PERE
7/29/03
3:30 PM

A UNIFIED FRAMEWORK FOR MEASURING PREFERENCES FOR SCHOOLS AND NEIGHBORHOODS*

Patrick Bayer
Department of Economics
Yale University

Fernando Ferreira
Department of Economics
University of California, Berkeley

Robert McMillan
Department of Economics
University of Toronto

July 2003

Abstract

Estimating the demand for non-marketed goods such as school quality poses challenging endogeneity and selection problems. To address these, this paper develops a comprehensive new approach for recovering the full distribution of preferences for school and neighborhood attributes from observed household choices. The framework has three key features: it recovers structural preference parameters, allowing for a great deal of preference heterogeneity; it deals with the non-random sorting of households across neighborhoods by modeling the sorting process directly, in a manner consistent with equilibrium considerations; and it allows this sorting to be influenced by unobservable choice characteristics, while overcoming significant endogeneity problems using a boundary fixed effects approach. We estimate the model using rich data on a large metropolitan area, drawn from a restricted version of the Census. The new estimates, the most reliable yet to appear in the literature, indicate that on average households are willing to pay an additional one percent in house prices when the average performance of the local school is increased by 5 percent, and also that there is considerable preference heterogeneity. Using our equilibrium framework to explore the general equilibrium implications of these new estimates, we show that the estimated heterogeneity gives rise to substantial variation in the capitalization of school quality in housing prices throughout the metropolitan region. Further, the full capitalization of school quality into housing prices is typically 70-75 percent greater than the direct effect, as the result of a social multiplier. Focusing on the direct effect only, the prior literature has neglected this significant general equilibrium mechanism, whereby increases in school quality also raise housing prices by attracting households with more education and income to the corresponding neighborhood.

* We are grateful to Pat Bajari, Steve Berry, Sandra Black, David Card, Ken Chay, David Cutler, Hanming Fang, David Figlio, David Lee, Tom Nechyba, Kim Rueben, Holger Sieg, Chris Taber, and Chris Timmins and seminar participants at Harvard University, the University of Florida, Yale University, the 2002 Urban School Finance Workshop at the University of Illinois-Chicago, and the 2003 Public Economic Theory Conference for providing many valuable comments and suggestions. We also thank Pedro Cerdan and Jackie Chou for help in assembling the data. This research was conducted at the California Census Research Data Center; our thanks to the CCRDC, and to Ritch Milby in particular. We gratefully acknowledge the financial support for this project provided by the National Science Foundation under grant SES-0137289 and by the Public Policy Institute of California.

1 INTRODUCTION

Economists have long been interested in estimating the demand for non-marketed goods such as school quality, and for good reason. From a policy perspective, recovering an accurate household valuation of school quality allows the direct benefit of education reforms to be quantified. More broadly, estimates of a wider range of underlying preference parameters play an important role in understanding the way that households sort in the housing market, which in turn determines the pattern of residential segregation and the matching of households to schools. Building on a large body of theoretical work studying the effects of household sorting on equilibrium in the housing market,[1] researchers have used equilibrium models more recently to simulate policy changes, notably choice-based education reforms, identifying important general equilibrium effects as households re-sort that do not appear in partial equilibrium.[2] But parameters in this body of research have typically been chosen through calibration or more arbitrarily. Based on direct estimation of demand parameters, more reliable and coherent sets of preference estimates can be used, with the potential to improve our understanding of policy reforms and the workings of the urban economy more generally.

Despite their usefulness, recovering demand parameters from observed choices in the housing market poses considerable challenges for estimation. Most obviously, data requirements are high, and even in the richest datasets, it is not possible to characterize fully the factors that make certain houses and neighborhoods particularly desirable – there will invariably be a component that is unobserved to the econometrician. Given that households sort on a non-random basis, in part influenced by these unobservable choice characteristics, important endogeneity problems arise as relevant choice characteristics that result from the sorting process, such as neighborhood sociodemographic composition, will be correlated with these unobservables. Underlying household preferences themselves are likely to be heterogeneous, depending on a wide range of household characteristics – their education, income, race, family structure, and more. Recovering mean preferences may be of limited value.

Of the two most prominent approaches to estimating demand in the literature, *hedonic price regressions* relate housing prices to housing and neighborhood attributes including school

[1] Much of the intuition for how household sorting affects the equilibrium in the housing market and consequently the matching of households with schools derives from a long line of theoretical work in local public finance. Important contributions to this literature date back to the work of Tiebout (1956) and include more recent research by Epple and Zelenitz (1981), Epple, Filimon, and Romer (1984, 1993), Benabou (1993, 1996), Fernandez and Rogerson (1996), and Nechyba (1997).

[2] Direct analyses of choice-based education reforms have been conducted by Epple and Romano (1998), Nechyba (1999, 2000, 2001), and Fernandez and Rogerson (2003), among others.

characteristics, while *traditional discrete choice models* estimate preferences by attempting to match the location decisions of the households in the data.[3,4] An attractive feature of the recent price regression literature has been its focus on addressing the likely correlation of school quality with unobserved housing and neighborhood quality, making use of a boundary fixed effects strategy.[5] However, it is difficult to determine how the estimates of a hedonic price regression relate to the fundamental preferences in the population, both because the equilibrium price function need not correspond to mean preferences [see Tinbergen (1956) and Rosen (1974)] and because this approach provides no clear way of estimating heterogeneous preferences.

While dealing with heterogeneity in preferences in a straightforward fashion, the discrete choice literature in urban and public economics has traditionally assumed away any systematic, unobserved differences in the quality of houses and neighborhoods. Thus when housing prices are included in the indirect utility function, the positive correlation of price with unobserved housing and neighborhood quality has been routinely ignored, leading to a severe understatement of price elasticities and to biases in the estimation of other taste parameters.[6] Moreover, the likely correlation of test scores with unobserved neighborhood quality has not been adequately addressed in the discrete choice literature.

The current paper makes two primary contributions. First, building on the strengths of the two main approaches in the literature, it sets out a new framework for estimating household preferences over a broad range of housing and neighborhood characteristics, some of which are determined by the way that households sort in the housing market. Here, we bring together in a unified framework the treatment of heterogeneity and selection that has been the focus of the discrete choice literature with the treatment of the correlation of school quality with unobserved neighborhood quality that has been the focus of recent work in the hedonic price regression literature. The framework generalizes the traditional discrete choice approach, allowing for

[3] This framework was first introduced for the study of housing markets by McFadden (1978). Examples related to estimating preferences for school quality include Quigley (1985), Nechyba and Strauss (1998), Bayer (1999), and Barrow (1999).

[4] *Hedonic demand models* studied by Rosen (1974), Epple (1987), Bajari and Benkhard (2002), and Heckman, Matzkin, and Nesheim (2003) among others provide another approach to estimating demand for non-marketed goods and attributes. These models have not been employed to date to estimate demand for schooling, although Nesheim (2003) proposes such an empirical exercise. The fundamental difference between hedonic demand and discrete choice models as applied to location choice problems relates to whether households are assumed to be able to select the level of consumption of each element of the bundle of attributes determined by the location decision in order to satisfy the first order condition associated with that element or whether households are constrained to choose among the set of choices (bundles) that exist in the data.

[5] See, for example, Black (1999), Clapp and Ross (2002), Kain, Staiger, and Samms (2003).

unobserved housing and neighborhood quality and adding a market clearing condition that equilibrates the market. The second contribution is to estimate the model using the richest data available, to obtain a broad range of underlying preference estimates that account the measurement difficulties already referred to. With these estimates in hand, we are then able to use our equilibrium framework to explore their general equilibrium implications, drawing attention to important effects neglected in the prior literature.

The framework nests both of the main approaches as direct restrictions. It is straightforward to show that when households have homogeneous preferences, estimation reduces to a hedonic price regression, with the unobserved quality term in the discrete choice utility function corresponding one-to-one with the error term in the price regression. When households have heterogeneous preferences, estimation proceeds in two stages. The first is a standard maximum likelihood estimation, which returns parameters that characterize heterogeneity in preferences. The second stage, which provides the estimates of mean preferences, consists of an adjusted price regression, where the adjustment (estimated in the first stage) essentially accounts for which individual in the distribution of tastes sets the price of a given attribute.

Consider, for example, the price of a house with a view of the Golden Gate Bridge. If such a view were extremely rare, the difference in price between houses with versus without a view would reflect the marginal willingness-to-pay (MWTP) of a household with a relatively strong taste for a view; consequently the equilibrium price of these houses would be exceedingly high relative to the MWTP of the mean household. In this case, the adjustment term estimated in the first stage of the analysis would decrease the price of houses with a view so that the adjusted price regression would accurately return mean preferences.

That the second stage of the estimation procedure is a modified price regression puts the analysis on a familiar footing relative to the hedonic price regression literature. It makes clear that the use of house price variation that provides the basis for identifying demand via hedonic price regressions also provides the fundamental basis for the recovery of mean preferences in the broader equilibrium choice model. It also makes the application of strategies for handling the endogeneity of school quality developed in the context of hedonic price regressions straightforward in this more general context. In particular, we incorporate the strategy developed in Black (1999), which uses sample of houses in close proximity to boundaries between school attendance zones, including boundary fixed effects in the analysis, thereby comparing the prices

[6] Recent papers by Bayer (1999), Bajari and Kahn (2002), and Bayer, McMillan,and Rueben (2002) include a term that captures the unobserved (to the econometrician) quality of houses in neighborhoods in the utility specification and address the endogeneity of price in this context.

of houses in similar neighborhoods but which fall on opposite sides of a boundary determining where children attend school.

The estimates provided in this paper are the first to appear in the literature addressing both the correlation of school quality and neighborhood sociodemographic characteristics with unobserved neighborhood quality. One difficulty that arises in the standard use of boundary fixed effects for isolating exogenous changes in school quality (and one that we demonstrate for our data) is that households sort with respect to these boundaries, with more highly educated and higher income households generally choosing the better school quality side of a boundary. Consequently, it is important to continue to control for neighborhood sociodemographic characteristics even when boundary fixed effects are employed. This sorting with respect to the boundaries also turns out to be an attractive feature of the data for dealing with the correlation of neighborhood sociodemographic characteristics and unobserved neighborhood quality.[7] By controlling for any unobserved aspects of housing and neighborhood quality that are constant across the boundaries, the use of boundary fixed effects absorbs much of the variation between sociodemographic composition and unobserved neighborhood quality, thereby providing estimates of demand for neighborhood sociodemographics that deal with this difficult endogeneity problem.

We estimate a number of specifications of the model in order to provide a direct comparison of preference estimates under different assumptions. Here, we benefit from using newly available, restricted-access Census microdata that provide the precise residential location of nearly a quarter of a million households in the San Francisco Bay Area. These Census data provide detailed information on a 1-in-7 sample of households, including each household member's race, education, income, age, immigration status, employment status, and job location and each household's residential and employment locations at the level of a Census block (a Census area with approximately 100 residents). Using these new Census data as a centerpiece, we have assembled an extensive data set characterizing the housing market in the San Francisco Bay Area, combining housing, employment, and neighborhood sociodemographic data drawn from the Census with neighborhood-level data on schools, air quality, climate, crime, topography, land use, and urban density.

The new estimates, the most reliable yet to appear in the literature, indicate that on average households are willing to pay an additional one percent in house price when the average

[7] Addressing this correlation is very difficult as one generally requires an instrument that is correlated with a neighborhood's sociodemographic distribution but not its unobserved quality and, to our knowledge, Bayer McMillan, and Rueben (2002) is the only paper in the literature that attempts to deal with this endogeneity problem.

performance of the local school is increased by 5 percent. The estimates also indicate considerable heterogeneity in preferences for school quality, with the presence of children, increased income and education significantly raising a household's willingness to pay for school quality. The inclusion of boundary fixed effects in the analysis significantly reduces the estimated willingness-to-pay measures for more educated, higher-income, and white neighbors. These reductions are exactly what one would expect if these variables in part proxy for unobserved housing and neighborhood quality. Despite these reductions, neighborhood composition continues to play a central role in driving the location decisions and there is substantial variation in preferences for neighborhood composition across the population.

Given the estimates of the full model, the real power of our equilibrium framework is that it provides a natural device for exploring the general equilibrium implications of the estimates we obtain. It offers both a complete characterization of the heterogeneity in preferences for schools and neighborhoods as well as a way of understanding how these aggregate to determine the equilibrium in the housing market. We show that the estimated heterogeneity gives rise to substantial variation in the capitalization of school quality in housing prices throughout the metropolitan region. Further, the full capitalization of school quality into housing prices is typically 70-75 percent greater than the direct effect, as the result of a social multiplier. Focusing on the direct effect only, the vast prior literature on capitalization has neglected this significant general equilibrium mechanism,[8] whereby increases in school quality also raise housing prices by attracting households with more education and income to the corresponding neighborhoods.

The rest of the paper is organized as follows: the next section sets out the model. Section 3 discusses estimation issues. The unique data used to estimate the model are described in Section 4. Estimation results are presented in Section 5, and results of general equilibrium simulations in Section 6. Section 7 concludes.

2 AN EQUILIBRIUM MODEL OF AN URBAN HOUSING MARKET

This section of the paper develops an equilibrium model of a self-contained, urban housing market. The model consists of two key elements: the household residential location decision problem and a market clearing condition. We model the residential location decision of each household as a discrete choice of a single residence. The utility function specification 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. Let

X_h represent the observable characteristics of housing choice h including characteristics of the house itself (e.g., size, age, and type), its tenure status (rented vs. owned), and the characteristics of its neighborhood (e.g., sociodemographic composition, school, crime, and topography). Let p_h denote the price of housing choice h. Each household chooses its residence h to maximize its indirect utility function V_h^i:

$$(1) \qquad \underset{(h)}{Max} \quad V_h^i = \alpha_X^i X_h - \alpha_p^i p_h + \xi_h + \varepsilon_h^i$$

The error structure of the indirect utility is divided into a correlated component associated with each house that is valued the same by all households, ξ_h, and an individual-specific term, ε^i_h. A useful interpretation of ξ_h is that it captures the unobserved quality of each house, including any unobserved quality associated with its neighborhood.

