Housing Consumption and Prices in a Unified Metropolitan Market with Heterogeneous Preferences*

Luis Quintero

Carnegie Mellon University

Nov 1, 2014.

*I would to thank Dennis Epple, Holger Sieg, Martin Gaynor, Karam Kang, Yaroslav Kryukov, Chris Cunningham, Christopher Timmins, Erwan Quintin, Daniele Coen-Pirani, Daniel Neill, Morris Davis, Ariel Pakes, Monika Piazzesi, Xun Tang and seminar participants at various universities and conferences for comments and discussions. Financial support for this project was provided by the NSF grant SES-0958705.

Abstract

We provide a new method for estimating hedonic functions in a metropolitan housing market for both rental rates and real estate asset prices. First, our method treats housing quality as unobserved by the econometrician. Second, it deals with the problem that implicit rental rates for owner-occupied housing are latent. Using a non-parametric matching approach, we show how to identify the rent-to-value ratio as a function of latent quality. Third, the paper provides a new estimation method for a generalization that incorporates heterogeneity in preferences. To our knowledge, this is the first paper that incorporates these in a unified treatment of metropolitan housing markets. We estimate the model with types obtained by using clustering techniques to learn categorization of households into types based on age and number of children. We obtain the robust result that the presence of children lowers the preference for housing quality and increases the welfare sensitivity to changes in income and price. The opposite happens as households age. Finally, we also estimated a joint model for the NYC and Chicago metropolitan areas to compare quality markets, and find that a compensating variation of approximately 20% of the annual income would be required to induce the median household in Chicago to move to NYC. Finally, we study the recent housing market crisis in the U.S., and obtain findings consistent with the view that changes in both credit market conditions and investor expectations were key contributors to the run-up in housing values.

Keywords: Hedonic Models, Asset Value of Housing, Non-parametric Identification, Semi-parametric Estimation, Housing Market Crisis, Heterogeneous preferences.

1 Introduction

A central challenge in estimating models with heterogeneous housing is separating quality from price. By its nature, house quality is not directly observed by the econometrician. One approach, from the hedonic literature, is to estimate a mapping from the observed characteristics of a house to its value. This approach assumes that housing can be characterized by a vector of characteristics and that unobserved heterogeneity is not systematically related to observables.[1] Measuring housing characteristics is in practice challenging and creates a variety of well-known identification and estimation problems due the potential of omitted variables. We bypass this step and provide a new method for estimating hedonic price functions treating housing quality as unobserved by the econometrician.

The pioneering work of Rosen (1974) transformed modeling of markets for differentiated products.[2] A great many fruitful applications have built on the Rosen's framework. We advance this agenda by developing a framework for estimating hedonic equilibrium models of metropolitan-wide housing markets. We distinguish between rental and asset markets for housing and show how to characterize and estimate both nonlinear equilibrium pricing functions. Key elements of our approach are the following. We threat the metropolitan area as a unified housing market. We show that rental and asset prices as a function of quality are separately identified for the entire distribution of housing in the metropolitan area, yielding an estimator of the quality distribution of housing in the entire market. We define quality in the broadest

[1]This has been an important agenda of hedonic theory and of associated empirical work linking house values to observed house characteristics. For a review of the recent literature see, for example, Kuminoff, Smith, and Timmins (2013).

[2]Closely related to the hedonic literature is also the work by Berry and Pakes (2007) on the pure characteristics model.

terms to incorporate not only structural housing characteristics per se, but also all associated natural and publicly provided amenities. Quality is latent so that it is not necessary to differentiate between observed and unobserved housing characteristics. Finally, we also show how to estimate period-to-period changes in housing supply as a function of quality.

Implementation of the approach is feasible with readily available data for metropolitan housing markets in the U.S. Our key simplifying assumption is that quality can be mapped onto a unidimensional index. In this respect, our approach is positioned between two widely employed characterizations of housing markets. One treats housing as a homogeneous and perfectly divisible commodity. The other treats housing as comprised of a fixed stock of heterogeneous housing types.[3] Our approach occupies the middle ground, with housing being continuous and unidimensional, as in the former approach, while being heterogeneous along the quality dimension and inelastically supplied within-period, as in the latter. In adopting this unidimensional characterization of housing, we forgo the potential of the multi-attribute framework for valuing individual elements of the bundle of housing attributes. We also avoid the severe challenges entailed in extending the multi-attribute framework to a multi-period setting; modeling changes in the stock of houses with varying attribute bundles and the associated equilibrium implicit prices of the attributes is difficult if not intractable. In exchange for our simplifying assumption, we gain a tractable framework that is ideally suited to study the distributions of quality and price in metropolitan housing markets and associated changes over time while exploiting the full set of equilibrium conditions at each point in time.

We show that there exists a new flexible parametrization of Rosen's model that

[3]The former dates to the classic works of Muth (1960, 1969) and Mills (1972) while the later was pioneered by Dunz (1989) and Nechyba (1997, 2000).

exploits generalized forms of the log-normal distributions Vianelli (1983) and that yields tractable solutions for the equilibrium pricing function. When quality is latent there is an obvious identification problem since there is no inherent scale for housing quality. For every non-linear pricing model, there exists a transformation of the utility function such that this transformation is observationally equivalent to the original model and pricing is linear. We can, therefore, normalize housing quality by setting it equal to the rental price in a baseline period and identify preferences for housing from the observed income expansion paths. We need data for more than one time period or multiple spatial markets to identify non-linearities in pricing of housing.[4]

In addition, we must overcome one more identification problem, that is common in this literature, but has not received proper attention. For rental units, we observe rental prices, but not housing values. For owner-occupied units, we observe housing values, but not rental prices. As a consequence, both rental rates and values are only partially observed by the econometrician. This imposes the need to identify the equilibrium rent-to-value ratio as a function of quality. We consider investors who trade real estate assets and model the equilibrium in these competitive asset markets. We show that the value of the house is given by the expected net present value of the discounted stream of rental income. Housing values and rents are, therefore, closely linked in equilibrium. One cannot analyze values separately from rents. At each point of time, the proportionality between rents and values can be captured by a time varying quality dependent rent-to-value function. We show that these functions are non-parametrically identified by characterizing the set of households that are indifferent between owning and renting for a given level of housing quality.

[4]Exploiting variations among multiple markets is also a useful strategy to obtain identification if characteristics are observed as discussed in Bartik (1987) and Epple (1987). Alternative strategies for identification are discussed in Ekeland, Heckman, and Nesheim (2004), Bajari and Benkard (2005), and Heckman, Matzkin, and Nesheim (2010) and Bajari, Fruehwirth, Kim, and Timmins (2012).

Our econometric findings provide valuable new evidence about metropolitan housing markets. We provide two applications. We study changes in price across the quality distribution in a metropolitan area (Miami) that experienced dramatic housing prices during the recent housing *bubble* . We find housing prices relative to annualized rents increased over the entire quality spectrum, but with especially pronounced increases at the lower end of qualities. These findings accord well with the widespread reporting of eased access to credit, especially for lower-income buyers, during the housing price run-up.[5]

In our second application, we estimate the model simultaneously for two different large metropolitan areas, New York and Chicago. The results provide a comparison of the quantity of housing of each quality for the two metropolitan areas. We estimate the compensating variation arising from the difference in quality and price a household of a given income would obtain in a larger relative to a smaller metropolitan area. This welfare measure is of interest in its own right and provides valuable insights into agglomeration economics. In particular, the aggregate of the compensating variations across the entire set of households occupying the larger metropolitan area provides a measure of the aggregate compensating variation foregone by households choosing to reside in the larger metropolitan area, and hence a measure of the minimum agglomeration benefits that must be provided by the larger metropolitan area. We find that for a household at the 50th income percentile in Chicago, a CV of approximately 20% of income is required to induce to move that household to New York.

[5]Our modeling approach is a closely related to recent work on nonlinear pricing in housing markets by Landvoigt, Piazzesi, and Schneider (2011) who find that the relaxation of credit constraints had a large impact on house prices in San Diego. Their findings are broadly consistent with our findings. While their results are based on a more sophisticated dynamic model than the one we employ, our approach focuses on empirical tractability based on formally derived identification results and feasible semi-parametric estimation. In that sense both approaches are complimentary.

2 Asset Prices and Rental Rates

We consider the determination of asset prices and rental rates for houses with heterogenous quality. Our model distinguishes between housing services, defined as the period flow of housing consumption, and housing assets. Housing values or prices for real estate assets depend on prevailing and expectations about future interest rates, costs of homeownership, property taxes and rental rates for housing services. Real estate values are determined in asset markets. Housing services can be rented in frictionless markets that allow for nonlinear pricing of housing quality. Current and future rental rates partially determine housing values, while housing values partially determine the supply of new housing units. As a consequence, both markets cannot be studied in isolation.

2.1 Asset Markets

First, we consider the asset markets for housing. Housing units differ by quality, which can be characterized by a one-dimensional ordinal measure denoted by h. For each level of housing quality h, there is an asset market in which investors can buy and sell houses at the beginning of each period. Let $V_t(h)$ denote the asset price of a house of quality h at time t.

Assumption 1 *Private investors are risk neutral.*

Note that we can weaken this assumption. We only need to require that the marginal investor is risk neutral. Investors can borrow capital in short term bond markets. The one-period interest rate is denoted by i_t. Investors are also responsible for paying property taxes to the city. The property tax rate is given by τ_t^p. Finally owners have additional costs due to appreciation and maintenance that occurs with rate δ_t.

Assumption 2 *Asset markets are competitive.*

The expected profits, Π_t, of buying a house with quality h at the beginning of period t and selling it at the beginning of the next period is then given by:

$$E_t[\Pi_t(h)] = E_t\left[-V_t(h) + v_t(h) + \frac{V_{t+1}(h)(1-\tau^p_{t+1}-\delta_{t+1})}{1+i_t}\right] \tag{1}$$

where the first term reflects the initial investment, the second term the flow profits from rental income at time t, and the last term the discounted expected value of selling the asset in the next period.[6]

Since investors are risk neutral and entry into the profession is free, expected profits for investors must be equal to zero. Hence housing values or asset prices must satisfy the following no-arbitrage condition:

$$0 = E_t\left[-V_t(h) + v_t(h) + \frac{V_{t+1}(h)(1-\tau^p_{t+1}-\delta_{t+1})}{1+i_t}\right] \tag{2}$$

Solving for $V_t(h)$, we obtain the following recursive representation of the asset value at time t:

$$V_t(h) = v_t(h) + \frac{(1-\tau^p_{t+1}-\delta_{t+1})}{(1+i_t)} E_t\left[V_{t+1}(h)\right]$$

By successive forward substitution of the preceding, we obtain:

$$V_t(h) = v_t(h) \; + \; E_t \sum_{j=1}^{\infty} \; \beta_{t+j} \; v_{t+j}(h) \tag{3}$$

where the stochastic discount factor is given by:

$$\beta_{t+j} = \prod_{k=1}^{j} \frac{(1-\tau^p_{t+k}-\delta_{t+k})}{(1+i_{t+k-1})} \tag{4}$$

This demonstrates that the asset value of a house of quality h is the the expected discounted flow of future rental income. The discount factors β_{t+j} depend on interest

[6]For analytical convenience, we are assuming that property taxes and maintenance expenditures are due at the beginning of the next period.

rates, property tax rates and depreciation rates. An alternative instructive way of writing this expression is as follows. Let $1+\pi_t(h) = \frac{v_{t+j(h)}}{v_{t+j-1(h)}}$ denote the rate of housing inflation at date t. Define $\tilde{\beta}_{t+j}$ as follows:

$$\tilde{\beta}_{t+j}(h) = \prod_{k=1}^{j} \frac{\left(1-\tau^p_{t+k}-\delta_{t+k}\right)\left(1+\pi_{t+k}(h)\right)}{\left(1+i_{t+k-1}\right)} \tag{5}$$

Then:

$$V_t(h) = \frac{v_t(h)}{c_t(h)} \tag{6}$$

where $c_t(h)$ is the user cost of capital:

$$c_t(h) = \frac{1}{1+E_t\sum_{j=1}^{\infty}\ \tilde{\beta}_{t+j}(h)} \tag{7}$$

Consider the time-invariant case studied by Poterba (1984, 1992):

$$E_t \prod_{k=1}^{j} \frac{(1-\tau^p_{t+k}-\delta_{t+k})(1+\pi_{t+k}(h))}{(1+i_{t+k-1})} = \left[\frac{(1-\tau^p-\delta)(1+\pi(h))}{1+i}\right]^j \tag{8}$$

When τ^p, δ, π, and i are small, the preceding closely approximates the continuous time solution of Poterba (1984): $u(h) = (i+\tau^p+\delta-\pi(h))$.

