NBER WORKING PAPER SERIES

ESTIMATING HOUSING DEMAND
WITH AN APPLICATION TO EXPLAINING
RACIAL SEGREGATION IN CITIES

Patrick Bajari
Matthew E. Kahn

Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities
Patrick Bajari and Matthew E. Kahn
NBER Working Paper No. 9891
July 2003
JEL No. R0, C5

ABSTRACT

We present a three-stage estimation procedure to recover willingness to pay for housing attributes.

In the first stage, we estimate a non-parametric hedonic home price function. Second, we recover

each consumer's taste parameters for product characteristics using first order conditions for utility

maximization. Finally, we estimate the distribution of household tastes as a function of household

demographics. As an application of our methods, we compare alternative explanations for why

blacks choose to live in center cities while whites suburbanize.

Patrick Bajari   Matthew E. Kahn
Department of Economics   Fletcher School
Duke University   Tufts University
305 Social Science Building   Medford, MA 02155
Box 90097   matt.kahn@tufts.edu
Durham, NC 27708-0097
and NBER
bajari@econ.duke.edu

1 Introduction.

Housing accounts for a major fraction of consumer spending and wealth. Shelter is the largest of the seven categories that comprise the CPI market basket, accounting for 32 percent of the total index. Housing equity represents the largest source of household wealth. Based on the 1998 Survey of Consumer Finances, the value of stocks owned by households was 7.8 trillion while the value of primary residences was 9.4 trillion. About two-thirds of American households own their homes while less than half of households own stocks. The majority of those who own both stocks and a home still have more wealth in home equity than in stocks (Di 2001).

In this paper, we study housing demand by applying a three-step procedure for estimating the demand for differentiated products described in Bajari and Benkard (2002). First, we estimate a hedonic house price function using a non-parametric technique based on the local polynomial methods described in Fan and Gijbels (1996). Second, we infer household specific preference parameters for continuous product characteristics using a first order condition for utility maximization. Finally, we model individual specific taste coefficients as a function of household demographics and household specific preference shocks. The methods that we propose are computationally straightforward and can be estimated using standard statistical packages.

This approach has a number of novel features that differ from existing studies of housing demand such as Palmquist (1984), Coulson and Bond (1990), Cheshire and Sheppard (1998), Epple and Sieg (1999) and Bayer, McMillan and Ruben (2002). First, the recent literature on discrete choice models emphasizes the importance of accounting for product characteristics observed by the consumer but not by the economist. Ignoring such attributes biases estimated price elasticities towards zero. See, for instance, Berry (1994), Berry, Levinsohn and Pakes (1995), Petrin (2002) and Ackerberg and Rysman (2002). In our analysis, we derive preferences for both observed and unobserved product characteristics. Second, our model has a flexible specification of consumer preferences. We use random coefficients and household level demographics

to model taste heterogeneity. In most previous models with random coefficients, such as Berry, Levinsohn and Pakes (1995) and McCollough, Polson and Rossi (2000), the random coefficients are typically assumed to be independently and normally distributed conditional on household demographics. In our analysis, we recover a non-parametric distribution of the random coefficients for continuous characteristics, and we only impose parametric assumptions for discrete product characteristics since they are required for identification. Finally, the methods that we propose allow for both discrete and continuous product characteristics.

As an application of our methods, we explore why the average white household lives in the suburbs while the average black household lives in the center city. We estimate our housing demand model for a sample of white and black migrants in three major cities to explore the merits of three potential explanations. The first hypothesis is that whites and blacks have different demands for the physical attributes of housing. Suburban homes are larger, newer and more likely to be single detached owner occupied units relative to center city housing units. Blacks may have a lower demand for such products because their household incomes are lower than whites. The second hypothesis is that racial segregation is inextricably tied to local labor markets and the disutility from commuting. If minority employment is disproportionately located in center cities and if the disutility from commuting is high, then blacks will urbanize.[2] The final hypothesis focuses on peer group selection. Communities differ with respect to their racial composition and their human capital levels. A household will recognize that by choosing a particular community it will determine who will be their neighbors and who their children will go to school with. The average suburban community has more college graduates and fewer minorities than the average urban community. Blacks and whites may differ with respect to their willingness to pay for these attributes. Estimates of willingness to pay for peers provides an alternative approach for recovering peer effects than the more standard "production function" approach as exemplified by Case and Katz (1991) and Crane (1991).

Manski (1993) has argued that a "production function" based approach for measuring peer effects often suffers from a reflection problem. By regressing a variable on its average, the peer group effect is not identified. Since we focus on estimating willingness to pay, if the hedonic price coefficients can be consistently estimated, then the preferences for peer groups can be identified from revealed preference under

---

[2] We recognize that another hypothesis is that black households seeking suburban housing products are discriminated against (Yinger 1986, Munnell et. al. 1996). Detecting and accounting for perceived or actual discrimination is beyond the scope of this paper (Heckman 1998).

our modeling assumptions. While consistently estimating implicit prices from hedonic coefficients can be difficult, the most important problem in practice is limited data and not a fundamental lack of identification as in Manski (1993).

The paper is organized as follows. In Section 2, we describe our micro data. The available information plays a key role in determining our modeling strategy. In Section 3, we present the three step structural housing demand model. In Section 4, we present details about the model and we thoroughly contrast our approach with recent methods used to estimate the demand for differentiated products. In Section 5, we present our empirical application and in Section 6 we conclude.

2 Data.

The micro data in our empirical application comes from the 1990 Census of Population and Housing Integrated Public Use Microdata Series (IPUMS) 1% unweighted sample. We study housing demand in three major metropolitan areas: Atlanta, Chicago, and Dallas.[3] Based on the Cutler, Glaeser, and Vigdor (1999) 1990 disimilarity index, Chicago is highly segregated, while Dallas features a low segregation score and Atlanta is in the middle of the distribution. We have intentionally chosen to not include high priced cities such as Los Angeles, New York and San Francisco because home prices are top coded in the Census data. The units for all dollar amounts in this paper are 1989 pre-tax dollars.

The Census data provides rich information on a household's demographic characteristics. For each member of the household roster, the Census reports age, race, education, employment status, income, place of residence and place of work. The household's owner or renter status is given. In addition, the Census publishes detailed housing unit characteristics such as the unit's age, number of rooms and whether the unit is single detached or part of a multi-unit building. Home prices and rents are self-reported. Following the convention in the urban literature, we define annual housing expenditure for owners as the reported home price multiplied by 7.5% (see Gyourko and Tracy 1991).

In the data, the place of residence identifiers are Public Use Micro Areas (PUMAs). PUMAs are intended to be similar areas containing 100,000 people or more. While PUMAs generally are aggregations of census tracts and urban places, they do not reflect the boundaries of political jurisdictions. The data set also

---

[3]   The Chicago metropolitan area consists of Chicago-Gary Lake, Aurora Elgin Gary-Hammond, East Chicago, Joliet, Illinois Lake County. Atlanta isthe Atlanta PMSA and Dallas consists of Dallas/Ft-Worth, Ft-Worth, and Arlington, Texas.

includes place of work identifiers that are more spatially aggregated than PUMA identifiers. We create measures of community characteristics by calculating PUMA specific means such as a community's percent of college graduates and percent of black households.

We focus on the housing choices of black and white migrant households. Migration status is determined by whether the household head lived in a different home in 1990 than in 1985. Unlike recent migrants, households who do not move may not be consuming their optimal housing bundle due to the transactions costs of moving. This makes it difficult to infer the preferences of non-movers from their current bundle.

3   A Model of Housing Choice.

In this section, we describe a method for estimating housing demand which builds on the techniques proposed in Bajari and Benkard (2002). In the model, there are $m = 1, ..., M$ metropolitan areas, $i = 1, ..., I_m$ individuals in metropolitan area m and $j = 1, ..., J_m$ housing units in metropolitan area m. In what follows, we shall suppress m in much of our notation, since we treat each city as a separate market and do not pool data from across cities in our estimation.

A home is a bundle of three types of attributes: physical attributes, community attributes, and attributes that are observed by the consumer but are not observed by the econometrician. The physical attributes include the number of rooms, the age of the unit, an ownership dummy and a dummy variable indicating that the unit is a single detached dwelling. The community attributes in our model are the percentage of black households in the PUMA, the percentage of college educated households in the PUMA and whether the PUMA is located in the center city. We let x_j denote the physical and community attributes of housing unit j, the unobserved product attribute is modeled as a scalar ξ_j and the price of the housing unit is p_j.