Each household's valuation of choice characteristics is allowed to vary with its own characteristics, Z^i, including education, income, race, employment status, and household composition. Specifically, each parameter associated with housing and neighborhood characteristics and price, α^i_j, for $j \in \{X, p\}$, is allowed to vary with a household's own characteristics,

$$(2) \qquad \alpha_j^i = \alpha_{0j} + \sum_{r=1}^{R} \alpha_{rj} Z_r^i ,$$

and equation (2) describes household i's preference for choice characteristic j.

Given the household's problem described in equations (1)-(2), household i chooses house h if the utility that it receives from this choice exceeds the utility that it receives from all other possible house choices - that is, when

$$(3) \qquad V_h^i > V_k^i \quad \Rightarrow \quad W_h^i + \varepsilon_h^i > W_k^i + \varepsilon_k^i \quad \Rightarrow \quad \varepsilon_h^i - \varepsilon_k^i > W_k^i - W_h^i \qquad \forall \quad k \neq h$$

where W^i_h includes all of the non-idiosyncratic components of the utility function V^i_h. As the inequalities depicted in (3) imply, the probability that a household chooses any particular choice

[8] This has a long history of study in the literature. See, for example, Oates (1969), Kain and Quigley (1975), Hayes and Taylor (1996), Bogart and Cromwell (2000), Black (1999), Figlio and Lucas (2000), Clapp and Ross (2002), and Kane, Staiger, and Samms (2003).

depends in general on the characteristics of the full set of possible house choices. Thus the probability P^i_h that household i chooses housing choice h can be written as a function of the full vectors of house/neighborhood characteristics (both observed and unobserved) and prices {**X**, **p**, ξ}:

$$(4) \qquad P^i_h = f_h(Z^i, \mathbf{X}, \mathbf{p}, \xi)$$

as well as the household's own characteristics Z^i.[9]

Equilibrium[10,11]

We assume that each household observed in our sample represents a continuum of households with the same observable characteristics, letting the measure of this continuum be μ. When the set of draws $\{\varepsilon^i_h\}$ for each household observed in the data is interpreted as unobserved heterogeneity in preferences for each house, this assumption is equivalent to working with the choice probabilities defined in equation (4) when deriving the conditions required for equilibrium. These choice probabilities depict the distribution of location decisions that would result for the continuum of households with a given set of observed characteristics as each household responds to its particular unobserved preferences.[12]

Aggregating the probabilities in equation (4) over all households yields the predicted number of households that choose each house h, $\hat{N}_h$:

$$(5) \qquad \hat{N}_h = \mu \bullet \sum_i P^i_h$$

where again μ represents the measure of the continuum of households with the same observable characteristics as household i. In order for the housing market to clear, the number of households

[9] The h subscript on the function f simply indicates that we are writing the probability that household i chooses house h, not that the form of the function itself varies with h.

[10] For more details on the assumptions underlying the equilibrium concept used in this analysis see Bayer, McMillan, and Rueben (2002), which extends the equilibrium analysis to include social interactions. For clarity of exposition, we ignore such interactions in presenting the model and equilibrium properties in this paper.

[11] The equilibrium concept developed here treats the supply of housing as fixed. This is done for expositional simplicity. A more generic housing supply function could be incorporated in the analysis.

[12] This assumption concerning the distribution of households requires an analogous assumption about the set of housing choices observed in the sample. In particular, we assume that each house observed in the

choosing each house h must equal the measure of the continuum of houses that each observed house represents:[13]

$$(6) \qquad \hat{N}_h = \mu, \ \ \forall h \ \ \Rightarrow \ \ \sum_i P_h^i = 1, \ \ \forall h$$

That the probabilities add to one for each house simply implies that supply must equal demand for each type of housing in each location. So, while it might seem strange to require equation (6) to hold for a particular house observed in the sample, the market clearing condition makes more sense given the interpretation that each house observed in the sample represents a continuum of like houses.[14]

It is a straightforward extension of the central proof in Berry (1994) to show that a unique vector of housing prices clears the market.[15] Writing this market-clearing vector of prices as $\mathbf{p}^*(\mathbf{Z}, \mathbf{X}, \xi)$, the probability that household i chooses house h can be written:

$$(7) \qquad P_h^i = f_h\left(Z^i, \mathrm{X}, \mathrm{p}^*(\mathrm{Z}, \mathrm{X}, \xi), \xi\right)$$

where the notation $\mathbf{p}^*(\mathbf{Z}, \mathbf{X}, \xi)$ indicates that the set of market-clearing prices is generally a function of the full matrix of the household $\{\mathbf{Z}\}$ and house and neighborhood characteristics $\{\mathbf{X}, \xi\}$ that are treated as the primitives of the sorting model. We define an equilibrium as any set of choice probabilities in equation (7) along with the vector prices, $\mathbf{p}^*$, such that the market clearing condition shown in equation (6) holds. Since a unique set of prices clears the housing market, the sorting equilibrium will also be unique when the model does not include any social interactions.[16]

3 ESTIMATION

sample represents a particular type of housing in a particular location, and that the continuum of such identical houses also has measure μ.

[13] Note that the measure μ drops out of the market-clearing condition depicted in equation (6) and, consequently, simply serves as a rhetorical device for understanding the use of the continuous choice probabilities shown in equation (4) rather than the actual discrete choices of the individuals observed in the data in defining equilibrium.

[14] It is important to stress that the market clearing assumption is not required for estimation.

[15] See Bayer, McMillan, and Rueben (2002) for related proofs.

[16] Again for a more detailed discussion of the equilibrium definition and properties with social interactions see Bayer, McMillan, and Rueben (2002).

Estimation with Homogeneous Preferences

Before turning to issues involved with the identification and estimation of the full model, it is helpful to examine a restricted version of the model. In particular, consider a specification of the utility function in which all households share the same value for each house up to an idiosyncratic error term:

$$(8) \qquad U_h^i = \alpha_{0X} X_h - \alpha_{0p} p_h + \xi_h + \varepsilon_h^i$$

where ε_h^i is i.i.d. across households and choices. In this case, the probability function shown in equation (4) is identical for all households. Consequently, the market clearing condition implies that prices adjust so that the non-idiosyncratic utility provided by each alternative is identical:

$$(9) \qquad \alpha_{0X} X_h - \alpha_{0p} p_h + \xi_h = K \qquad \Rightarrow \qquad p_h = {}^{\alpha_{0x}}\!/_{\alpha_{0p}} X_h + {}^{1}\!/_{\alpha_{0p}} \xi_h$$

Equation (9) is a standard hedonic price regression. This equivalence makes clear that a hedonic price regression returns the mean valuation of housing and neighborhood attributes when the underlying assumptions of the sorting model specified above (which include the assumption of a fixed stock of housing) are combined with the additional assumption that households have identical preferences for houses and locations.[17]

Estimation with Heterogeneous Preferences

We now describe the estimation of the sorting model when households have heterogeneous preferences.[18] We begin by introducing some notation that simplifies the exposition. In particular, we rewrite the indirect utility function as:

$$(10) \qquad V^i_h = \delta_h + \mu^i_h + \varepsilon^i_h.$$

where:

$$(11) \qquad \delta_h = \alpha_{0X} X_h - \alpha_{0p} p_h + \xi_h$$

[17] This condition holds no matter what assumption is made concerning the distribution of the idiosyncratic error term and in the absence of such idiosyncratic preferences.

(12) $$\lambda_h^i = \left(\sum_{k=1}^{K} \alpha_{kX} Z_k^i\right) X_h - \left(\sum_{k=1}^{K} \alpha_{kp} Z_k^i\right) p_h$$

In equation (12), k indexes household characteristics and δ_h captures the portion of the utility provided by housing choice h that is common to all households. When the household characteristics included in the model are constructed to have mean zero, δ_h is the *mean indirect utility* provided by housing choice h. The unobservable component of δ_h, ξ_h, captures the portion of unobserved preferences for housing choice h that is correlated across households, while ε_h^i represents unobserved preferences over and above this shared component.

The estimation of the model begins by maximizing the probability that each household chooses its observed housing choice. In particular, for any combination of the heterogeneous parameters in λ and mean indirect utilities, δ_h, the model predicts the probability that each household i chooses house h. When ε_h^i is assume to be drawn from the type 1 extreme value distribution, for example, this probability can be written:

(13) $$P_h^i = \mathbf{exp}(\delta_h + \hat{\lambda}_h^i) \Big/ \sum_k \mathbf{exp}(\delta_k + \hat{\lambda}_k^i)$$

Maximizing the probability that each household makes its correct housing choice, conditioning on the full set of observed household characteristics Z^i and choice characteristics $\{X_h, p_h\}$, gives rise to the following log-likelihood function:

(14) $$\ell = \sum_i \sum_h I_h^i \ln(P_h^i)$$

where I^i_h is an indicator variable that equals 1 if household i chooses house h in the data and 0 otherwise. The first step of the estimation procedure consists of searching over the parameters in λ and vector of mean indirect utilities to maximize ℓ.

Notice, however, that the set of observed residential choices provides no information that distinguishes the components of δ. Given the estimate of δ obtained from fitting the observed individual location decisions, equation (11) provides an estimating equation that bears more than

[18] Complete details of the estimation procedure including methods for simplifying the computation and the asymptotic properties of the estimator can be found in Bayer, McMillan, and Rueben (2002).

a passing resemblance to the hedonic price regression shown in equation (9). In particular, moving price to the left-hand side of equation (11) yields:

$$(15) \qquad p_h + \frac{1}{\alpha_{0p}}\delta_h = \frac{\alpha_{0x}}{\alpha_{0p}} X_h + \frac{1}{\alpha_{0p}}\xi_h$$

Consequently, in the presence of heterogeneous preferences, the mean indirect utility δ_h estimated in the first stage of the estimation procedure provides an adjustment to the hedonic price equation so that the price regression accurately returns mean preferences.

In the context of our model, therefore, equation (15) provides the intuition for why the equilibrium price function differs from the mean marginal willingness-to-pay when households have heterogeneous preferences. Figure 1 provides this intuition in graphical form for a simple example in which households value a single, discrete characteristic of a house, such as a view of the Golden Gate Bridge.[19] If such a view were extremely rare, as represented by H_1 in Figure 1, the difference in price between houses with versus without a view would reflect the marginal willingness-to-pay (MWTP) of a household with a relatively strong taste for a view, as indicated by p_1^* in Figure 1. If such a view were plentiful, however, the price of the view would generally reflect the MWTP of someone much lower in the distribution of tastes for a view, as indicated by p_2^*. In the first case, the price of these houses would be exceedingly high relative to the marginal willingness-to-pay of the mean household, which is indicated by p_M^*, while in the second case, the price of a view in equilibrium would more closely resemble mean MWTP. In essence, the equilibrium model controls for which individual in the distribution of tastes sets the price of a given attribute given the supply of that attribute. This provides an adjustment that reflects the difference between this household's valuation and that of the mean household so that the adjusted hedonic price regression accurately reflects mean preferences. In the first case, the mean indirect utility is negative, which effectively reduces the left-hand side of equation (15) so that it reflects what the mean household would be willing to pay for a view.[20] Without incorporating the adjustment for the difference in mean utilities in these two cases, a hedonic price regression would clearly return different estimates in the two scenarios.

The second stage of the estimation procedure involves estimating equation (15). In general, both p_h and δ_h are endogenous variables and consequently we describe an instrumental

[19] For this example, we ignore the idiosyncratic preference term for expositional simplicity.

[20] Notice that the coefficient on δ_h in equation (15) essentially converts utility to prices for the mean household.

variables estimator below. Before describing this IV estimator, we first relate equation (14) back to the two fundamental sources of variation in the data that we described in the introduction.

Intuitively two sources of variation in the observed housing data help to identify differences in preferences across households: (i) *choice variation* – the characteristics and price of houses chosen versus not chosen by households of a particular type and (ii) *within-type price variation* – the variation in price and housing/neighborhood characteristics among the set of houses chosen by a particular type of households. While this latter form of variation seemingly provides a natural way to learn about the trade-offs that households of a particular type make on average, equation (15) makes clear that using this form of variation while at the same time ignoring the underlying choice variation gives rise to an important form of selection bias.

To see this more clearly, consider a simple example with two types of households: t_1 and t_2. Letting the interaction terms in the utility specification above represent the additional preferences of households of type t_2 versus type t_1, δ_h simply represents the indirect utility provided by housing choice h to households of type t_1. With utility defined in this way, houses chosen more often by households of type t_1 naturally must have relatively high values of δ_h. In light of this fact, consider estimating a hedonic price regression using only the set of houses chosen by households of type t_1 (i.e., using only within-type price variation). As equation (15) makes clear, this is essentially a case of sample selection on the dependent variable. Because the set of houses used in such an estimation have relatively high values of δ_h on average, these observations necessarily have relatively low values of p_h conditional on the observed choice characteristics X_h. In this way, a hedonic price regression estimated using only houses chosen by households of a particular type naturally introduces a negative correlation between X_h and ξ_h, leading to a downward bias in the estimation of the willingness of households of this type to pay for observable housing and neighborhood characteristics X_h. In essence, the first stage of the estimation procedure outlined above provides the adjustment to the hedonic price regression so that it accurately returns the preferences of the baseline household category (type t_1 in our example or mean utility when household characteristics are constructed so as to have mean zero).

This simple example makes clear two issues related to the estimate of heterogeneous preferences. First, it is necessary to use choice variation in order to properly estimate heterogeneous preferences – i.e., the use of type-specific price variation alone leads to biased preference estimates. Second, while the estimation of discrete choice models superficially appears to be based entirely on choice variation, when traditional discrete choice models are augmented to allow for terms that capture the unobserved quality of alternatives, the estimation of the full model does indeed use the same form of price variation that forms the basis for hedonic price regressions, as illustrated by the second-stage regression equation shown in (15).