Our model does not necessarily assume that investors have correct expectations about housing rental appreciation. It is possible that expectations of rental price increases prove to be greater than the actual rates of increase that are realized.

2.2 Rental Markets

To complete our model of asset prices, we need to derive the equilibrium rent function that prevails in the market for housing services. One could follow a number of different approaches to derive this function. Here we follow the hedonic literature and allow for non-linear pricing in a rental market for housing services. There is a continuum of renters with mass equal to N_t. We normalize the population at the initial date to be one ($N_1 = 1$) and treat $\{N_t\}_{t=1}^{\infty}$ as an exogenous process.

In our baseline model, renters differ in income denoted by y. We can extend this model and allow for additional sources of heterogeneity among households. The main advantage of the baseline model is that we can obtain a closed form solution to the equilibrium rental price function, as we will see below, which is helpful to establish the basic identification results. However, all of the key results go through in more general class of models discussed in Section 2.3.

Let $F_t(y)$ be the metropolitan income distribution at time t. Renters have preferences defined over housing services h and a composite good b. Let $U_t(h, b)$ be the utility of a household at time t.

Since housing quality is ordinal, housing quality is only defined up to a monotonic transformation. Given such a normalization, we can define a mapping $v_t(h)$ that denotes the period t rental price of a house that provides quality h. The transactions cost in the rental market are zero. Hence, renters can costlessly change its housing consumption on a period-to-period basis as rental rates change. It follows that a renter's optimal choice of housing to rent at each date t maximizes its period utility at date t:

$$\begin{aligned} & \max_{h_t, b_t} \; U_t(h_t, c_t) \\ s.t. \;\; y_t \;\; &= \;\; v_t(h_t) + c_t \end{aligned} \tag{9}$$

where c_t denotes expenditures on a composite good.

The first-order condition for the optimal choice of housing consumption is:

$$m_t(h_t, y_t - v_t) \;\equiv\; \frac{U_h(h_t, y_t - v_t)}{U_c(h_t, y_t - v_t)} \;=\; v_t'(h_t) \tag{10}$$

Solving this expression yields the housing demand $h_t(y_t, v_t(h))$. Integrating over the income distribution yields the aggregate housing demand $H_t^d(h|v_t(h))$:

$$H_t^d(h|v_t(h)) = \int_0^\infty 1\{h_t(y, v_t(h)) \ \leq h\} \, dF_t(y) \tag{11}$$

where $1\{\cdot\}$ denotes an indicator function. Thus $H_t^d(h|v_t(h))$ is the fraction of renters whose housing demand is less than or equal to h.

To characterize household sorting in equilibrium, we impose an additional restriction on household preferences.

Assumption 3 *The utility function satisfies the following single-crossing condition:*

$$\frac{\partial m_t}{\partial y}\Big|_{U_t(h, y - v(h)) = \bar{U}} > \ 0 \tag{12}$$

Assumption 1 states that high-income renters are willing to pay more for a higher quality house than low-income renters – a weak restriction on preferences. The single-crossing condition implies the following result.

Proposition 1 *If $F_t(y)$ is strictly monotonic, then there exists a monotonically increasing function $y_t(v)$ which is defined as*

$$y_t(v) \ \ = \ \ F_t^{-1}(G_t(v)) \tag{13}$$

Note that $y_t(v)$ fully characterizes household sorting in equilibrium.

Finally, we need to consider housing supply to close the model. Let $q_t(h)$ denote the density of housing of quality h at date t. The distribution of housing quality is thus fixed at time. However, it can change over time. These changes are captured by the following law of motion:

$$q_t(h) = s(q_{t-1}(h), V_t(h), V_{t-1}(h)) \tag{14}$$

Supply of quality h at date t thus depends on the quantity of that housing quality the previous period, the values of houses of that quality in the previous and current periods. This formulation reflects the fact that home builders produce and sell dwellings and hence are concerned about the market value of the dwelling, $V_t(h)$, and not implicit rent. Including lagged values of quantity and price serves to capture potential adjustment costs. To estimate the model we will later use a parametric version of this function.

Assumption 4 *We adopt the following constant-elasticity parametric form for this supply function:*

$$q_t(h) = \frac{1}{k_t}\, q_{t-1}(h) \left(\frac{V_t(h)}{V_{t-1}(h)} \right)^{\zeta} \tag{15}$$

where

$$k_t = \int_0^{\infty} q_{t-1}(h) \; \left(\frac{V_t(h)}{V_{t-1}(h)} \right)^{\zeta} \; dh \tag{16}$$

While this function is not explicitly derived from a specification of a cost function for the producer, it has attractive properties. It is parsimonious; it introduces only one additional parameter, ζ. Equation (15) also implies that the stock of housing of quality h does not change from date $t-1$ to date t if the the asset price of that quality of housing does not change. If the price of housing type h rises, the quantity rises as a constant elasticity function of the proportion by which the price increases. If the price of housing type h falls, the quantity declines reflecting depreciation and reduced incentive to invest in maintaining the housing stock. The magnitude of the response depends on the elasticity ζ.

In period one, we take the housing stock, $R_1(h)$, as given. The market clearing condition for the housing market in period one is then:

$$G_1(v_1(h)) = R_1(h) \tag{17}$$

Consider periods $t > 1$. The distribution of housing supply in period t is:

$$R_t(h) = \int_0^h k_t \; q_{t-1}(x) \; \left(\frac{V_t(x)}{V_{t-1}(x)} \right)^{\zeta} \; dx \tag{18}$$

We thus obtain a recursive relationship governing the evolution of the supply of housing over time. Market clearing in the housing market at date t requires:

$$G_t(v_t(h)) = R_t(h) \tag{19}$$

The "number" of renters of income y at date t is given by:

$$n_t^y(y) = N_t f_t(y) \tag{20}$$

where $f_t(y)$ is the income density. Similarly, the number of houses at rental v is:

$$n_t^v(v) = N_t g_t(v) \tag{21}$$

where $g_t(v)$ is the income density. Single-crossing implies that, in equilibrium, the house rental expenditure at date t by income y must satisfy:

$$N_t F_t(y) = N_t G_t(v) \tag{22}$$

or $F_t(y) = G_t(v)$. In equilibrium rental markets must clear for each value of h. We can define an equilibrium in the rental market for each point of time as follows:

Definition 1 *A hedonic housing market equilibrium is an allocation of housing consumption for each renter and price function $v_t(h)$ such that*
a) Renter behave optimally given the price function;
b) Housing markets clear, i.e. for each level of housing quality h, we have:

$$H_t^d(h| \; v_t(h)) = R_t(h) \tag{23}$$

An equilibrium exists under standard assumptions discussed in the hedonic literature.

To obtain a closed form solution for the equilibrium pricing function, we impose additional functional form assumptions.

Assumption 5 *Income and housing are distributed generalized log-normal with location parameter (GLN4).*[7]

$$\begin{aligned} \ln(y_t) &\sim GLN4(\mu_t, \sigma_t^{r_t}, \beta_t) \\ \ln(v_t) &\sim GLN4(\omega_t, \tau_t^{m_t}, \theta_t) \end{aligned} \tag{24}$$

We will show below that these functions are sufficiently flexible to fit the housing value and income distributions in all metro areas and all time periods that we consider in the empirical analysis.

Imposing the restriction that $r_t = m_t$ permits us to obtain a closed-form mapping from house value to income. We then establish that the further assumption that $\theta_t - \beta_t$ is time invariant permits us to obtain a closed-form solution to the hedonic price function.[8]

Proposition 2 *If $r_t = m_t \ \forall t$, the income housing value locus is given by the following expression:*

$$y_t = A_t \, (v_t + \theta_t)^{b_t} - \beta_t \tag{25}$$

with $a_t = \mu_t - \frac{\sigma_t}{\tau_t}\omega_t$, $A_t = e^{a_t}$, and $b_t = \frac{\sigma_t}{\tau_t}$.

[7] The four-parameter distribution for income simplifies to the standard two-parameter lognormal when the location parameter, β_t, equals zero and the parameter $r_t = 2$. Similarly for the house value distribution. See Appendix B

[8] We impose both of these restrictions when estimating our model.

For our discussion of identification below, it is useful to note that all of parameters of the sorting locus, $a_t = \mu_t - \frac{\sigma_t}{\tau_t}\omega_t$, $A_t = e^{a_t}$, $b_t = \frac{\sigma_t}{\tau_t}$, and θ_t can be estimated directly from the data. In addition, it will be useful below to note that if $b_t > 1$, this function is convex.

To obtain a closed form solution for the equilibrium price function, we adopt the following functional form for household preferences.

Assumption 6 *Let utility given by:*

$$U = c_t(h) + \frac{1}{\alpha}\ln(y_t - v_t(h) - \kappa) \tag{26}$$

with $c_t(h) = \ln(1 - \phi(h+\eta)^\gamma)$*, where* $\alpha > 0$*,* $\gamma < 0$*,* $\phi > 0$*, and* $\eta > 0$.[9]

In addition to yielding a closed-form solution for the hedonic price function, this utility function proves to be relatively flexible in allowing variation in price and income elasticities. The conventionally defined income and price elasticities are obtained when the hedonic function is linear, i.e.,when $v(h) = ph$. The price elasticity of demand is then given by:

$$\frac{dh}{dp}\frac{p}{h} = \frac{(-\alpha\phi\gamma h + (h+\eta)((h+\eta)^{-\gamma} - \phi))}{(-\alpha\phi\gamma + (1-\gamma)(h+\eta)^{-\gamma} - \phi)}\frac{1}{h} \tag{27}$$

and the income elasticity of demand is given by:

$$\frac{dh}{dy}\frac{y}{h} = \frac{-\alpha\phi\gamma}{p}\left[\frac{1}{-\alpha\phi\gamma + (1-\gamma)(h+\eta)^{-\gamma} - \phi}\right]\frac{\kappa - \frac{p}{\alpha\phi\gamma}\left[-\alpha\phi\gamma h + (h+\eta)^{1-\gamma} - \phi(h+\eta)\right]}{h} \tag{28}$$

We will show that this specification of household preferences yields plausible price and income elasticities.

Given this parametric specification of the utility function, we have the following result:

[9] This utility function requires the following two assumptions be satisfied $1 - \phi(h+\eta)^\gamma > 0$ and $y_t - v_t - \kappa > 0$.

Proposition 3 *If $b_t > 1$ ($\sigma_t > \tau_t$) and $\kappa = \theta_t - \beta_t$ $\forall t$, the hedonic pricing function is well defined and given by:*

$$v_t(h) \;=\; \left(A_t\left[1-(1-\phi(h+\eta)^{\gamma})^{\alpha(b_t-1)}\right]\right)^{\frac{1}{1-b_t}} - \theta_t \qquad (29)$$

for all $h > (\frac{1}{\phi})^{\frac{1}{\gamma}} - \eta$

Note that $\frac{\sigma_t}{\tau_t} > 0$ is required for the price function to be increasing with h.

