ESTIMATING EQUILIBRIUM MODELS OF SORTING ACROSS LOCATIONS[1]

Patrick Bayer
Yale University
and
NBER

and
Christopher Timmins
Duke University

January 2005

Abstract

With the growing recognition of the role played by geography in all sorts of economic problems, there is strong interest in measuring the size and scope of local spillovers (i.e., simple anonymous agglomeration or congestion effects, or more complicated interactions between individuals or firms of specific types). It is well-understood, however, that such spillovers cannot be distinguished from unobservable local attributes using just the observed location decisions of individuals or firms. We propose an empirical strategy for recovering estimates of spillovers in the presence of unobserved local attributes for a broadly applicable class of equilibrium sorting models. This approach relies on an instrumental variables strategy derived from the internal logic of the sorting model itself. We show practically how the strategy is implemented, provide intuition for our instrumental variables, discuss the role of effective choice-set variation in identifying the model, and carry-out a series of Monte Carlo experiments to demonstrate the instruments' performance in small samples.

Key Words: Local Spillovers, Location Choice, Economic Geography, Natural Advantage, Social Interactions, Network Effects, Endogenous Sorting, Discrete Choice Models, Agglomeration, Congestion

---

[1] We wish to thank Dan Ackerberg, Pat Bajari, Steve Berry, William Brock, Greg Crawford, Steven Durlauf, Penny Goldberg, Vernon Henderson, Han Hong, Michael Keane, Robert McMillan, Stephen Morris, Marc Rysman, Steven Stern, Chris Udry and two anonymous referees for their helpful comments and suggestions. We gratefully acknowledge financial support from the NSF under grant SES-0137289.

1 INTRODUCTION

Models of location choice – whether of firms or households, within or across cities – have long been central to regional and urban economics. From the inter-jurisdictional sorting models of Tiebout (1956) to the models of segregation developed by Schelling (1969, 1971) to the "new economic geography" of Fujita, Krugman, and Venables (2000), a central feature has been the role of *local interactions* or *spillovers*, whereby the payoffs from choosing a location depend in part on the number or attributes of other individuals or firms that choose the same or nearby locations in equilibrium. In some cases, these local spillovers operate through anonymous channels, with payoffs depending upon simply the number of other individuals or firms selecting the same location, while in other circumstances, the attributes of one's neighbors (e.g., race, income, or education in the case of individuals, and industry classification in the case of firms) might matter as well. It is the interplay between these sorts of spillovers and the *natural advantages* embedded in the landscape of alternative locations that can explain, at a regional level, the geographic and size distribution of cities, and at an urban level, the stratification of households across communities on the basis of income, education, and race, neighborhood density patterns, ethnic enclaves, ghettos, and problems of inner-city decay and suburban sprawl.

Ultimately, local spillovers must derive from some underlying mechanism. For example, households may desire to live in large metropolitan areas because of the size and scope of the labor market or the urban amenities that cities provide. At the same time, congestion operating through increased travel times and the increased price of land may detract from the welfare of individuals in large cities. While distinguishing the precise role of each of these mechanisms may be of interest, often the central problem in an empirical application is simply that of distinguishing the aggregate behavioral effect of local spillovers from that of fixed natural advantages that are tied to locations, particularly when the latter are not observed by the researcher. Recent empirical work attempting to distinguish the magnitude of local interactions has focused on subjects as diverse as crime in cities [Glaeser, Sacerdote, and Scheinkman (1996)], racial segregation [Bayer, McMillan, and Rueben (2002)], interjurisdicational sorting related to schooling [Epple and Sieg (1999), Bayer, Ferreira, and McMillan (2003)], human capital spillovers in the labor market [Morretti(2002)], the general equilibrium effects of environmental policy [Sieg et al. (2003), Timmins (2003)], welfare participation [Bertrand et al. (2000)], unemployment spells [Topa (2001)], development economics [Deichmann et al. (2002), Krugman (1995)] and agglomeration economies in firm locations and investment [Henderson (1999)], among many others.

It is well-understood that there is nothing in, for example, just the observation of many people residing in New York City or numerous high-tech firms locating in Silicon Valley that can distinguish between local spillovers and a distribution of underlying natural advantages across locations [Glaeser and Scheinkman (2002)]. Behavioral data alone provide no guidance as to whether such examples of clustering of individuals or firms should be interpreted as evidence of a strong agglomerating force or the inherent desirability of these locations. Ellison and Glaeser (1997) formalize this as an observational equivalence theorem, stating that "the relationship between mean measured levels of concentration and industry characteristics is the same regardless of whether concentration is the result of spillovers, natural advantage, or a combination of the two." In essence, this result is based on the fact that the observed, aggregate decisions in any model of sorting across locations can be entirely accounted for by a vector of location-specific fixed effects, which intermingle the influence of both natural advantages and local spillovers.

In light of this observational equivalence, we propose a novel solution to the problem of distinguishing local spillovers that uses an instrumental variables estimator to decompose this vector of location-specific fixed effects into components attributable to the inherent features of locations (including those that are unobserved by the researcher) and local spillovers. Drawing on empirical techniques originally developed to model differentiated product demand in Industrial Organization applications, we model the location decision of an individual or firm with a discrete choice framework that enables us to cast the problem of distinguishing local spillovers as a standard endogeneity problem in a familiar regression context. The tendency in these models (that is, if OLS were used to estimate this regression) is to overstate the size of estimated agglomeration effects (or to understate the magnitude of congestion effects), mis-attributing the role of desirable unobservable fixed features. This mis-attribution can have important implications when predicting the new equilibrium distributions of individuals or firms after a significant policy change, explicitly measuring the value of agglomeration in cities, looking for evidence of production "clusters" [Porter (2000)], or valuing local public goods or amenities while controlling for the consequences of agglomeration or congestion.

The requirement for an appropriate instrument in this context is a variable that is correlated with the fraction of individuals (or firms) that selects a given location, but which is not correlated with the unobserved fixed attributes of that location. Because the demand for a particular location is affected by not only its own features, but also by the way these features fit into the broader landscape of the available locations, the logic of the choice model itself implies that a function of the fixed attributes of *other* locations ought to serve as an appropriate instrument for the share of individuals that choose a given location. Moreover, the power of this instrument will increase with the variation in the choice set that the researcher observes in the data – variation that will arise naturally when geography plays an important role in individuals' preferences.

Section 2 begins with a brief discussion of the class of equilibrium sorting models to which our estimation strategy applies, laying-out the notation of the random utility framework and proving the existence of a sorting equilibrium. In Section 3, we describe our estimation algorithm and instrumental variables strategy, while in Section 4, we provide additional intuition for these instruments. We highlight the role of choice-set variation in enhancing the information content of our instruments, and provide heuristic examples that demonstrate how this sort of variation can discern the presence of local spillovers in a broad class of sorting models. In Section 5, we provide Monte Carlo evidence on the performance of our instrumental variables strategy. While we apply these instruments in other papers to particular empirical problems [Bayer, McMillan, and Rueben (2002), Timmins (2005)], the purpose of this paper (and of the Monte Carlo application in particular) is to demonstrate that these instruments work well in a "controlled-data" environment, and to provide some intuition for why this is the case. Section 6 concludes.

2 AN EQUILIBRIUM MODEL OF LOCATION CHOICE WITH LOCAL SPILLOVERS

This section sets out an equilibrium model of residential location choice with local spillovers.[2] Consider a setting in which each individual i chooses a location (indexed by j) in order to maximize utility, $U_{i,j}$ given by:[3]

$$U_{i,j} = X'_j \beta_i + \alpha_i \sigma_j + \xi_j + \varepsilon_{i,j} \tag{2.1}$$

where each location j is described by (i) an observable vector of attributes, X_j, (ii) the share of individuals who choose this location j, σ_j, and (iii) a location-specific unobservable ξ_j, which we assume to be invariant to the location decisions made by the individuals in the model. The taste parameters in equation (2.1) may vary with observable individual characteristics, Z_i:

$$\begin{aligned} \alpha_i &= \alpha_0 + Z'_i \alpha_1 \\ \beta_i &= \beta_0 + Z'_i \beta_1 \end{aligned} \tag{2.2}$$

and individuals can have unobserved preferences (over and above the shared component ξ_j) for location j, $\varepsilon_{i,j}$, which are assumed to be distributed independently across individuals according to the distribution $F(\boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma}$ is defined over locations.