The Correlation of School Quality with Unobserved Neighborhood Quality

In attempting to estimate preferences for school quality, an important endogeneity issue has been raised in the literature by Black (1999), Clapp and Ross (2002), and Kane *et al.* (2003) among others. These authors point out that the quality of local schools is likely to be positively correlated with unobserved housing and neighborhood quality. While each of these authors make this point in the context of a hedonic price regression, the same point holds in the context of our broader sorting model, in which ξ_h represents unobserved housing/neighborhood quality. In this context, assuming that school quality is uncorrelated with ξ_h is likely to overstate willingness-to-pay for school quality, misattributing preferences for unobserved house and neighborhood quality as tastes for schools quality. To address this issue, we follow the identification strategy developed in Black (1999). Using a sample of houses near school attendance zone boundaries, Black estimates a hedonic price regression that includes boundary fixed effects. By including boundary fixed effects, this strategy essentially compares the prices of houses in otherwise similar neighborhoods, but that fall on opposite sides of a boundary determining where students will attend school. Any differences in prices not associated with housing characteristics are then interpreted as the marginal willingness-to-pay for school quality.

In order to incorporate boundary fixed effects in the estimation, we assign each house to a region r. When a house is close to a boundary between two school districts, it will fall into a *boundary region* and when a house is more centrally located within a school district, it will fall into a *central region*. Letting ψ_r be a region fixed effect for the region r to which house h belongs, we can re-write the utility function shown in equation (1) as:

$$(16) \qquad \underset{(h)}{Max} \quad V_h^i = \alpha_X^i X_h - \alpha_p^i p_h + \psi_r + \xi_h + \varepsilon_h^i$$

Having accounted for a region fixed effect, ξ_h now represents the unobserved quality associated with the particular housing unit h within region r.

In extending this identification strategy to the broader sorting model, an additional issue concerns the treatment of houses not near a school district boundary. In essence, while we seek to use only the variation in the data at the boundaries to estimate preferences for school quality, the logic of the choice model developed in Section 2 requires the use of all houses in the choice set. Notice, however, that given the specification of equation (16), equations (11)-(12) become

$$(17) \qquad \delta_h = \alpha_{0X} X_h - \alpha_{0p} p_h + \psi_r + \xi_h$$

$$(18) \qquad \lambda_h^i = \left(\sum_{k=1}^{K} \alpha_{kX} Z_k^i \right) X_h - \left(\sum_{k=1}^{K} \alpha_{kp} Z_k^i \right) p_h$$

That is, the boundary fixed effect appears only in the choice-specific constant regression. Thus, the first stage of the estimation procedure remains unchanged, returning estimates of the interaction parameters and the choice-specific constants. In the second stage of the estimation procedure, (i.e., the estimation of equation (17)), we use only an appropriately weighted sample of houses in boundary versus central regions. In this way, the estimation of the interaction (heterogeneity) parameters in the utility function shown in equation (18) is based on the full sample of houses, while the estimation of the mean preference parameters (those in equation (17)) is based only on across-boundary variation in prices. And, when preferences are restricted to be homogeneous, our equilibrium model reduces to the analysis in Black (1999).

It is worth emphasizing that the utility specification shown in equation (17) as well as in the original specification allows household characteristics to affect the coefficient on price. So households with more income, for example, are allowed to have a greater marginal willingness-to-pay for unobserved housing and neighborhood quality. This specification thus allows for the proper estimation of heterogeneity in preferences for school quality even in circumstances where school quality is correlated with unobservable housing and neighborhood attributes valued more strongly by some households versus others. If, for example, high-income households have relatively strong preferences for both good schools and homes with a view of the San Francisco Bay, our baseline analysis will properly estimate the relative preferences of high-income households for school quality, as long as the coefficient governing the interaction of household income and price captures the fact that higher income households have a greater willingness to pay for unobserved quality (the view of the Bay).

We employ the boundary fixed approach methodology because it provides an attractive way of controlling for much of the correlation between unobserved housing/neighborhood quality and school quality. There are important reasons to expect, however, that such boundary fixed effects do not control for all of this correlation First, school district boundaries may influence more than the school that a child attends, influencing who children and parents interact with more generally. To address this issue, in the analysis below we control for the sociodemographic characteristics of the households that reside on the same side of the boundary. A second reason to expect that boundary fixed effects do not control completely for unobserved housing/neighborhood quality is the fact that housing and school quality are both likely to be normal goods. Consequently, for both their own consumption and for the future re-sale value of their homes, home-owners on the better school quality side of a boundary are likely more likely to invest in improving the quality of their housing unit. This introduces a positive bias between housing and school quality that is very difficult to address given the fact that many of these improvements are likely to be unobserved in the data.[21] In sum, while the use of boundary fixed effects should control for much of the correlation of school quality with unobserved housing and neighborhood quality, we still expect the estimated preferences for school quality to be slightly overstated.

4 DATA

Having laid out many of the issues concerning the equilibrium model of sorting, estimation, and identification, we now describe the data used in our analysis as well as some empirical issues related to these data. Our analysis is facilitated by access to restricted Census microdata for 1990. These restricted Census data provide the detailed individual, household, and housing variables found in the public-use version of the Census, but unlike the public-use data, also include information on the location of individual residences and workplaces at a very disaggregate level of geography. In particular, while the public-use data specify the PUMA (a Census region with approximately 100,000 individuals) in which a household lives, the restricted data specify the Census block (a Census region with approximately 100 individuals), thereby identifying the local neighborhood that each individual inhabits as well as the characteristics of each neighborhood far more accurately than has been previously possible with such a large-scale data set.

We use data from six contiguous counties in the San Francisco Bay Area: Alameda, Contra Costa, Marin, San Mateo, San Francisco, and Santa Clara. We focus on this area for two main reasons. First, it is reasonably self-contained. Examination of Bay Area commuting patterns in 1990 reveals that a very small proportion of commutes originating within these six counties ended up at work locations outside the area; and similarly, a relatively small number of commutes to jobs within the six counties originated outside the area. And second, the area is

[21] We do, however, show evidence below that the observed housing and neighborhood characteristics are not all that different on the high versus low side of the school district boundaries used in the analysis.

sizeable along a number of dimensions, including over 1,100 Census tracts, and almost 39,500 Census blocks, the smallest unit of aggregation in the data. The sample consists of about 650,000 people in just under 244,000 households. The Census provides a wealth of data on the individuals in the sample – race, age, educational attainment, income from various sources, household size and structure, occupation, and employment location.[22]

The Census data also provide a variety of housing characteristics: whether the unit is owned or rented, the corresponding rent or owner-reported value, property tax payment, number of rooms, number of bedrooms, type of structure, and the age of the building. In constructing neighborhood characteristics, we begin by characterizing the stock of housing in the neighborhood surrounding each house. Using the Census data, we also construct neighborhood racial, education and income distributions based on the households within the same Census block group, a Census region containing approximately 500 housing units. We merge additional data describing local conditions with each house record, constructing variables related to crime rates, land use, local schools, topography, and urban density. For each of these measures, a detailed description of the process by which the original data were assigned to each house is provided in a Data Appendix. The list of the principal housing and neighborhood variables used in the analysis, along with means and standard deviations is given in the first column of Table 1.

Refining the House Price Variables Provided in Census

For a variety of reasons, the house price variables reported in the Census are ill suited for our analysis. House values are self-reported and top-coded, and rents may reflect substantial tenure discounts. Moreover, because we have implicitly defined the model and developed its equilibrium properties in terms of a single price variable for both owner-occupied and rental properties, we must relate house values to rents in some way.[23] Consequently, we make four adjustments to the housing price variables reported in the Census aiming to get a single measure for each unit that reflects what its monthly rent would be at current market prices. We describe

[22] Throughout our analysis, we treat the household as the decision-making agent and characterize each household's race as the race of the 'householder' – typically the household's primary earner. We assign households to one of four mutually exclusive categories of race/ethnicity: Hispanic, non-Hispanic Asian, non-Hispanic Black, and non-Hispanic White.

[23] This requirement may seem more restrictive than it actually is. Note that we treat ownership status as a fixed feature of a housing unit in the analysis - whether a household rents or owns is endogenously determined within the model by its house choice. We allow households to have heterogeneous preferences for home-ownership (a positive interaction between household wealth and ownership, for example, will imply that wealthier households are more likely to own their housing unit, as we find below) and other house characteristics. Thus the use of a single house price variable does not impose any serious restrictions on the model.

the reasoning behind each adjustment briefly here, leaving a detailed description of the methodology for the Data Appendix.

Because house values are self-reported, it is difficult to ascertain whether these prices represent the current market value of the property, especially if the owner purchased the house many years earlier. Fortunately, the Census also contains other information that helps us to examine this issue and correct house values accordingly. In particular, the Census asks owners to report a continuous measure of their annual property tax payment. The rules associated with Proposition 13 imply that the vast majority of property tax payments in California should represent exactly one percent of the transaction price of the house at the time the current owner bought the property or the value of the house in 1978. Thus, by combining information about property tax payments and the year that the owner bought the house (also provided in the Census in relatively small ranges), we are able to construct a measure of the rate of appreciation implied by each household's self-reported house value. We use this information to modify house values for those individuals who appear to be reporting values much closer to the original transaction price rather than current market value.

A second deficiency of the house values reported in the Census is that they are top-coded at $500,000, a top-code that is often binding in California. Again, because the property tax payment variable is continuous and not top-coded, it provides information useful in distinguishing the values of the upper tail of the distribution.

The third adjustment that we make concerns rents. While rents are presumably not subject to the same degree of misreporting as house values, it is still the case that renters who have occupied a unit for a long period of time generally receive some form of tenure discount. In some cases, this tenure discount may arise from explicit rent control, but implicit tenure discounts generally occur in rental markets even when the property is not subject to formal rent control. In order to get a more accurate measure of the market rent for each rental unit, we utilize a series of local hedonic price regressions in order to estimate the discount associated with different durations of tenure in each of over 40 sub-regions within the Bay Area.

Finally, we construct a single price vector for all houses, whether rented or owned. In order to make owner- and renter-occupied housing prices as comparable as possible, we seek to determine the implied current annual rent for the owner-occupied housing units in our sample. Because the implied relationship between house values and current rents depends on expectations about the growth rate of future rents in the market, we estimate a series of hedonic price regressions for each of over 40 sub-regions of the Bay Area housing market. These regressions return an estimate of the ratio of house values to rents for each of these sub-regions and we use these ratios to convert house values to a measure of current monthly rent. Again, the procedure is described in detail in the Data Appendix.

School Characteristics

While we have an exact assignment of Census blocks to school districts, we have only been able to attain precise maps that describe the way that city blocks are assigned to schools in 1990 for Alameda County. In the absence of information about within-district school attendance areas, we employ four alternative approaches for linking each house to a school. The crudest procedure assigns average school district characteristics to every house falling in the school district. A refinement on this makes use of distance-weighted averages. For a house in a given Census block group, we calculate the distance between that Census block group and each school in the school district. We then construct weighted averages of each school characteristic, weighting by the reciprocal of the distance-squared as well as enrollment. As a third approach we simply assign each house to the closest school within the appropriate school district. In our fourth and preferred approach, we adjust this closest school assignment procedure to ensure that the predicted enrollment of each school as calculated by summing over the school-aged children in each Census block group assigned to a school equals the actual enrollment of that school. We describe this procedure in detail in the Data Appendix. In practice, all four methods of defining school characteristics yielded very similar results, with the estimates based on school district averages revealing a small amount of aggregation bias. To keep the exposition of the results manageable below, we simply report results for the fourth method described here.

As our measure of school quality, we use the average test score for each school, averaged over two years. Averaging over two years helps to reduce any year-to-year noise in the measure. When variables that characterize the sociodemographic composition of the school or surrounding neighborhood are included in the analysis, the estimated coefficient on average test score picks up what households are willing to pay for an improvement in average student performance at a school holding the sociodemographic composition constant. While the average test score is an imperfect measure of school quality, it has the advantage of being easily observed by both parents and researchers and consequently has been used in most analyses that attempt to measure demand for school quality.

Boundary Fixed Effects

A number of empirical issues arise in incorporating boundary fixed effects into our analysis. The first issue concerns the choice of jurisdiction for which the boundaries are defined.

While Black (1999) uses school attendance zones within a school district, in the analysis presented in this paper we use boundaries between school districts in the Bay Area.[24] A central feature of local governance in California helps to eliminate some of the problems that naturally arise with the use of school district boundaries, as Proposition 13 ensures that the vast majority of school districts within California are subject to a uniform effective property tax rate of one percent. A second issue concerns the width of the boundaries. While a narrow band makes the assumption that unobserved neighborhood quality is the same on opposite sides of the boundary more accurate, a wider band allows the use of more data. We experimented with a variety of distances and report the results for 0.25 miles, as these were far more precise due to the larger sample size.

Table 1 displays descriptive statistics for various samples related to the boundaries. The first two columns report means and standard deviations for the full sample; the third column reports means for the sample of houses within 0.25 miles of a school district boundary, the fourth and fifth columns report means on the high versus low average test score side of the school district boundary; the sixth column reports weighted means for the sample of houses within 0.25 miles of a school district boundary (we describe the weight below).

Comparing the first column to the third column of the table, it is immediately obvious that the houses near school district boundaries are not fully representative of those in the Bay Area as a whole. Houses near boundaries tend to be slightly more expensive, more often owner-occupied, and larger in neighborhoods that are less dense and have lower crime rates than the sample as a whole. To address this problem, we create sample weights for the houses near the boundary according to the following procedure: Using a logistic regression, we first regress a dummy variable indicating whether a house is in a boundary region on the vector of housing and neighborhood attributes. Fitted values from this regression provide an estimate of the likelihood that a house is in the boundary region given its attributes. We use the inverse of this fitted value as a sample weight in subsequent regression analysis conducted on the sample of houses near the boundary. Column 6 of Table 1 shows the resulting weighted means. As the numbers clearly demonstrate, using these weights makes the sample near the boundary much more representative of the full sample, column 6 typically being closer to column 1 than column 3.

The principal issue arising from Table 1 concerns differences across school district boundaries, which are displayed in columns 4 and 5. Comparing the average characteristics of houses with 0.25 miles of the boundary on the high versus low school quality side reveals that

[24] This difference implies that our results are not directly comparable to Black (1999). It is important to keep this distinction in mind throughout our discussion of the boundary fixed effects.

houses on the high school quality side cost \$53 more per month and are assigned to schools with a 43-point average test score increase. The difference is highly statistically significant. The standard deviation of the test score measure is 74, so this translates into a raw mean marginal willingness to pay for a one standard deviation increase in school quality of approximately \$91 in monthly rent or \$24,100 in house value.[25] The remaining differences in average housing characteristics on opposite sides of the school district boundary are fairly small (relative to the overall standard deviation of each variable), but the effect of these differences on housing prices, however, is generally in the direction of increasing prices on the high school quality side. Houses on the high school quality side of the boundary are more often owner-occupied and have slightly lower crime rates.