Our analysis of rental markets, therefore, provides an analytical characterization of the the rental price of housing, $v_t(h)$, as a function of house quality, h. The market fundamentals determining $v_t(h)$ are the quality of the housing stock and the demand for housing services arising from the distribution of income in the metropolitan population. Equilibrium also depends indirectly on the equilibrium in asset markets since supply depends on asset values..

2.3 Extensions

In this section we show how to extend this model to allow for a richer set of heterogeneity among households. Let us assume that there exist $i = 1, ..., I$ different types of households. For examples, households may differ by age, number of children, level of education, race, or ethnicity. The fraction of each type i at time t is given by s_{it}. Each type of household has a utility function that depends on its type $U_i(h, b)$. For example, a straight forward extension of the parametric utility function used in the previous section is the following utility function:

$$U_i(h, b) = \ln(1 - \phi_i(h + \eta_i)^{\gamma_i}) + \frac{1}{\alpha_i}\ln(b - \kappa_i)$$

Moreover, let $F_{i,t}(y)$ denote the income distribution for each type.

In principle, we can proceed as before and derive the demand of each household type as above, treating housing as a continuous good. The main drawback of this approach is that we can no longer obtain analytical solutions for the equilibrium price functions for almost all reasonable specifications of the model. As a consequence, we have to rely on numerical solution methods. Anticipating the need to rely on numerical solution algorithms, it useful to develop the extension of our model using a discretized approximation of the housing stock.[10]

Given a grid of values $(h_1, ..., h_J)$, we can use a discrete distributions to approximate the continuous distributions characterizing housing demand and supply. Moreover, we can index the pricing function accordingly and let $v_{jt} = v_t(h_j)$. In the discretized version of the model, there will no longer be a one to one mapping between between income and housing consumption conditional on type. Instead there will exist cut-off points $\hat{y}_{i,j,t}$ such that

$$U_i(h_j, \hat{y}_{i,jt} - v_{jt}) \;\; = \;\; U_i(h_{j+1}, \hat{y}_{i,j,t} - v_{j+1,t}) \tag{30}$$

As a consequence all households of type i with $\hat{y}_{i,j-1,t} \leq yl\hat{y}_{i,j,t}$ will consume housing quality j assuming that preferences for each type satisfy a single-crossing condition.[11] For the parametrization of the utility function discussed above, these cut-off levels are given by:

$$\hat{y}_{i,j,t} \;\; = \;\; \frac{v_{j,t} - e^{(M_{i,j+1} - M_j)\alpha_i} \; v_{j+1,t}}{1 - e^{(M_{i,j+1} - M_j)\alpha_i}} + \kappa_i \tag{31}$$

where $M_{i,j} \; = \; \ln(1 - \phi_i(h_j + \eta_i)^{\gamma_i})$. Given these income cut-offs, the demand of household type i for houses with quality h_j is given by:

$$H^d_{i,j,t}(v_{1t}, ..., v_{Jt}) \;\; = \;\; F_{i,t}(\hat{y}_{i,j,t}) \; - \; F_i(\hat{y}_{i,j-1,t}) \tag{32}$$

[10] Another advantage of the the discrete approach is that we can easily relax the functional form assumption for the utility function.

[11] Notice the similarity between this model and household sorting model estimated in Epple and Sieg (1999).

where $F_i(y)$ is the income distribution function of type i. Summing over all types then yields the total demand. The market clearing conditions for housing can, therefore, be written as a system of J nonlinear equations in each period t:

$$\sum_{i=1}^{I} s_{i,t} \; H_{i,j,t}(v_{1,t}, ..., v_{J,t}) \quad = \quad r_{j,t} \;\; \forall j \tag{33}$$

where $r_{j,t}$ is the fraction of units in each quality bin h_j

$$r_{j,t} = R_t(h_j) - R_t(h_{j-1}) \tag{34}$$

It is straightforward then to extend the definition of equilibrium to our extended model.

3 Identification

We consider identification of the model assuming that a) we have access to data for one market; and b) h is not observed. Moreover, we first consider the model with one type for which we have an analytical solution of the equilibrium price function. Since housing quality is ordinal and latent, there is no well-defined unit of measurement for housing quality. This imposes some obvious limits on identification, as formalized by the following proposition.

Proposition 4 *For every model with equilibrium rental price function $v(h)$, there exists a monotonic transformation of h denoted by h^* such that the resulting equilibrium pricing function is linear in h^*, i.e. $v(h^*) = h^*$.*

We can use arbitrary monotonic transformations of h and redefine the utility function accordingly. Proposition 4 then implies that if we only observe data in one housing market and one time period, we cannot identify $u_1(h)$ separately from $v(h)$. A corollary of Proposition 4 is that we can normalize housing quality by setting $h = v_t(h)$ in some baseline period t. As we show in the proof of Proposition 5, this allows us to establish identification of the preference parameters. If, in addition, we make the standard assumption that per-period preferences are not changing over time, we can also identify the price functions in all subsequent periods.

Assumption 7 *The utility function is time invariant.*

Assumption 7 implies that the following result.

Proposition 5 *The parameters of our utility function and the price function in all periods $t + s$, $s > 1$ are identified.*

Once we normalize quality in the baseline period, the parameters of the utility function are identified of the observed income-expansion paths in the baseline period. Conditional on knowing the utility function, the rental price functions in all subsequent periods only depend on the observed joint distribution of rents and income. The proof of Proposition 5 provided in the appendix formalizes this result.

Thus far we have implicitly assumed that the distribution of rents is observed by the econometrician. Here, we discuss how to relax this assumption and account for the fact that rents are not observed for owner-occupied housing and need to be imputed.

One key implicit assumption of our model is that households are indifferent between renting from themselves (i.e. living in owner occupied housing) or from a third person. Households consumption is only a function of income holding prices fixed. Hence households with income y consume the same number of housing independently whether they live in a rental unit, for which observe, $v_t(y)$, or live in an owner-occupied unit, for which we observe $V_t(y)$. By varying income y we can trace out the equilibrium locus $V_t(v)$. As a consequence, we have the following result:

Proposition 6 *There exists an equilibrium locus $v_t = v_t(V_t)$ which characterizes the rent of any housing unit as a function of its asset price. Moreover, this function is non-parametrically identified.*

Proposition 6 implies that we can impute rents for owner-occupied using the rent-value functions. In practice, our data are more noisy since rents and values are not perfectly correlated with income as predicted by our model. However, we can use $E[v_t|y]$ and $E[V_t|y]$ to estimate the two sorting loci, $v_t(y)$ and $V_t(y)$, and proceed as discussed above. As a consequence the rent-to-value function is non-parametrically identified.

Once we have identified the rent-to-value function, it is also straight forward to identified the housing supply function based on the market clearing condition in periods $t \geq 2$. We have the following result

Proposition 7 *The parameters of housing supply function are identified if we observe the equilibrium for, at least, two periods.*

Note that the key insights of the identification strategy cary over to the model with multiple discrete types. Proposition 4 is still valid. Our approach for identifying rent-to-value functions described in Proposition 6 also generalizes to models in which households are characterized by an observed vector of characteristics. The key assumption is that the average quality of housing consumption conditional on observed characteristics is the same for owners and renters, i.e. there is no sorting on unobservables into home ownership. The non-parametric matching algorithm extends to more general demand models in which demands depends on a vector of observed state variables.[12]

The main problem is to extend the results in Proposition 5. Note that the proof of Proposition 5 presented in the appendix relies on the analytical solution of the equilibrium price function. But the basic ideas behind the proof of Proposition 5 carry over. Given the normalization of quality in terms of values in the baseline period, the condition for identification is that there exists a unique value of the parameters of the utility function that is consistent with the J market clearing conditions in the based line period. Loosely speaking, the parameters of the utility function have to be identified of the observed income expansion path in the first period for each household type. If these conditions are met, the parameters of the utility function can be recovered from the observed equilibrium in the baseline period. Equilibrium prices for all subsequent periods are given by the market clearing conditions in all subsequent periods.

4 Estimation

The proofs of identification are constructive and can be used to define a three step estimator for our model. First, we estimate the rent-to-value functions using a non-parametric matching estimator. We estimate the rent-value function for each time period allowing for changes in the user-cost functions across time periods. This approach captures changes in credit market conditions and investor expectations in a flexible non-parametric way. Second, we impute rents for owner-occupied housing and estimate the joint aggregate distribution of rents and income for each time period.

[12]An alternative strategy to identify and estimate the rent-to-value function is discussed in Bracke (2013), who uses observations of houses that were both rented and sold within a short period.

Third, we estimate the structural parameters of the rental model using an extremum estimator which matches quantiles of the income and value distributions while imposing the parameter constraints in Propositions 2 and 3 and the housing market equilibrium restriction that $R_{t+j}(h) = G_{t+j}(v_{t+j}(h))$ for $j \geq 1$.

Let $\tilde{F}_{t,j}^N$ denote the jth percentile of empirical income distribution at time t that is estimated based on a sample with size N. Similarly, let $\tilde{G}_{t,j}^N$ denote the jth percentile of empirical housing value distribution at time t that is estimated based on a sample with size N. Moreover, let $F_t(y_{t,j};\psi)$ and $G_t(y_{t,j};\psi)$ denote the theoretical counterparts of the quantiles predicted by our model. Our extremum estimator is then defined as:

$$\hat{\psi}^N = \text{argmin}_{\psi \in \Psi}\, L^N(\psi) \tag{35}$$

subject to the structural constraints. The objective function is:

$$L^N(\psi) = (1 - W)\ (l_y^N(\psi) + l_r^N(\psi))\ +\ W\ l_h(\psi)$$

for some weight $W \in [0, 1]$ and:

$$\begin{aligned}
l_y^N(\psi) &= \sum_{t=1}^{T}\sum_{j=1}^{J}([F_t(y_{t,j};\psi) - F_t(y_{t,j-1};\psi)] - [\tilde{F}_{t,j}^N - \tilde{F}_{t,j-1}^N])^2 \\
l_r^N(\psi) &= \sum_{t=1}^{T}\sum_{j=1}^{J}[G_t(v_{t,j};\psi) - G_t(v_{t,j-1};\psi)] - [\tilde{G}_{t,j}^N - \tilde{G}_{t,j-1}^N])^2 \\
l_h^N(\psi) &= \sum_{t=2}^{T}\sum_{j=1}^{J}([G_t(v_t(h_j;\psi) - R_t(h_j;\psi))^2
\end{aligned}$$

Note that W is the weight that is assigned to the market clearing conditions. We use a standard bootstrap procedure to estimate the standard errors.[13]

This estimator can be extended to estimate the model with multiple observed types discussed in Section 2.3. There are two differences. First, we need to estimate

[13]We find that our estimates do not depend significantly on the choice of this parameter.

the rent to value loci separately for each type to convert housing values into rent. That gives us the aggregate rent distributions for each type. Second, we do not have an analytical solution to the hedonic price function, but need to compute the equilibrium prices numerically using the J market clearing conditions in the discretized version of the model. With these two changes, the modified estimator algorithm follows the steps discussed above for the simpler model.

5 Empirical Results

5.1 The Housing Market of Chicago

We have obtained the data from the American Housing Survey, the most comprehensive national housing survey in the United States. It is conducted in the field from May 30 through September 8. There is a national and a metropolitan version, and, in selected years, also an extended metropolitan component for some metropolitan areas in the national version. We use the latter for Chicago in 199 and 2003 to illustrate the methods developed in this paper. There are surveys conducted every year, but the metropolitan areas covered in the metropolitan version change in each year. There is no fixed interval over which the same metropolitan area is surveyed again. The unit of observation in the survey is the housing unit together with the household. The same housing unit is followed through time, but the sample of households may change. Database limitations make it inaccurate to determine if the same household resides in the unit in different surveyed years, and impossible to track households that moved. [14] The Chicago metro area is defined by the Census in 1999 and 2003

[14] The sample is selected from the decennial census. Periodically, the sample is updated by adding newly constructed housing units and units discovered through coverage improvement. The survey data is weighted because there is incomplete sampling lists and non response. The weights are

to consist of the following counties: Cook, Du Page, Kane, Lake, McHenry, and Will.