In the model, the price of housing unit j is a function $\mathbf{p}_m$, which maps the characteristics of a housing unit into prices as follows:

$$p_j = \mathbf{p}_m(x_j, \xi_j). \tag{1}$$

The hedonic is determined in equilibrium by the interaction of a large number of buyers and sellers. Household utility is a function of housing characteristics (x_j, ξ_j) and consumption of a composite commodity c, with a price normalized to one dollar (pre-tax). The utility that consumer i receives for product j, u_{ij}, can

5

therefore be written as

$$u_{ij} = u_i(x_j, \xi_j, c). \tag{2}$$

Households are rational, utility maximizers who choose their preferred bundle given their income y_i. Product $j^*(i)$ is utility maximizing for household i if

$$j^*(i) = \arg\max_j \ u_i(x_j, \xi_j, y_i - \mathbf{p}_m(x_j, \xi_j)). \tag{3}$$

Note that in equation (3) we have substituted the budget constraint directly into the utility function. Suppose that product characteristic k, $x_{j,k}$, is a continuous variable and that product j^* is utility maximizing for household i, then the following first order condition must hold

$$\frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial x_{j,k}} - \frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial c} \frac{\partial \mathbf{p}_m(x_{j^*}, \xi_{j^*})}{\partial x_{j,k}} = 0 \tag{4}$$

$$\frac{\frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial x_{j,k}}}{\frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial c}} = \frac{\partial \mathbf{p}_m(x_{j^*}, \xi_{j^*})}{\partial x_{j,k}}. \tag{5}$$

Equation (5) is the familiar condition that the marginal rate of substitution between a continuous characteristic and the composite commodity is equal to the partial derivative of the hedonic.

In the theory of revealed preference, if one observes a household's choices under all possible budget sets, then it is possible to recover the entire weak preference relation for that household under well known regularity conditions (see Mas-Colell (1977)). In the Census data, however, a single cross section of households is observed, and therefore global identification of preferences is not possible. From equation (5), it is clear that household preferences will only be identified locally. Therefore, we use the following specification for consumer preferences:

$$\begin{aligned} u_{ij} \ = \ & \beta_{i,1} \log(room_j) + \beta_{i,2} \log(age_j) + \beta_{i,3} \log(\xi_j) + \beta_{i,4} own_j + \beta_{i,5} single_j + \\ & \beta_{i,6} \log(mblack_j) + \beta_{i,7} \log(mba_j) + \beta_{i,8} city_j + c \end{aligned} \tag{6}$$

6

where

$$\beta_i = f(demo_i) + \eta_i \tag{7}$$

$$E(\eta_i | demo_i) = 0 \tag{8}$$

In the parametric model (6)-(8), each household i has a unique set of taste parameters $\beta_{i,1} - \beta_{i,8}$ for the product characteristics. Since utility functions are identified only up to monotonic transformations, without loss of generality, we normalize the coefficient on the composite commodity to one.

If we make the functional form assumption (6), equation (4) becomes

$$\frac{\beta_{i,k}}{x_{j^*,k}} = \frac{\partial \mathbf{p}_m(x_{j^*}, \xi_{j^*})}{\partial x_{j,k}} \tag{9}$$

$$\beta_{i,k} = x_{j^*,k} \frac{\partial \mathbf{p}_m(x_{j^*}, \xi_{j^*})}{\partial x_{j,k}} \tag{10}$$

Using equation (10), we note that $\beta_{i,k}$ is just identified given $x_{j^*,k}$ and the hedonic price function (1). Note that from equation (10), the population distribution of $\beta_{i,k}$ is non-parametrically identified. For a more complete discussion on non-parametric identification of tastes and of the hedonic, including hedonics where the error term enters in a general, non-linear fashion, see Bajari and Benkard (2002). In equation (7) and (8), household i's tastes are modeled as a function, f, of demographic characteristics, d_i, and a household specific residual η_i. In our model, d_i corresponds to age of the head of household, annual household income, household size, and dummy variables for whether the head of household is male, married, black, a college graduate and a center city worker.

We will not be able to identify the taste coefficients, $\beta_{i,k}$ corresponding to discrete product characteristics. Without parametric assumptions, we can only recover a bound on these parameters. For instance, suppose that household i chooses product j^*. Define $\hat{x}_i$ as a vector of observed characteristics with $single = 1$ and all elements correspond to x_{j^*}. Next, define $\overline{x}_i$ similarly, except $single = 0$. The implicit price faced by household i for single detached housing in the market is then $\frac{\Delta p_m}{\Delta single} = \mathbf{p}_m(\hat{x}_i, \xi_j) - \mathbf{p}_m(\overline{x}_i, \xi_j)$. It can easily be shown that the following inequalities bound household i's preferences for single detached housing

7

$$[single = 1] \implies \left[\beta_{i,5} > \frac{\Delta p_m}{\Delta single} \right], \tag{11}$$

$$[single = 0] \implies \left[\beta_{i,5} < \frac{\Delta p_m}{\Delta single} \right]. \tag{12}$$

That is, if household i lives in single detached housing we can infer that i's preference parameter exceeds the implicit price for this characteristic. Analogous conditions can be derived for the other discrete characteristics.

4   Estimation.

Our approach to estimation involves three steps. In the first step, we estimate the housing hedonic using flexible, non-parametric methods based on the techniques described in Fan and Gijbels (1996). Second, we use the first order condition (10) to infer household level preference parameters for continuous product characteristics. Third, we estimate (7) and (8) by regressing household level preference parameters on the demographic characteristics d_i. We also discuss how to estimate preferences for discrete characteristics by maximum likelihood.

4.1   First Step: Estimating the Hedonic.

We have experimented with several different methods for estimating the hedonic, including high order polynomials and kernel regression. The method that produced the most appealing estimates are based on local polynomial modeling. To estimate the implicit prices faced by household i, who chooses $j^*(i)$, we suppose that locally, the hedonic satisfies

$$\begin{aligned}
p_j &= \alpha_{0,j^*} + \alpha_{1,j^*} \log(room_j) + \alpha_{2,j^*} \log(age_j) + \alpha_{3,j^*} \log(\xi_j) + \alpha_{4,j^*} own_j + \\
&\quad \alpha_{5,j^*} single_j + \alpha_{6,j^*} \log(mblack_j) + \alpha_{7,j^*} \log(mba_j) + \alpha_{8,j^*} city_j + \varepsilon
\end{aligned}$$

In our notation, we denote the hedonic coefficients as $\alpha_{0,j^*} - \alpha_{8,j^*}$ to emphasize the fact that our estimates will be local in the sense that a unique set of implicit prices is estimated for each value of $j = 1, ..., J$. Following Fan and Gijbels (1996), we use weighted least squares to estimate $\alpha_{0,j^*} - \alpha_{8,j^*}$.

8

$$\alpha_{j^*} = \arg\min_{\alpha} (\mathbf{p} - \mathbf{X}\alpha)' \mathbf{W} (\mathbf{p} - \mathbf{X}\alpha) \tag{13}$$

$$\mathbf{p} = [p_j], \ \mathbf{X} = [x_j], \mathbf{W} = diag\{K_h(x_j - x_{j*})\} \tag{14}$$

In equation (13) and (14), $\mathbf{p}$ is the vector of all prices for all products j in a given metropolitan area, $\mathbf{X}$ is a vector of regressors, which for each product includes an intercept and eight characteristics and $\mathbf{W}$ is a matrix of kernel weights. Notice that the kernel weights are a function of the distance between product j^* and product j. Thus, the local linear regression assigns greater importance to observations near j^*. Local linear methods have the same asymptotic variance and a lower asymptotic bias than the Nadaraya-Watson estimator. Also, the Gasser-Mueller estimator has the same asymptotic bias and a higher asymptotic variance than local linear methods.

Our estimates of equation (13) and (14) allow us to recover an estimate of the unobserved product characteristic

$$\xi_{j^*} = p_{j^*} - \left(\begin{array}{c} \alpha_{0,j^*} + \alpha_{1,j^*} \log(room_j) + \alpha_{2,j^*} \log(age_j) + \alpha_{3,j^*} \log(\xi_j) + \alpha_{4,j^*} own_j + \\ \alpha_{5,j^*} single_j + \alpha_{6,j^*} \log(mblack_j) + \alpha_{7,j^*} \log(mba_j) + \alpha_{8,j^*} city_j \end{array} \right) \tag{15}$$

In equation (15), the unobserved product characteristic ξ_{j^*} is estimated as the residual to our hedonic regression. While there are certainly other interpretations of the hedonic residual, we believe that this interpretation is the most important in our data. Given the lack of appropriate instruments, we maintain the standard hedonic assumption that the unobserved product characteristics are independent of the observed product characteristics.