When this model is applied to the population as a whole, the inclusion of σ_j allows for anonymous local spillovers that can have a positive (agglomeration) or negative (congestion) effect on utility.[4] When the model is applied to a specific group of individuals, it permits

[2] Developing a model of firm location decisions with local spillovers would require only modest changes to the model specified here. We focus on households to keep the exposition as straightforward as possible.
[3] The basic form of this utility function is based on the random utility model developed in McFadden (1978) and the specification of Berry, Levinsohn, and Pakes (1995), which includes choice-specific unobservable characteristics. We use the linear form for utility to make the issues of estimation and identification as clear as possible, broader forms of the utility function (especially with respect to the treatment of income and the form of the local spillovers) certainly fit within the scope of the estimation procedure presented in this paper.
[4] This utility specification bears a strong resemblance to the class of discrete choice models used by Brock and Durlauf (2001) to identify social interactions. Our specification, however, differs from theirs in an important

self-segregating preferences among the individuals in that group. Ultimately, local spillovers must derive from some underlying mechanism. For example, households may desire to live in large metropolitan areas because of the size and scope of the labor market or the urban amenities that large cities provide. At the same time, the increased congestion may detract from the utility provided by large versus small cities. When such mechanisms are observed in the data, they can be included directly in the utility function. In many empirical settings, however, the mechanisms through which local spillovers operate are more numerous, more difficult to characterize, or less easily measured, and the inclusion of σ_j in the utility function distinguishes the collective magnitude of these local spillovers. By framing these spillovers in the context of the utility specification shown in equation (2.1), the goal of this paper is to provide a general estimation strategy for circumstances where it is not possible to structurally characterize and measure all of the mechanisms through which local spillovers operate. It is also important to emphasize that while we adopt a simple linear form for the role of local spillovers throughout this paper, non-linear forms could also easily be handled within this framework.

Because of the emphasis that we will place below on the role of geography in helping to identify local spillovers, it is important to point out how exactly geography might enter the utility function. While the specific role of geography will vary with the application in question, one way in which it will naturally enter is as a preference for a geographic "bliss point". In models of sorting within a metropolitan area, for example, the geographic

---

way that makes the empirical settings in which the two models can be applied almost completely distinct. In particular, unlike the model Brock and Durlauf (2001) the model written here explicitly includes unobservable location-specific characteristics, ξ_j. As we show below, the inclusion of this error term implies that the model is no longer identified by observed choice behavior alone – a lack of identification that is not surprising given the conclusions of Glaeser and Scheinkman (2002) and Ellison and Glaeser (1997). Consequently, it is in providing a strategy for estimating sorting models in the presence of unobservable fixed attributes of locations that we contribute to the literature concerned with the econometric identification of social interactions. It is also worth noting that, as discussed in the introduction and demonstrated below, the problem of identifying the model when unobservable fixed attributes are included reduces to that of decomposing a choice-specific fixed effect. Thus, the asymptotic properties of the estimation procedure that we propose rely on the sample growing large in the dimension of the number of alternatives. In this way, the types of applications that we have in mind for the framework developed here are those where the researcher observes economic agents selecting one of many possible alternatives (as is often the case in geographic sorting). Consequently, the economic environments we consider will be very different from those that fit within the framework of Brock and Durlauf (2001), which can be applied to models with as few as two alternatives.

distribution of employment serves as a natural starting point for evaluating household preferences over locations, with a measure of commuting distance or time entering the utility function. In models of sorting across metropolitan areas with costly migration, an individual's birthplace defines a natural geographic bliss point, with some function of distance from individual *i*'s birthplace to location *j* entering the utility function. In such cases, geography can enter utility as an interaction between a household characteristic (e.g., birth location) and the location itself. This sort of interaction is easily included in our framework,[5] and can play an important role in identifying our model.

Finally, it is important to note two simplifying assumptions that we maintain throughout this paper. First, we assume that an individual's utility from selecting location *j* is affected only by the characteristics of that location, including the share of individuals who also choose alternative *j*. In general, the model and estimation strategy can be extended to account for the possibility of spillovers across location (i.e., where the attributes of nearby alternatives enter directly into the utility received from choosing location *j*). Second, while it is straightforward to include other endogenous variables in the analysis (the most important of which is a price associated with each location), we ignore the role of prices and other endogenous variables in the analysis presented in this paper in order to focus attention on the key issues concerning the estimation of local spillovers.[6]

Equilibrium

Throughout our analysis, we assume that individual *i*'s vector of unobserved

---

[5] Such a utility function might be given by: $U_{i,j} = X'_j\beta_i + \alpha_i\sigma_j + f(\ell_i, \ell_i, \theta) + \xi_j + \varepsilon_{i,j}$ where $f(\ell_i, \ell_j, \theta)$ is a function of the geographic location of alternative *j*: ℓ_j and the geographic bliss point of individual *i*: ℓ_i.

[6] There are two approaches to dealing with endogenously determined prices as choice attributes. If prices differ both spatially and with observable attributes of decision-makers, their unbiased effect on utility can be recovered in the first (i.e., maximum likelihood) stage of the two-stage estimator described in Section 3, at which point unobserved local attributes are controlled for non-parametrically with local fixed effects. Alternatively, if prices do not vary with individual attributes, their utility effect can be disentangled from that of any unobservable attributes in the second stage of that procedure by making use of additional instruments (e.g., determinants of labor demand in the case of endogenously determined local wages, or local construction costs in the case of endogenously determined housing prices). The situation is similar to that encountered in many hedonic applications with endogenously determined attributes. See Bayer, McMillan, and Rueben (2002) and Timmins (2005) for examples of these approaches to dealing with endogenous prices.

preferences $\bar{\varepsilon}_i$ is observed by all of the other individuals in the model, and that agents play a static simultaneous-move game according to a Nash equilibrium concept. Moreover, we assume that a continuum of individuals with different unobserved preferences exists for each vector of observed characteristics Z_i that occurs in the world. This assumption (which is essentially that the number of agents is sufficiently large to avoid integer problems) ensures that the unobserved components of preferences can be integrated out.[7] The resulting choice probabilities depict the distribution of location decisions that would result from a continuum of individuals with a given set of observed characteristics Z_i, each responding to its particular unobserved preferences.

Given the utility specification described in equation (2.1), the probability $P_{i,j}$ that individual i chooses alternative j can be written as a function of the full vectors of choice characteristics (both observed and unobserved) and individual i's observed characteristics Z_i:

$$(2.3) \quad P_{i,j} = g_{i,j}(Z_i, X, \sigma, \xi) \quad \forall\ i, j$$

Aggregating these probabilities over all individuals yields the share of individuals choosing location j:

$$(2.4) \quad \sigma_j = \int g_{i,j}(Z_i, X, \sigma, \xi)\, dh(Z) \quad \forall\ j$$

which can be summarized in vector notation as:

$$(2.5) \quad \sigma = g(h(Z), X, \sigma, \xi; \theta)$$

[7] Note that it is possible to incorporate other assumptions concerning the nature of idiosyncratic preferences and the equilibrium concept within this framework. We could, for example, treat each household's idiosyncratic preferences as private information and relax the assumption that each household observed in the data represents a continuum of other households. In this case, the choice probabilities correspond to the expected decisions of other agents (possibly masking important elements of strategic interactions between households). Seim (2001) uses this interpretation of the error structure, along with a Bayesian-Nash equilibrium concept, in estimating a model of entry in retail markets. In developing the theoretical properties of the equilibrium, the estimation procedure, and the identification strategy, we work with the interpretation of ε_i specified above.

where $h(\mathbf{Z})$ is the density of individual characteristics in the population. This system of equations implicitly defines the vector of population shares $\boldsymbol{\sigma}$, and maps $[0,1]^J$ into itself, where J is the total number of alternatives in the discrete choice set In this perfect information setting, a *sorting equilibrium* is defined to be a set of individual location decisions that are each optimal given the location decisions of all other individuals in the population. In Bayer and Timmins (2005), we demonstrate that such an equilibrium always exists and describe conditions under which a unique equilibrium obtains.[8] In general, conditional on the other elements of the model, multiple equilibria arise in the presence of a large enough agglomeration effect, thus there is no way to demonstrate *a priori* that a unique equilibrium obtains in a given empirical environment unless one is willing to assume a congestion effect.