	1999	**2003**
No Children	0.567	0.560
Children	0.432	0.434
Owners	0.611	0.594
Renters	0.388	0.405
White	0.650	0.684
Non-white	0.349	0.315

Table 1: Percentages of Households by Characteristics.

Table 1 shows some descriptive statistics of the survey results. Initially, we will define household types as defined by k-means clustering. This is a standard method in data mining, originally from signal processing[15]. The method partitions the points in a multidimensional data matrix into k clusters. An iterative partitioning minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances. The results presented here are for squared Euclidean distances. Similar results were obtained with cosine distance and are available from the authors.

Overall, we are interested in capturing the interaction of 2 very important variables that we believe influence greatly the economics decisions of households: age and number of children. They are very important variables for policy (i.e. affect rent control for older households) as well. Households are being classified with respect to

designed to match independent estimates of the total number of homes. Under-coverage and nonresponse rate is approximately 11 percent. Compared to the level derived from the adjusted Census 2000 counts, housing unit under-coverage is about 2.2 percent.

[15]Neill (2006) provide a clear treatment of this topic. Nath (2007) shows an interesting application of this method to detection of crime hot spots.

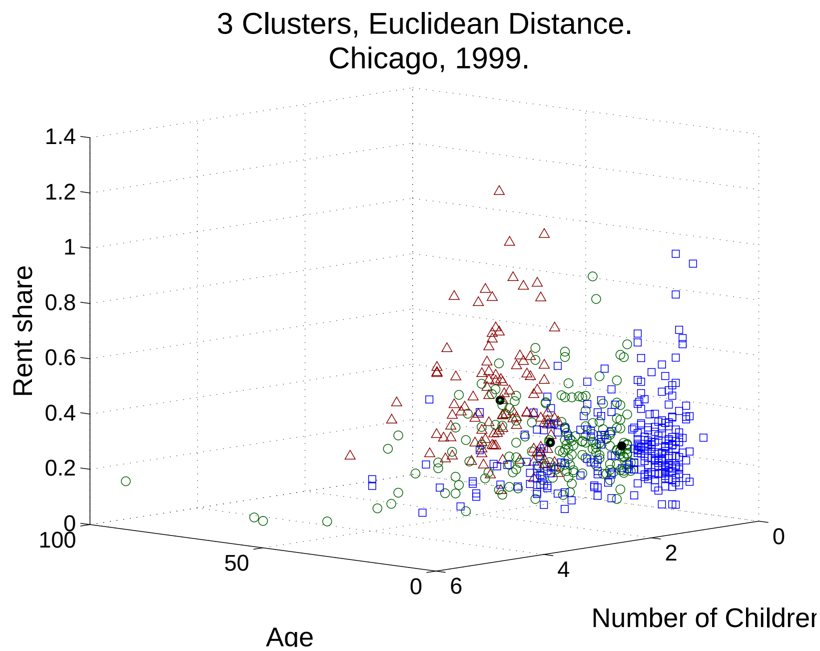

Figure 1: k-means clustering Chicago SMA, 1999.

their behavior in the housing market, captured by the share of income they spend on rent.

Table 2 show the estimated clusters (in color) and centroids (black dots). Figure (1) shows a sample clustering. Using Euclidian distance, we can think of the centroids as the typical household in each particular cluster. We are choosing the clusters to be 3. This number is suggested by measures of the optimal number of clusters calculated with our sample data. As we will see, our estimation supports the claim that these types do present significantly different preferences. We find a rather intuitive classification of the algorithm into 3 groups that could be described as: young households with few or no children, middle age with more than one child, and older households with no children. We expect the last cluster to be comprised of a combination of households whose children have left their household already and a minority of those

Table 2: k-means clustering centroids. Chicago

	Cluster	# Children	Age	Share of Population
	1	0.29	29.03	0.261
1999	2	1.45	45.60	0.465
	3	0.24	72.31	0.273
	1	0.29	29.19	0.246
2003	2	1.43	45.75	0.484
	3	0.269	73.10	0.269

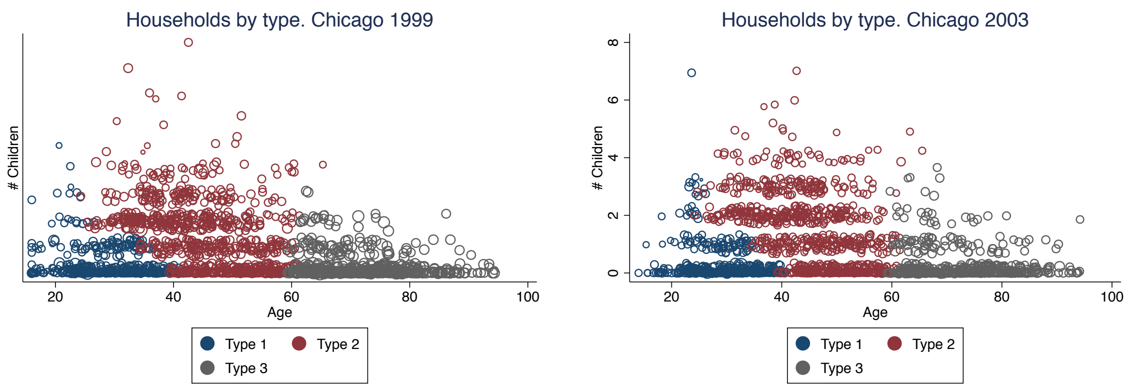

Figure 2: # of Children for each household type obtained through k-means clustering.

that never had children, but we tell the two apart from our data.

Figure (2) shows that age is an important factor defining the clustering. However, there are also important differences in the number of children in each group, the second group showing significantly more children in the household. In the figure, the size of the circles represent the number of households represented by each plotted point.

Across types, younger households (type 1) are renters more often than households

Figure 3: Percentage of owners and renters for each household type obtained through k-means clustering.

in the other 2 types, but differences change gradually (figure 3). Figure 4 shows that households in the 1st and 2nd type have overall higher incomes. When we look at values (Figure 15) we see that type 2 households have overall more expensive houses as expected. However, higher rents are being paid by households of type 1. This corresponds to common knowledge about younger compared to older households.

As mentioned in section 2.1, the model incorporates both owners and renters in a single market, separating the consumption from the investment decision. The mapping between rents and values provided by the user cost allows our pricing functions to be consistent across both markets. In particular, our optimization implies that households with equal incomes will consume equal quality of housing whether they rent or own[16]. Using this result, we estimate the user cost at each income level, use the estimated demand functions to express it as a function of quality, and calculate the implied rents for owner occupied housing across the quality distribution. The obtained user cost for both periods is between 0.058 and 0.061 for 1999, and 0.051

[16] Given that our data is noisy, we estimate this equivalence by calculating the mean rents and values by income. These means give us the user cost $u(h)$.

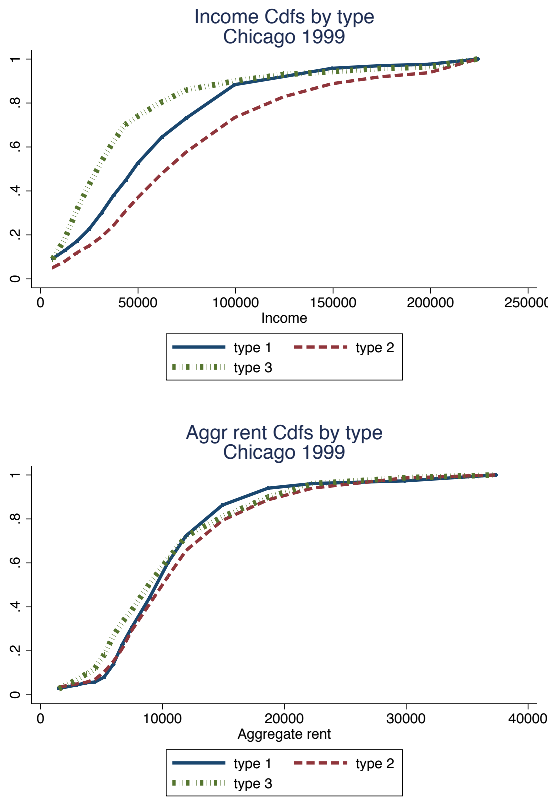

Figure 4: Aggregate rents and Income distributions for each household type obtained through k-means clustering.

and 0.062 for 2003, across qualities. Figure 4 shows the resulting aggregate rents that include owners and renter occupied housing units in the market. Also, figure 14 in Appendix C show the model does a good job of fitting observed distributions, which is a central part of our estimation strategy.

Table 3 shows the estimated structural parameters. The inverse of α is related the preference for non-housing consumption. Our estimates show that the preference for this consumption is highest in type 2. This could reflect that expenditures related to the presence of children can increase the desired demand for other goods and services brought by needs particular to these demographics. Inversely, this parameter implies type 3 exhibits a stronger relative preference for housing quality.

Another important parameter that seems to be driving most of the heterogeneity in preferences is η. η is related to the utility observed at very low qualities of housing. In this case, this direct utility from very low qualities is highest for type 2, then for type 1, and finally for type 3. Similarly, the lowest quality for which a household would consume any positive level of housing follows the same ordering across types (lowest for type 2, followed by type 1 and then type 3). This suggests that households of type 2 are more inclined to participate in the market (or, less likely to join other households). This can reflect a higher availability of an alternative housing option for households of type 3 and 1 in that order (i.e. stay with their extended family or attend a nursing home), and therefore a higher opportunity cost of consuming very low qualities.

Table 3: Estimates 1999-2003 for Chicago with types defined by k-means clustering.

Type	α	ϕ	η	γ	ζ
Type 1	0.5364	12.4928	14.7588	-1.3299	
	(0.0023)	(0.2144)	(0.0156)	(0.1233)	
Type 2	0.3584	11.5265	32.4983	-1.1388	0.065
	(0.0422)	(0.1328)	(0.0312)	(0.0629)	(0.013)
Type 3	0.6815	12.8727	10.6241	-1.3627	
	(0.0246)	(0.211)	(0.0124)	(0.0245)	

Our estimate for the annual supply elasticity is 0.065. Recall that the changes in supply stock of a certain quality over time depend on changes in values and the estimated elasticity ζ through the supply equation (16). The implied supply growth by our ζ estimate is around between 3.8% and 6% for the 4 year period, which correspond to annualized changes of approximately 1.3%. Figure 16 shows the relationship

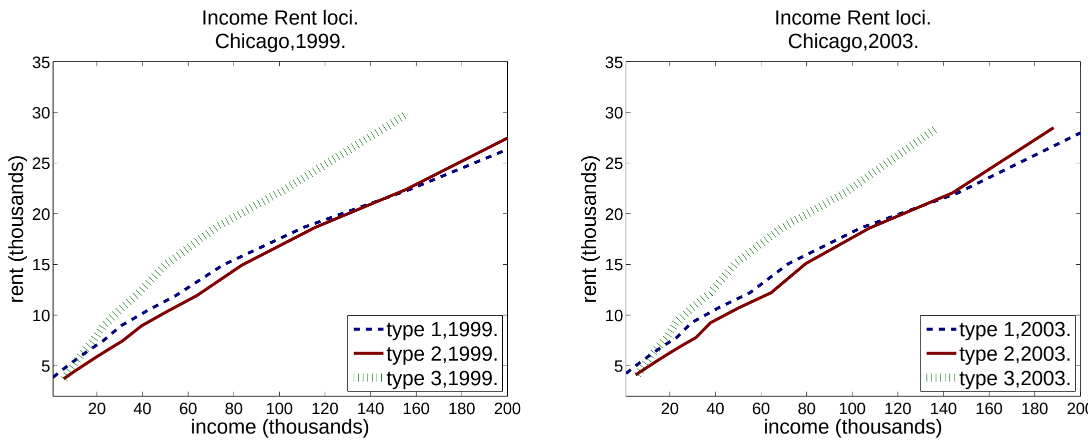

Figure 5: Income loci using types obtained through k-means clustering.

between the pricing functions, the value functions, and its effect on quality growth across time. There is an estimated growth of 4.6% in the total number of units supplied for the 4 year period, with the largest number of additional units being created in the qualities located around the middle of the distribution. Overall, the changes are along the lines of the estimates of supply elasticities summarized in Glaeser (2004).