In local linear regressions, the choice of kernel and bandwidth is extremely important. We chose the following normal kernel function with a bandwidth of 3

$$K(z) = \Pi_{k=1}^{7} N(z_k/\widehat{\sigma}_k^2) \tag{16}$$

$$K_h(z) = K(z/h)/h \tag{17}$$

In equation (16), the function K is a product of standard normal distributions which we denote using N. In equation (16), for the kth characteristic, we evaluate the normal distribution at $z_k/\widehat{\sigma}_k^2$, where $\widehat{\sigma}_k^2$

9

is the sample standard deviation of characteristic k.[4] Fan and Gijbels (1996) describe asymptotically appropriate methods for choosing the bandwidth. However, in our application, since the hedonics depend on 8 covariates such asymptotic approximations are not likely to be very reliable. Based on visual inspection of the estimates, we choose the bandwidth equal to 3.

Given the large number of covariates, we should not interpret these estimates as "non-parametric". However, compared to other flexible functional forms, local linear methods appeared to give much more plausible estimates of the implicit prices. In all three cities, the estimated implicit prices have intuitively plausible signs and magnitudes.

4.2 Second Step: Applying the First Order Conditions.

After estimating the implicit prices, we next estimate the preferences for continuous characteristics. If household i chooses product j^*, equation (10) must hold. The left hand side of equation (10) is the agent's unobserved preference parameter. The right hand side depends only on $x_{j^*,k}$, agent i's choice of the continuous characteristic k and the implicit price of characteristic k faced by agent i in the market. Using estimates of implicit prices obtained from the first step and the observed choice of $x_{j^*,k}$, $\beta_{i,k}$ can be recovered. Since the preference parameter for every individual can be recovered using our first order condition, the population distribution of tastes can be estimated. In the next step, we describe how to estimate the joint distribution of tastes and demographics.

4.3 Third Step: Modeling the Joint Distribution of Tastes and Demographics.

After household level preference parameters are recovered, we then estimate (7)-(8). We could easily do this using very flexible methods as in our estimates of the hedonic. However, for presentation purposes, it is more convenient to model the joint distribution of tastes and demographic characteristics using a linear regression model.

Equation (11) and (12) demonstrate that preference parameters for discrete characteristics are not identified. In what follows, we shall assume that $\beta_{i,k}$ is normally distributed with a mean that depends on demographic characteristics and an unknown variance. If a household i chooses to live in a detached unit,

---

[4] We need to normalize the units of the different characteristics in equation (16). Otherwise, age, which takes on values from 1 to 70 would have a much greater influence on the weights than the percentage of college educated heads of households in the PUMA, which only takes on values from 0 to 1.

then equation (11) implies that $\beta_{i,k} > \frac{\Delta p_m}{\Delta single}$, and if i lives in an attached unit, then $\beta_{i,k} < \frac{\Delta p_m}{\Delta single}$. Conversely, if i lives in an attached unit, then the valuation must be less than the implicit price for detached housing. Suppose that for discrete characteristic k

$$\beta_{i,k} = h(d_i, \theta_k) + \eta_{i,k} \tag{18}$$

$$h(d_i, \theta_k) = \theta_0 + \sum_s \theta_s d_{i,s} \tag{19}$$

where $d_i = (d_{i,s})$ is a vector of household i's demographic characteristics, θ_k is a vector of parameters and $\eta_{i,k}$ is an i.i.d. taste shock to household i for the discrete characteristic k. If $\eta_{i,k}$ is normally distributed with mean zero and standard deviation σ, then by equation (11) and (12), then the probability that household $i = 1, ..., I$ chooses to live in single detached housing is

$$(1 - N(h(d_i; \theta_k) - \frac{\Delta p_m}{\Delta single}; \sigma)) \tag{20}$$

The likelihood function for the population distribution of tastes for single detached housing can be written as

$$L(\theta, \sigma) = \prod_{i=1}^{I} N(h(d_i; \theta_k) - \frac{\Delta p_m}{\Delta single}; \sigma)^{1 - single_{j*(i)}} (1 - N(h(d_i; \theta_k) - \frac{\Delta p_m}{\Delta single}; \sigma))^{single_{j*(i)}}. \tag{21}$$

where in equation (21), $single_{j*(i)}$ is an indicator variable that is equal to one if household i purchases single detached housing and zero otherwise. We estimate the model above using maximum likelihood. In principal, we could model correlation between the taste coefficients for all discrete product characteristics using a multivariate normal distribution and estimate a more flexible model of how tastes for characteristics are correlated. However, for expositional clarity, we estimate the tastes for each product characteristic independently. An alternative approach, which does not require assuming that tastes lie in a parametric family, is to use the bounds approach described in Bajari and Benkard (2002).

4.4 Discussion.

Since the methods we are employing are somewhat unique, it is useful to compare our approach to methods that have been used to estimate the demand for housing attributes and the demand for differentiated products

more generally. First, the recent literature on estimating discrete choice models has emphasized that price elasticities tend to be biased towards zero if the economist fails to account for ξ_j, product characteristics that are observed by the consumer but not by the economist. See, for instance, Berry (1994), Berry, Levinsohn and Pakes (1995), Petrin (2002) and Ackerberg and Rysman (2002). This problem is ignored in most previous studies of housing demand.[5] We expect this problem to be especially important in housing research based on standard Census data where researchers commonly find that the observed covariates explain half or less of the variation in prices.

Second, our model allows for a flexible specification of the joint distribution of tastes and demographic characteristics. In our model, we estimate the distribution of 9 random coefficients, 8 random coefficients for the observed product characteristics and one for unobserved product characteristics, as a function of 8 demographic variables. The random coefficients in our model can easily be estimated non-parametrically, except in the case of discrete product characteristics where we impose functional form assumptions for identification. In most previous papers that estimate discrete choice models, such as Berry, Levinsohn and Pakes (1995) or McCollough, Poulson and Rossi (2000), the random coefficients are typically assumed to be independently and normally distributed, conditional on demographics. Given that these models are computationally intensive, a parsimonious parametric specification is often needed. By contrast, our procedure is extremely easy to compute and can be quickly estimated using standard statistical packages.

The analyst does face a couple of trade-offs, however, in using our approach as compared to the random coefficient logit model estimated by Berry, Levinsohn and Pakes (1995) or the random coefficient probit estimated by McCollough, Poulson and Rossi (2000). The first is that in characteristic models such as ours, not all product are strong gross substitutes. The number of products with positive cross price elasticities is proportional to the number of characteristics used in the analysis. (See Anderson, DePalma and Thisse (1995) for a complete discussion.) In our example, since there are 8 product characteristics, we do not believe that this is particularly problematic, however, it can be an issue in data sets with fewer product characteristics. Second, the instrumental variable strategies proposed by Berry, Levinsohn and Pakes (1995) cannot be implemented in our model in a straightforward fashion.

Third, while our approach to estimating preferences is similar to the hedonic two-step of Rosen (1974)

---

[5] An important exception is Bayer, McMillan and Ruben (2002).

and Epple (1987), there are some important distinctions. In our model, household level preferences are only locally identified from a single cross section of data. While we do regress preferences on household level demographics to obtain an estimate of how preferences change with demographic characteristics of the household, we do not attempt to globally estimate preferences as in the standard second stage hedonic regression. As a result, our procedure does not involve a step that requires the economist to regress product characteristics on non-linear functions of product characteristics, which has been criticized as a limitation of second stage (Brown and Rosen 1982). Epple (1987) presents a formal model of the necessary conditions for achieving identification of structural parameters using the hedonic two step. Finding valid instruments to implement this procedure is often quite challenging. Hedonic researchers have instead attempted alternative identification strategies such as pooling multiple segmented markets at a point in time and arguing that differences in equilibrium prices and quantities across markets reflect differences in supply conditions, and this variation traces out "the" demand curve (Palmquist 1984).

Our paper is probably most closely related to the important recent contributions of Epple and Sieg (1999) and Bayer, McMillan and Ruben (2002). Our approach differs from Epple and Sieg (1999) in that we do not use a vertical model of product differentiation (see Bresnahan 1987). A limitation of our approach compared to their research, however, is that we do not exploit market clearing conditions in our analysis and that we do not present an equilibrium model of the entire housing market which could be used to compute interesting policy counterfactuals. Bayer, McMillan and Ruben (2002) apply techniques similar to Berry, Levinsohn and Pakes (1995) to estimate a model of housing demand using a rich data set from the San Francisco Bay Area. Unlike Bayer, McMillan and Ruben (2002), we only impose parametric assumptions on the random coefficients in our model when they are required for identification. A weakness of our analysis, compared to this paper is that Bayer, McMillan and Ruben (2002) have a richer data set on community characteristics. We attempt to compensate for this limitation of our data by checking the robustness of our results on data from multiple cities.

Finally, one appealing feature of our analysis is that it is very transparent. Hedonics is a well understood and popular technique in housing research. It makes clear, using simple econometric methods, exactly what needs to be assumed about the joint distribution of observed and unobserved product characteristics in order to correctly calculate the implicit prices. Then, from the hedonics, we simply use the implicit prices

13

to engage in revealed preference.