More significantly, houses on the high school quality side of the boundary are more likely to be inhabited by white households and households with more education and income. These across-boundary differences in sociodemographic composition are exactly what one would expect if household sort on the basis of preferences for school quality, thereby leading those with stronger tastes or increased ability to pay for school quality to choose the higher school quality side of the boundary. In proceeding to the empirical analysis below, we draw two main conclusions from Table 1: first, that the use of boundary fixed effects is likely to control for much of the correlation of school quality with unobserved features of housing and neighborhood quality unrelated to household sorting; second, there appears to be a significant amount of sorting with respect to school district boundaries on the basis of race, education, and income. This sorting suggests that an exogenous change in school quality both increases housing prices and changes the sociodemographic composition of the corresponding neighborhood. Because this latter effect is also likely to affect housing prices, the direct effect of school quality on prices is likely to be much lower than the \$91 per month for a one standard deviation increase in school quality obtained from a simple examination of the average price difference across boundaries.

5 RESULTS

Hedonic Price Regressions

We begin our primary empirical analysis by presenting results of a series of hedonic price regressions. As described in Section 2, these regressions correspond to estimating the equilibrium

[25] As described above, we construct a single price vector for all houses, whether rented or owned. Because the implied relationship between house values and current rents depends on expectations about the growth rate of future rents in the market, we estimate a series of hedonic price regressions for each of over 40 sub-regions of the Bay Area housing market. These regressions return an estimate of the ratio of house values

model of sorting under the assumption that households have homogeneous preferences. We begin with these hedonic price regressions for two reasons. First, they allow us to explore a number of alternative assumptions about the use of boundary fixed effects, the sample, and the effect of including various additional controls in a setting where it is easy to compare results across models. Second, these regressions facilitate comparison to the results of the prior literature, Black (1999) in particular.

Table 2 presents the results from six hedonic price regressions. The dependent variable in these regressions is our constructed monthly house price measure, which equals monthly rent for renter-occupied units and an imputed monthly rent for owner-occupied units (see Section 4 for more details). Results are reported for the full sample and for sample of houses within 0.25 miles of school district boundaries, with and without including fixed effects. For each of these three specifications, results are reported for a specification that does and does not include the five neighborhood sociodemographic variables used in our earlier analysis: average income, percent of households with a college degree, percent Asian, percent Black, and percent Hispanic, each measured at the Census block group level.

No clear bias emerges when the sample is reduced to only those houses near a school district boundary. When neighborhood sociodemographic characteristics are excluded, for example, the implied marginal willingness-to-pay for a one standard deviation increase in school quality moves from $146 in monthly rent ($38,500 in house value) to $145 in monthly rent ($38,200 in house value), when the sample is restricted to houses near a boundary. The estimated mean MWTP for a one standard deviation in school quality declines to $101 ($26,700) when the boundary fixed effects are included in the analysis. This more or less mirrors the raw MWTP of $91 ($24,100) calculated using the descriptive statistics in Table 1. Thus, controlling for a host of fixed housing and neighborhood characteristics does very little to this estimate. Including neighborhood sociodemographic characteristics in the analysis, however, dramatically reduces the estimated MWTP for school quality to $28 in monthly rent ($7,400 in house value), and this estimate is not sensitive to whether boundary fixed effects are included or not.

When sociodemographic variables are included in the analysis, the interpretation of the coefficient on average test score represents the effect of an increase in test score holding sociodemographic composition constant. Conversely, the coefficients on the sociodemographic variables pick up the valuation of these characteristics holding average test score constant - in other words, the value of these characteristics unrelated to their effect on average test score. One

to rents for each of these sub-regions and we use the average of these ratios for the Bay Area, 264.1, to convert monthly rent to house value for the purposes of reporting results at the mean.

concern is that the average test score used in the analysis may be a noisy measure of student performance and, consequently, that neighborhood sociodemographic differences partially capture unobserved differences in school quality as well. To address this issue, we compare the results presented above, which average a school's average test score over two years to results based on a test score for only the first year of these two years of data. We find no difference in the estimate when the test score is based on a one- versus two-year average.[26]

That the estimated mean MWTP for a one standard deviation increase in school quality of \$28 in monthly rent or \$7,400 in house value is relatively small may relate to the fundamental informational problem that households face in attempting to distinguish the quality of a school. The ability of households to glean from average test score data the quality of a school as opposed to increased performance due directly to its sociodemographic composition may be very difficult indeed; empirical researchers know that to do this well requires an immense amount of data. Consequently, to the extent that households have trouble measuring school quality, one would not expect them to value it much when making their residential location decision. Put another way, to the extent that households instead use the posted average test score or the sociodemographic composition of the neighborhood as proxies for school quality when making their location decision, the location decision will be driven much more heavily by neighborhood sociodemographic variables rather than by school quality itself.

To facilitate comparison, the estimates reported in the previous paragraph can be put into the same terms employed by Black (1999), who reports the average percentage increase in house price associated with a 5 percent increase in average test score, evaluated at the mean. Because we use a different definition of boundary fixed effects than in the Black study, we find it sensible to only compare the final estimates for the hedonic price regressions that include both neighborhood sociodemographic characteristics and boundary fixed effects. In particular, our final estimate implies that a 5 percent increase in average test score would increase house prices at the mean by approximately 0.9 percent or \$2,700 in house value as opposed to 1.7 percent or \$3,100 in house value as found in Black (1999).[27] These numbers seem similar enough to us to be plausible given the broad differences in the sample and study area used in these analyses.[28]

---

[26] Kane, Staiger and Samms (2003) point out that the *change* in average test scores form year to year exhibits a great deal of noise unrelated to actual school quality. Not surprisingly, the level exhibits far less noise.

[27] The estimate of 1.7 percent is based on the specification reported in Column 4 of Table V in Black (1999), which includes neighborhood sociodemographic characteristics measured at the Census block group level. It is worth noting that Black's handling of neighborhood sociodemographic composition may be somewhat problematic due to the use of school attendance zone boundaries. Because houses on both sides of attendance zone boundaries often fall into the same Census block group, the neighborhood sociodemographic variables will be identical by construction for many boundaries in Black's analysis. This

One final important result from Table 2 relates to the estimated coefficients on the five neighborhood sociodemographic variables, with and without including boundary fixed effects. One of the central issues in estimating models of location choice (whether hedonic or discrete choice) concerns the correlation of neighborhood sociodemographic characteristics with unobserved housing and neighborhood quality. For example, houses with a view of the San Francisco Bay should tend to command higher prices and draw higher-income households to the neighborhood. Thus, if the view is unobserved in the data, the hedonic price regression will tend to attribute the additional price associated with these houses to the average income of the neighborhood. The boundary fixed approach used here provides a convenient way of controlling for fixed neighborhood unobservable characteristics such as a view of the Bay. At the same time, because there is significant sorting (as shown in Table 1) with respect to the boundaries themselves, the neighborhood sociodemographic composition of the Census block groups on one side of the boundary versus the other can be quite different.

Comparing the coefficients on the neighborhood sociodemographic characteristics with and without the inclusion of boundary fixed effects (columns 5 and 6 of Table 2) yields exactly what one would expect if the boundary fixed effects control for unobserved components neighborhood quality unrelated to the sorting of households across the boundary. In particular controlling for fixed effects reduces the coefficient on percent Black from –\$114 to -\$41; increases the coefficient on percent Hispanic from \$147 to \$240; changes the sign of the coefficient on percent Asian from \$-76 to \$201; reduces the coefficient on the percent of households with a college degree from \$192 to \$92; and reduces the coefficient on average neighborhood income (/\$10,000) from \$111 to \$101. Thus, while the boundary fixed effects do not seem to be effective in controlling for differences in the sociodemographic composition across neighborhoods, they do seem to be effective in controlling for more fixed aspects of

---

will make it impossible to evaluate whether there is significant sorting with respect to school attendance zone boundaries based on Census block group data.

[28] There are many potential differences, of course, between the study areas and samples used in the analysis: Black uses owner-occupied homes in suburban Boston, Massachusetts, while we use both owner- and renter-occupied housing for the full San Francisco Bay Area. One might expect, and this is also implied by the estimates from heterogeneous preferences model reported below, that suburban home-owners are simply willing to pay more for school quality than other households in a metropolitan region. Another intriguing potential explanation for the difference relates to California's school finance system, which gives school districts almost no control over the property tax rate or the level of education expenditures. While this lack of fiscal control on the part of local jurisdictions lends more credence to the use of school district boundaries in our analysis, it also suggests that schools in California (which are restricted to roughly equal levels of expenditures per student) may not differ much in ways that affect a school's quality as perceived by parents other than through differences in peer composition. Thus, the variation in school quality unrelated to a school's sociodemographic composition may be much greater in

unobserved neighborhood quality, and thus provide a way of properly estimating preferences for neighborhood sociodemographic characteristics in the presence of an important endogeneity problem.

Estimates of Model with Heterogeneous Preferences

We now turn to estimates of the model with heterogeneous preferences. Table 3 reports estimates of the interaction parameters for a specification that does not include the variables that characterize neighborhood sociodemographic composition. As the table suggests, this specification simultaneously controls for the effect of each of a series of household characteristics (income, education, race, work status, age, and household structure) on a household's marginal willingness-to-pay for each of a series of housing and neighborhood attributes. While the numbers reported in Table 3 are not directly interpretable in dollar terms (we make this conversion in Table 8 below), the estimates of the coefficients reveal significantly positive interactions of household income, education, and the age of the householder with school quality and significantly negative interactions with school quality if a household has children or is Asian, Black, or Hispanic versus White. The remaining pattern of signs and magnitudes in the table are what one would expect in every important case, with, for example, utility declining rapidly in commuting distance and the interaction of income and price revealing a positive income elasticity of demand for housing, especially for home-ownership and larger houses.

Table 4 extends the analysis of Table 3 by including a series of neighborhood sociodemographic characteristics. The inclusion of variables characterizing the sociodemographic composition reduces the magnitude of the interaction of income with school quality by 60%, of education with school quality by almost 75%, and of Black and Hispanic with school quality by 80-90%. These reductions are not surprising in the presence of strong sorting social interactions. If, for example, more highly educated households want to live together and each have a relatively high taste for school quality, households will stratify on the basis of education in equilibrium, with more highly educated households living in better quality school districts. If one did not account for the fact that part of the corresponding higher prices in these neighborhoods was due to the willingness of these households to pay to live with other highly educated households, one would expect to severely overstate the preference of highly educated households for school quality, as in Table 3. The results for household structure also appear more

---

Massachusetts and consequently, a 5 percent increase in average test score holding neighborhood sociodemographics constant could amount to a more meaningful increase.

reasonable in the specification of Table 4 versus Table 3, as households with children now have a significantly positive interaction with school quality rather than a significantly negative one.

Signing the Bias in the Hedonic Price Regression

Having estimated the vector of mean indirect utilities in the first stage of the estimation procedure, the second stage of the estimation procedure outlined above concerns the estimation of the mean preferences parameters shown in equation (11), which is re-written in equation (15). For general forms of the utility function, both housing price, p_h, and mean utility, δ_h, will be correlated with the unobserved housing/neighborhood quality, ξ_h, in equilibrium. In this case, the estimation of equation (11) requires an additional variable that is correlated with p_h but not with unobserved housing/neighborhood quality, ξ_h.

Before constructing an appropriate instrumental variable, however, it is important to point out that we can sign the direction of the bias in the hedonic price regression simply under the assumption that price enters indirect utility negatively ($\alpha_{0p} > 0$). Re-writing equation (15) as:

$$(19) \qquad p_h = {}^{\alpha_{0X}}\!/_{\alpha_{0p}} X_h - {}^{1}\!/_{\alpha_{0p}} \delta_h + {}^{1}\!/_{\alpha_{0p}} \xi_h$$

it is immediately obvious that the hedonic price regression includes a negative function of the mean indirect utility that a house provides in the error term. Thus, to the extent that an included regressor is positively correlated with the vector of mean indirect utilities, the hedonic price coefficient is biased downwards and, consequently, understates mean preferences for the attribute in question. Table 5 shows the partial correlation of key choice attributes with δ for various specifications of the sample and inclusion of boundary fixed effects. In each case, the reported partial correlation conditions on the full set of covariates used in the hedonic price regression including neighborhood sociodemographic characteristics. Focusing on the final column, which includes the boundary fixed effects, the negative correlation between δ. And most of the variables, especially the percentage of a block group that is Black and the average income of a block group, implies that a hedonic price regression will tend to *overstate* mean preferences for these characteristics. The correlation of δ with average test is only slightly negative and, consequently, we expect the mean MWTP for average test score that we estimate using the heterogeneous preferences model to be slightly smaller than that estimated via the hedonic price regression. The one partial correlation that is positive in the final column of Table 5 is that with the percentage of the block group that has a college degree or more. Consequently, this coefficient will increase in moving from the hedonic price regression to the estimated mean preferences obtained from the equilibrium choice model.

Forming an Instrument for Price

To estimate equation (11) we need an instrument for price. The type of instrument that we propose rises naturally out of the sorting model when households value only the features of their own house and attributes of the surrounding neighborhood – where the size of this neighborhood potentially could be quite large. That is, as long as households do not value the features of housing and neighborhoods beyond some threshold distance from their own home when making their residential location decision, the exogenous attributes of houses and neighborhoods that are positioned beyond this threshold make suitable instruments for housing price. In developing this type of instrument, we exploit an inherent feature of the sorting process – that the overall demand for houses in a particular neighborhood is affected by not only the features of the neighborhood itself, but also by the way these features relate to the broader landscape of houses and neighborhoods in the region. Thus we assume that the exogenous attributes of houses and neighborhoods a sizeable distance away from a house influence the equilibrium in the housing market, thereby affecting prices, but have no direct effect on utility.