The figures for income rent loci (figure 5) and share of income spent on housing (figure 6) provide further insight on the differential behaviors across types. Type 3 households, for instance, pay higher rents at each level income, which implies a higher share of income spent on housing. On the other side, type 2 households spend the lowest share of income in housing. This behavior is determined depend mainly by the estimated elasticities, which depend on a combination of the different preferences and income distributions for each type.

Finally, elasticities show that the type 2's demand for housing quality is more sensitive to changes in both price and income. These results support the discussion of the differences in the preferences for housing quality among types above. Types 1 and 3 show similar elasticities, with type 1 being more sensitive to changes in prices.

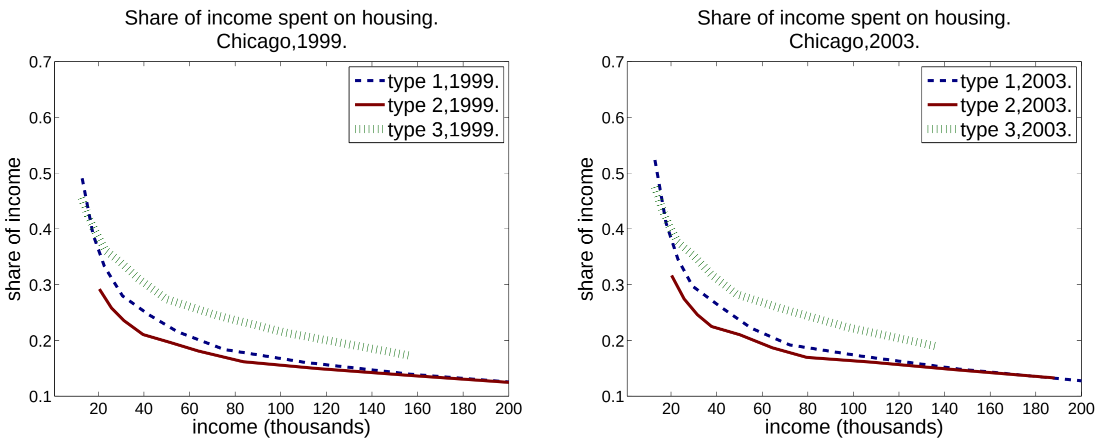

Figure 6: Share of income using types obtained through k-means clustering.

m

5.2 Multiple Metropolitan Areas: Chicago versus New York

One of the most advantageous features of our model is its capacity to separate quality from price by tracing the changes in prices for different levels of the quality distribution through time. One natural generalization is the incorporation of multiple metropolitan areas. We exploit similar assumptions to those in the single metropolitan area to obtain identification of the preference parameters and the pricing functions. In particular, we extend the time invariance assumption for preferences and assume preferences are invariant across metropolitan areas as well. Proposition 5 continues to characterization identification in the multiple metropolitan area context.[17] Hence, we obtain identification by adopting a normalization in one metropolitan area and time period.

[17]Identification proofs are similar to those already presented and are available upon request from the authors.

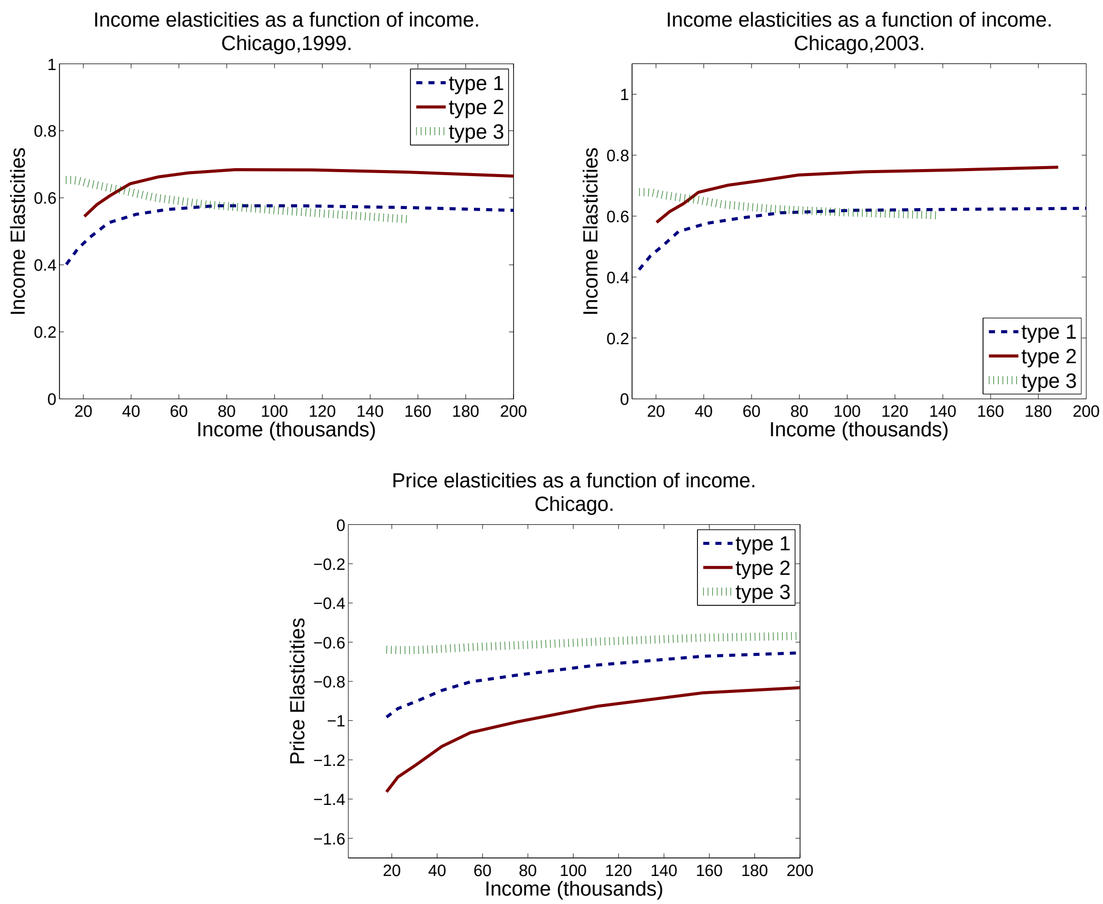

Figure 7: Elasticities using types obtained through k-means clustering.

Fortunately, the AHS provides data for concurrent years, 1999 and 2003, for two major metropolitan areas, Chicago and New York. Hence, as our second application, we estimate our model for these two periods for these two metropolitan areas. AHS definitions of each of these metropolitan areas is unchanged across these two periods. We use New York metropolitan area in 1999 as the base for our normalization.

Estimation follows the same three-step procedure. We estimate rent-to-value functions for all the quality distribution for each metropolitan area. Then, we compute the corresponding aggregate rent distributions.[18] Finally, we obtain the structural parameters using an estimator that includes orthogonality conditions for predicted and observed income and rent percentiles for each metropolitan area.[19]

The fit we obtain is good. In the interest of space, we do not report a detailed listing of the parameter estimates. It is of interest to note, however, that the estimated utility function parameters imply price and income elasticities similar to those obtained for the Metropolitan area. Income elasticities for New York decline slightly over the range of income from .7 to .6, and price elasticities increase slightly from -.7 to -.6 over the income range.

To illustrate the implications of the model, we begin with comparisons of housing prices and housing stocks by quality. We then turn to an analysis that provides insight for comparing agglomeration economies across the two metropolitan areas.The population of the New York metropolitan area is roughly twice as large as that of the Chicago metropolitan area. Hence, we would expect higher prices in New York at every quality level. The hedonic price functions for the two metropolitan areas, shown

[18]The years used in this analysis largely precede the "bubble" period in US housing markets. Our parameter estimates for this multiple-market analysis are essentially unchanged if we treat user cost as constant within each metropolitan area at each date.

[19]Preference parameters must simultaneously satisfy structural constraints for all metropolitan areas considered.

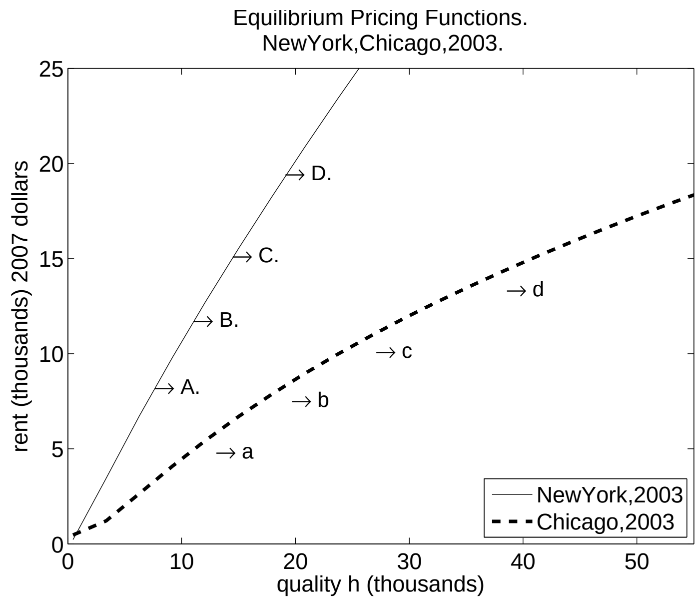

Figure 8: Points a, b, c, and d show the quality consumption and rent paid in Chicago by households at the 20^{th}, 40^{th}, 60^{th}, and 80^{th}-percentiles of the income distribution. Points A, B, C, and D represent the household's behavior with the same incomes in New York.

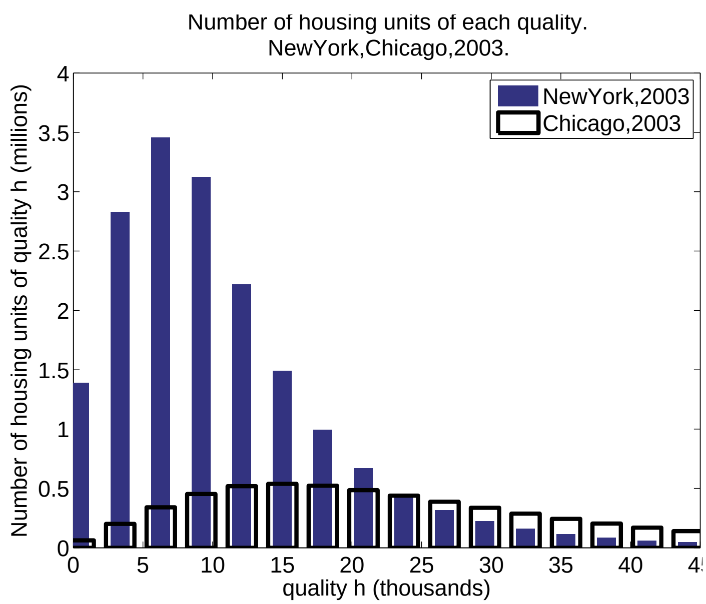

Figure 9: New York has a larger number of housing units, as expected from its larger population. When similar qualities are priced equally, the large differences in the value of the housing stock are reduced.

for 2003 in Figure 2 confirm this expectation. Points a, b, c, and d show house quality and annualized rent paid by households at the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentiles of the income distribution in Chicago at the optimally chosen housing consumption levels for those households. The corresponding upper-case values A, B, C, and D show qualities and prices that households with those incomes would optimally choose if they were located in New York. At each income level, households pay more in New York than Chicago, and consume lower quality housing in New York than Chicago. In considering these comparisons, it is important to keep in mind that our house quality measure is comprehensive and includes all locational amenities in addition to the housing structure itself. Differences in cultural or environmental amenities are embodied in our quality measure.