3 ESTIMATION

We now describe a procedure for estimating the local spillovers model just specified and relate applicable asymptotic results developed elsewhere in the literature. We begin by introducing some additional notation that simplifies the exposition, summarizing the portion of the value provided by location j that is shared equally by all individuals as a choice-specific constant δ_j (which includes the location-specific unobservable, ξ_j). In this way, the utility function can be rewritten as:

(3.1) $$U_{i,j} = \delta_j + X_j' Z_i \beta_1 + \sigma_j Z_i \alpha_1 + \varepsilon_{i,j}$$

where

(3.2) $$\delta_j = \beta_0 X_j + \alpha_0 \sigma_j + \xi_j$$

[8] We present these results elsewhere rather than including them here in order to maintain the focus on the

The choice-specific constant is essentially a fixed or random effect (depending on the specific application) and, consequently, the observed location decisions alone do not contain enough information to identify the components of δ_j.[9] This is the essence of the observational equivalence result established by Ellison and Glaeser (1997) and, consequently, in writing the likelihood of observing the location decisions in the data, we treat the vector of choice-specific constants as parameters to be estimated.

As we described in the previous section, uniqueness is not a generic feature of the sorting equilibrium. Consequently, the likelihood function L is not well-defined without an additional assumption about how an equilibrium is chosen in the presence of a multiplicity of equilibria. To this end, we write the probability that the equilibrium with shares $\boldsymbol{\sigma}$ arises conditional on the exogenously given data $\{\mathbf{X}, \mathbf{Z}\}$ and parameters $\{\boldsymbol{\delta}, ,\alpha_1, \beta_1, \boldsymbol{\Sigma}\}$ as $P(\boldsymbol{\sigma} \mid \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ represents the conditioning set for the likelihood. The full likelihood function, then, can be written as the product of this equilibrium selection probability and the probability that each individual chooses the location that he is observed to select in the data, conditional on the share of individuals that chooses each location in the selected equilibrium:

$$(3.3) \qquad L = P(\sigma \mid \Omega)\prod_i \prod_j \left(P_i(j \mid \sigma, \Omega)\right)^{I_{i,j}}$$

$I_{i,j}$ is an indicator variable that equals one if individual i chooses alternative j in the data and zero otherwise, and $P_i(j|\boldsymbol{\sigma}, \boldsymbol{\Omega})$ is the probability that individual i chooses location j conditional on the exogenous variables and the equilibrium share of individuals that select each location.

In describing a sorting equilibrium in Section 2, we assumed an economic environment with a large number of individuals, so that each individual could effectively

---

estimation and identification of the equilibrium model in this paper.

[9] It is worth noting that if all choice characteristics relevant for utility were observed (i.e., $\xi_j = 0 \; \forall \; j$), the model would be identified by the observed choice behavior. It is the problem of distinguishing social interactions from such unobserved factors, however, that provides the basis for the Ellison and Glaeser result as

integrate out over the preferences of others when making his or her own location decision. In this way, the probability that a given equilibrium is selected conditional on $\mathbf{\Omega}$ is not affected by any individual's particular tastes but is instead only a function of the full distribution of individual characteristics and tastes. Consequently, the two components of the likelihood function are orthogonal to one another, as the particular individual location decisions that constitute the second component have no effect on the set $\mathbf{\Omega}$ upon which the first component of likelihood function is conditioned. The orthogonality of these two components ensures that it is possible to estimate the parameters of the model using only the second component of the likelihood function, which is simply the likelihood of observing the individual location decisions in the data conditional on both the exogenous characteristics and the observed vector of location population shares, $\boldsymbol{\sigma}$. Put another way, estimation can be based on the assumption that each individual behaves optimally given the collective choices made by other individuals in the equilibrium that has arisen. An added benefit of this strategy is that the procedure does not require the explicit calculation of the model's equilibrium, which significantly reduces the computational burden.[10]

To estimate the full set of the model's parameters, we propose a two-stage estimation procedure. In the first stage, we estimate the vector of choice-specific constants $\boldsymbol{\delta}$, the interaction parameters $\{\alpha_I, \beta_I\}$, and the parameters of the covariance matrix, $\mathbf{\Sigma}$, by maximizing the second component of the likelihood function shown in equation (3.3).[11] The second step in the estimation procedure uses the estimated vector of choice-specific constants, along with a set of appropriate instruments, to estimate equation (3.2) via

well as out analysis.

[10] When the number of economic agents is small and perfect information is assumed (e.g., in a model of firm entry), it is incorrect to estimate the model based on the optimality conditions for each agent conditional on the equilibrium decisions of the other agents. In this case, because the unobservable associated with each agent actually affects the equilibrium that arises, the two components of the likelihood function shown in equation (3.3) are not independent.

[11] When geographic preferences are incorporated explicitly in the utility function such that each individual economic agent has a geographic bliss point that affects the utility associated with choosing a given location, the parameters associated with this form of an interaction between individual and location attributes will be estimated as part of the first stage of the estimation procedure. For example, for the utility specification described in footnote 5 the parameters of the geographic preference function $f(\ell_i, \ell_j, \theta)$ would be estimated in the first stage of the full estimation procedure described here.

instrumental variables regression.[12] It is in estimating this regression that we confront the fact that the share of individuals that select each location (as well as other variables such as the price of each alternative) is determined endogenously as part of the sorting equilibrium.

Instrumenting for Local Spillovers

Given the estimate of the vector $\boldsymbol{\delta}$ obtained from maximizing the probability that each individual chooses the correct alternative, equation (3.2) is simply a regression equation. The logic of the choice process itself, however, ensures that σ_j and ξ_j are correlated, as an increase in the unobserved quality of a location mechanically raises the demand for that location. In order to fully estimate the parameters of the sorting model, therefore, it is necessary to find an instrument for σ_j.[13]

The instruments that we construct arise naturally out of the sorting model when individuals value only the characteristics of their chosen location.[14] In this case, the fixed attributes of other locations, particularly those that are close substitutes in geographic space for the given location, make ideal instruments for the share of individuals that choose location j. In developing this set of instruments, we exploit an inherent feature of the sorting process – that the overall demand (as well as relative demand of different types of individuals) for a particular location is affected by not only the features of the location itself, but also by the way these features fit in to the broader landscape of available alternatives. In particular, the exogenous attributes of other locations influence the sorting equilibrium (and, thereby, the share of individuals who choose a given location), but have no direct effect on utility. In this way, it is natural to use functions of the set of exogenous attributes of location

[12] In this way, the first stage of our estimation procedure is equivalent to standard discrete choice estimation, with the parameters chosen to best fit the observed choices made by individuals in the data, while the second stage corresponds to a linear regression. Consequently, the parameters of the model are globally identified under the regularity conditions that are generally required for standard discrete choice and linear regression models.

[13] While the second stage of our analysis can be viewed as an instrumental variables regression, this does not imply that its parameters are non-parametrically identified. In particular, note that the left-hand-side variable in this regression, $\boldsymbol{\delta}$, is estimated rather than observed in the data. It is in recovering $\boldsymbol{\delta}$ in the first stage of the analysis that the parametric assumption concerning the distribution of ε binds.

[14] This strategy can be extended to allow individuals to value the features of locations in a neighborhood surrounding the chosen location, as long as the geographic extent of this 'neighborhood' is reasonably small.

j and the exogenous attributes of other locations, $f(X_j, X_{-j})$, as instruments for the share of individuals that choose location *j*.

The logic of constructing an instrument in this way is particularly compelling when geography is explicitly incorporated in the analysis. Consider, for example, a specification of geographic preferences whereby individuals have a distinct geographic bliss point, with utility falling in the distance from that point. In this case, each individual will tend to substitute among locations near the bliss point and, consequently, the share of individuals that choose any particular location will be shaped not only by the exogenous characteristics of a particular location, but also by how these compare to other locations close in geographic space. In this way, two locations with the same fixed attributes X_j will attract different numbers of individuals if these locations are surrounded by other locations with very different levels of exogenous characteristics, and forming an instrument for σ_j based on the features of these nearby locations is intuitively appealing.