In practice, the precision of the estimation is improved significantly when the logic of this IV strategy is used to construct a single instrument for price that approximates the optimal instrument. The optimal instrument for p_h in the choice-specific constant regression (equation (11)) is given by:

$$(20) \qquad E\left({\partial \xi_h} \big/ {\partial \alpha_{0p}} \right) = E\left(p_h \mid \Omega\right)$$

that is, the expected value of p_h conditional on the information set Ω, which contains the full distribution of *exogenous* choice (X_h) and individual characteristics (Z^i). Notice that this instrument implicitly incorporates the impact of the full distribution of the set of choices in exogenous characteristic space as well as information on the full distribution of observable household characteristics into a single instrument for price.

For computational purposes, we use a well-defined instrument that maintains the inherent logic of this optimal instrument while being straightforward to compute. This 'quasi-' optimal instrument is based on the predicted vector of market-clearing prices calculated for an initial estimate of the parameter values with the vector of unobserved characteristics ξ set identically equal to 0 and using only the exogenous features of locations. This condition corresponds to using the prediction at the mean instead of the expected value. Operationally, the estimation proceeds as follows:

1. Include a full set of variables in the model that accounts for housing and neighborhood attributes in the region that households value directly when making their location decision – for the analysis conducted below we assume households care about the housing stock and land use within five miles of their house.

2. Using a conjecture of the model's parameters, setting $\xi_h=0$ for all h, and including only *exogenous* choice characteristics in X, calculate the vector of housing prices that clears the market, $\hat{p}^*(X_h, Z^i)$. In practice, we make a reasonable conjecture on the price coefficient and then simply run equation (15) via OLS. In calculating the vector of market clearing prices, we use only variables that describe housing and land use – not those related to neighborhood sociodemographic composition, tests scores, or crime.

3. Using $\hat{p}^*$ as an instrument for p, estimate the choice-specific constant regression.[29]

Like the optimal instrument, the instrument that we propose provides a measure of the way that the full landscape of possible choices affects the demand for each house/neighborhood. In essence, this instrument extracts additional information from Ω than that which is contained in the vectors of choice characteristics X already appearing in estimating equation (11). In the regressions reported below, we include a full set of controls for the characteristics of the house itself and its neighborhood as well as five variables that described land use[30] and six variables that describe the housing stock[31] in each of 1, 2, 3, 4, and 5 mile rings around the house. In sum, the additional information embedded in our instrument derives from the exogenous features of the housing stock and land use in a region beyond five miles from the house in question. Importantly, this information is collapsed into a single instrument that uses this information in a concise manner that is consistent with the logic of the sorting model.

Table 6 reports first stage price regressions analogous to the hedonic price regression reported in Table 2 with the inclusion of the optimal price instrument. In each specification, the optimal price instrument is strongly predictive of price, over and above the set of variables

[29] In practice, we repeat Steps 2 and 3 of this procedure using the estimated parameters from step 3 to construct a new price instrument in step 2 for the next iteration. While this iterative process is not necessary to ensure consistency, it does ensure that the final estimates are not sensitive to our initial conjecture of the coefficient on price. For this reason, we believe that this iterative procedure is likely to be more efficient than applying the procedure once, but do not have a proof of this.

[30] Percent industrial; Percent commercial; Percent residential; Percent open space; Percent other.

[31] Percent owner-occupied single family homes with 7 rooms or more; Percent owner-occupied single family homes with less than 7 rooms; Percent renter-occupied single family homes; Percent renter-occupied units in large apartment buildings; Percent of units in small apartment buildings; Percent other.

included in *X*, increasing the R^2 of each regression by approximately 2-4 percentage points. So the price instrument, which is derived entirely from the exogenous characteristics of the alternatives and the distribution of household characteristics in the population, adds significantly to the predictive power of these regressions.[32]

Estimates of the Adjusted Hedonic Price Regression

Using the optimal price instrument as an instrument for price, we report in Appendix Table 1 the results of six specifications of the mean indirect utility regression that correspond to the six specifications reported in Table 2. For these regressions, we use the estimated δ vector from the specification shown in Table 3 when sociodemographic characteristics are excluded and in Table 4 when they are included. To facilitate comparison with the hedonic price regressions, Table 7 reports the implied measures of the mean willingness to pay for school quality that result from these size specifications. These estimates are generated by dividing the coefficient associated with each choice characteristic in the δ regressions by the coefficient on price. Standard errors are calculated using the delta method.

As the first two rows of Table 7 make clear, the estimated mean MWTP for a one standard deviation increase in school quality as estimated by the equilibrium model with heterogeneous preferences is generally slightly less than the estimate obtained from a hedonic price regression (the equilibrium model with homogeneous preferences).[33] For the specification that excludes both boundary fixed effects and neighborhood sociodemographic characteristics, the mean MTWP for a one standard deviation increase in school quality estimated using the model with heterogeneity is $123 in monthly rent as compared to the $145 estimate obtained using the standard hedonic price regression. When only boundary fixed effects are included, the estimated mean MWTP is $82 as compared with $101 in the hedonic price regression. And, finally, when neighborhood sociodemographic characteristics are included in the model the estimated mean MWTP falls from $28 in monthly rent ($7,400 in house value) as estimated in the hedonic price regression approach to $20 ($5,300) when boundary fixed effects are excluded and $26 ($6,900) when they are included in the analysis.

As in the hedonic price regressions, the inclusion of boundary fixed effects influences the estimates of the coefficients on the neighborhood sociodemographic variables in a pattern that

---

[32] As a side note, it is also important to point out that the coefficients on the other characteristics do not have much meaning as they represent the effect of these characteristics on price controlling for the estimated market-clearing price given only the exogenous attributes of the set of alternatives.

[33] This is as expected given the negative partial correlations between δ and average test score shown in table 5.

suggests that the boundary fixed effects control well for unobserved neighborhood differences not affected by the household sorting. In this case controlling for fixed effects reduces the magnitude of the mean MWTP associated with percent Black from –\$319 to –\$267; increases the estimate on percent Hispanic from \$18 to \$139; changes the sign of the coefficient on percent Asian from \$-96 to \$155; reduces the coefficient on the percent of households with a college degree from \$206 to \$138; and reduces the coefficient on average neighborhood (/10,000) from \$96 to \$88. Again, because school quality is positively correlated with average neighborhood income and education and the percentage of white households, the inclusion of boundary fixed effects, by lowering the implied MWTP measures associated with these neighborhood sociodemographic variables, actually has the effect of increasing the estimate of the mean MWTP for school quality.

Comparing the hedonic price regressions reported in Table 2 to the mean MWTP estimates derived from the heterogeneous preferences model in Table 7, it is worth exploring why a difference arises between these specifications for certain housing and neighborhood attributes and not others. Comparing results of our preferred specification (column 6), which includes both boundary fixed effects and neighborhood sociodemographic characteristics, reveals that the estimates related to housing characteristics, school quality, and crime tend to be slightly overstated in the hedonic price regression, while those related to neighborhood sociodemographic composition and race in particular change dramatically. Here, the analysis of Figure 1 is helpful in understanding why this is the case. Consider, for example, the estimated mean coefficient on Percent Black, which is -\$267 in the heterogeneous preferences model as opposed to only -\$41 for the hedonic price regression. For simplicity, assume that neighborhoods are completely segregated, so that the equilibrium price of a Black neighborhood is driven by the MWTP of the Black household with the least MWTP for a Black neighborhood (or, alternatively, the White household with the greatest MWTP). Here, the hedonic price regression returns the MWTP of the household on the *margin* between choosing a Black versus White household, which in this case is substantially greater than the MWTP of the *mean* household, which is estimated in the heterogeneous preferences specification. Put another way, a much lower differential in price between Black and White neighborhoods is required to equilibrate the housing market than would be required to make the mean household indifferent between these neighborhoods.

The case of school quality is also worth discussing in more detail. Because the Bay Area contains over 700 schools, the equilibrium difference in housing prices between each of the neighborhoods associated with each school is more appropriately characterized by Figure 2, which again simplifies the problem to one dimension of choice characteristic space. In this case, the equilibrium difference in price between each pair of schools ranked according to quality is the

MWTP of the household on the corresponding boundary between schools. These equilibrium prices are represent by the p_j^* terms on the vertical axis. If there are roughly an equal number of students in each school, averaging the equilibrium price over all of the houses in the sample corresponds roughly to the mean MWTP and, consequently, there is only a slight difference between the estimates returned from the model with heterogeneous preferences and the hedonic price regression.

Thus, in general, when we can view the choice problem as single-dimensional, one would expect the hedonic price regression to diverge from mean preferences for choice characteristics (especially those in fixed or limited supply) for which the preferences of the marginal household differ systematically from those of the mean household in the population. Which household is on the margin depends explicitly on the set of alternatives and the attributes available in the market as well as on the distribution of households and their preferences. Consequently, the valuation of attributes returned by the hedonic price regression will depend on these distributions, especially for certain types of attributes. It is important to stress that the heterogeneous preferences model that we use here explicitly accounts for the distribution of characteristics in the population as well as in the set of alternatives.

Finally, the change from the hedonic regression for the percentage of neighbors with a college degree is also interesting to examine. In this case, the estimated mean MWTP is greater than the corresponding coefficient in the hedonic price regression. This suggests that the mean household would generally want to increase its consumption of college-educated neighbors at the equilibrium price. In practice, however, because college-educated households themselves demand more housing and neighborhood attributes, this may be difficult to do without also increasing the consumption of these other goods. Consequently, because the single residential location decision determines a full bundle of goods, households may not be able to perfectly satisfy their preferences for any particular element of this bundle even if they are willing to pay the implied marginal price for that element, especially when the elements of a bundle are correlated – in this case, increased *percent college educated* often implies increases in other housing and neighborhood attributes.[34]

Heterogeneity in Willingness-to-Pay

Table 8 reports the implied estimates of the heterogeneity in MWTP for school quality and neighborhood sociodemographic characteristics across households with different characteristics for our preferred specification, which includes both neighborhood sociodemographic characteristics and boundary fixed effects. Focusing first on the heterogeneity in tastes of school quality, a household's willingness-to-pay increases with income, the presence of children, education, employment, and age. Black households have a significantly lower willingness to pay for school quality relative to White households, although this may result in part from unobservable difference in, for example, wealth that are not included in this analysis.

The presence of children increases demand for house size and school quality, but decreases demand for owner-occupied and newer housing, both of which might proxy in part for housing quality. That the presence of children generally decreases demand for housing quality is most likely due to the fact that disposable income declines as a result of having children. The increased demand for house size and school quality is especially noteworthy given this decline in disposable income. More income and education increases demand for all aspects of housing and school quality as well as for more educated and higher income neighbors. College-educated households in particular have a strong preference to live near other college-educated households.

Finally, Table 8 reveals strong, segregating racial interactions, with households of each race preferring to live near others of the same race. The $97 estimate listed in the fourth row and fifth column of the table, for example, implies that Black versus White households are willing to pay $97 more per month to live in a neighborhood that has 10 percent more Black versus White households. It is important to point out that this is a difference between the *positive* MWTP of Black households for this change and the *negative* MWTP of White households. It is also important to point out that these interactions pick up any direct preferences for living near others of the same race (e.g., a recent immigrant from China may want to interact with neighbors who also have immigrated from China) as well as any unobservable neighborhood or housing amenities valued more strongly by households of this group (e.g., recent immigrants from China may have similar tastes for shops, restaurants, and other neighborhood amenities). As discussed above, it is these strong, segregating racial interactions that cause the large difference between the estimates on the neighborhood racial characteristics in the hedonic price regressions and the mean MWTP estimates derived from the broader choice model.

6 SIMULATIONS

While the framework developed in this paper is useful as a way of reliably estimating the distribution of preferences for school quality and neighborhood sociodemographic composition,

---

[34] The fact that households are selecting a bundle of goods and that the set of available bundles may not span important subspaces of the consequent multi-dimensional attribute space is the main argument for

the real power of the equilibrium framework developed above comes after the model is estimated. At that stage, it can be used as a tool for exploring a series of economic and policy questions related to the equilibrium in the housing market. In this section, we illustrate the power of the equilibrium framework by exploring the capitalization of an exogenous change in the average test score for each school throughout the metropolitan region. For each school, we calculate both a 'partial equilibrium' increase in house value that accompanies a rise in school quality holding neighborhood sociodemographic measures constant and a 'general equilibrium' increase that accounts for the way that neighborhood sociodemographic characteristics would change in moving to a new sorting equilibrium. The differences between the partial and general equilibrium estimates inform us as to the importance of a social multiplier in the overall elasticity of demand faced by each school. Such a social multiplier arises because the exogenous change in school quality induces higher income and more educated to sort into the corresponding neighborhood, thereby leading to a further increase in house prices.

Simulation Details

Each of the simulations that we conduct begins by raising the average test score of a given school by one standard deviation (74 points). In this counterfactual environment, we calculate a new equilibrium for the model. The conditions required for an equilibrium consist of a set of location decisions for each household and a set of housing prices such that (i) each household's decision is optimal given the decisions of all other households, and (ii) the set of housing prices clears the market. The basic structure of the simulations consists of a loop within a loop. The outer loop calculates the sociodemographic composition of each neighborhood, given a set of prices and an initial sociodemographic composition of each neighborhood. The inner loop calculates the unique set of prices that clears the housing market given an initial sociodemographic composition for each neighborhood. Thus for any change in the primitives of the model, we first calculate a new set of prices that clears the market; as discussed in Section 2, Berry (1994) ensures that there is a unique set of market clearing prices. Using these new prices and the initial sociodemographic composition of each neighborhood, we then calculate the probability that each household makes each housing choice, and aggregating these choices to the neighborhood level, calculate the predicted sociodemographic composition of each neighborhood. We then replace the initial neighborhood sociodemographic measures with these new measures and start the loop again – i.e., calculate a new set of market clearing prices with these updated neighborhood sociodemographic measures. We continue this process until the neighborhood

using a discrete choice versus hedonic model of demand.

sociodemographic measures converge. The set of household location decisions corresponding to these new measures along with the vector of housing prices that clears the market then represents the new equilibrium.[35]

It is important to point out that because the model itself does not perfectly predict the housing choices that individuals make, the neighborhood sociodemographic measures initially predicted by model, $\overline{Z}_n^{PREDICT}$, will not match the actual sociodemographic characteristics of each neighborhood, $\overline{Z}_n^{ACTUAL}$. Consequently, before calculating the new equilibrium for any simulation we first solve for the initial prediction error associated with each neighborhood *n*:

$$(21) \qquad \omega_n = \overline{Z}_n^{ACTUAL} - \overline{Z}_n^{PREDICT}$$

In solving for the new equilibrium, we add this initial prediction error ω_n to the sociodemographic measures calculated in each iteration before substituting these measures back into the utility function.