We next turn to a comparison of the distributions of housing qualities in the two metropolitan areas. As we have just seen, at every income level, a household in New York consumes lower quality than the corresponding household in Chicago. The effect of this difference in consumption levels is to shift the distribution of quality in New York to the left relative to that in Chicago. This effect is augmented by differences in the income distributions in the two cities. The income distribution in Chicago first-order stochastically dominates the income distribution in New York for each of our two time periods. This relatively higher concentration of low-income households in New York accentuates the leftward shift in the quality distribution in New York relative to Chicago. The distributions of housing (numbers of housing units) by quality are shown in Figure 2 . Chicago has *relatively* more high quality housing. Given its much higher population, however, New York has a larger number of housing units at almost at all quality levels than Chicago.

We find the annualized housing supply elasticity to be .055. This is a larger elasticity than we found for Miami, but recall that the elasticity for Miami was estimated

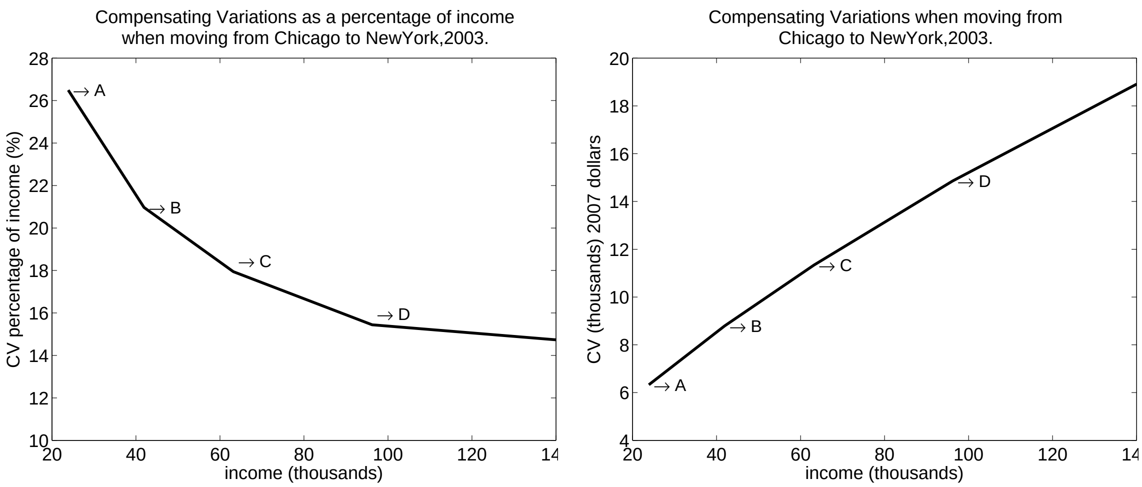

Figure 10: Given higher prices in New York City, a household with income y will consume a higher quality h in Chicago. A positive compensating variation would be required to compensated households if moved from the Chicago to the New York housing market.

over a period with dramatic run-up in housing prices. For smaller price changes, such as those for our sample period in New York and Chicago, the Miami supply elasticity would likely be higher.

We next turn to an investigation of agglomeration economies in New York relative to Chicago. In Figure 3, we plot compensating variation that would be required for household of a given income in Chicago to be equally well off in New York. For a household at the 20th income percentile in Chicago, compensating variation of approximately 25% of the household's income ($6,000) would be required. For a household at the 80th percentile, CV of approximately 16% of income ($15,000) would be required. These CV values can be thought of the minimum additional compensation in New York that would be needed to compensate households for the higher prices in New York relative to prices in Chicago. Put differently, productivity, and hence earnings, in New York would need to be higher by these amounts to compensate a household for the differences in housing price functions between the two metropolitan areas.

5.3 The Housing Market Bubble: The Case of Miami

Our final application focuses on Miami (FL) using samples from 1995, 2002 and 2007. We divide this period into two sub-periods: a) the pre-bubble period from 1995 - 2002; b) the bubble period from 2002 -2007.[20] Since our main interest is on analyzing the

[20]The Miami Metropolitan Area is defined by the Census in 1995 and 2002 to consist of Broward and Miami-Dade counties. In 2007, Palm Beach county is added to the definition of the Miami Metropolitan Area. In order to keep a constant definition of the metropolitan area across periods, we use micro data to construct the aggregates for 2007, so that only data for Broward and Miami-Dade counties are used in every period. Also, all dollar values are in 2007 dollars in this and our subsequent application.

chnage in housing markets during the bubble period we only report estimates of the model with one type.

First, we estimate the rent-to-value functions using our non-parametric matching estimator for the three periods in our data set.[21] Figure 4 plots the estimated functions for the three time periods.

We find that the rent-to-value ratio was ranging between 0.07 and 0.06 in 1995. The function changed little between 1995 and 2002 with user costs slightly decreasing at the low and high end of the quality distribution. In contrast, we see large changes in the rent-value function during the bubble period between 2002 and 2007. The range of the function is from 0.035 to 0.046.

Note that the average 30 year mortgage rate was 7.95 percent in 1995, 6.54 percent in 2002 and 6.34 percent in 2007. Hence, credit became cheaper during the time period. In addition, there is widespread evidence that credit also became more available during the the bubble period, especially for first time home owners at the lower end of the distribution. Figure 2 shows that the largest changes in the rent-value ratio are for lower quality houses which is consistent with the notion that demand for these asset may have increased more strongly due to changes in credit markets.

Another possible explanation for the change in the user-cost function is that investor expectations about future appreciation in rental rates and, thus, housing values changed during that time period. Our estimates indicate the average expectations of annual real rental appreciation must have been on the order of 2 percentage points. Our non-parametric matching approach does not allow us to distinguish between the hypothesis that changes in the rent-value ratio were driven by changes in credit market conditions or by changes in investor expectations.

[21] Note that it is difficult to estimate the locus outside a range of 50 and 500 thousand dollars due to sparseness of data in the sample outside this range.

Figure 11:

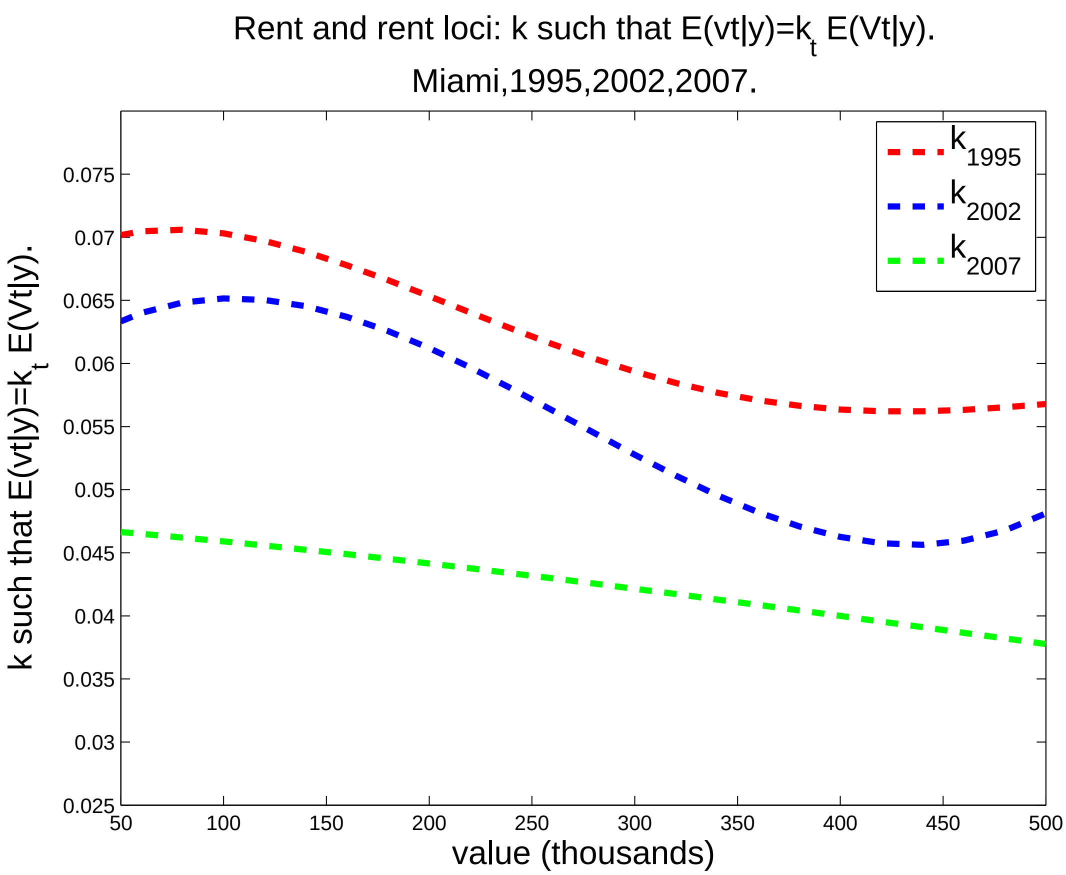
Rent and rent loci: k such that E(vt|y)=k_t E(Vt|y).
Miami,1995,2002,2007.
k such that E(vt|y)=k_t E(Vt|y).
value (thousands)
0.025
0.03
0.035
0.04
0.045
0.05
0.055
0.06
0.065
0.07
0.075
50
100
150
200
250
300
350
400
450
500
k_{1995}
k_{2002}
k_{2007}

Figure 12: Growth in income and rent across periods.

Based on our estimates of the rent-value ratios, we can convert housing values into imputed rents. Merging the observe rent distribution and the imputed rent distribution for owner occupied housing we obtain the aggregate rent distribution for each time period. We then estimate the remaining parameters of the model using the estimator discussed in Section 4. Our main interest in characterizing the change in the rental functions. Figure 5.3 shows that rents were fairly stable during the pre-bubble period between 1995 and 2002. They increased by moderate amounts and decreased at the lower end of the quality distribution. Rents increased at a faster rate during the bubble period with 5 year increases ranging between 2 and 6 percent. Our model of rental prices accounts for changes in real income, housing supply and population growth that occurred during the period from 1995-2007. Our findings are consistent with research by Sommer, Sullivan, and Verbrugge (2011) who also report that there was no “bubble ”in rental rates for housing.