As this example suggests, many forms of $f(X_j, X_{-j})$ can potentially serve as appropriate instruments for σ_j. In choosing a specific form to use in estimating the model, we are guided by the optimal instrument for σ_j, given in this case by the expected share conditional on the full distribution of exogenous choice and individual characteristics $\{\mathbf{X}, \mathbf{Z}\}$. Because the equilibrium in the sorting model is not generically unique, this conditional expectation is not well defined.

Yet, the general logic of optimal instruments points to a strategy for constructing an instrument that summarizes the impact of the distribution of alternatives in exogenous characteristic space into a single instrument for share. We propose as an instrument the predicted share of each alternative at an estimate of the parameter values with both (i) the vector of unobserved characteristics $\boldsymbol{\xi}$ and (ii) the local spillover parameters α set equal to zero. This corresponds to using the predicted share of individuals that chooses a location based only on the observed, exogenous choice and individual characteristics used in the model, ignoring the role of local spillovers. This instrument provides a measure of the way that the full landscape of possible choices impacts the demand for each alternative, combining this information in a concise manner that is consistent with the economic behavior governed by the choice model itself.

To see the logic of this proposed instrumental variables strategy, consider again a setting in which households have geographic preferences such that each individual has a distinct geographic bliss point. In this case, the nature of the sorting problem ensures that the share of individuals that choose any particular location is shaped primarily by both the characteristics of this location and those of other locations close in geographic space. Our proposed instrument captures exactly this logic but uses only that portion of the share of individuals predicted to choose a particular location based on the *observable, exogenous* features of locations. In this way, given an even spread of geographic bliss points in the population, the magnitude of the instrument for two locations with identical exogenous characteristics X_j will be greater for the location that is surrounded by fewer high-quality alternative locations. Clustering in the distribution of geographic bliss points or variation in the exogenous characteristics of individuals across bliss points only serves to increase the variation in our proposed instrument across locations.[15]

Many of the preference parameters (in particular those associated with geography), are recovered in the first stage of the overall estimation procedure. Consequently, specifying values for these parameters when developing instruments is straightforward. The only other parameters that are needed in order to construct our proposed instrument are the parameters of equation (3.2), which govern household preferences for exogenous location attributes, β_0. Importantly, any initial guess for these parameters can be used to derive valid instruments. Because this initial guess is not likely to be very accurate, however, an iterative procedure that forms new instruments with each iteration based on the estimate of β_0 obtained in the previous iteration can be used to improve efficiency.

Asymptotic Properties of the Estimator

The asymptotic distribution theory for this estimator is developed in Berry, Linton,

[15] While we use individual heterogeneity in geographic preferences as an important leading example of the type of variation that gives rise to effective choice-set variation throughout our analysis, it is important to emphasize that it is not geographic preferences *per se* but rather variation in the interaction of individual characteristics with locational characteristics that is required to generate the effective choice set variation that induces variation

and Pakes (2004). In general, there are two dimensions in which a sample can grow large: I, the number of individuals (micro-data) or simulated individuals (aggregate data), and J, the total number of alternatives.[16] For any sample of alternatives of finite size J the consistency and asymptotic normality of the first-stage estimates $(\boldsymbol{\delta}, \alpha_1, \beta_1)$ follows directly as long as $I \Rightarrow \infty$. If the true vector $\boldsymbol{\delta}$ were used in the second stage of the estimation procedure, the consistency and asymptotic normality of the second-stage estimates (α_0, β_0) would follow as long as $J \Rightarrow \infty$.[17] In practice, ensuring the consistency and asymptotic normality of the second-stage estimates is complicated by the fact the vector $\boldsymbol{\delta}$ is estimated rather than known. Berry, Linton, and Pakes (2002) show that the consistency of the second-stage estimates follows as long as I grows fast enough relative to J such that $J\, log(J)/I$ goes to zero, while asymptotic normality at the rate $\sqrt{J}$ follows as long as J^2/I is bounded. Intuitively, these conditions ensure that the noise in the estimate of $\boldsymbol{\delta}$ becomes inconsequential asymptotically, thereby ensuring that the asymptotic distribution of (α_0, β_0) is dominated by the randomness in $\boldsymbol{\xi}$ as it would be if $\boldsymbol{\delta}$ was known.

4 SOURCES OF VARIATION IN THE DATA AND THE PROPOSED IV STRATEGY

A natural concern with this type of structural estimation procedure is that the parameters of the model, and especially those associated with local spillovers, might be mechanically shaped by the researcher's assumptions concerning the functional form of the utility function and the distribution of the error term. In this section, we demonstrate that, while structure does play a role in the estimation, under appropriate circumstances, it is underlying variation in the data that drives the identifying variation in our proposed instrument.

---

in our proposed instrument.

[16] If data are observed from multiple markets as in the Monte Carlo simulations below, J indicates the total number of alternatives. This means that the methodology presented here is applicable in settings where individuals choose from a large set of alternatives in a single market or a smaller number of alternatives in each of many distinct markets or choice environments.

Consider first an empirical setting in which a researcher possesses only *aggregate* shares for a single cross section of data in which all individuals have identical geographic preferences. In this case, there is no hope of distinguishing heterogeneity in tastes from the idiosyncratic term $\varepsilon_{i,j}$ and, consequently, we consider only how the following restricted version of the model is identified:

$$(4.1)\qquad U_{i,j} = X_j'\beta + \alpha\sigma_j + \xi_j + \varepsilon_{i,j} = \delta_j + \varepsilon_{i,j}$$

Without variation in the choice set or any other form of individual heterogeneity (such as variation in geographic preferences), it is impossible to estimate any aspect of the distribution of $\varepsilon_{i,j}$ using only observed location decisions and, consequently, a researcher following the two-step estimation procedure that we propose above would be required to specify $\varepsilon_{i,j}$'s exact distribution. Consider the case in which the researcher assumes that $\varepsilon_{i,j}$ is distributed according to the Weibull distribution, giving rise to the multinomial logit model. In this model, the share of individuals that choose each alternative is given by:

$$(4.2)\qquad \sigma_j = \frac{\exp\left(X_j'\beta + \alpha\sigma_j + \xi_j\right)}{\sum_k \exp\left(X_k'\beta + \alpha\sigma_k + \xi_k\right)}$$

and for an initial guess of $\hat{\beta}$ our proposed instrument is given by:

$$(4.3)\qquad \tilde{\sigma}_j = \frac{\exp\left(X_j'\hat{\beta}\right)}{\sum_k \exp\left(X_k'\hat{\beta}\right)}$$

Since the denominator in equation (4.3) is identical for all alternatives, the useful information in the instrument will come simply from the non-linear transformation of the exogenous

[17] This statement assumes a number of additional regularity conditions [see Berry, Linton, and Pakes (2002)].

characteristics, $\exp\left(X'_j\hat{\beta}\right)$. Consequently, the only information contained in the instrument over and above the exogenous choice characteristics X_j derives from the assumed distribution of $\varepsilon_{i,j}$; i.e., no other source of variation in the data helps to estimate local spillovers. An obvious concern, then, in attempting to estimate the model without any effective variation in the choice set or in geographic preferences is that the parameter estimates are not robust to mis-specification of the utility function (e.g., the inclusion of higher-order terms) or the distribution of the error term.

In our opinion, therefore, it is only reasonable to estimate local spillovers when one observes some form of effective variation in the choice set. Effective variation in the choice set can arise in data drawn from multiple geographically-distinct markets, a single market observed over many periods, or variation in the orientation of individuals within a single market (we discuss this last form of variation in greater detail below). In this way, we agree with the implication of the observational equivalence result of Ellison and Glaeser (1997) when one observes only a single cross-section of data with no effective choice set variation. As we now demonstrate, however, our proposed instrument has empirical content that goes beyond the non-linear transformation implied by the distribution of $\varepsilon_{i,j}$ when the researcher observes effective variation in the choice set

Consider, for example, the case in which the researcher has aggregate data from multiple geographically-distinct markets. A researcher might observe, for example, households' choices of community or neighborhood in each of a number of distinct metropolitan areas (see, for example, Bajari and Kahn (2001)). Imagine again attempting to identify the utility function:

(4.4) $$U_{i,j}^m = X'^m_j\beta + \alpha\sigma_j^m + \xi_j^m + \varepsilon_{i,j}^m$$

where the superscript indicates that individual i is choosing among alternatives in market m. Again for simplicity, consider the case in which the researcher assumes that $\varepsilon_{i,j}{}^m$ is distributed according to the Weibull distribution. In this case, the share of individuals in market m that chooses each alternative is given by:

(4.5) $$\sigma_j^m = \frac{\exp\left(X_j'^m \beta + \alpha\sigma_j^m + \xi_j^m\right)}{\sum_{k\in m}\exp\left(X_k'^m \beta + \alpha\sigma_k^m + \xi_k^m\right)}$$

and for an initial guess of $\hat{\beta}$ our proposed instrument is given by:

(4.6) $$\tilde{\sigma}_j^m = \frac{\exp\left(X_j'^m \hat{\beta}\right)}{\sum_{k\in m}\exp\left(X_k'^m \hat{\beta}\right)}$$

Notice that, unlike in the single market case, the denominator in equation (4.6) varies across markets and, consequently, rather than reducing to a simple non-linear transformation of X_j governed by the assumed distribution of $\varepsilon_{i,j}$, the instruments in this case are also affected by the particular set of choices available within each market The variation in our proposed instrument is therefore determined in part by non-linearities implied by the assumption about the error distribution, but also by variation in the choice set across markets.[18] With enough variation in the choice set, then, the majority of the additional variation in the instrument over and above the included exogenous characteristics X_j will be driven by differences in the choice set across individuals.[19, 20]

A particularly intriguing form of effective choice set variation given the geographic

[18] It is important to point out that our identification strategy requires that at least three alternatives with distinct values of X be available in some markets. This is because only differences in the utility provided by the choices in a given market are identified in discrete choice models (i.e., individual choice behavior is invariant to the addition of a constant to the utility provided by all alternatives).

[19] Again it is worth noting in the context of equation (4.6) that because the variation in the denominator is a function of the other choices available in market m, our proposed instrument has information content over and above a non-linear transformation of the variables included in the utility function even if the utility function itself were specified as a non-linear function of the underlying exogenous and endogenous characteristics (X_j and σ_j) for alternative j: $U_{i,j} = f(X_j, \beta_i) + g(\sigma_j, \alpha_i) + \xi_j + \varepsilon_{i,j}$.

[20] The form of the instrument in equation (4.6) makes it clear that there is a special case in which the denominator does not vary given variation in choice sets across markets – i.e., when the choice set changes in such a way as to hold the denominator constant. Thus, when only aggregate shares are observed for choices across markets, our proposed identification strategy requires effective choice set variation in the form of variation in the denominator. For almost any data generating process this will occur with probability one.

nature of many empirical location choice problems derives from variation in the orientation of individuals within a single market or among a single set of alternatives. This form of variation arises, for example, when individuals choose from an identical set of alternatives but have geographic preferences governed by a distinct geographic bliss point which varies across individuals.[21] In this case, the geographic distribution of bliss points generates effective variation in the choice set, as each individual views the set of alternatives conditional on his or her own perspective as governed by the bliss point. In this way, when geographic preferences can be accounted for explicitly in the location choice problem, they will provide a source of variation in the data needed to distinguish local spillovers from unobservable location-specific attributes and, consequently, the logic of the underlying sorting problem itself holds the key to solving the central endogeneity problem that arises in a broad class of sorting models.

A Pair of Heuristic Examples

We conclude this discussion with a pair of examples demonstrating that, given enough data with varying choice sets, one could distinguish local spillovers from *any* general form of the utility function and alternative distributional assumptions on unobserved tastes. To provide the intuition for how agglomeration effects might be distinguished, we begin by considering an example akin to the 1992 US presidential election in which a location should be interpreted as a position in policy, rather than geographic, space. Consider a political election with two candidates, A and B. If candidate A is more popular in the absence of local spillovers, the presence of an agglomeration effect (which here can be interpreted as a preference for picking, or putting one's financial resources behind, the winner) will tend to increase his share of the vote. Now consider the introduction of a third candidate, C. In the presence of the agglomeration effect, the introduction of C can actually increase the share of individuals that choose B. This is especially likely when C is a close substitute for A, thereby drawing most of his share away from the originally more popular candidate A and,

[21] Again it is important to emphasize that it is variation in the interaction of individual characteristics with locational characteristics that is required to generate the effective choice set variation. Individual heterogeneity

consequently, reducing the pull of the agglomeration effect that initially enhanced candidate A's share of the votes relative to B. Such an increase in the share of individuals selecting a particular alternative is impossible in a standard random utility model, *no matter what the functional form of utility or the assumed error structure*. In this way, the presence of an agglomeration effect produces substitution patterns that are distinguishable from higher order forms of the components of the utility function (e.g., introducing quadratic terms) and from alternative assumptions about the error distribution (e.g., random coefficients).

Showing how the presence of a congestion effect can lead to a violation of this basic property of random utility models requires a more subtle example. In particular, consider a residential choice between two locations, A and B, with characteristics (X_1, X_2). Suppose location A is relatively abundant in the characteristic X_1, while B has relatively more of X_2. Imagine also that there are an equal number of individuals of two types, those that have a strong preference for X_1 relative to X_2 and vice-versa. In the absence of any local spillover, one would generally expect individuals to sort across locations by type in such a way that there will be an equal number of individuals selecting each. The presence of a pure congestion effect would have very little impact on this initial allocation, as the roughly equal proportion of individuals selecting each location would leave little room for individuals to spread more evenly in response to their distaste for congestion. Now consider the introduction of a close substitute for location A, which, as in the previous example, draws most of its residents away from A in the absence of local spillovers. In the absence of a congestion effect, one would expect little change in the choices of individuals with a strong preference for X_2 relative to X_1. In the presence of a congestion effect, however, the shrinking share of individuals choosing location A will tend to draw some individuals from location B. This, as in the previous example, is a clear violation of the strict substitutability of alternatives in a classical random utility model, as the introduction of a new location in this case increases the (type-specific) share of individuals that choose an existing location.

Given a correctly specified model, the non-linearities implied by the discrete choice problem ensure that our model will be parametrically identified. From a practical point-of-

---

in geographic preferences is a leading example of such variation.

view, however, the identification of a particular model specification based on such non-linearities always invites questions concerning model mis-specification. Instead, one would like to know the source of variation in the data that ties down the parameter estimates. These examples demonstrate that the presence of local spillovers can lead to substitution patterns that cannot arise in the classical random utility model, no matter what the form of the utility function or the distribution of unobservable tastes. This suggests that it is generally possible to distinguish local spillovers from a classical random utility framework as long as the researcher is able to learn about substitution patterns – that is, as long as the researcher observes enough variation in the choice set. Our proposed instrument is intended to capture this variation. In the following section, we provide Monte Carlo evidence on the relative bias in the estimation of local spillovers under different data generating processes and degrees of variation in the choice set.

5 MONTE CARLO EVIDENCE

In this section, we conduct a series of Monte Carlo experiments designed to evaluate the quality of our proposed instrumental variables strategy in small sample settings, both according to an objective set of criteria and relative to alternative strategies for estimating local spillovers employed elsewhere in the literature. We consider a sorting model with M markets and J locations in each market. Each location is described by a pair of exogenous and observable attributes, $(X_{1,j}^{m}, X_{2,j}^{m})$, and an attribute which is unobserved by the econometrician but known to the individuals engaged in the hypothetical decision-making process, ξ_j^m. $[X_{1,j}^{m}, X_{2,j}^{m}, \xi_j^m]$ are distributed identically and independently across markets and choices according to:

$$(5.1) \qquad \begin{bmatrix} X_{1,j}^{m} \\ X_{2,j}^{m} \\ \xi_j^m \end{bmatrix} \sim \quad i.i.d\, N\left[\begin{pmatrix} \pi_1 \\ \pi_1 \\ \pi_1 \end{pmatrix};\begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{pmatrix}\right]$$

For our initial Monte Carlo exercises, we set $\pi_1 = \pi_2 = \pi_\xi = 0$ and $\sigma_1^2 = \sigma_2^2 = \sigma_\xi^2 = 2$, but alternative values (as well as non-zero off-diagonal elements in the variance-covariance matrix) will also be considered.