Adjusting Crime Rates and Average Test Scores

Because some neighborhood attributes, such as crime rates and average test scores, depend in part on the sociodemographic composition of the neighborhood, it is natural to expect these neighborhood characteristics to adjust as part of the movement to a new sorting equilibrium. Getting precise measures of the impact of neighborhood sociodemographic characteristics on crime rates and test scores is, of course, an exceedingly difficult exercise, as selection problems abound. For example, an OLS regression of crime rates on neighborhood sociodemographic characteristics almost certainly overstates the role of these characteristics in producing crime as it ignores the fact that households sort non-randomly across neighborhoods. As a result, we take an approach that seeks to provide bounds for the characteristics of the new equilibrium that results for each of our simulations. For one bound, we calculate a new equilibrium without allowing crime rates and average test scores to adjust with the changing neighborhood and school sociodemographic compositions. For the other bound, we calculate a new equilibrium, adjusting crime rates and average test scores according the adjustments implied by an OLS regression of

[35] While this procedure always converges to an equilibrium, the model does not guarantee that this equilibrium is generically unique. In all of the calculations presented in this paper, we report results that start from the initial equilibrium and follow the procedure summarized here. In general, the counterfactual simulations conducted do not change the overall economic environment very much at all, and consequently, we believe that this procedure yields reasonable results.

the crime rate and average test score on neighborhood and school sociodemographic composition. These simple production functions are shown in Appendix Table 2, with all of the variables constructed to have mean zero and standard deviation one. The first bound will understate the impact of sociodemographic shifts on the implied crime rate and average test score in each neighborhood, while the second bound will tend to overstate the impact of these sociodemographic shifts. As the results below indicate, these bounds provide a tight range for the predictions from our simulations.

Partial and General Equilibrium Capitalization

Table 9 reports the distribution of average house price changes in the neighborhood corresponding to each school throughout the Bay Area that result from increasing the average test score of that school by 74 points (1 standard deviation).[36] The partial equilibrium results reveal a mean capitalization estimate of $27.10 per month ($7,200 in house value) and a median estimate of $26.40 per month. Not surprisingly, these numbers closely resemble the estimated mean MWTP in our preferred specification of $26 per month and in the hedonic price regression of $28 per month. Thus, the hedonic price regression comes close to measuring the mean partial equilibrium capitalization of a marginal change in school quality. The general equilibrium results reveal a mean capitalization estimate of $46-48 per month ($12,100-$12,600 in house value) and a similar median estimate. Thus, the full general equilibrium capitalization of school quality is 70-75 percent greater than the direct (partial equilibrium) effect of school quality on housing prices.

The table also shows changes in socio-demographics for the neighborhoods where test scores rose.[37] Increasing test scores leads to an increase in the proportion of high-income households in the neighborhood and a reduction in the proportion of poor households. The change is magnified comparing the general equilibrium change to that in partial equilibrium, where the socio-demographics are constrained not to fully adjust. For instance, while the change in the proportion of high income households rises by 0.8 percentage points in partial equilibrium, it rise by 1.3 percentage points in general equilibrium. In line with this, a one standard deviation increase in its test score would raise average incomes in a school's vicinity by an average of $1050 in partial equilibrium, and over $1770 in general equilibrium. A similar pattern is apparent for education: the proportion of highly educated households in the neighborhood rises, with the effect being more pronounced in general equilibrium. And in terms of race, the proportion of

[36] In each simulation, average house prices in the Bay Area as a whole are constrained to equal the pre-simulation level.

white households rises – by over one percentage point in general equilibrium – while the proportion of black households falls – by over 0.8 percentage points.

The other significant feature of these simulation results is the heterogeneity in capitalization across the metropolitan area. In general equilibrium, the price increase accompanying a one standard deviation increase in school quality is roughly twice as large at the 90th percentile (\$56.80-\$59.10) than at the 10th percentile (\$28.80-\$29.64). These numbers, which reflect both the underlying heterogeneity in the population in preferences for neighborhood sociodemographic characteristics as well as school quality, again emphasize the importance of using the broader heterogeneous preferences model when exploring questions related to demand for school quality.

7 CONCLUSION

This paper has set out a new approach for recovering underlying preferences for school and neighborhood attributes that addresses a series of challenging selection and endogeneity problems. The framework itself builds on the strengths of both the hedonic price and traditional discrete choice approaches, and we show that it nests both of these as direct restrictions. It has several appealing features: first, the approach allows for a great deal of heterogeneity in household preferences. Then, to account for non-random sorting, it models the household location choice decision directly, allowing household choices to be influenced by choice-specific unobservables. Because certain neighborhood characteristics determined through the sorting process are likely to be correlated with these unobservables, we account for the endogeneity of these characteristics using a boundary fixed effects approach. This allows us to obtain consistent estimates of preferences for both school and neighborhood attributes.

Estimating our equilibrium model makes significant data requirements. We need to know the household characteristics of a very large set of households drawn from a large metropolitan area, the actual housing choices they make, and detailed information about the local neighborhoods they live in. In this regard, we are fortunate to have access to newly available restricted Census data that allow us to construct a unique data set well-suited to the study of household location choices. Using these data, we estimated the model to obtain a large set of preferences parameters - the most reliable yet to appear in the literature. And we used the equilibrium model to explore the general equilibrium consequences of these estimates, drawing attention to the operation of an important social multiplier neglected in the prior capitalization literature.

[37] This part of the table has not yet been disclosed from the Census Data Center.

The estimates of the equilibrium model of sorting also serve as a platform for analyzing a number of demand-related applications in the future, i.e., in addition to the analysis of capitalization presented in this paper. The estimated model provides a well-defined characterization of the relative importance of schooling versus other housing, neighborhood, and geographic factors in driving the location decisions of the heterogeneous households of a major metropolitan area. This combination is extremely powerful for conducting economic and policy research involving the interplay of household mobility/stratification and schools, making it possible, for example, to calculate the elasticity of neighborhood house prices and rents as well as the sociodemographic composition of the local neighborhood and school with respect to school quality for each school in the metropolitan region. These 'demand' elasticities provide a series of measure of the competitiveness of a school's local environment and can be used to explore many aspects of the interplay between household mobility and school competition. Moreover, a slightly extended version of the model provides a way of calculating the strength of preferences for school quality (on the basis of both observed and unobserved characteristics) among the households that select into a particular school. This permits the researcher to control directly for the non-random sorting of households across schools and school districts which leads to a form of selection bias (often referred to as Tiebout bias in the local public finance literature) in the estimation of education production functions, voting models, or other models that condition implicitly on the set of households in a particular school or jurisdiction.

REFERENCES

Bajari, Patrick and Lanier Benkhard, (2002), "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," unpublished manuscript, Stanford University.

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

Barrow, Lisa (2003), "School Choice through Relocation: Evidence from the Washington, D.C. area," *Journal of Public Economics*, forthcoming.

Bayer, Patrick, (1999), *An Empirical Analysis of the Equilibrium in the Education Market,* Stanford University Dissertation.

Bayer, Patrick, Robert McMillan, and Kim Rueben, (2002), "An Equilibrium Model of an Urban Housing Market: A Study of the Causes and Consequences of Residential Segregation," unpublished manuscript, Yale University.

Benabou, Roland, (1993), "The Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, 108(3), pp.619-652.

Benabou, Roland, (1996), "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance," *American Economic Review*, Vol. 86, No. 3., pp. 584-609.

Berry, Steven, (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, Vol. 25, pp. 242-262.

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

Black, Sandra (1999) "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, May 1999.

Bogart, William T. and Brian A. Cromwell (2000) "How Much is a Neighborhood School Worth?" *Journal of Urban Economics*, 47, 280-305.

Brock, William and Steven Durlauf (2001), "Discrete Choice With Social Interactions" *The Review of Economic Studies* 68, No. 2, pp. 235-260.

Brock, William and Steven Durlauf (2001), "A Multinomial Choice Model of Neighborhood Effects, *American Economic Review* **92**, No. 2, pp. 298-303.

Clapp, John M., and Stephen L. Ross, (2002) "Schools and Housing Markets: An Examination of School Segregation and Performance in Connecticut,' unpublished manuscript, University of Connecticut.

Epple, Dennis, (1987), "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *Journal of Political Economy*, 107: 645-81.

Epple, D., R. Filimon, and T. Romer, (1984), "Equilibrium Among Local Jurisdictions: Towards an Integrated Approach of Voting and Residential Choice," *Journal of Public Economics*, Vol. 24, pp. 281-304.

Epple, D., R. Filimon, and T. Romer, (1993), "Existence of Voting and Housing Equilibrium in a System of Communities with Property Taxes," *Regional Science and Urban Economics*, Vol. 23, pp. 585-610.

Epple, Dennis and Holger Sieg, (1999), "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, Vol. 107, No. 4., pp. 645-681.

Epple, Dennis and Richard Romano, (1998), " Competition Between Private and Public Schools, Vouchers, and Peer Group Effects," *American Economic Review* 88(1): 33-62.

Fernandez, Raquel and Richard Rogerson, (1996), "Income Distribution, Communities, and the Quality of Public Education." *Quarterly Journal of Economics*, Vol. 111, No. 1., pp. 135-164.

Figlio, David and Maurice Lucas (2000) "What's in a Grade? School Report Cards and House Prices," NBER Working Paper 8019, Cambridge, MA.

Hayes, Kathy J. and Lori L. Taylor, (1996) "Neighborhood School Characteristics: What Signals Quality to Homebuyer?" *Economic Review: Federal Reserve Back of Dallas* (4), 2-9.

Heckman, James, Rosa Matzkin, and Lars Nesheim, (2003), "Simulation and Estimation of Hedonic Models," unpublished manuscript, University of Chicago.

Kane, Thomas, Douglas Staiger, and Gavin Samms, (2003) "School Accountability Ratings and Houise Values," *Brookings-Wharton Papers on Urban Affairs*, forthcoming.

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

Nechyba, Thomas J., (1997), "Existence of Equilibrium and Stratification in Local and Hierarchical Tiebout Economies with Property Taxes and Voting," *Economic Theory*, Vol. 10, pp. 277-304.

Nechyba, Thomas J., (1999), "School Finance Induced Migration and Stratification Patterns: the Impact of Private School Vouchers," *Journal of Public Economic Theory*, Vol. 1.

Nechyba, Thomas J., (2000), Mobility, Targeting and Private School Vouchers," *American Economic Review* 90(1), 130-46.

Nechyba, Thomas J., and Robert P. Strauss, (1998), "Community Choice and Local Public Services: A Discrete Choice Approach," *Regional Science and Urban Economics*, Vol. 28, pp. 51-73.

Oates, Wallace "The Effects of Property Taxes and Local Public Spending on Property Values: An empirical Study of Tax Capitalization and the Tiebout Hypothesis," *Journal of Political Economy* 77(6): 957-71.

Quigley, John M., (1985), "Consumer Choice of Dwelling, Neighborhood, and Public Services," *Regional Science and Urban Economics*, Vol. 15(1).

Rosen, Sherwin, (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82: 34-55.