There are at least three limitations to the model of housing demand presented in the previous section. The first is the standard exogeneity assumptions employed in hedonics. In general, one would not expect unobserved product characteristics to be independent of observed product characteristics. In our opinion, biases in implicit prices are primarily a data issue. With sufficiently detailed data, one can recover plausible estimates of implicit prices. However, to the best of our knowledge, there are no publicly available data sources in which we can merge the detailed demographic data used in the census with detailed data on home prices and characteristics. To address this problem, we checked our estimates of implicit prices against those found by other researchers to make sure that they appeared reasonable. This problem is not unique to hedonics, as most discrete choice models similarly assume that product characteristics are exogenous.

Second, the results of our analysis will depend to some extent on the functional form of the utility function. Given that we only have a single cross section of data, it is not possible to globally identify preferences. We note, however, that we impose the minimal functional form assumptions that are required for identification. Also, many of the results that we present will be based on the willingness to pay for fairly small changes in the bundle consumed by a household. Since the marginal rate of substitution for continuous characteristics is identified without functional form assumptions, many of our results will remain robust for a large set of alternative specifications for preferences.

Third, we recognize that observed differences in mean willingness to pay for housing attributes between whites and blacks may represent other factors than "preferences". While we control for household head educational attainment and household income, race may still proxy for permanent income and wealth accumulation (see Duca and Rosenthal 1994). Despite the limitations of this exercise, we believe that flexibly modeling the joint distribution of tastes and demographic characteristics is an important step in better understanding the determinants of housing demand.

5 Results.

Our empirical application focuses on the housing choices of migrants ages 18-60 who work and commute by private vehicle to work. In all three cities, the average black and white migrant head of household both

works and commutes by private vehicle to work. Our sample restriction allows us to focus on the housing demands of middle class households independent of job search concerns. The unemployed and those not working in the legal work force may choose a residential community because of the expected economic opportunities from living there. Since it is quite difficult to measure the accessibility of public transit by community, and we do not want to jointly model the choice of commuting mode and choosing a housing product, we focus solely on private vehicle commuters.[6]

Table 1 provides some basic summary statistics concerning white and black migrants in our sample. In each city, we compare migrants to all household heads who work, who are 18-60 who commute by car to work. A large share of the household heads move over a five year period. In Chicago, 54% of all whites switched homes between 1985 and 1990 while for blacks 46% migrated. Migrants are better educated but have lower household incomes than the entire population. Across the three cities, both black and white migrants are more likely to rent their home, and live in a smaller housing unit. In all three cities, black migrants are more likely to live in the suburbs than the average black household.

To ease the computational burdens in estimating the flexible hedonic specification, a random sample of 2000 migrants was drawn for each city. These three samples of 2000 observations each will form the basis for all of the subsequent empirical work.[7] Table 2 presents the sample demographic means for white and black migrants included in our structural estimates across the three metropolitan areas. Across the three cities, the average white household's income is $15,000 higher than the average black's household. White household heads are much more likely to be married than black household heads. This marriage differential is 14 percentage points in Atlanta, 20 percentage points in Chicago and 17 percentage points in Dallas. Averaged across the three cities, white household heads are 18 percentage points more likely to be a college graduate than black household heads.

5.1 Hypothesis #1 The Demand for Physical Housing Attributes and Housing Type

This section tests the hypothesis that whites are moving to the suburbs because they have a greater demand for larger, newer housing units. The suburban housing stock features homes that are larger, newer, and more

---

[6]   Glaeser, Kahn and Rappaport (2000) report evidence that access to public transit is a major urbanizing force for households in poverty. In this paper, we are focusing on people who work and almost all people who work have income above the poverty line.

[7]   In our local linear regressions, however, we use all of the households in the sample to estimate the implicit prices in equations (13) and (14).

likely to be single detached relative to the center city housing stock. The standard explanation for this fact is that the housing stock is durable and land's price per square foot falls with distance from the Central Business District to compensate for the longer commute downtown (Alonso 1964). We test whether whites and blacks have different tastes for such attributes and whether average demographic differences between races such as differences in income, educational attainment and marriage rates explains differences in the demand for physical attributes of the housing unit.

Reduced form evidence documents the differences in housing consumption across demographic groups for the migrant samples. In Table 3, we present OLS regression results from 21 separate regressions. In each regression, we control for the household income and household size, and the household head's education, race, age, sex and marital status. The top panel of the Table reports results for Atlanta. Column (6) reports the regression where the dependent variable is the housing unit's number of rooms. We find that married households live in homes that have .68 more rooms in Atlanta and .7 more rooms in Chicago than non-married households. Across the three cities, married migrants are over 20 percentage points more likely to own a home and to live in a single detached suburban home. An extra $10,000 in household income increases the propensity to own a home by 3 percentage points in each of the cities.

For each city, we estimate the structural model presented in Section 3. This yields for each housing attribute, a preference parameter for each migrant. The discrete choice literature has modeled such random coefficients as normally distributed. In Figure One, we present a histogram of the taste parameters for rooms for the 2000 Chicago migrants. The important point conveyed by this figure is that tastes are right skewed and are not normally distributed. Tastes are also correlated across product attributes. Table 4 reports, by city, the correlation matrix of tastes across attributes. Across all three cities, the taste for rooms and the taste for communities with high levels of human capital are positively correlated. Surprisingly, the taste for communities with high levels of minorities is positively correlated with the taste for high human capital communities.

To test hypothesis 1, we use the random coefficient estimates to construct willingness to pay for an increase in rooms from 4 to 6. Let $WTPROOMS_i$ denote household i's willingness to pay to increase the number of rooms from 4 to 6. By equation (6), it must satisfy:

$$WTPROOMS_i = \beta_{1,i}(\log(6) - \log(4)) \tag{22}$$

This measure of willingness to pay for structure is regressed on household demographics. Table 5 reports these estimates of willingness to pay for 2 more rooms for each of the three cities. Across all three cities, richer college educated married migrants in larger households are willing to pay more for more space. Holding all of these demographics constant, blacks are willing to pay $399 per year less in Chicago for the extra living space. In Atlanta, all else equal, blacks are willing to pay $172 less for extra living space.[8]

But, all else is not equal. As shown in Table 2, Blacks have lower incomes, lower college graduate rates and lower marriage rates. Based on the Atlanta coefficient estimates, if blacks caught up to the average white for these three demographic categories their willingness to pay for rooms would significantly increase. Demographic differences matter as much as "taste differences" in explaining the demand for physical attributes across races. The differential demand for housing unit size between whites and blacks is our strongest housing structure evidence for why black urbanization takes place. For other physical attributes of housing we do not find such major differences in willingness to pay between groups. Table 6 presents our estimates for how much migrants are willing to pay to live in a house that is 35 years old versus a house that is 10 years old. Across all three cities, migrants are willing to pay for newer housing but the differences in willingness to pay across demographic groups are small. For example, in Chicago a college graduate is willing to pay $2.76 more than an identical non-college graduate to live in older housing.

Table 7 presents estimates of the household willingness to pay to own based on equation (21). Note that the price coefficient has been normalized to equal -1. In Chicago, the college educated are willing to pay $72 more per year. An extra $10,000 in household income increases willingness to pay to own by $44 per year. Only in Atlanta, do we reject the hypothesis that whites and blacks have equal willingness to pay to own.[9]

Table 8 presents the willingness to pay differentials for single detached housing. In Atlanta and Chicago, blacks are willing to pay less for such housing. In Chicago, all else equal, blacks are willing to pay $410 less per year for such housing than whites. Similar to the case of rooms, married people are willing to pay

---

[8] We choose not to present estimates of (7) since by equation (22), the dependent variables in the regressions presented in Tables 5-7 and Tables 10-11 are equal to a constant times the dependent variable in equation (7).

[9] It is important to note that with our one cross-section of data, we are estimating a static model of housing demand. We recognize that the tenure choice decision has a dynamic component.

more to live in single detached housing. Since blacks have lower marriage rates, this contributes to the total white/black gap with respect to demand for physical attributes. While household income and educational attainment sharply increase demand for detached housing in Chicago, such income effects have a much smaller impact in Atlanta and Dallas.

5.2 Hypothesis #2: Commuting and Place of Work

Employment suburbanization has tracked residential suburbanization for the last 50 years in the United States (Garreau 1992, Glaeser and Kahn 2001). As workers have suburbanized many jobs have followed them to the suburbs. As jobs migrate to the suburbs, suburban residents have shorter commutes and land intensive firms can purchase more land cheaper than in the CBD. Such "job sprawl"could increase the black/white suburbanization gap if households do not like commuting and blacks tend to work in highly urbanized industries.