Finally, we compute the implied predicted annualized capital gains during the pre-bubble and bubble periods. The predicted capital gains combine our estimates of the

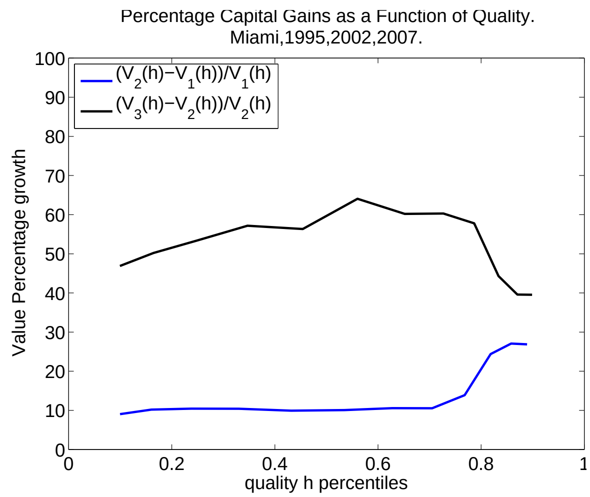

Figure 13: Capital Gains

rent-to-value functions with the predicted equilibrium hedonic rent functions. The results are illustrated in Figure 6. First consider the pre-bubble period. Our estimates predicts 7-year capital gains were approximately 10 percent for most houses. Houses in the upper two deciles of the quality distribution had larger gain of up to 20 precent. During the bubble period, our model predicts 5-year capital gains ranging between 30 and 60 percent. We see larger increases at the low and middle levels of quality than at the high end (with the exception of the highest percentiles). This pattern of gains is consistent with the notion that loosening of credit market constraints played a role in explaining the run-up in housing markets. We would expect that this change in lending policies would primarily affect lower and middle quality housing units. However, our findings are also consistent with the hypothesis that some investors did not have realistic expectations about future price levels.

6 Conclusions

We have developed a new approach for estimating hedonic price functions for rents and values treating housing quality as a latent index. Our method has a number of advantages. First, it does not require any a priori assumptions about the characteristics that determine house quality. Second, it is easily implementable using metropolitan-level data on the distribution of house values and rents, as well as the distribution of household income. Third, it provides a straightforward summary of the changes in prices across the house quality distribution. In particular, we do not need to collapse the change in the distribution of prices into one number, as, for example, the Case-Shiller index. Fourth, it is comprehensive in incorporating all location-specific amenities in addition to services provided by the dwelling. Fifth, it gives new insights into the mechanism that generates housing price changes. The estimation of these models fully identifies preferences, which would allow us to perform counterfactual and policy analyses.

We introduce an application in which we use $k-$means to cluster households into 3 types, which can be described roughly as older households with no children, younger households with no children, and middle aged households with children. We fully identify the structural preference parameters for these groups. We find a stronger preference for housing quality in the group of younger households, and interesting contrasting elasticities across groups. We also find that middle age households with children are more willing to enter the housing market even at very low qualities. In particular, parameters and elasticities estimated suggest that older households welfare is more strongly affected by upward movements in price. This estimation gives pricing functions that are consistent in both the rental and owner's market, as well as giving plausible housing supply elasticities.

We have applied our methods to study the housing markets of Miami. We find that our model of nonlinear pricing is consistent with the data observed in Miami before and during the bubble period. We find that rent-to-value functions are highly non-linear in quality. Moreover, rent-to-value ratios dropped by up to 50 percent from the pre-bubble levels during the bubble period. Rents only increased moderately. These facts then account for the large run-up in housing values and the much smaller change in rents that we observed in Miami during the bubble period. Our findings are broadly consistent with the view that changes in credit market conditions for low income houses together with changes in investor expectations about future housing prices may have accounted for the large run-up in housing values.

We have also estimated the model for New York and Chicago simultaneously, and thereby provided a contrast of hedonic price functions and the house quality distributions across the two metropolitan areas. We have estimated compensating variations arising from the difference in hedonic price functions between the two metropolitan areas. We find that a CV of approximately 20% of the annual income would be required to induce the median household in Chicago to move to NYC. The CV for a household of a given income provides a measure of the welfare costs of housing in New York relative to Chicago, and hence a measure of the minimum agglomeration benefits via increased household earnings that would be required to make that household as well off in the larger as in the smaller metropolitan area.

References

Bajari, P., & Benkard, L. (2005). Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, *113*, 1239–77.

Bajari, P., Fruehwirth, J., Kim, K., & Timmins, C. (2012). A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution. *American Economic Review*, *102*, 1898–1926.

Bartik, T. (1987). The Estimation of Demand Parameters in Hedonic Price Models. *Journal of Political Economy*, *95*, 81–88.

Berry, S., & Pakes, A. (2007). The Pure Characteristics Demand Model. *International Economic Review*, *48 (4)*, 1193–1225.

Bracke, P. (2013). House Prices and Rents: Micro Evidence from a Matched Dataset in Central London.. Working Paper.

Dunz, K. (1989). Some Comments on Majority Rule equilibria in Local Public Goods Economies. *Journal of Economic Theory*, *47*, 228–234.

Ekeland, I., Heckman, J., & Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, *112(1)*, S60–S101.

Epple, D. (1987). Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products. *Journal of Political Economy*, *95*, 59–80.

Glaeser, E. (2004). Housing Supply.. *NBER Reporter: Research Summary Spring.*

Heckman, J., Matzkin, R., & Nesheim, L. (2010). Nonparametric Identification and Estimation of Nonadditive Hedonic Models. *Econometrica*, *78(5)*, 1569–1591.

Kuminoff, N., Smith, V., & Timmins, C. (2013). The New Economics of Equilibrium Sorting and Its Transformational Role for Policy Evaluation. *Journal of Economic Literature*.

Landvoigt, D., Piazzesi, M., & Schneider, M. (2011). The Housing Market(s) of San Diego. Working Paper.

Mills, E. (1972). *Studies in the Structure of the Urban Economy*. John Hopkins Press. Baltimore.

Muth, R. (1960). The Demand for Non-Farm Housing. In Harberger, A. (Ed.), *The Demand for Durable Goods*. University of Chicago Press.

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

Nath, S. (2007). *Crime Pattern Detection Using Data Mining*. Springer.

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

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

Neill, D. (2006). *Detection of Spatial and Spatio-Temporal Clusters*. Ph.D. thesis, Carnegie Mellon University.

Poterba, J. (1984). Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach.. *Quarterly Journal of Economics*, *99*, 729–52.

Poterba, J. (1992). Taxation and Housing: Old Questions, New Answers. *American Economic Review*, *82 (2)*, 237–242.

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

Sommer, K., Sullivan, P., & Verbrugge, R. (2011). Run-up in the House Price-Rent Ratio: How Much Can Be Explained by Fundamentals?. BLS Working Papers.

Vianelli, S. (1983). The family of normal and lognormal distributions of order r. *Metron*, *41*, 3–10.

A Proofs

Proof 1 *The single-crossing condition implies that there is stratification of households by income in equilibrium. Stratification implies that there exists a distribution function for house values $G_t(v)$ such that:*

$$F_t(y) = G_t(v) \tag{36}$$

Hence there exists a monotonic mapping between income and housing value. If F_t is strictly monotonic, it can be inverted, and hence F_t^{-1} exists. Q.E.D.

Proof 2 *Equating the quantiles for income and value distributions, i.e. setting $F_t(y_t(v)) = G_t(v)$ for $y_t > \exp(\mu_t) - \beta_t$, and $v_t > \exp(\omega_t) - \theta_t$, yields:*

$$\frac{\int_0^{[(\ln(y_t+\beta_t)-\mu_t)/\sigma_t]^{r_t}/r_t} e^{-t} t^{1/r_t - 1} dt}{2r\Gamma(1+1/r_t)} = \frac{\int_0^{[(\ln(v_t+\theta_t)-\omega_t)/\tau_t]^{m_t}/m_t} e^{-t} t^{1/m_t - 1} dt}{2m\Gamma(1+1/m_t)} \tag{37}$$

Assuming $r_t = m_t$ in each period, the quantiles are equal when

$$\frac{\ln(y_t+\beta_t) - \mu_t}{\sigma_t} = \frac{\ln(v_t+\theta_t) - \omega_t}{\tau_t} \tag{38}$$

Similar steps lead to the same conclusion when $y_t < \exp(\mu_t) - \beta_t$, and $v_t < \exp(\omega_t) - \theta_t$. Solving (38) yields:

$$y_t = e^{(\mu_t - \frac{\sigma_t}{\tau_t}\omega_t)}(v_t + \theta_t)^{\frac{\sigma_t}{\tau_t}} - \beta_t \tag{39}$$

Q.E.D.

Proof 3 *The household's FOC is:*

$$\alpha u'(h) \cdot dh = \frac{dv}{(y_t - v_t - \kappa)} \tag{40}$$

Substituting the income loci (25):

$$\alpha u'(h) dh = \frac{dv}{A_t(v_t+\theta_t)^{b_t} - \beta_t - v_t - \kappa}$$

Since $\kappa = \theta_t - \beta_t \ \ \forall t$, the FOC becomes:

$$\alpha_i u_i'(h) dh = \frac{dv}{A_t(v_t+\theta_t)^{b_t} - (v_t+\theta_t)} \tag{41}$$

Integrating the right hand side yields:

$$\int \frac{dv}{A_t(v+\theta_t)^{b_t} - (v+\theta_t)} = \frac{1}{b_t-1}\left(\ln\left(-\frac{1}{A_t}\left(v_t+\theta_t - A_t\left(v+\theta_t\right)^{b_t}\right)\right) - b_t \ln v + \theta_t\right) + c_t$$

which implies:

$$\alpha u(h) = \frac{1}{b_t-1}\ln\left(1 - \frac{\left(v_t+\theta_t\right)^{1-b_t}}{A_t}\right) + c_t \tag{42}$$

Notice that integrating the left hand side recovers the original function $u(h)$. Using the utility function we get

$$\alpha \ln(1-\phi(h+\eta)^{\gamma}) = \frac{1}{b_t-1}\ln\left(1 - \frac{\left(v+\theta_t\right)^{1-b_i}}{A_t}\right) + c_t \tag{43}$$

Solving for v_t

$$(1-\phi(h+\eta)^{\gamma})^{\alpha(b_t-1)} = \left(1 - \frac{\left(v_t+\theta_t\right)^{1-b_t}}{A_t}\right) e^{c_t} \tag{44}$$

and hence

$$v_t = \left(A_t\left[1 - \frac{(1-\phi(h+\eta)^{\gamma})^{\alpha(b_t-1)}}{e^{c_t}}\right]\right)^{\frac{1}{1-b_t}} - \theta_t \tag{45}$$

Normalizing the constant of integration to $c = 0$ gives the result. Q.E.D.

Proof 4 *We can write the household's optimization problem as:*

$$\max_h \ \ u_1(h) + u_2(y - v(h)) \tag{46}$$

The FOC of this problem with respect to h is given by:

$$u_1'(h) - u_2'(y - v(h))\ v'(h) = 0 \tag{47}$$

Now define $h^* = v(h)$ and hence $h = v^{-1}(h^*)$. The decision problem associated with this model is then

$$\max_{h^*} \ \ u_1(v^{-1}(h^*)) + u_2(y - h^*) \tag{48}$$

and the FOC with respect to h^* is

$$u_1'(v^{-1}(h^*))\, v^{-1'}(h^*) - u_2'(y - h^*) = 0 \tag{49}$$

Now $h = v^{-1}(h^*) = v^{-1}(v(h))$ and hence $v^{-1'}(h^*)\, v'(h) = 1$. Hence we conclude that the two models are observationally equivalent. In the first case, we have non-linear pricing and in the second case we have linear pricing. Q.E.D.