We maintain the linear functional form described in the previous section for the utility of individual i, participating in market m by choosing location j:

$$(5.2) \quad U_{i,j}^m = \left(\beta_{01} + \beta_{11} Z_i^m\right) X_{1,j}^m + \left(\beta_{02} + \beta_{12} Z_i^m\right) X_{2,j}^m + \alpha\sigma_j^m + \xi_j^m + \varepsilon_{i,j}^m$$

where we treat Z_i^m as a scalar, and assume that it is drawn from a log-normal distribution with underlying N(0, ½), distributed identically and independently across individuals and markets as well as independently of $[X_{1,j}^m, X_{2,j}^m, \xi_j^m]$. For brevity's sake, we consider a single vector of preferences for exogenous attributes, $(\beta_{01}, \beta_{02}, \beta_{11}, \beta_{12})$ = (1.0, 2.0, 0.3, 0.4), but allow α to take values representing the marginal disutility of congestion ($\alpha < 0$), the marginal utility of agglomeration ($\alpha > 0$), and the case in which no local spillovers are, in fact, present ($\alpha = 0$).

For a given number (M) and size (J) of markets,[22] we calculate the probability distribution of choices over the set of locations for 10,000 individuals, each with a randomly drawn value of Z_i^m. Recall that the asymptotic properties of the estimator hold as the total number of alternatives in the data J x M grows large. In this way, each simulated individual is assumed to represent a continuum of individuals with the same Z_i^m, but who differ in their idiosyncratic unobservables, $\varepsilon_{i,j}^m$, which are assumed to be drawn from a Weibull distribution. Simulating these decisions involves an iterative process whereby we (i) start

[22] We consider two sets of data dimensions, which resemble different empirical contexts but which share a common number of total observations: (J, M) = (100, 10) and (10, 100). The first case, in which individuals choose over a wide array of alternatives but are only observed in a limited set of markets, approximates the data dimensions in most regional economics applications and in Industrial Organization applications where individuals in different years or countries choose over a large number of products (e.g., automobiles), accounting for the decisions of their neighbors out of a desire to conform (i.e., to gain access to a network of qualified repairmen), or to be unique. The second case describes the opposite extreme, in which individuals are spread over many markets and choose amongst a limited set of alternatives. This corresponds to models of choice over alternative information networks [Rysman (1999)], as well as individuals choosing to live in one of a small set of sub-communities in many different SMSA's. Given the arguments made in Section 4, we expect identification to be stronger in the cases with data taken from greater numbers of markets.

with an initial guess at the distribution of individuals across choices in each market (e.g., uniform), (ii) solve for the optimal choices of all individuals given the distribution assumed in (i), (iii) determine new shares for each alternative as defined by these choices, and (iv) return to (ii), inserting the share calculated in (iii) into the utility function. We iterate over these steps until the share of individuals choosing each alternative that enters into the utility function is the same as that which comes out of the aggregate choice process. We repeat this iterative procedure for each of the $m = 1, 2, ..., M$ markets, and then repeat the entire process five-hundred times, once for each simulation run in our Monte Carlo analysis.

Section 3 described our proposed instrument for σ_j^m and how it is to be used in an iterative estimation routine. Restating, we begin with an initial guess (indicated by the superscript 0) at the values of the parameters ($\beta_{01}{}^0$, $\beta_{02}{}^0$, $\beta_{11}{}^0$, $\beta_{12}{}^0$) (e.g., their estimated values if we were to ignore the endogeneity of σ_j^m), and use them to calculate aggregate choice shares, ignoring unobserved choice attributes (ξ_j^m) and local spillovers ($\alpha\sigma_j^m$):

$$(5.3) \qquad \tilde{\sigma}_j^{m,0} = \int \frac{\exp\left(\left(\beta_{01}^0 + \beta_{11}^0 Z_i^m\right) X_{1,j}^m + \left(\beta_{02}^0 + \beta_{12}^0 Z_i^m\right) X_{2,j}^m\right)}{\sum_{k \in m} \exp\left(\left(\beta_{01}^0 + \beta_{11}^0 Z_i^m\right) X_{1,k}^m + \left(\beta_{02}^0 + \beta_{12}^0 Z_i^m\right) X_{2,k}^m\right)} f\left(Z_i^m\right) dZ_i^m$$

We use the vector of $\tilde{\sigma}_j^{m,0}$ as instruments for σ_j^m in the second-stage estimation of equation (3.2), yielding a vector of consistent estimates of all parameters, ($\beta_{01}{}^1, \beta_{02}{}^1, \beta_{11}{}^1, \beta_{12}{}^1$). We then use these parameter estimates to form new fitted shares according to (5.3) and repeat the estimation process.

Tables 1 and 2 describe the outcome of this process for $(J, M) =$ (100, 10) and (10, 100), respectively. For the purpose of comparison, we also report the results of three alternative estimation procedures: (i) *OLS* – a procedure that is identical to ours in every respect, except for ignoring the endogeneity of σ_j^m in the estimation of equation (3.2), (ii) *No Spillovers* – a procedure that simply ignores the presence of local spillovers in equation (3.2) (i.e., constraining $\alpha_0 = 0$), and (iii) *ML Logit* – a procedure that estimates *all* of the model's parameters in a single maximum likelihood routine, treating the Weibull errors, $\varepsilon_{i,j}{}^m$, as the only unobservables (i.e., ignoring the presence of the unobserved attribute, ξ_j^m).

In addition to reporting the mean and standard deviation of the parameters recovered from the five-hundred simulated estimations, we summarize the quality of the estimated local spillover parameter with its mean squared error (MSE), and by reporting the percentage of estimates for which we fail to reject the true underlying value of α at a 5% significance level.[23]

Table 1 summarizes the outcomes of the Monte Carlo procedure applied to the data set containing 10 markets and 100 choices per market in each simulation run. The first thing to note about these results is the precision with which the coefficients on the interactions between individual and choice attributes (i.e., β_{11} and β_{12}) are estimated in the models that allow for an unobservable choice attribute (i.e., the *No Spillovers*, *OLS*, and *IV* models) when such an attribute is, in fact, present.[24] The bias from ignoring the presence of the unobserved attribute is even more apparent when considering the coefficients on the common utility terms (i.e., β_{01} and β_{02}), with the *ML Logit* estimates consistently underestimating the true values of these parameters by more than 68% of their values. Of the three models that allow for an unobserved attribute, the mean *IV* point estimates are quite accurate, and exhibit very little variation across simulation runs.

Turning next to the estimates of the local spillover term, the two models that include the interaction but fail to control for its endogeneity (i.e., *ML Logit* and *OLS*) show strong upward biases in its effect on utility, as expected. The size of this bias, moreover, is greater in the case of congestion than in the case of agglomeration. For each value of α_0, the *IV* estimate is very close to its true value and has a MSE that is typically an order of magnitude smaller than that resulting from the other models.[25] The final column of Table 1 reports the percentage of the simulation runs for which the confidence interval around the point estimate

[23] Unlike the MSE, this second criterion incorporates both the point estimate of α and its standard error in each simulation run. According to this criterion, an estimate of α may still be "good" even if it is far from the true value, if it has a correspondingly large standard error that reflects underlying noisy data.

[24] Note that the first stage of each of these estimation procedures (i.e., in which β_{11} and β_{12} are recovered) is identical. The only difference comes in the second-stage decomposition of δ_j according to equation (3.2), depending upon whether local spillovers are ignored, included but treated as an exogenous variable, or included and instrumented in the manner described above.

[25] Only in the case of the OLS model applied to $\alpha = 3$ does the MSE approach that of the IV estimates.

of α_0 contains that parameter's true value. Note first that, given the way in which the simulated share data were constructed (i.e., based on 10,000 different individuals, each of whom represents a continuum of observationally identical individuals who differ in their values of $\varepsilon_{i,j}{}^{m}$), the *ML Logit* estimates are effectively based on an infinite number of individual observations and thus have zero standard errors.[26] Given that none of the *ML Logit* estimates of α approach the parameter's true value, the percentage of confidence intervals satisfying the criterion in the final column of Table 1 is zero. For the *OLS* estimates, a small number of simulation runs typically satisfy the criterion (particularly when $\alpha \geq 0$). In this case, α is identified by the regression in (3.2) that decomposes δ_j into the utility effects of choice attributes, and its precision depends upon the number of choices and markets rather than on the number of individuals. Increased standard errors relative to the *ML Logit* model mean that some confidence intervals contain the true value of α, but the biases in the *OLS* estimates that are apparent in their MSE's ultimately limit the number of simulation runs where the criterion is satisfied. In the case of the *IV* model, however, this bias is greatly reduced, increasing the number of simulation runs that fail to reject the true value of α to at least 90% in every case.