Atlanta, Chicago and Dallas all have a significant amount of employment located outside of their center cities. In our migrant Atlanta sample, 59% of workers work in the suburbs and 53% of black workers work in the suburbs. In Chicago, 70% of workers are in the suburbs and 42% of black workers work in the suburbs. In Dallas, 62% work in the suburbs and 54% of blacks work in the suburbs.

The commuting hypothesis posits that blacks live in the city because they work in the city. We test this hypothesis in Table 9 by estimating the demand for living in the city, as in equation (21). For each of the three cities, we estimate two separate models. We first estimate how much blacks are willing to pay to live in the city holding all other migrant demographics and housing choice attributes constant. In Atlanta and Dallas, all else equal, blacks are willing to pay $192 and $1154 dollars a year more than whites to live in the city. To test how much of this black differential willingness to pay is driven by place of work, we include in the specification a dummy variable indicating whether the migrant works in the center city. We recognize that working in the city is an endogenous regressor. Migrants may work in the city because they found out about a job opportunity from a city neighbor. In such a job search model, living in the city would raise your probability of working in the city. We thus instrument for whether a migrant works in the city using the worker's industry information (dummy variables for whether the migrant works in manufacturing, wholesale, retail, services, construction, or for the government). These dummies will be valid instruments if workers have invested in industry specific human capital and industries differ with

respect to their suburbanization propensities (see Neal 1995). For example, land intensive industries such as manufacturing tend to suburbanize.[10]

For each city, we present our results where we include the work in city indicator and instrument for this variable. In all three cities, including this variable sharply reduces black willingness to pay to urbanize. In Dallas, black willingness to pay falls from 1154 to 355 when place of work is controlled for and in Chicago the hypothesis that the black coefficient is zero cannot be rejected. It is important to note that the first stage IV results have an R2 of .02 for each city. Still across the three cities, we find that black employment urbanization helps to explain black residential urbanization.

While not directly relevant for answering our paper's question, our structural estimates provide some insights into migrant value of time. This is relevant for judging the plausibility of our parameter estimates. Chicago migrants who work in the city are willing to pay $381 more per year to live in the city than migrants who work in the suburbs. Migrants who commute by car and work and live in city have an 11 minute shorter one way commute than migrants who live in the city and work in the suburbs. This suggests that migrants are willing to pay $4.15 to avoid an hour of commuting. Many travel cost studies assume that people value their commute time at 50% of their hourly wage. Our Chicago estimate is roughly in line with this assumption.

5.3   Hypothesis #3: The Demand for Peers

Each day we interact with our neighbors. Thus, their attributes affect our well being and our human capital accumulation. Hedonic studies have measured the benefit of having high skilled neighbors by studying land rents in high and low skill areas (Rauch 1993, DiPasquale and Kahn 1999). Other empirical studies have measured peer group effects by regressing outcome measures such as a person's employment status on the average of this outcome measure for one's peer group (Case and Katz 1991, Crane 1991). The practical problem of measuring what are social externalities associated with peer exposure has led to an econometrics literature focused on the reflection problem (Manski 1993).   An alternative to the "production function" method for recovering the benefits of peers is to measure the willingness to pay to live among them. If a household greatly benefit from living in a particular community, we should observe this household being willing to pay a great deal to live there.

---

[10]   Glaeser and Kahn (2001) provide a detailed discussion of what types of industries demand to be centrally located.

In all urban studies, researchers face the challenge of defining what is the geography of the community and what is a parsimonious set of community attributes that matter. With regard to defining community, we use the census' PUMA definition. While we are well aware that communities can differ block by block in this car age people come into contract with a wide variety of people who live in an area around them. We model community attributes using the PUMA's percent college graduates and the community's percent black. A community with more college graduates will have a richer variety of stores and restaurants. An emergent property of having an active educated set of neighbors is that the local schools will be better and crime will be lower. This community will feature an active set of role models who regularly go to work and they will be vocal stakeholders in preserving the quality of the community. In terms of cross-community variation, it is true that communities with more college graduates have fewer black residents but the correlations across the three cities are not extremely large. The correlation between a PUMA's percent black and percent college graduate in Atlanta, Chicago and Dallas is -.54, -.40, and -.54 respectively. We are thus confident that we can disentangle the separate demands for these attributes.

Table 10 reports our structural estimates of the willingness to pay to live in a community whose percent of adults who are college graduates increases from 10% to 35%. Across all three cities, the college educated are willing to pay much more to live in the high human capital areas. In Chicago, college graduates are willing to pay $1225 more per year than non-college graduates to live in such a community. This indicates that among migrants there is a significant degree of sorting. High skilled individuals are sorting into high skilled communities. Across all three cities, holding demographics constant, blacks are willing to pay less to live in this high human capital community. In Atlanta, blacks are willing to pay $585 less and in Chicago they are willing to pay $1154 less. While richer households are willing to pay more to live in higher educated communities, the slope is small. In Chicago, an extra $10,000 in household income increases willingness to pay by $185 per year. Since the suburbs feature higher levels of college graduates, the evidence in Table 10 supports the hypothesis that differential sorting into high skill communities explains part of the urbanization puzzle.

Whites and blacks may have different preferences over their favorite community racial composition. Using a hedonic 1st stage housing price regression in each decade from 1950 to 1990, Cutler, Glaeser and Vigdor (1999) concluded that whites are willing to pay more than blacks to live in white communities.

20

We study their claim using our structural approach that explicitly recovers estimates of willingness to pay. Table 11 presents regression estimates of willingness to pay for living in a community that is 35% black versus a community that is 10% black (holding everything else constant). The most interesting result in the table is that across the three cities, blacks are willing to pay less than whites to live in such a community. In Chicago, black migrants are willing to pay $684 less per year than whites to live in such a community. Unlike Cutler, Glaeser and Vigdor (1999), we find no evidence for the hypothesis that whites are willing to pay more to live with other whites.

Whites are willing to pay more to live in college educated communities and these are often white communities but controlling for community human capital levels, whites actually exhibit a taste for integration. The migrant black households who make up our sample are willing to pay more than observationally identical whites to live in whiter communities. It is important to note that our sample consists of working migrant households and this trims from the sample many minorities who reside in ghetto poverty.

6   Conclusion

Studying housing demand is a natural application of recently developed techniques for flexibly modeling the demand for differentiated products. We have adapted the empirical strategy presented in Bajari and Benkard (2002) to model housing. The three stage model relaxes some commonly made assumptions in the literature such as that random coefficients are normally distributed and independent across housing attributes. In our application, these assumptions appear to be inappropriate.

Our empirical application focused on the important urban question of why blacks choose to live in cities while whites suburbanize. Based on our structural estimates, we conclude that white suburbanization is driven by their greater demand for larger single detached housing units and their greater demand for living in high human capital communities. Black demand for suburban housing products would significantly increase if their incomes, educational attainment and marriage rates were as high as whites. Black urbanization is also explained by their propensity to work in center cities. All else equal, the disutility from commuting provides an incentive to choose urban residential communities. We reject the hypothesis that white demand for the suburbs is driven by the desire to live among other whites.

We have focused on the utility maximizing choices of migrants without investigating the social conse-

quences of their choices. One important externality generated by suburbanization of the well to do is the increased isolation of the urban minority poor. We have shown that highly educated, wealthy black migrants have a high demand for suburban housing products. Wilson (1987) has argued that such selective migration of wealthier blacks to the suburbs reduces the center city poor's exposure to "role models" and this in turn reduces the next generation's educational attainment and commitment to working in the legal sector. This dynamic process would explain why Cutler and Glaeser (1997) conclude using 1990 data that ghettos are bad for blacks (based on teenage educational attainment and idle status) while Collins and Margo (2000) found that blacks who lived in segregated ghettos in the 1950s did not suffer such adverse effects. The social consequences of privately optimal housing choices merits further research.

7 References.

[1]   Ackerberg, Dan and Rysman, Mark (2002), Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects, Mimeo, UCLA and Boston College.

[2]   Alonso, William (1964), Location and Land Use. Harvard University Press.

[3]   Anderson, S., DePalma, A., and J. Thisse (1995), Discrete Choice Theory of Product Differentiation, Cambridge, MIT Press.

[4]   Bajari, Patrick and C. Lanier Benkard (2002), Demand Estimation with Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach, Stanford University Working Paper.

[5]   Bayer, Patrick, McMillan and Kim Ruben, (2002) The Causes and Consequences of Residential Segregation: An Equilibrium Analysis of Neighborhood Sorting. Working Paper, Yale University.

[6]   Berry, Steven (1994), Estimating Discrete Choice Models of Product Differentiation, Rand Journal of Economics, 25(2), 242-62.