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

Figure 1: Demand for a View of the Golden Gate Bridge

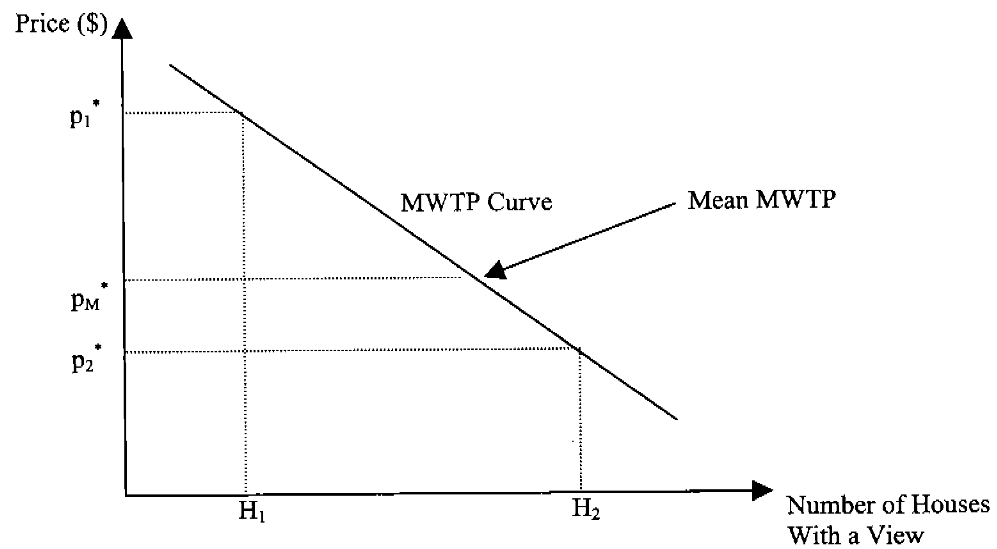

Figure 2: Demand for School Quality

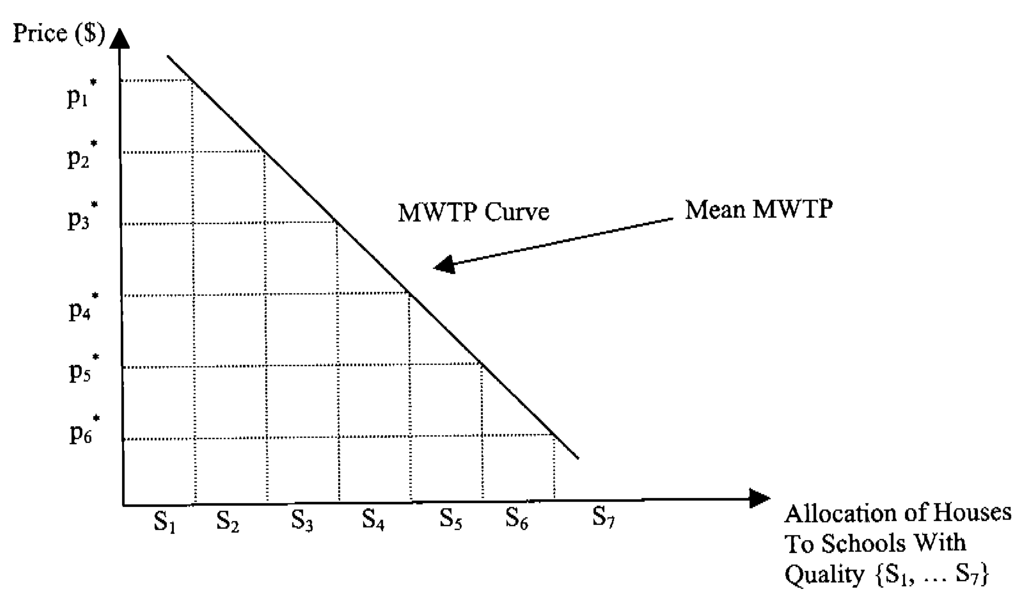

Table 1. Comparing the Sample Near School District Boundaries

Sample	**full sample**		**within 0.25 miles of boundaries**			
Boundary/Weights			**actual sample**	**high test score side**	**low test score side**	**sample weights**
Observations	**242,100**		**27,958**	**13,348**	**14,610**	**27,958**
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	S.D.	Mean	Mean	Mean	Mean
Housing/Neighborhood Characteristics						
monthly house price	1,087	755	1,130	1,158	1,105	1,098
average test score	527	74	536	558	515	529
1 if unit owned	0.597	0.491	0.629	0.632	0.626	0.616
number of rooms	5.114	1.992	5.170	5.207	5.134	5.180
1 if built in 1980s	0.143	0.350	0.108	0.118	0.099	0.148
1 if built in 1960s or 1970s	0.391	0.488	0.424	0.412	0.437	0.406
elevation	210	179	193	194	192	212
population density	0.434	0.497	0.352	0.349	0.355	0.374
crime index	8.184	10.777	6.100	6.000	6.192	7.000
% census block group white	0.681	0.232	0.704	0.712	0.686	0.676
% census block group black	0.081	0.159	0.071	0.065	0.076	0.080
% census block group Hispanic	0.110	0.114	0.113	0.107	0.119	0.117
% census block group Asian	0.122	0.120	0.112	0.110	0.113	0.121
% block group college degree or more	0.438	0.196	0.457	0.463	0.451	0.433
average block group income	54,744	26,075	57,039	58,771	55,457	55,262
Household Characteristics						
household income	54,103	50,719	56,663	58,041	55,405	55,498
1 if children under 18 in household	0.333	0.471	0.324	0.322	0.325	0.336
1 if black	0.076	0.264	0.066	0.062	0.070	0.076
1 if hispanic	0.109	0.312	0.111	0.102	0.119	0.115
1 if asian	0.124	0.329	0.112	0.114	0.110	0.121
1 if white	0.686	0.464	0.706	0.717	0.696	0.682
1 if less than high school	0.154	0.361	0.141	0.134	0.147	0.152
1 if high school	0.184	0.388	0.176	0.177	0.175	0.183
1 if some college	0.223	0.417	0.222	0.222	0.223	0.225
1 if college degree	0.291	0.454	0.294	0.295	0.294	0.286
1 if more than college	0.147	0.354	0.166	0.172	0.161	0.155
age (years)	47.607	16.619	47.890	48.104	47.699	47.660
1 if working	0.698	0.459	0.705	0.702	0.709	0.701
distance to work (miles)	8.843	8.597	8.450	8.412	8.492	8.490

Table 2: Hedonic Price Regressions

	Without Neighborhood Scoiodemographics			With Neighborhood Scoiodemographics		
Sample	full sample	within .25 mile of boundaries		full sample	within .25 mile of boundaries	
Boundary Fixed Effects	No	No	Yes	No	No	Yes
Observations	242,100	27,958	27,958	242,100	27,958	27,958
	(1)	(2)	(3)	(4)	(5)	(6)
average test score (in standard deviations	**145.62**	**144.96**	**101.17**	**28.76**	**27.63**	**28.43**
	(1.57)	**(3.85)**	**(4.15)**	**(1.70)**	**(4.76)**	**(5.88)**
1 if unit owned	151.59	124.67	140.84	125.59	112.21	123.91
	(3.12)	(8.77)	(8.47)	(2.94)	(8.16)	(8.16)
number of rooms	153.00	155.79	143.40	121.88	122.48	121.15
	(0.79)	(2.24)	(2.20)	(0.76)	(2.16)	(2.16)
1 if built in 1980s	129.28	91.28	130.12	89.48	92.79	109.41
	(3.89)	(11.01)	(11.22)	(3.71)	(10.36)	(11.01)
1 if built in 1960s or 1970s	22.98	8.82	93.30	4.13	4.33	6.96
	(2.85)	(8.20)	(8.05)	(2.69)	(7.74)	(8.20)
elevation (/100)	31.80	3.50	48.34	2.53	-17.42	13.29
	(0.79)	(2.53)	(3.25)	(0.76)	(2.43)	(4.85)
population density	-78.26	-155.27	-145.41	46.14	61.17	33.54
	(4.29)	(15.63)	(18.97)	(4.31)	(15.79)	(19.43)
crime index	0.75	2.16	6.84	1.41	0.49	4.97
	(0.19)	(0.71)	(1.79)	(0.20)	(0.81)	(1.85)
% census block group Black				-71.56	-113.70	-40.74
				(10.21)	(32.21)	(38.37)
% census block group Hispanic				128.63	146.56	240.31
				(14.42)	(46.21)	(60.74)
% census block group Asian				-1.38	-76.33	200.60
				(11.25)	(37.85)	(53.62)
% block group college degree or more				286.57	192.17	91.51
				(10.16)	(30.31)	(43.25)
average block group income (/10000)				100.32	110.84	101.26
				(16.14)	(55.33)	(60.03)
F-statistic for boundary fixed effects			23.345			8.754
R^2	0.37	0.38	0.44	0.44	0.47	0.49

Note: All regressions shown in table also include controls for land use (% industrial, % residential, % commercial, % open space, % other) in 1, 2, 3, 4, and 5 mile rings around location and six variables that characterize the housing stock in each of these rings. The dependent variable is monthly house price, which equals monthly rent for renter-occupied units and a monthly price for owner-occupied housing calculated as described in the text. Standard errors are in parentheses.

Table 3. Interaction Parameter Estimates - Model Without Neighborhood Sociodemographics

		House Characteristics					Neighborhood Attributes			
Household Characteristics	**Average Test Score (+1.s.d.)**	Monthly House Price (/1000)	Owner Occupied	Number of Rooms	Built in 1980s	Built in 1960-1979	Elevation (/100)	Population Density	Crime Index	Distance to Work
household income	**0.050**	0.121	0.305	0.074	0.142	0.038	0.016	0.028	-0.001	-0.004
(/10,000)	**(0.004)**	(0.003)	(0.010)	(0.002)	(0.011)	(0.009)	(0.001)	(0.013)	(0.001)	(0.001)
1 if children under 18 in household	**-0.190**	0.063	-0.102	0.544	-0.316	0.146	0.010	-0.740	0.015	0.036
	(0.047)	(0.065)	(0.094)	(0.025)	(0.112)	(0.083)	(0.022)	(0.101)	(0.005)	(0.005)
1 if black	**-1.395**	-0.941	-0.510	0.152	0.004	0.401	-0.062	-1.285	0.110	-0.023
	(0.080)	(0.127)	(0.167)	(0.044)	(0.211)	(0.144)	(0.041)	(0.159)	(0.007)	(0.011)
1 if hispanic	**-0.642**	0.168	-0.036	-0.268	-0.180	-0.157	-0.104	-0.155	0.050	0.014
	(0.072)	(0.122)	(0.130)	(0.036)	(0.164)	(0.115)	(0.040)	(0.136)	(0.007)	(0.007)
1 if asian	**-0.167**	0.315	1.765	-0.503	1.037	0.686	-0.015	0.941	0.030	0.003
	(0.062)	(0.080)	(0.122)	(0.031)	(0.145)	(0.108)	(0.028)	(0.095)	(0.006)	(0.007)
1 if college degree or more	**0.787**	0.917	-0.032	-0.012	0.489	-0.045	0.225	-0.007	0.031	-0.006
	(0.053)	(0.071)	(0.108)	(0.029)	(0.135)	(0.093)	(0.024)	(0.111)	(0.006)	(0.006)
1 if working	**0.007**	0.244	0.563	0.032	0.641	0.406	-0.048	-0.437	-0.027	-0.858
	(0.049)	(0.067)	(0.103)	(0.027)	(0.125)	(0.086)	(0.025)	(0.097)	(0.005)	(0.008)
age (years)	**0.015**	0.010	0.090	0.004	-0.034	-0.009	0.003	-0.006	0.001	-0.001
	(0.001)	(0.002)	(0.003)	(0.001)	(0.004)	(0.003)	(0.001)	(0.003)	(0.000)	(0.000)

Note: The parameters shown describe the elements of the utility function that interact household characteristics, shown in row headings, with choice characteristics, shown in column headings. Standard errors are in parentheses.

Table 4. Interaction Parameter Estimates - Model With Neighborhood Sociodemographics

Household Characteristics	Average Test Score (+1 s.d.)	House Characteristics: Monthly House Price (/1000)	Owner Occupied	Number of Rooms	Built in 1980s	Built in 1960-1979	Neighborhood Attributes: Elevation (/100)	Population Density	Crime Index	Neighborhood Sociodemographics: % Block Group Black	% Block Group Hisp	% Block Group Asian	% Blk Group College	Blk Group Avg Income	Distance to Work (miles)
household income	**0.020**	0.121	0.303	0.076	0.144	0.028	0.010	0.011	-0.001	-0.223	0.113	-0.009	0.385	0.012	-0.004
(+10,000)	**(0.005)**	(0.004)	(0.011)	(0.003)	(0.012)	(0.009)	(0.002)	(0.017)	(0.001)	(0.060)	(0.064)	(0.039)	(0.034)	(0.002)	(0.001)
1 if children under 18 in household	**0.102**	0.231	-0.238	0.582	-0.399	0.095	0.051	-0.947	0.002	1.594	2.294	1.857	-2.171	0.055	0.027
	(0.058)	(0.075)	(0.103)	(0.028)	(0.122)	(0.092)	(0.025)	(0.127)	(0.006)	(0.416)	(0.527)	(0.387)	(0.331)	(0.016)	(0.005)
1 if black	**-0.282**	0.143	-1.006	0.002	0.027	0.577	-0.068	-1.106	0.045	14.874	7.082	7.371	2.607	-0.023	-0.010
	(0.116)	(0.170)	(0.205)	(0.053)	(0.253)	(0.184)	(0.052)	(0.228)	(0.009)	(0.560)	(0.888)	(0.747)	(0.680)	(0.035)	(0.013)
1 if hispanic	**-0.077**	0.204	-0.138	-0.246	-0.147	-0.248	-0.067	-0.128	0.005	4.435	12.471	2.757	0.830	0.011	0.012
	(0.089)	(0.139)	(0.147)	(0.041)	(0.185)	(0.131)	(0.045)	(0.169)	(0.008)	(0.568)	(0.620)	(0.587)	(0.492)	(0.022)	(0.008)
1 if asian	**0.072**	0.558	1.633	-0.571	0.612	0.457	-0.006	-0.053	0.006	4.236	3.330	14.060	-0.016	-0.022	0.012
	(0.078)	(0.095)	(0.138)	(0.035)	(0.166)	(0.123)	(0.033)	(0.132)	(0.007)	(0.562)	(0.721)	(0.429)	(0.449)	(0.022)	(0.007)
1 if college degree or more	**0.200**	0.501	0.428	0.006	0.588	0.106	0.031	0.486	0.022	1.279	-0.638	-1.935	8.986	0.009	0.009
	(0.065)	(0.079)	(0.118)	(0.032)	(0.148)	(0.101)	(0.027)	(0.134)	(0.007)	(0.504)	(0.607)	(0.450)	(0.366)	(0.020)	(0.007)
1 if working	**0.093**	0.272	0.604	0.021	0.897	0.425	0.023	-0.515	-0.019	-0.712	-0.335	-0.434	-1.931	0.033	-0.896
	(0.062)	(0.074)	(0.113)	(0.030)	(0.138)	(0.096)	(0.028)	(0.125)	(0.007)	(0.444)	(0.563)	(0.436)	(0.350)	(0.016)	(0.009)
age (years)	**0.013**	0.011	0.097	0.003	-0.033	-0.010	0.003	-0.011	0.001	-0.022	-0.085	-0.005	-0.018	0.001	-0.001
	(0.002)	(0.002)	(0.003)	(0.001)	(0.004)	(0.003)	(0.001)	(0.003)	(0.000)	(0.013)	(0.016)	(0.013)	(0.010)	(0.001)	(0.000)

Note: The parameters shown describe the elements of the utility function that interact household characteristics, shown in row headings, with choice characteristics, shown in column headings. Standard errors are in parentheses.