Turning to Table 2, which summarizes the results derived from the data containing 100 markets and 10 choices in each market, the same trends found in Table 1 are evident. The salient difference between Tables 1 and 2 appears in the increased precision with which the effect of local spillovers are estimated by the *IV* model, as evidenced by the 92% reduction in MSE for every value of α. This confirms the intuition presented in Section 4 for how local spillovers can be distinguished from other choice models by looking at variation in substitution patterns across markets. Data taken from more markets increases this source of variation, with more precise estimates as a result. In the following sub-section, we demonstrate the performance of our estimation strategy in the opposite data environment.

[26] Stated more formally, these estimates are based on a number of individual observations that corresponds to the double machine precision with which shares are measured in Fortran 90. In discrete choice models based on aggregate share data, standard errors are adjusted to reflect the number of individual decision-makers underlying the measured shares (see Berry, Levinson, and Pakes (1995) for an application to automotive

When Does The Estimator Perform Poorly?

The results presented in Tables 1 and 2 highlight the performance of our proposed instrumenting strategy versus simpler estimators that do not properly account for the correlation between σ_j and ξ_j. The results also demonstrate that, given enough variation in the choice set across markets, the parameters estimated using our method match the true parameters very closely. In this section, we focus on the performance of our estimator in less ideal data environments.

This exercise has two purposes. First, we seek to examine the intuition developed in Section 4 concerning the role of functional form versus effective choice set variation in identifying the model's parameters. In that discussion, we anticipated that our estimator would perform poorly in circumstances in which the data exhibited little effective choice set variation. We examine this proposition explicitly here. Second, we highlight the role of functional form by showing how our instrumenting strategy performs when functional form assumptions are incorrect. Together, these simulations will prove valuable in ensuring that our proposed method is not mis-applied to circumstances where it might yield poor results.

Table 3 reports the mean (μ), standard deviation (σ), and mean squared error (MSE) of the estimated coefficient on the spillover term (α) for a number of simulation environments. All of the specifications reported in the table consider data with 100 markets, 10 choices in each market, and a true value of $\alpha = 3$.[27] The table is divided so as to contrast the results (i) with and without choice set variation and (ii) when the model is correctly specified and mis-specified in a variety of ways.

The left-hand columns of the first row repeat the corresponding baseline IV results from Table 2. In the right-hand columns, we report the results for a correctly specified model but constrain the observed individual (Z) and choice (X) attributes to be identical across markets (so that only the draws for ξ_j and $\varepsilon_{i,j}$ vary across markets).[28] As expected, our estimator performs very poorly when cross-market variation in observable, exogenous choice

purchase survey data). In our application, that number is effectively infinite.

[27] The results for the alternative values of α and alternative choice set sizes yielded qualitatively similar conclusions.

[28] Note that we allow individual and choice attributes to vary randomly across choices within a market, but

attributes is eliminated. While the mean of the estimates (μ = 4.21) remains closer to the true value α = 3 than the ML Logit estimates reported in Table 2 (μ = 4.77), the standard deviation and, consequently, the MSE are very large in comparison to those of the baseline specification.

In the remaining rows of Table 3, we contrast the performance of the estimator with and without cross-market variation in X under a variety of alternative specifications for the data generating process. In each, we maintain the baseline model for estimation, implying that the estimation model is mis-specified.

We first consider measurement error in the X's.[29] With cross-market variation in the X's, this has a predictable effect on the estimates of the β parameters, biasing them toward zero (not shown in Table 3). Not surprisingly, the impact on the estimated spillover coefficient goes in the opposite direction, compensating for the change in the contribution made by the X's by increasing the mean estimate of α to μ = 3.03. As in our baseline example, however, our estimator performs substantially better with cross-market variation in the choice set. The increased vulnerability of the estimator to mis-specification in the absence of choice set variation is certainly not surprising, given the complete reliance on functional form for identification in that case.

We next consider an alternative specification in which ξ_j is drawn from a distribution with fatter tails than the normal distribution (e.g., Student's t-distribution with three degrees of freedom). With choice set variation, results are similar to those of the baseline specification. Without choice set variation, however, the MSE grows dramatically. This is to be expected, since the second-stage of the estimation algorithm assumes only that ξ_j is mean-zero, which is still the case. The true variance of ξ_j is increased, however, relative to the variation in the X's. This reduces the precision of the instrumental variables strategy and leads to the larger MSE's.

Our third model mis-specification considers an alternative distribution for the

---

we restrict that random variation to be the same across markets.

[29] In particular, we add an i.i.d. N(0, 0.4) random draw to each X before it is incorporated into the estimation algorithm The variance in each X in the data generating process is 2.0, implying a signal-to-noise ratio of 5.

idiosyncratic error, $\varepsilon_{i,j}$. In particular, we draw this error from the Student's t-distribution with two degrees of freedom, while the estimation model continues to assume that it is taken from the Weibull distribution. This results in a clear upward bias in the estimate of α, even in the presence of cross-market variation in the X's. The standard deviation of these point estimates is small (0.15); combined with the upward bias, this leads to an MSE of 1.38. Without cross-market variation in the choice set, the point estimates remain upwardly biased but become more dispersed, leading to an MSE of 19.35.

Finally, we consider two alternative specifications for the data generating process that replace the linear $X'_j\beta_i$ with non-linear functions: $\frac{(X'_j\beta_i)^3}{200}$ and $\frac{X'_j\beta_i}{\sqrt{1+(X'_j\beta_i)^2}}$ (i.e., mis-specified functions #1 and #2, respectively). The estimation algorithm, however, continues to assume that utility is linear in the X's. Not surprisingly, this has the effect of undermining the performance of the instrumental variables estimator, increasing the mean of the estimates with choice set variation to $\mu = 5.01$ and raising the MSE to 7.08 in the case of mis-specification #1. In the case of mis-specification #2 (a more extreme form of non-linearity) the mean with choice set variation only rises to $\mu = 4.60$, but estimates are highly variable and the MSE explodes to 7,185. The impact is even more extreme in the absence of cross-market variation, in which case the mean and MSE rise to 17.31 and 18,836 in the case of mis-specification #1, and 23.73 and 22,403 in the case of mis-specification #2. Without knowing *a priori* whether any functional form assumption is really correct, this example clearly demonstrates the crucial role played by cross-market variation in our estimation strategy.

Considered together, the results of Table 3 highlight the crucial need for choice set variation in order for our instrumental variables strategy to be successful, confirming our intuition from Section 4. With the exception of the second mis-specified function of the X's, however, the instrumental variables strategy performs reasonably well even in the presence of model mis-specification. The same cannot always be said of the alternative estimation strategies; for example, the mean ML Logit estimate of α in the presence of measurement error in the X's is 5.16 with an MSE of 4.67, even with cross-market variation.

6 CONCLUSION

It is well-established that individuals' or firms' location decisions alone are insufficient to distinguish the behavioral effects of spillovers (i.e., anonymous agglomeration or congestion effects, type-specific social interactions amongst individuals, and industry- or sector-specific spillovers amongst firms) from those of local natural advantages. That is, what a naïve model interprets as an agglomeration effect may simply be the effect of desirable unobservable choice attributes reflected in the decisions of others. At the same time, determining the separate roles of spillovers and natural advantages is at the heart of many questions in regional and urban economics (not to mention labor, public finance, development, environmental economics, and industrial organization). This paper proposes an empirical strategy for recovering these separate determinants of behavior in a broadly applicable class of equilibrium sorting models. That strategy re-casts the problem as one of an endogenous variable in a familiar regression context and relies upon the behavioral model itself to derive instruments based on alternatives' isolation in exogenous attribute space. We provide intuition for that instrumental variables strategy, describe practically how it is implemented, and use a series of Monte Carlo exercises to show that it performs well in a variety of empirical settings and in comparison to other approaches that have been used for similar problems.

While our model specification does rely on a distributional assumption about unobserved tastes, a virtue of our strategy is that the information content of our instruments (and, hence, the power of the model in identifying spillovers) is enhanced by data exhibiting effective variation in the choice set across individuals. We demonstrate this with our Monte Carlo exercises and with a pair of heuristic examples that show that, with sufficient effective choice-set variation, our estimation strategy can distinguish spillovers from any other type of random utility model without relying on any other modeling assumptions. This is an important feature if we are to have confidence in predictions based on our model, even in light of inevitable model mis-specification.