[7]   Berry, Steven; James Levinsohn, and Ariel Pakes (1995), Automobile Prices in Market Equilibrium, Econometrica, 63(4), 841-90.

[8]   Berry, Steven; James Levinsohn, and Ariel Pakes. "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy", American Economic Review 89 (3), 400-430.

[9]   Bresnahan, Timothy (1987) Competition and Collusion in the American Automobile Industry. Journal of Industrial Economics, vol. XXXV (no. 4, June–Special Issue) 457-482.

[10] Brown, James N and Harvey S. Rosen (1982), On the Estimation of Structural Hedonic Price Models. Econometrica, 50(3), 765-768

[11] Case, Anne and Lawrence Katz (1991), The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths. NBER Working Paper 3705.

[12] Cheshire, Paul and Stephen Sheppard (1998), Estimating the Demand for Housing, Land and Neighbourhood Characteristics, Oxford Bulletin of Economics and Statistics, 60(3) 357-382.

[13] Collins, William and Robert A. Margo (2000), Residential Segregation and Socioeconomic Outcomes: When Did Ghettos Go Bad? Economics Letters 69(2) 239-243.

[14] Costa, Dora (1998), The Evolution of Retirement. University of Chicago Press.

[15] Coulson, Edward and Eric Bond (1990), A Hedonic Approach to Residential Succession, Review of Economics & Statistics. 72(3). 433-444.

[16] Crane, Jonathan (1991), The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out of High School and Teenage Childbearing, American Journal of Sociology 96(5) 1126-1159.

[17] Cutler, David., Edward Glaeser (1997), Are Ghettos Good or Bad?, Quarterly Journal of Economics, 112(3) 827-872.

[18] Cutler, David., Edward Glaeser, and Jacob Vigdor (1999), The Rise and Decline of the American Ghetto, Journal of Political Economy. 107(3) 455-506.

[19] Di, Zhu Xio (2001,) The Role of Housing as a Component of Household Wealth. W01-6. Joint Center for Housing Studies. Harvard University.

[20] DiPasquale, Denise and Matthew E. Kahn (1999), Measuring Neighborhood Investments: An Examination of Community Choice, Real Estate Economics 27(3) 389-424.

[21] Duca, John and Stuart Rosenthal (1994), Borrowing Constraints and Access to Owner Occupied Housing, Regional Science and Urban Ecoonomics 24(3) 301-322.

[22] Epple, Dennis (1987), Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products, Journal of Political Economy, 95(1), 59-80.

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

[24] Fan, Jinqing and Irene Gijbels, Local Polynomial Modelling and Its Applications (Monographs on Statistics and Applied Probability, (66), CRC Press.

[25] Garreau, Joel. Edge City: Life on the New Frontier. New York: Anchor Books, 1992.

[26] Gaspar, Jess and Edward L. Glaeser (1998), Information Technology and the Future of Cities Journal of Urban Economics, 43(1) 136-156.

[27] Glaeser, Edward L and Matthew E. Kahn (2001), Decentralized Employment and the Transformation of the American City Brookings-Wharton Papers on Urban Affairs, vol. 0, no. 0, 1-47.

[28] Glaeser, Edward L, Matthew E. Kahn, and Jordan Rappaport (2000), Why Do the Poor Live in Cities? NBER Working Paper W7636.

[29] Gyourko, Joesph. and Joesph Tracy (1991), The Structure of Local Public Finance and the Quality of Life, Journal of Political Economy. 91(4), 774-806.

[30] Heckman, James (1998), Detecting Discrimination, Journal of Economic Perspectives, 12, 101-116.

[31] Manski, Charles (1993), Identification of Endogenous Social Effects: The Reflection Problem, Review of Economic Studies, 60, 531-542.

[32] Mas-Colell (1977) The Recoverability of Consumer Preferences From Market Demand Behavior, Econometrica, 45(6), 1409-1430.

[33] McCulloch, Robert E; Nicholas G Polson, and Peter E. Rossi (2000), A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters. Journal of Econometrics, 99(1), 173-193.

[34] Munnell, Alicia, G. Tootell, Lynn Browne and J. McEneaney (1996), Mortgage Lending in Boston: Interpreting HMDA Data. American Economic Review, 86, 25-53.

[35] Neal, Derek (1995), Industry-Specific Human Capital: Evidence from Displaced Workers Journal of Labor Economics, 13(4), 653-677.

[36] Palmquist, Raymond (1984), Estimating the Demand for the Characteristics of Housing, Review of Economics and Statistics, 66(3), August, 394-404.

[37] Petrin, Amil (2002), Quantifying the Benefits of New Products: The Case of the Minivan, Journal of Political Economy, 110(4) 705-729.

[38] Rauch, James (1993), Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities, Journal of Urban Economics, 34(3) 380-400.

[39] Rosen, Sherwin (1974), Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, Journal of Political Economy 82 (January/February) 34-55.

[40] Wilson, William Julius (1987). The Truly Disadvantaged. University of Chicago Press.

[41] Yinger, John (1986), Measuring Racial Discrimination with Fair Housing Audits. American Economic
      Review, 76 881-893.

Table 1: Demographic Means for Migrants

	Atlanta				Chicago				Dallas			
	Whites	White Migrants	Blacks	Black Migrants	Whites	White Migrants	Blacks	Black Migrants	Whites	White Migrants	Blacks	Black Migrants
Age	38.5639	35.2778	37.1946	34.4020	39.6639	35.4292	41.1371	36.8976	37.6311	34.1971	37.2549	33.8209
College Graduate	0.4346	0.4573	0.2320	0.2470	0.3237	0.3634	0.1935	0.2124	0.3631	0.3603	0.2071	0.2313
Household Income	55527.7000	51290.3100	36524.6400	34035.8600	54048.3200	50040.2700	40395.3800	36629.7200	48350.0500	42595.9000	31591.1400	29522.2400
Male	0.7596	0.7398	0.5808	0.5677	0.7931	0.7747	0.5974	0.5740	0.7628	0.7302	0.5783	0.5599
Household Size	2.7095	2.6115	2.8370	2.7291	3.0190	2.8301	3.1400	2.8989	2.7528	2.5751	2.8941	2.7461
Married	0.6276	0.5781	0.4515	0.4106	0.6730	0.6144	0.4724	0.4147	0.6131	0.5309	0.4246	0.3752
Migrant	0.6240		0.6777		0.5359		0.4556		0.6309		0.6645	
City	0.0662	0.0667	0.2232	0.1679	0.2009	0.1990	0.5565	0.4867	0.2051	0.2162	0.4864	0.4513
Home Owner	0.6721	0.5391	0.4398	0.2958	0.6909	0.5597	0.5179	0.3590	0.5550	0.3709	0.3824	0.2102
Single Detached House	0.6943	0.5821	0.4809	0.3456	0.6065	0.4833	0.4407	0.3338	0.6369	0.4920	0.4864	0.3357
Rooms	6.1452	5.7907	5.3120	4.9404	5.8226	5.4847	5.3525	4.9975	5.3415	4.9065	4.6223	4.2849
Age of House	15.5529	12.9190	19.6377	16.9919	31.4790	28.3798	38.0910	33.6947	17.7749	15.4060	19.8946	16.8385

The raw data is the 1990 U.S Census of Population and Housing. This sample consists of all heads of households ages 18-60 who work and who commute by car to work

Table 2: Summary Statistics for the Migrants Included in the Structural Estimation

	Atlanta		Chicago		Dallas	
	Black Migrants	White Migrants	Black Migrants	White Migrants	Black Migrants	White Migrants
Age	33.7750	35.3790	37.1579	35.3341	33.2857	34.5416
College Graduate	0.2333	0.4651	0.1886	0.3685	0.2239	0.3487
Household Income	35005.6100	50580.4900	36071.4900	51226.5100	28024.2700	43254.9600
Male	0.5833	0.7329	0.5175	0.7737	0.5251	0.7490
Household Size	2.6646	2.5934	2.9167	2.8595	2.8533	2.5876
Married	0.4354	0.5730	0.3991	0.6027	0.3552	0.5399
Work in City	0.4729	0.3901	0.5789	0.2748	0.4556	0.3716
Live in City	0.1500	0.0671	0.5044	0.1992	0.4479	0.2148
Community % College	0.2546	0.3283	0.1946	0.2809	0.2346	0.2702
Community % Black	0.4432	0.1334	0.4649	0.0687	0.2350	0.1069
Home Owner	0.3125	0.5283	0.3202	0.5751	0.2162	0.3779
Detached House	0.3542	0.5803	0.3158	0.4808	0.3282	0.4871
Rooms	5.0729	5.7382	4.8860	5.5248	4.2703	4.8960
Age of Unit	15.9594	13.4191	32.3114	28.6332	17.4556	14.8601
Annual Housing Expenditure	5822.2220	8030.6000	5705.1700	8817.6560	4661.2630	6584.6170

This table reports sample means for 2000 observations for each city drawn at random from the set of migrants.