Table 5. Partial Correlations Between Choice Variables and Choice Specific Constant (δ)

Sample	full sample	within .25 miles of boundaries	within .25 miles of boundaries
Boundary Fixed Effects	No	No	Yes
Observations	242,100	27,958	27,958
	(1)	(2)	(3)
<u>Housing/Neighborhood Variables</u>			
monthly house price	-0.057	-0.057	-0.056
	(0.001)	(0.002)	(0.002)
average test score	-0.016	-0.015	-0.003
	(0.001)	(0.002)	(0.002)
% census block group black	-0.078	-0.050	-0.038
	(0.001)	(0.002)	(0.002)
% census block group Hispanic	-0.030	-0.023	-0.012
	(0.001)	(0.002)	(0.001)
% census block group Asian	-0.034	-0.004	-0.004
	(0.001)	(0.002)	(0.001)
% block group college degree or more	-0.001	0.003	0.031
	(0.001)	(0.002)	(0.001)
average block group income	-0.054	-0.066	-0.045
	(0.001)	(0.002)	(0.002)
number of rooms	0.001	-0.009	-0.013
	(0.001)	(0.002)	(0.002)

Note: Figures in table are partial correlations conditional on all other covariates. Other covariates include all housing and neighbrohood characteristics shown in Table 2 as well as controls for land use (% industrial, % residential, % commercial, % open space, % other) in 1, 2, 3, 4, and 5 mile rings around location and six variables that characterize the housing stock in each of these rings. Standard errors are in parentheses.

Table 6: First Stage Price Regressions

	Without Neighborhood Scoiodemographics			With Neighborhood Scoiodemographics		
Sample	full sample	within .25 mile of boundaries		full sample	within .25 mile of boundaries	
Boundary Fixed Effects	No	No	Yes	No	No	Yes
Observations	242,100	27,958	27,958	242,100	27,958	27,958
	(1)	(2)	(3)	(4)	(5)	(6)
average test score (in standard deviations)	25.92	18.57	-9.51	4.96	5.29	11.02
	(1.83)	(5.06)	(6.27)	(1.69)	(4.68)	(5.81)
1 if unit owned	44.16	6.62	37.09	37.57	22.32	33.86
	(3.16)	(8.93)	(8.77)	(3.02)	(8.47)	(8.47)
number of rooms	37.36	33.20	38.65	31.27	30.20	29.18
	(1.24)	(3.41)	(3.41)	(1.20)	(3.45)	(3.41)
1 if built in 1980s	19.24	-34.31	13.81	14.84	5.18	17.91
	(3.90)	(10.79)	(11.44)	(3.72)	(10.57)	(11.22)
1 if built in 1960s or 1970s	-0.98	-19.80	-8.97	-4.10	-9.59	-11.29
	(2.78)	(8.05)	(8.36)	(2.64)	(7.58)	(7.74)
elevation (/100)	6.72	-18.47	33.57	1.70	-14.55	15.99
	(0.79)	(2.53)	(4.89)	(0.75)	(2.36)	(4.77)
population density	-59.53	-113.08	-102.61	10.70	32.79	16.09
	(4.18)	(15.18)	(18.67)	(4.24)	(15.48)	(19.12)
crime index+A28	-0.77	-0.11	4.01	0.53	-0.84	2.57
	(0.18)	(0.70)	(1.74)	(0.20)	(0.80)	(1.81)
% census block group Black				-36.28	-50.69	0.47
				(6.88)	(31.74)	(37.90)
% census block group Hispanic				9.14	26.41	138.64
				(10.19)	(46.21)	(59.42)
% census block group Asian				-6.50	-9.46	239.71
				(11.55)	(37.22)	(50.47)
% block group college degree or more				163.75	16.06	-47.61
				(16.58)	(30.12)	(42.45)
average block group income (/10000)				280.60	29.12	21.17
				(14.24)	(55.33)	(60.03)
Optimal price instrument	**0.738**	**0.771**	**0.663**	**0.736**	**0.740**	**0.731**
	(0.006)	**(0.018)**	**(0.018)**	**(0.008)**	**(0.022)**	**(0.022)**
constant	175.28	158.49	204.63	180.23	187.33	226.98
	(1.90)	(7.39)	(17.87)	(2.17)	(7.99)	(17.92)
F-statistic for price instrument	13973	1843	1411	9398	1151	1130
R^2	**0.40**	**0.42**	**0.47**	**0.46**	**0.49**	**0.51**

Note: All regressions shown in table also include controls for land use (% industrial, % residential, % commercial, % open space, % other) in 1, 2, 3, 4, and 5 mile rings around location and six variables that characterize the housing stock in each of these rings. The dependent variable is monthly house price, which equals monthly rent for renter-occupied units and a monthly price for owner-occupied housing calculated as described in the text. Standard errors are in parentheses.

Table 7: Implied Mean MWTP Measures

	Without Neighborhood Sociodemographics			With Neighborhood Sociodemographics		
Sample	full sample	within .25 mile of boundaries		full sample	within .25 mile of boundaries	
Boundary Fized Effects	No	No	Yes	No	No	Yes
Observations	242,100	27,958	27,958	242,100	27,958	27,958
	(1)	(2)	(3)	(4)	(5)	(6)
average test score (in standard deviations)	**126.08**	**122.89**	**81.53**	**20.17**	**20.19**	**26.22**
	(1.96)	**(5.36)**	**(0.77)**	**(1.72)**	**(4.77)**	**(6.13)**
1 if unit owned	209.76	178.37	184.54	165.38	150.77	161.05
	(3.29)	(8.99)	(1.14)	(3.19)	(8.76)	(9.24)
number of rooms	148.98	149.36	138.71	122.03	121.12	118.93
	(1.51)	(4.24)	(0.55)	(1.48)	(4.23)	(4.40)
1 if built in 1980s	129.93	74.74	106.17	99.69	85.50	95.55
	(3.94)	(10.87)	(1.44)	(3.79)	(10.69)	(11.84)
1 if built in 1960s or 1970s	28.48	9.46	15.39	13.79	7.40	4.50
	(2.78)	(8.03)	(1.05)	(2.67)	(7.71)	(8.51)
elevation (/100)	21.09	-4.82	46.46	-1.06	-18.04	12.83
	(0.81)	(2.48)	(0.64)	(0.75)	(2.46)	(5.04)
population density	-100.43	-153.53	-133.08	19.41	41.68	30.33
	(4.23)	(15.64)	(2.38)	(4.30)	(15.76)	(20.09)
crime index	-2.95	-2.30	1.78	0.00	-1.39	1.96
	(0.18)	(0.70)	(0.22)	(0.20)	(0.81)	(1.91)
% census block group black				-324.67	-318.83	-267.08
				(10.14)	(32.15)	(39.84)
% census block group Hispanic				-4.42	18.06	138.95
				(14.35)	(46.87)	(63.13)
% census block group Asian				-97.39	-96.22	155.27
				(11.15)	(37.39)	(55.73)
% block group college degree or more				286.02	206.02	137.71
				(10.50)	(30.58)	(44.53)
average block group income				87.08	96.11	87.61
				(1.25)	(3.86)	(4.00)
F-statistic for boundary fixed effects			5.349			4.162

Note: Specification shown in table also include controls for land use (% industrial, % residential, % commercial, % open space, % other) in 1, 2, 3, 4, and 5 mile rings around location and six variables that characterize the housing stock in each of these rings. MWTP measures are reported in terms of a monthly house price. Standard errors are in parentheses.

Table 8. Heterogeneity in Marginal Willingness to Pay for Select Housing/Neighborhood Attributes

		House Characteristics			Neighborhood Sociodemographics				
	Average Test Score +1 s.d.	Own vs. Rent	+1 Room	Built in 1980s vs. pre-1960	+10% Black vs. White	+10% Hisp vs. White	+10% Asian vs. White	+10% College Educated	Blk Group Avg Income + $10,000
Mean MWTP	**26.22**	161.05	118.93	95.55	-26.71	13.90	15.53	13.77	87.61
	(6.13)	(9.24)	(4.40)	(11.84)	(3.98)	(6.31)	(5.57)	(4.45)	(4.00)
Household Income (+$10,000)	**1.57**	21.84	6.12	10.51	-1.53	0.77	-0.05	2.62	1.54
	(0.35)	(0.71)	(0.17)	(0.76)	(0.39)	(0.41)	(0.25)	(0.22)	(0.11)
Children Under 18 vs.	**7.10**	-12.87	40.06	-24.52	10.38	15.03	12.17	-14.18	5.05
No Children	**(3.78)**	(6.67)	(1.80)	(7.94)	(2.70)	(3.41)	(2.51)	(2.15)	(1.06)
Black vs. White	**-18.05**	-63.55	1.56	2.95	96.82	46.13	48.02	16.99	-0.45
	(7.50)	(13.25)	(3.40)	(16.38)	(3.62)	(5.75)	(4.84)	(4.40)	(2.27)
Hispanic vs. White	**-4.64**	-6.44	-14.14	-8.07	28.89	81.36	18.01	5.43	2.07
	(5.80)	(9.53)	(2.63)	(12.00)	(3.68)	(4.01)	(3.81)	(3.19)	(1.41)
Asian vs. White	**5.79**	113.65	-32.92	43.94	27.74	21.95	92.49	-0.05	1.99
	(5.08)	(8.96)	(2.27)	(10.77)	(3.64)	(4.67)	(2.78)	(2.91)	(1.41)
College Degree or More vs.	**14.12**	33.83	4.50	42.06	8.34	-4.16	-12.70	59.29	3.66
Some College or Less	**(4.24)**	(7.67)	(2.05)	(9.57)	(3.27)	(3.94)	(2.91)	(2.37)	(1.29)
Householder Working vs.	**6.63**	42.72	3.69	60.60	-4.71	-2.17	-2.81	-12.62	3.88
Not Working	**(4.02)**	(7.31)	(1.94)	(8.92)	(2.88)	(3.65)	(2.82)	(2.27)	(1.04)
Age (+10 years)	**0.86**	6.49	0.30	-2.07	-0.15	-0.56	-0.03	-0.12	0.11
	(0.11)	(0.21)	(0.06)	(0.25)	(0.08)	(0.10)	(0.08)	(0.06)	(0.03)

Note: The first row of table reports the mean marginal willingness-to-pay (MWTP) for the change reported in the column heading. The remaining rows report the difference in willingness to pay associated with the change listed in the row heading holding all other factors equal. Standard errors are in parentheses.

Table 9. The Capitalization of School Quality: Distribution of Housing Price Changes

	Partial Equilibrium	General Equilibrium	
		Unadjusted	Adjusted
Percentile of Simulated Distribution	Monthly Price Increase	Monthly Price Increase	Monthly Price Increase
90	$33.95	$56.83	$59.05
50	$26.38	$45.80	$47.32
10	$19.45	$28.83	$29.64
Mean	$27.10	$46.00	$47.70

Note: The figures shown in this table report the mean and distribution of changes in monthly housing prices for a corresponding catchement area following an increase in a school's average test score by 74 points (1 standard deviation). The first column shows partial equilibrium results, which do not account for any subsequent changes to the neighborhood sociodemographic distribution. The second and thrid columsn report general equilibrium results, which account for sociodemographic changes to the neighborhood. In this case, the second column (unadjusted) reports results of simulatiosn that hold crime and the average test score and pre-simulation levels, while the third column adjusts crime and average test scores according to production functions estimated via OLS reported in Appendix Table 2.

Appendix Table 1: Choice Specific Constant Regressions

	Without Neighborhood Sociodemographics			With Neighborhood Sociodemographics		
Sample	full sample	within .25 mile of boundaries		full sample	within .25 mile of boundaries	
Boundary Fized Effects	No	No	Yes	No	No	Yes
Observations	242,100	27,958	27,958	242,100	27,958	27,958
monthly housing price (/1000)	-10.23	-9.73	-11.34	-15.94	-15.97	-16.19
	(1.39)	(1.13)	(1.36)	(1.71)	(1.56)	(1.69)
average test score (in standard deviations)	1.29	1.20	0.92	0.32	0.32	0.42
	(0.02)	(0.05)	(0.01)	(0.03)	(0.08)	(0.01)
1 if unit owned	2.15	1.74	2.09	2.64	2.41	2.61
	(0.03)	(0.09)	(0.01)	(0.05)	(0.14)	(0.01)
number of rooms	1.52	1.45	1.57	1.95	1.93	1.93
	(0.02)	(0.04)	(0.01)	(0.02)	(0.07)	(0.01)
1 if built in 1980s	1.33	0.73	1.20	1.59	1.37	1.55
	(0.04)	(0.11)	(0.02)	(0.06)	(0.17)	(0.02)
1 if built in 1960s or 1970s	0.29	0.09	0.17	0.22	0.12	0.07
	(0.03)	(0.08)	(0.01)	(0.04)	(0.12)	(0.01)
elevation (/100)	0.22	-0.05	0.53	-0.02	-0.29	0.21
	(0.01)	(0.02)	(0.01)	(0.01)	(0.04)	(0.01)
population density	-1.03	-1.49	-1.51	0.31	0.67	0.49
	(0.04)	(0.15)	(0.03)	(0.07)	(0.25)	(0.03)
crime index	-0.03	-0.02	0.02	0.00	-0.02	0.03
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)
% census block group black				-5.18	-5.09	-4.32
				(0.16)	(0.51)	(0.06)
% census block group Hispanic				-0.07	0.29	2.25
				(0.23)	(0.75)	(0.10)
% census block group Asian				-1.55	-1.54	2.51
				(0.18)	(0.60)	(0.09)
% block group college degree or more				4.56	3.29	2.23
				(0.17)	(0.49)	(0.07)
average block group income				1.39	1.53	1.42
				(0.02)	(0.06)	(0.01)
F-statistic for boundary fixed effects			4.545			3.963

Note: Specification shown in table also include controls for land use (% industrial, % residential, % commercial, % open space, % other) in 1, 2, 3, 4, and 5 mile rings around location and six variables that characterize the housing stock in each of these

Appendix Table 2: OLS Crime and Education Production Functions

	Production Function	
Dependent Variable	crime index	average test score
Observations	242,100	242,100
$\mathbf{R^2}$	0.33	0.41
Percent Black	0.285	-0.188
	(0.005)	(0.005)
Percent Hispanic	0.099	-0.074
	(0.004)	(0.003)
Percent Asian	0.088	-0.041
	(0.003)	(0.003)
Percent College Degree or More	0.017	0.127
	(0.004)	(0.004)
Average Income	-0.071	0.311
	(0.046)	(0.043)

Note: This table shows the results of the OLS estimation of simple crime and education production functions. These functions are used in the simulations that adjust crime and school quality with changing neighborhood sociodemographic composition. We use these 'adjusted' results to provide a bound on the simulation results. Standard errors are provided below estimated coefficient. All variables are normalized to have mean zero and standard deviation one.