REFERENCES

Bajari, Patrick, and Matthew Kahn (2001), "Why Do Blacks Live in Cities and Whites Live in Suburbs?" Mimeo, Stanford University.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2003). "A Unified Approach for Measuring Preferences for School Quality," Mimeo, Yale University.

Bayer, Patrick, Robert McMillan, and Kim Rueben (2002). "The Causes and Consequences of Residential Segregation: An Equilibrium Analysis of Neighborhood Sorting." Mimeo, Yale University.

Bayer, Patrick and Christopher Timmins (2005). "On the Equilibrium Properties of Locational Sorting Models," *Journal of Urban Economics*. 57:462-477.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile Prices in Market Equilibrium." *Econometrica*. 63:841-890.

Berry, Steven, Oliver Linton, and Ariel Pakes (2002). "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems." *Cowles Foundation Working Paper*. Yale University.

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan, (2000), "Network Effects and Welfare Cultures," *Quarterly Journal of Economics,* August: 1019-1055.

Brock, William and Steven Durlauf (2001). "Discrete Choice with Social Interactions." *Review of Economic Studies*. 68(2):235-260.

Deichmann, Uwe, Marianne Fay, Jun Koo, and Somik Lall (2002). "Economic Structure, Productivity, and Infrastructure Quality in Southern Mexico." Mimeo, The World Bank.

Ellison, Glenn and Edward Glaeser (1997). "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy*. 105(5):889-927.

Epple, Dennis and Holger Sieg (1999). "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy*. 107(4):645-681.

Fujita, Masahisa, Paul Krugman, and Anthony Venables (1999). *The Spatial Economy,* MIT Press: Cambridge.

Glaeser, Edward and Jose Scheinkman (2002). "Non-Market Interactions." Mimeo, Harvard University.

Glaeser, Edward, Bruce Sacerdote, and Jose Scheinkman (1996). "Crime and Social Interactions," *Quarterly Journal of Economics*. 106: 507-548.

Henderson, Vernon (1999). "Marshall's Economies," NBER Working Paper #7358.

Krugman, Paul (1995). *Development, Geography, and Economic Theory*. The MIT Press. Cambridge, MA.

McFadden, Daniel, (1978). "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*. eds. Karlquist, A., et al. Elsevier North-Holland, New York.

Moretti, Enrico (2002). "Human Capital Externalities in Cities." Forthcoming chapter in *Handbook of Urban and Regional Economics*. North-Holland / Elsevier Science.

Porter, Michael (2000). "Location, Competition, and Economic Development: Local Clusters in a Global Economy." *Economic Development Quarterly.* 14(1):15-34.

Rysman, Marc (1999). "Competition Between Networks: A Study of the Market for Yellow Pages." Mimeo, Boston University.

Schelling, Thomas C. (1969). "Models of Segregation." *American Economic Review*. 59(2), 488-93.

________ (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology*. 1:143-186.

Seim, Katja (2001). "Geographic Differentiation and Market Structure." Yale Ph. D. Dissertation.

Tiebout, Charles M. (1956). "A Pure Theory of Local Expenditures." *Journal of Political Economy*. 64: 416-424.

Timmins, Christopher (2005). "If You Can't Take the Heat, Then Get Out of the Cerrado: Estimating the Equilibrium Amenity Costs of Non-Marginal Climate Change in Brazil." Mimeo, Duke University Department of Economics.

Topa, Giorgio, (2001), "Social Interactions, Local Spillovers, and Unemployment." *Review of Economic Studies*. 68:261-295.

Table 1
Monte Carlo Simulation Results
Data Generation: J=100, M=10

True α	Estimation Method	β_{11} (0.3)		β_{12} (0.4)		β_{01} (1.0)		β_{02} (2.0)		α		MSE α	% Fail to Reject True α at 5% Significance
		μ	σ	μ	σ	M	σ	μ	σ	μ	σ		
-3	ML Logit	0.23	0.05	0.24	0.05	0.17	0.15	0.56	0.15	4.87	1.04	63.06	0
	No Interaction	0.30	0.01	0.40	0.02	0.77	0.03	1.71	0.03				
	OLS	0.30	0.01	0.40	0.02	0.97	0.03	1.93	0.04	2.06	1.19	27.00	0
	IV	0.30	0.01	0.40	0.02	1.01	0.03	2.00	0.03	-3.20	1.38	1.95	94
0	ML Logit	0.21	0.05	0.21	0.06	0.20	0.17	0.63	0.18	5.05	0.71	25.98	0
	No Interaction	0.30	0.02	0.40	0.03	1.00	0.03	2.00	0.03				
	OLS	0.30	0.02	0.40	0.03	0.98	0.03	1.96	0.04	3.07	0.82	10.09	3
	IV	0.30	0.02	0.40	0.03	1.00	0.03	2.00	0.04	-0.14	1.06	1.15	95
3	ML Logit	0.22	0.18	0.26	0.32	0.15	0.26	0.53	0.42	6.09	0.96	10.44	0
	No Interaction	0.31	0.11	0.43	0.17	1.27	0.03	2.36	0.03				
	OLS	0.31	0.11	0.43	0.17	0.99	0.08	1.96	0.13	4.50	0.46	2.45	8
	IV	0.31	0.11	0.43	0.17	1.19	0.05	2.00	0.07	2.84	0.95	0.93	90

Table 2
Monte Carlo Simulation Results
Data Generation: J=10, M=100

True α	Estimation Method	β_{11} (0.3)		β_{12} (0.4)		β_{01} (1.0)		β_{02} (2.0)		α		MSE α	% Fail to Reject True α at 5% Significance
		μ	σ	M	σ	μ	σ	μ	σ	μ	σ		
-3	ML Logit	0.22	0.02	0.23	0.03	-0.04	0.03	0.11	0.04	3.71	0.12	44.98	0
	No Interaction	0.30	0.01	0.40	0.02	1.13	0.03	1.75	0.04				
	OLS	0.30	0.01	0.40	0.02	0.84	0.04	1.71	0.05	0.38	0.31	11.49	0
	IV	0.30	0.01	0.40	0.02	1.00	0.04	2.00	0.05	-3.01	0.40	0.16	96
0	ML Logit	0.21	0.03	0.20	0.04	-0.01	0.04	0.18	0.06	3.92	0.10	15.40	0
	No Interaction	0.30	0.02	0.40	0.03	1.27	0.03	2.00	0.04				
	OLS	0.30	0.02	0.40	0.03	0.89	0.04	1.80	0.05	2.05	0.22	4.24	0
	IV	0.30	0.02	0.40	0.03	1.00	0.04	2.00	0.05	-0.01	0.30	0.09	96
3	ML Logit	0.22	0.17	0.24	0.31	0.12	0.22	0.37	0.35	4.77	0.55	3.43	0
	No Interaction	0.31	0.07	0.41	0.11	1.44	0.04	2.30	0.04				
	OLS	0.31	0.07	0.41	0.11	0.92	0.06	1.87	0.09	4.20	0.15	1.47	0
	IV	0.31	0.07	0.41	0.11	0.99	0.06	1.99	0.08	2.98	0.26	0.07	91

Table 3
Monte Carlo Simulation Results
Estimates Of α Under Alternative Data Generation Scenarios and Model Mis-Specifications
J = 10, M = 100, True α = 3

	Model	Data Generating Process ($X_{1,j}$, $X_{2,j}$)					
		With Choice Set Variation			No Choice Set Variation		
		μ	σ	MSE	μ	σ	MSE
	Correctly Specified Model	2.98	0.26	0.07	4.21	28.79	829.22
Mis-Specified Models	Measurement Error in ($X_{1,j}$, $X_{2,j}$)	3.03	0.68	0.46	0.71	45.82	2102.43
	Alternative Distribution for ξ_j	2.83	1.02	1.06	1.67	17.84	319.50
	Alternative Distribution for $\varepsilon_{i,j}$	4.17	0.15	1.38	4.56	4.12	19.35
	Mis-Specified Function of ($X_{1,j}$, $X_{2,j}$) #1	5.01	1.74	7.08	17.31	136.58	18835.96
	Mis-Specified Function of ($X_{1,j}$, $X_{2,j}$) #2	4.60	84.79	7185.44	23.73	148.31	22403.08