Table 3: Descriptive OLS Regressions of Migrant Housing Choice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Atlanta	Community % College	Community % Black	Live in Suburbs	Home Owner	Unit Detached	Rooms	Age of Unit
Household Income	0.0008	-0.0005	-0.0005	0.0027	0.0020	0.0172	-0.0403
College Graduate	0.0510	0.0057	-0.0327	0.0788	0.0327	0.2701	1.2412
Black	-0.0570	0.3043	-0.0910	-0.0967	-0.1236	-0.2065	1.8960
Household Size	-0.0091	-0.0012	0.0162	0.0255	0.0598	0.3853	0.5388
Age	-0.0020	0.0029	-0.0008	0.0122	0.0106	0.0377	0.1347
Married	-0.0158	-0.0334	0.0790	0.2249	0.2874	0.6806	-3.8962
Male	-0.0198	0.0032	-0.0045	0.0652	0.0777	0.0082	0.3601
Constant	0.3803	0.0728	0.9202	-0.3195	-0.2866	2.0125	10.6838
R2	0.1630	0.3357	0.0527	0.2901	0.3617	0.4317	0.0274
Chicago	Community % College	Community % Black	Live in Suburbs	Home Owner	Unit Detached	Rooms	Age of Unit
Household Income	0.0008	-0.0002	0.0006	0.0032	0.0020	0.0152	-0.1049
College Graduate	0.0733	0.0000	-0.0413	0.0493	0.0272	0.4533	-0.2234
Black	-0.0668	0.3870	-0.2903	-0.1721	-0.0813	-0.2755	1.6540
Household Size	-0.0099	0.0079	-0.0184	0.0230	0.0493	0.3562	1.6582
Age	-0.0004	-0.0005	0.0011	0.0101	0.0074	0.0243	-0.0464
Married	-0.0124	-0.0222	0.0938	0.2207	0.2805	0.7002	-4.2258
Male	-0.0191	-0.0078	-0.0199	-0.0005	0.0298	-0.1011	1.8590
Constant	0.2752	0.0922	0.7577	-0.1621	-0.2249	2.3581	32.0950
R2	0.2063	0.3934	0.0659	0.2426	0.2440	0.3802	0.0405
Dallas	Community % College	Community % Black	Live in Suburbs	Home Owner	Unit Detached	Rooms	Age of Unit
Household Income	0.0004	-0.0004	0.0003	0.0036	0.0030	0.0186	-0.0554
College Graduate	0.0239	-0.0069	-0.0332	0.0913	0.0340	0.5427	-2.6528
Black	-0.0282	0.1156	-0.2013	-0.0483	-0.0655	-0.2073	1.1757
Household Size	-0.0076	0.0118	-0.0153	0.0170	0.0690	0.2442	1.4448
Age	-0.0006	0.0008	-0.0004	0.0085	0.0099	0.0404	0.0936
Married	-0.0035	-0.0374	0.1623	0.2228	0.2064	0.8001	-2.3924
Male	-0.0126	0.0179	-0.0132	-0.0033	0.0467	-0.2949	2.0705
Constant	0.2943	0.0756	0.7561	-0.2655	-0.3242	1.6652	10.9509
R2	0.0858	0.1394	0.0607	0.2684	0.3172	0.4161	0.0481

This table reports 21 reduced form OLS regressions; seven for each city. The dependent variables "Live in Suburbs", "Home Owner, and "Unit Detached" are dummy variables. In each regression, housing consumption measures are fitted as a function of personal attributes. The omitted category is a white, non-college graduate who is female and not married. Household Income is measured in 1,000s of 1989 $.
There are 2000 observations in each regression.

Figure One

Chicago Migrant Preference Heterogeneity Over Rooms
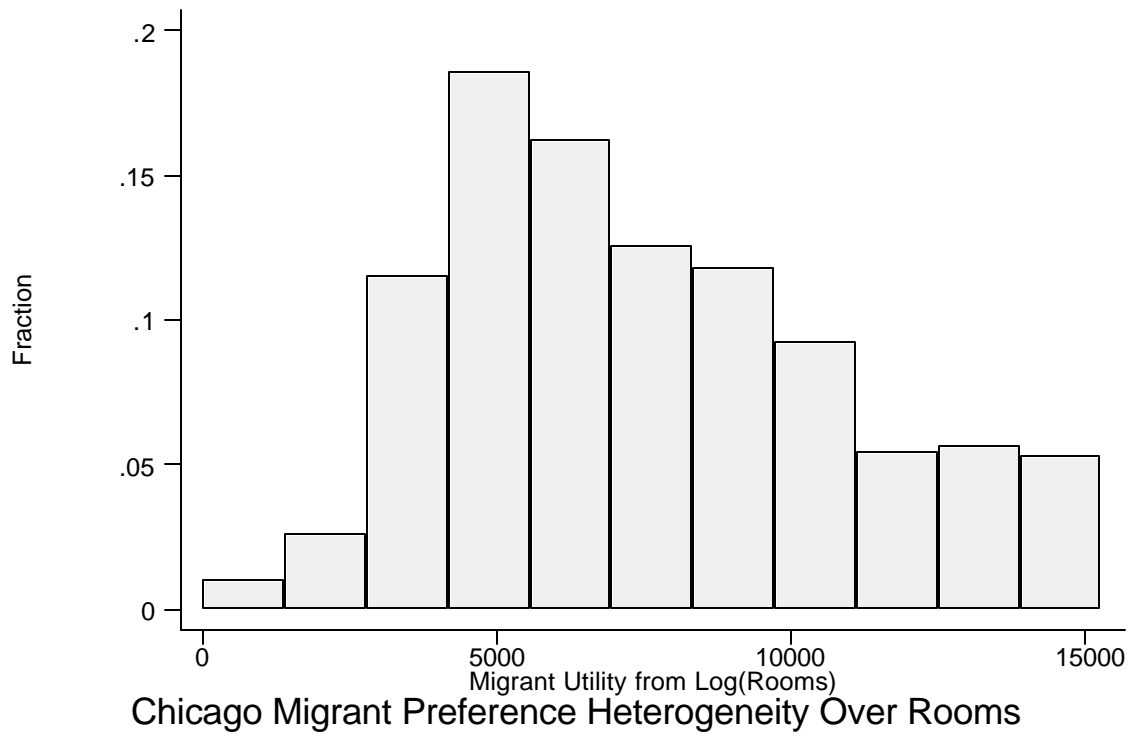
Table 4: Correlation Matrix of Marginal Utilities by Metropolitan Area

Atlanta

	Rooms	Age of Unit	Community % BA
Age of Unit	0.0749		
Community % BA	0.1347	-0.0475	
Community % Black	0.0219	0.2264	0.4577

Chicago

	Rooms	Age of Unit	Community % BA
Age of Unit	0.1231		
Community % BA	0.4126	0.1783	
Community % Black	0.1259	0.1414	0.3687

Dallas

	Rooms	Age of Unit	Community % BA
Age of Unit	0.0414		
Community % BA	0.2501	0.1458	
Community % Black	0.1476	0.2561	0.4633

Table 5: Estimates of The Willingness to Pay for Rooms

	Atlanta	Chicago	Dallas
Age	21.9402	19.4257	22.8256
	(2.3616)	(2.7855)	(2.0161)
College Graduate	184.7462	378.0459	316.2760
	(41.5490)	(55.0933)	(40.9990)
Black	-172.3309	-398.7179	-140.3617
	(44.1881)	(67.9734)	(55.7413)
Household Income	10.2370	12.4079	11.0030
	(0.8263)	(1.0465)	(0.6090)
Household Size	209.7445	217.7264	132.4001
	(17.9231)	(28.7072)	(14.8875)
Male	30.3560	-53.2453	-149.3462
	(47.5370)	(59.4864)	(47.9940)
Married	411.3082	576.3986	467.0055
	(54.0722)	(69.0660)	(50.3100)
Constant	688.9144	748.4642	659.6018
	(88.5487)	(110.2144)	(80.1425)
R2	0.4578	0.4106	0.4323
Observations	2000	2000	2000

Each column of the table presents a separate OLS regression.
The dependent variable is a migrant's willingness to pay per year for an increase from 4 to 6 rooms in a housing unit, holding all other housing product attributes constant.
Standard errors are reported in parentheses.
The omitted category is a white, non-college graduate who is female and not married.
Household Income is measured in 1,000s of 1989 $.

