007. Cities, workers, and wages: a structural analysis of the urban wage premium. Rev. Econ. Stud. 74, 477–506.
Hansen, L.P., Singleton, K., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50, 1269–1286.
Hastings, J., Kane, T., Staiger, D., 2006. Paternal preferences and school competition: evidence from a public school choice program, Working paper.
Heckman, J., MaCurdy, T., 1980. A life cycle model of female labour supply. Rev. Econ. Stud. 47, 47–74.
Henderson, J.V., Thisse, J.F., 2001. On strategic community development. J. Polit. Econ. 109, 546–569.
Holmes, T.J., 2005. The location of sales offices and the attraction of cities. J. Polit. Econ. 113, 551–581.
Holmes, T., 2011. The diffusion of Wal-Mart and economies of density. Econometrica 79, 253–302.
Holmes, T., Stevens, J., 2014. An alternative theory of the plant size distribution, with geography and intra- and international trade. J. Polit. Econ. 122, 369–421.
Hotz, J., Miller, R., 1993. Conditional choice probabilities and estimation of dynamic models. Rev. Econ. Stud. 60, 497–529.
Judd, K., 1998. Numerical Methods in Economics. MIT Press, Cambridge.
Keane, M., Wolpin, K., 1997. The career decisions of young men. J. Polit. Econ. 105, 473–523.
Kennan, J., Walker, J., 2011. The effect of expected income on individual migration decisions. Econometrica 79, 211–251.
Lucas Jr., R.E., 1976. Econometric policy evaluation: a critique. In: Brunner, K., Meltzer, A. (Eds.), The Phillips Curve and Labor Markets, Carnegie-Rochester Conference Series on Public Policy, vol 1. American Elsevier, New York, pp. 19–46.
Lucas Jr., R.E., Rossi-Hansberg, E., 2002. On the internal structure of cities. Econometrica 70, 1445–1476.
Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. 60, 531–542.
McFadden, D., 1974. The measurement of urban travel demand. J. Public Econ. 3, 303–328.
McFadden, D., 1978. Modelling the choice of residential location. In: Karlqvist, A., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Planning Models. Elsevier North-Holland, Amsterdam, pp. 531–552.
Murphy, A., 2013. A dynamic model of housing supply, Working paper.
Nechyba, T., 1997. Local property and state income taxes: the role of interjurisdictional competition and collusion. J. Polit. Econ. 105, 351–384.
Nevo, A., 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. J. Econ. Manag. Strateg. 9, 513–548.
Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. Elsevier, Amsterdam, pp. 2111–2245.

Neyman, J., 1923. On the application of probability theory to agricultural experiments: essay on principles. Transl. Stat. Sci. 5, 465–472.
Ortalo-Magne, F., Rady, S., 2006. Housing market dynamics: on the contribution of income shocks and credit constraints. Rev. Econ. Stud. 73, 459–485.
Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1988. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge.
Redding, S., Sturm, D., 2008. The costs of remoteness: evidence from German division and reunification. Am. Econ. Rev. 98, 1766–1797.
Rosenthal, S., Strange, W., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam, pp. 2119–2171.
Rothstein, J., 2006. Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. Am. Econ. Rev. 96, 1333–1350.
Rust, J., 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. Econometrica 55, 999–1033.
Rust, J., 1994. Structural estimation of Markov decision processes. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. Elsevier, Amsterdam, pp. 3081–3143.
Scotchmer, S., 1986. The short-run and long-run benefits of environmental improvement. Public Econ. 30, 61–81.
Sieg, H., Smith, V.K., Banzhaf, S., Walsh, R., 2002. Interjurisdictional housing prices in locational equilibrium. J. Urban Econ. 52, 131–153.
Sieg, H., Smith, V.K., Banzhaf, S., Walsh, R., 2004. Estimating the general equilibrium benefits of large changes in spatially delineated public goods. Int. Econ. Rev. 45, 1047–1077.
Suzuki, J., 2013. Land use regulation as a barrier to entry: evidence from the Texas lodging industry. Int. Econ. Rev. 54, 495–523.
Tiebout, C., 1956. A pure theory of local expenditures. J. Polit. Econ. 64, 416–424.
Todd, P., Wolpin, K., 2006. Assessing the impact of a school subsidy program in Mexico: using a social experiment to validate a dynamic behavioral model of child schooling and fertility. Am. Econ. Rev. 96, 1384–1417.
Tra, C., 2010. A discrete choice equilibrium approach to valuing large environmental changes. J. Public Econ. 94, 183–196.
Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.
Walsh, R., 2007. Endogenous open space amenities in a locational equilibrium. J. Urban Econ. 61, 319–344.
Weintraub, G., Benkard, C.L., Van Roy, B., 2008. Markov perfect industry dynamics with many firms. Econometrica 76, 1375–1411.
Westhoff, F., 1977. Existence of equilibrium in economies with a local public good. J. Econ. Theory 14, 84–112.
Wu, J., Cho, S., 2003. Estimating households' preferences for environmental amenities using equilibrium models of local jurisdictions. Scott. J. Polit. Econ. 50, 189–206.
Yoon, C., 2012. The decline of the Rust Belt, Working paper.