A New Approach to Estimating Hedonic Pricing Functions for Metropolitan Housing Markets*

Dennis Epple

Carnegie Mellon University and NBER

Luis Quintero

Johns Hopkins University

Holger Sieg

University of Pennsylvania and NBER

September 19, 2015

*We would like to thank Pat Bayer, Markus Berliant, Moshe Bushinsky, Morris Davis, Uli Doraszelski, Fernando Ferreira, Matt Kahn, Lars Nesheim, Henry Overman, Ariel Pakes, Monika Piazzesi, Martin Schneider, Paolo Somaini, Xun Tang, Frank Wolak 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

Two central challenges in estimating models of differentiated products are separating quality from price and separating observed components of quality from unobserved components. There are important applications where the latter decomposition is not needed, and it is then possible to undertake more general treatment of the former. We develop a method that allows researches to identify and estimate hedonic price functions for housing when quality is treated as latent. We illustrate the usefulness of our new technique using two applications. First we estimate the model jointly for the Chicago and New York metropolitan areas, obtaining estimated hedonic price functions for the two metropolitan areas using a common latent quality scale. This permits comparison of housing prices for each quality level between the two metropolitan areas, comparison of housing stocks of each quality level, and calculation of differences in agglomeration economies between the two metro areas. Second, we gain new insights into the causes and effects of the recent housing market crisis focusing on Miami. Using data for three time periods, we estimate changes in rents and prices across the quality spectrum during the dramatic run-up in housing prices in the years preceding the financial crisis.

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

1 Introduction

Two central challenges in estimating models of differentiated products are separating quality from price, and separating observed components of quality from unobserved components. A major focus of the hedonic literature has been estimation of models of differentiated products with the objective of delineating the demand and supply of individual product characteristics. The pioneering work of Rosen (1974) transformed this literature and inspired extensive research focused on applications and associated issues of identification and estimation.[1]

The typical hedonic approach utilizes observations for product prices and a vector of product characteristics available for one or more markets. The objective in implementation is then to develop identification and estimation procedures that yield consistent parameter estimates taking account of the potential presence of unobserved product characteristics. This is in practice quite challenging. In many applications, study of product characteristics is the central objective. For such problems, the challenge of separating observed and unobserved components cannot be sidestepped. There are other important applications, however, where the delineation of the role of individual product characteristics is not needed. For such applications, avoiding the complexities of separating the roles of observed and unobserved characteristics permits a less restrictive approach to estimation of the hedonic price function.

The key to this alternative approach is treatment of quality as latent. We develop this new strategy and demonstrate its value by applications to metropolitan-wide housing markets. In our application to housing markets, we make additional innovations. We provide an integrated treatment of rental rates and asset (property) values.

[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).

We show how to estimate the nonlinear equilibrium pricing functions for both rentals and property values. Key elements of our approach are the following. We develop a non-parametric matching approach to identify the implicit rent-to-value ratio for owner-occupied houses as a function of latent quality. 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 terms to incorporate not only structural housing characteristics per se, but also all publicly provided amenities as well as natural amenities. Finally, we also show how to estimate period-to-period changes in housing supply as a function of quality. Implementation of our new 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.[2] 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, such as modeling changes in the stocks and prices of houses with varying attribute bundles. In exchange for our simplifying assumption, we gain a tractable framework that is ideally suited to study

[2]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).

the distributions of quality and price in metropolitan housing markets, comparisons across metropolitan areas, and changes over time within and across markets.[3]

We show that there exists a new flexible parametrization of Rosen's model that exploits generalized forms of the log-normal distributions proposed by Vianelli (1983). This approach yields a closed-form solution for the equilibrium hedonic pricing function. The baseline specification incorporates variation in income across households that have a common preference function. We then generalize to consider multiple household types, with the income distribution and the preference function varying across types. An innovation of our approach for this multiple-type model is use of k-means clustering to group households into types.

When quality is latent there is an obvious identification issue 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 model 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]

[3]Ekeland, Heckman, and Nesheim (2004) establish non-parametric identification using data for a single market when marginal utility and marginal product functions are additive. Their paper demonstrates the payoff from exploiting all equilibrium implications of the hedonic structure. We follow their lead, exploiting all equilibrium information in identification and estimation of our model. Heckman, Matzkin, and Nesheim (2010) extend their analysis of non-parametric identification to non-additive models utilizing a unidimensional quality scale with multidimensional types. Analogously, we use a unidimensional index of housing quality with multiple household types. We adopt a flexible parametric specification of preferences and distribution of household characteristics, but an interesting question for future work is non-parametric identification of our latent-quality framework along the lines of Hechman, Matzkin, and Nesheim (2010).