Table 6: Estimates of The Willingness to Pay for an Older Housing Unit

	Atlanta	Chicago	Dallas
Age	-4.2261	0.1381	-3.9095
	(1.1981)	(1.8801)	(1.3866)
College Graduate	-10.6141	2.7577	111.8226
	(21.9891)	(39.1612)	(28.1964)
Black	-46.9915	62.5131	-38.9609
	(23.4239)	(47.4732)	(38.3352)
Household Income	1.7531	3.3548	2.2824
	(0.3908)	(0.5506)	(0.4188)
Household Size	-29.4079	-71.2458	-58.1456
	(8.9125)	(12.5904)	(10.2387)
Male	-9.6398	-59.0414	-78.6016
	(27.3633)	(44.1225)	(33.0071)
Married	115.1276	90.0929	79.6203
	(29.1504)	(44.1556)	(34.5999)
Constant	-453.4905	-1175.1670	-472.3534
	(48.7979)	(79.9419)	(55.1167)
R2	0.0337	0.0376	0.0531
Observations	2000	2000	2000

Each column of the table presents a separate OLS regression.
The dependent variable is a migrant's willingness to pay per year for an increase from 10 to 35 years in the age of a housing unit, holding all other housing product attributes constant.
Standard errors are reported in parentheses.
The omitted category is a white, non-college graduate who is female and not married.
Household Income is measured in 1,000s of 1989 $.

Table 7: Estimates of the Demand for Owning

	Atlanta	Chicago	Dallas
Age	5.7628	10.8808	0.9960
	(0.5457)	(1.4349)	(0.1411)
College Graduate	53.5150	72.3841	13.0556
	(9.8024)	(28.3355)	(2.8978)
Black	-38.1048	-48.5070	5.2907
	(10.6864)	(39.3679)	(4.0455)
Household Income	1.5823	4.4116	0.3716
	(0.1747)	(0.5000)	(0.0450)
Household Size	18.6377	50.4153	4.8466
	(4.1044)	(8.8029)	(1.0871)
Male	25.6344	12.8634	-2.1273
	(11.0205)	(31.8090)	(3.5123)
Married	64.2837	165.0626	8.5168
	(11.8269)	(30.9482)	(3.6587)
Price of Ownership	-1.0000	-1.0000	-1.0000
Constant	-315.7914	-301.8911	52.2557
	(22.9383)	(62.5935)	(6.1095)
observations	2000	2000	2000

Each column of the table presents a separate estimate of equation (21).
Standard errors are reported in parentheses.
The omitted category is a white, non-college graduate who is female, and not married.
Household Income is measured in 1,000s of 1989 $.

Table 8: Estimates of the Demand for Single Detached Housing

	Atlanta	Chicago	Dallas
Age	-0.0211	16.1145	0.1754
	(0.4314)	(2.5122)	(0.2191)
College Graduate	-0.6411	82.1002	-6.5952
	(7.3269)	(49.7589)	(4.4498)
Black	-36.5337	-409.6949	-5.5963
	(7.2683)	(76.1423)	(5.7880)
Household Income	0.3890	5.9420	0.0345
	(0.1682)	(0.7282)	(0.0850)
Household Size	3.7795	75.1032	-0.2434
	(2.9635)	(15.7585)	(1.3193)
Male	-6.0654	31.6040	3.2634
	(8.0832)	(61.4414)	(5.0320)
Married	40.8022	598.1848	10.2411
	(9.3722)	(57.0570)	(4.7934)
Price of Single Detached Unit	-1.0000	-1.0000	-1.0000
Constant	236.9936	-815.7973	-23.6736
	(25.4464)	(127.1599)	(9.8171)
Observations	2000	2000	2000

Each column of the table presents a separate estimate of equation (21).

Standard errors are reported in parentheses.
The omitted category is a white, non-college graduate who is female, and not married.
Household Income is measured in 1,000s of 1989 $.

Table 9: Estimates of The Demand for Living in the Center City

	Atlanta		Chicago		Dallas	
	MLE	IV	MLE	IV	MLE	IV
Work in City		1107.7020		381.0468		2284.1580
		(560.0689)		(160.7365)		(610.1467)
Age	6.2919	7.81	3.5841	2.5744	3.3249	-6.0978
	(2.8533)	(3.2849)	(0.9379)	(1.2506)	(7.2669)	(4.0110)
College Graduate	192.7268	184.0818	72.4209	59.5484	253.1178	74.6930
	(53.1446)	(64.2010)	(18.5754)	(22.9696)	(145.8183)	(72.9149)
Black	192.0734	127.0259	44.9930	-41.9344	1154.3430	355.3576
	(56.1501)	(81.7120)	(23.6132)	(47.8523)	(180.1841)	(103.0735)
Household Income	4.2080	3.5120	1.5404	1.0469	-2.0029	-1.7019
	(0.7983)	(1.0986)	(0.2468)	(0.3607)	(2.3551)	(1.1744)
Household Size	-45.0881	-10.8730	28.4383	30.0853	109.7375	67.3517
	(24.4813)	(33.4863)	(5.2666)	(5.9734)	(51.6970)	(25.4514)
Male	14.3055	111.4717	7.8602	-5.3234	70.9471	110.5087
	(53.1691)	(78.5761)	(20.5317)	(24.9219)	(160.4220)	(81.6356)
Married	-179.3233	-294.9678	36.3894	27.7950	-1059.3110	-452.7784
	(67.2694)	(85.4042)	(20.2460)	(23.2029)	(178.2471)	(88.6307)
Price of Living in the City	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
Constant	517.8559	-135.9546	606.8317	538.8738	-794.6178	-654.2031
	(211.7071)	(397.3694)	(51.0477)	(92.3438)	(903.1547)	(450.8096)
Observations	2000	2000	2000	2000	2000	2000

Each column of the table presents a separate estimate of equation (21).

Standard errors are reported in parentheses.
IV indicates that the model was estimated using an instrumental variable strategy.
The dummy variable "Work in City" equals one if a migrants works in the city. This variable is instrumented for
using dummy variables for what industry the migrant works in.
The omitted category is a white, non-college graduate who is female, not married and works in the suburbs.
Household Income is measured in 1,000s of 1989 $.

Table 10: Estimates of The Willingness to Pay to Live in a Highly Educated Community

	Atlanta	Chicago	Dallas
Age	-15.6400	6.3942	-1.5148
	(3.3339)	(4.9032)	(1.4405)
College Graduate	523.1925	1225.3590	178.6172
	(59.8207)	(106.3816)	(29.2932)
Black	-585.4029	-1154.4310	-221.9220
	(66.0349)	(130.7574)	(39.8264)
Household Income	9.4491	18.5349	3.8869
	(0.9991)	(1.7275)	(0.4351)
Household Size	-68.9991	-74.1159	-38.6235
	(24.6644)	(30.2752)	(10.6369)
Male	-182.6332	-264.7713	-87.6784
	(69.6553)	(114.0511)	(34.2910)
Married	-91.6378	124.8641	36.2052
	(76.5356)	(108.5081)	(35.9458)
Constant	3350.2190	3284.0540	1857.4340
	(131.1406)	(206.5099)	(57.2607)
R2	0.1745	0.2595	0.1102
Observations	2000	2000	2000

Each column of the table presents a separate OLS regression.
The dependent variable is a migrant's willingness to pay per year for an increase from 10% to 35%
of community members who are college graduates, holding all other housing product attributes constant.
Standard errors are reported in parentheses.
The omitted category is a white, non-college graduate who is female and not married.
Household Income is measured in 1,000s of 1989 $.

Table 11: Estimates of The Willingness to Pay for Living in a Black Community

	Atlanta	Chicago	Dallas
Age	-5.6254	0.8965	-1.8859
	(0.8075)	(0.7535)	(0.5645)
College Graduate	-20.0670	6.0702	9.1012
	(14.4892)	(13.8707)	(11.4790)
Black	-433.7638	-683.5248	-214.2852
	(15.9943)	(41.4719)	(15.6066)
Household Income	0.5206	0.2751	0.7133
	(0.2420)	(0.1768)	(0.1705)
Household Size	-2.8986	-20.0737	-23.6381
	(5.9740)	(5.4234)	(4.1682)
Male	-11.8165	4.3038	-31.7136
	(16.8712)	(19.0165)	(13.4374)
Married	28.8122	26.6823	65.5898
	(18.5377)	(18.2222)	(14.0859)
Constant	7.8506	-164.0204	-115.4729
	(31.7635)	(29.7271)	(22.4384)
R2	0.2972	0.3646	0.1343
Observations	2000	2000	2000

Each column of the table presents a separate OLS regression.

The dependent variable is a migrant's willingness to pay per year for an increase from 10% to 35% of community members who are black, holding all other housing product attributes constant.

Standard errors are reported in parentheses.

The omitted category is a white, non-college graduate who is female and not married.

Household Income is measured in 1,000s of 1989 $.