Proof 5 *Recall from our discussion following Proposition 2 that parameters A_t, b_t, θ_t can be estimated directly from data for income and house rent distributions. We show these are sufficient for identification of the utility function parameters. First consider the normalization $v_t(h) = h$. Recall that the equilibrium hedonic pricing function is given by:*

$$v_t = \left(A_t \left[1 - [1 - \phi(h+\eta)^\gamma]^{\alpha(b_t - 1)}\right]\right)^{\frac{1}{1-b_t}} - \theta_t \tag{50}$$

Setting

$$\alpha = \frac{1}{b_t - 1} \tag{51}$$

implies

$$v_t = (A_t \left[1 - [1 - \phi(h+\eta)^\gamma]\right])^{\frac{1}{1-b_t}} - \theta_t \ = \ (A_t \phi (h+\eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{52}$$

Setting

$$\phi = \frac{1}{A_t} \tag{53}$$

implies

$$v_t = ((h+\eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{54}$$

Setting

$$\gamma = 1 - b_t \tag{55}$$

implies

$$v_t = (h + \eta) - \theta_t \tag{56}$$

Finally, setting

$$\eta = \theta_t \tag{57}$$

implies.

$$v_t = h \tag{58}$$

That establishes identification of the parameters of the utility function. The price equation in period $t + s$ *is then given by:*

$$v_{t+s}(h) = \left(A_{t+s} \left[1 - [1 - \phi(h + \eta)^{\gamma}]^{\alpha(b_{t+s}-1)} \right] \right)^{\frac{1}{1-b_{t+s}}} - \theta_{t+s} \tag{59}$$

The parameters of joint value and income distribution in period t *nail down the parameters of the utility function. The assumption of constant utility then imply that* $v_{t+s}(h)$ *is fully identified by the parameters* b_{t+s}*,* A_{t+s}*, and* θ_{t+s}*. Q.E.D.*

Proof 6 *The result follows from the discussion in the text.*

Proof 7 *Given our normalizations, we have also identified the housing supply function in the first period since* $R_1(h) = G_1(v)$ *which then identifies the density of housing quality in the first period* $q_1(h)$*.*

Proposition 5 implies that $v_2(h)$ *is identified. As a consequence* $G_2(v_2(h))$ *is identified. Proposition 6 implies that* $V_1(h)$ *and* $V_2(h)$ *are identified. As a consequence* ζ *is identified of the market clearing condition:*

$$R_2(h) = k_2 \int_0^h q_1(x) \left(\frac{V_2(x)}{V_1(x)} \right)^{\zeta} dx \tag{60}$$

Q.E.D.

Note that this proof generalizes for more complicated parametric forms of the supply function.

B The Generalized Lognormal Distribution with Location (GLN4)

The generalized lognormal distribution with location GLN4 pdf is given by:

$$f(y) = \frac{1}{2(x+\beta)r^{\frac{1}{r}}\sigma\Gamma\left(1+\frac{1}{r}\right)} e^{-\frac{1}{r\sigma^r}|\ln(x+\beta)-\mu|^r} \tag{61}$$

The CDF of the GNL4 distribution is given by:

$$F_t(y) = \begin{cases} \dfrac{\Gamma\left(\frac{1}{r}, B(y+\beta)\right)}{2\Gamma(\frac{1}{r})} & \text{for } y < \exp(\mu) - \beta, \\ \frac{1}{2} & \text{for } y = \exp(\mu) - \beta, \\ \frac{1}{2} + \dfrac{\gamma\left(\frac{1}{r}, M(y+\beta)\right)}{2\Gamma(\frac{1}{r})} & \text{for } y > \exp(\mu) - \beta. \end{cases} \tag{62}$$

where

- $B(y) = \dfrac{\left[\frac{\mu - log(y+\beta)}{\sigma}\right]^r}{r}$, $M(y) = \dfrac{\left[\frac{log(y+\beta)-\mu}{\sigma}\right]^r}{r}$,
  and
- $\Gamma(s,z) = \in_z^\infty e^{-t}t^{v-1}dt$, $\gamma(v,z) = \int_0^z e^{-t}t^{v-1}dt$
  are the incomplete gamma functions.

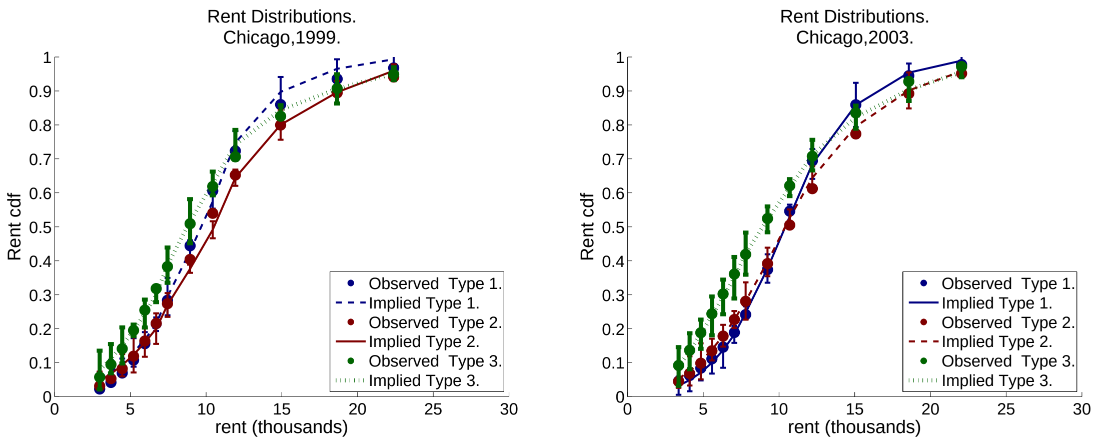

Figure 14: Fit of aggregate rent distributions with confidence intervals using types obtained through k-means clustering.

C Estimation Details: types defined by k-means clustering.

The estimated parameters imply a good fit of the observed distributions, which is an important part of our estimation approach. The bands show a 95% confidence interval in the estimation of each point in the distribution calculated with the standard errors obtained through parametric bootstrapping. Our discrete method estimates the price of each point in the partition of quality, and therefore provides different standard error for each price estimated. The following figures show the fit and other details of the estimation.

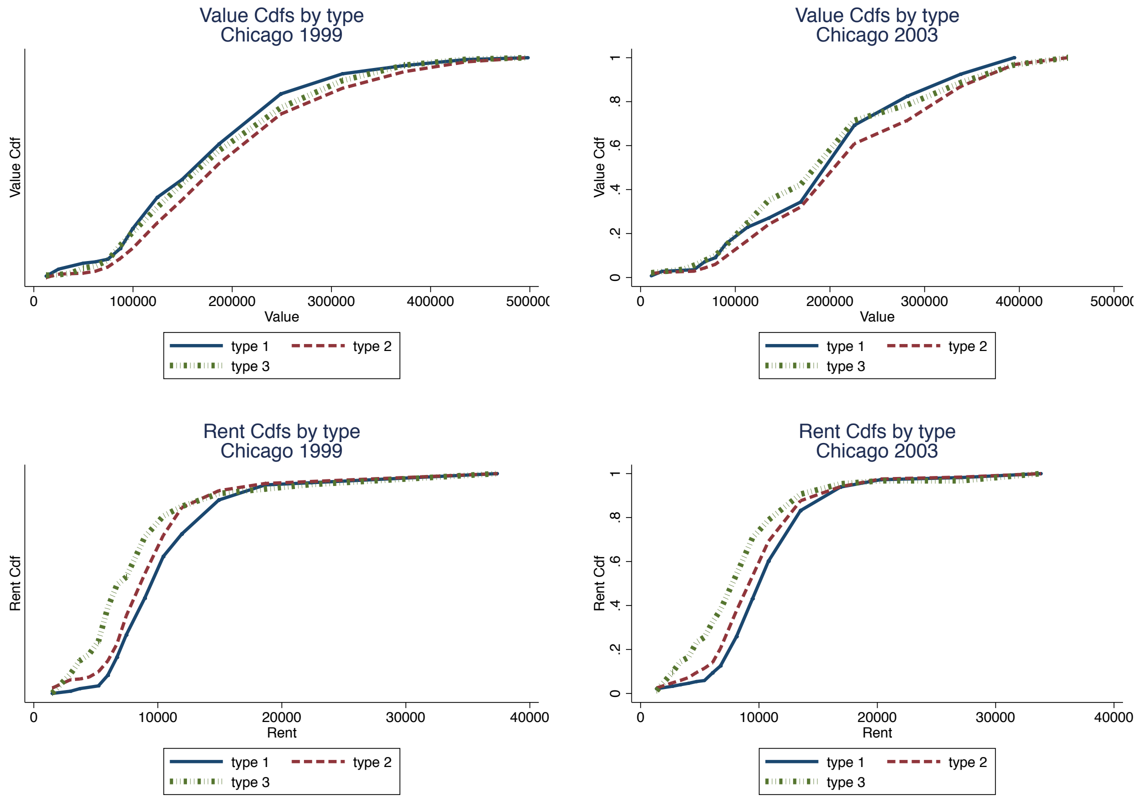

Figure 15: House values and rents distributions for each household type obtained through k-means clustering.

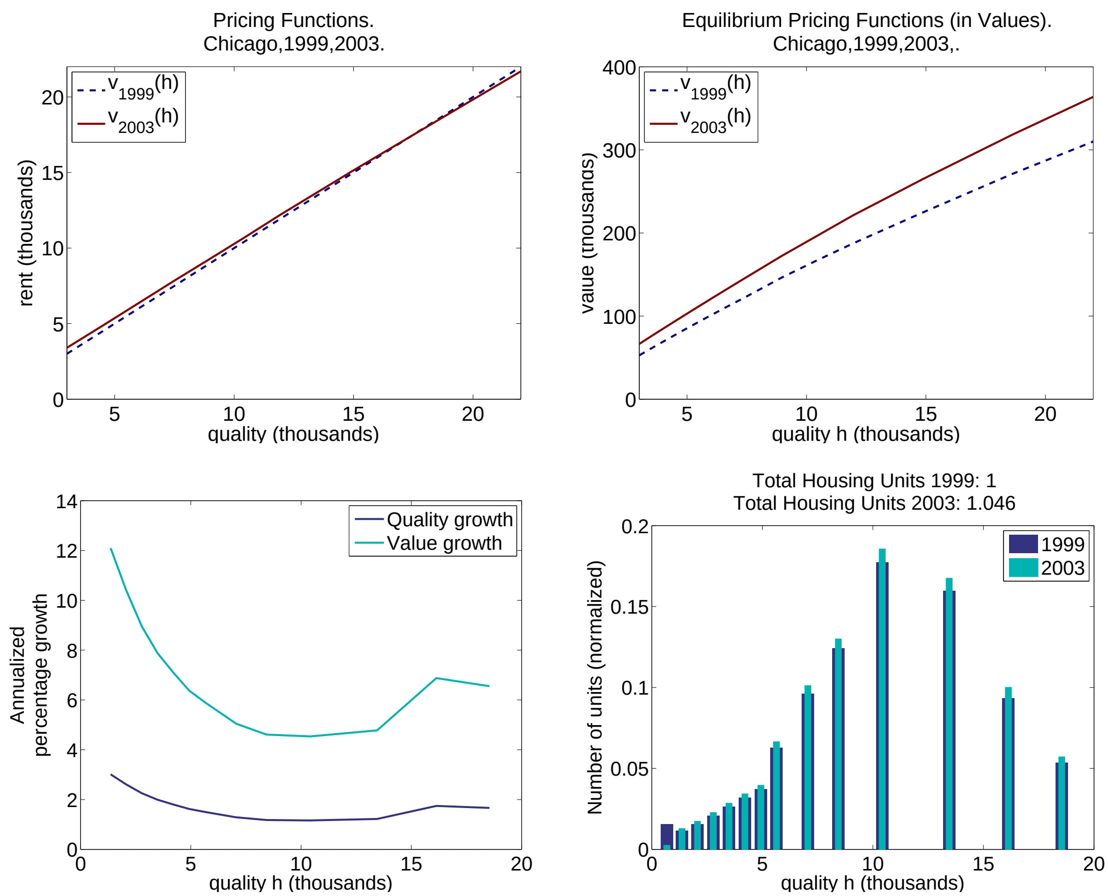

Figure 16: Estimated pricing functions, values and corresponding supply growth using types obtained through k-means clustering.