[4]Exploiting variations among multiple markets is also a useful strategy to obtain identification if

In addition, we must overcome one more identification problem. 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. Empirical hedonic studies typically resolve this problem by either only focusing on owner occupied housing or using a common estimate of the user cost that does not depend on quality. The former approach ignores rentals while the later approach may be inconsistent with the data. In both of our applications we can reject the null hypothesis that the user cost does not depend on 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.

Our empirical findings provide valuable new insights into metropolitan housing markets. We provide two applications. We estimate the model jointly using data for two large metropolitan areas, Chicago and New York. 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

---

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).

price a household of a given income would obtain in a larger relative to a smaller metropolitan area. We also compare compensating variations as a function of income across household types. These measures are of interest in their own right and also provide valuable insights into agglomeration economies. 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 compensating variation of approximately 20% of income is required to induce that household to move to New York.

In our second application, we study changes in price across the quality distribution in Miami which 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]

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,

[5]This part of our analysis build on the work of Landvoigt, Piazzesi, and Schneider (2015) who employ a latent unidimensional housing quality scale and the approach of equating CDF's of housing demand and supply to impose market clearing for each house quality type.

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. 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.[6]

Assumption 1 *Investors discount housing assets at a rate that reflects the perceived financial market risk of housing assets.*

Let the one-period risk-adjusted interest rate be 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 *The market for housing assets is 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:

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

[6] Our approach is closely related to that of Poterba (1984), Poterba (1992) and Poterba and Sinai (2008).

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.[7]

In equilibrium, expected profits for investors must be equal to zero. Hence housing values or asset prices must satisfy the following no-arbitrage condition:

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

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

$$V_t(h) = v_t(h) + \frac{(1 - \tau_{t+1}^p - \delta_{t+1})}{(1 + i_t)} E_t \left[V_{t+1}(h) \right] \tag{3}$$

By successive forward substitution of the preceding, we obtain:

$$V_t(h) = v_t(h) \ + \ E_t \sum_{j=1}^{\infty} \ \beta_{t+j} \ v_{t+j}(h) \tag{4}$$

where the stochastic discount factor is given by:

$$\beta_{t+j} = \prod_{k=1}^{j} \frac{(1 - \tau_{t+k}^p - \delta_{t+k})}{(1 + i_{t+k-1})} \tag{5}$$

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

$$\tilde{\beta}_{t+j}(h) = \prod_{k=1}^{j} \frac{(1 - \tau_{t+k}^p - \delta_{t+k}) \ (1 + \pi_{t+k}(h))}{(1 + i_{t+k-1})} \tag{6}$$

Then:

$$V_t(h) = \frac{v_t(h)}{c_t(h)} \tag{7}$$

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

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

$$c_t(h) = \frac{1}{1 + E_t \sum_{j=1}^{\infty} \ \tilde{\beta}_{t+j}(h)} \tag{8}$$

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

$$E_t \prod_{k=1}^{j} \frac{(1 - \tau^p_{t+k} - \delta_{t+k})(1 + \pi_{t+k}(h))}{(1 + i_{t+k-1})} = \left[\frac{(1 - \tau^p - \delta)(1 + \pi(h))}{1 + i}\right]^j \tag{9}$$

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 assume that investors have correct expectations about housing rental appreciation. There may be time periods, for example, where 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. We follow the hedonic literature in allowing 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 model, owner-occupants households make decisions about housing consumption using an implicit rental that equals the amount the dwelling would command on the rental market. Hence, for simplicity in the presentation in this section, we refer to all housing consumers as renters.

In our baseline model, renters differ in income denoted by y. We then 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. We show that key results go through in the more general class of models discussed below.

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, a household can costlessly change its housing consumption on a period-to-period basis as rental rates change. It follows that a household's optimal choice of housing at each date t maximizes its period utility at date t:

$$\begin{aligned} & \max_{h_t, b_t} \; U_t(h_t, c_t) \\ s.t. \; y_t &= v_t(h_t) + c_t \end{aligned} \tag{10}$$

where c_t denotes expenditures on a composite good.

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

$$m_t(h_t, y_t - v_t) \equiv \frac{U_h(h_t, y_t - v_t)}{U_c(h_t, y_t - v_t)} = v_t'(h_t) \tag{11}$$

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

$$H_t^d(h|v_t(h)) = \int_0^\infty 1\{h_t(y, v_t(h)) \le h\} \, dF_t(y) \tag{12}$$

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:*

$$\frac{\partial m_t}{\partial y}\Big|_{U_t(h,y-v(h))=\bar{U}} > 0 \tag{13}$$

Assumption 3 states that high-income households are willing to pay more for a higher quality house than low-income households – 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*

$$y_t(v) \;=\; F_t^{-1}(G_t(v)) \tag{14}$$

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:

$$q_t(h) = s(q_{t-1}(h), V_t(h), V_{t-1}(h)) \tag{15}$$

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:*

$$q_t(h) = \frac{1}{k_t}\, q_{t-1}(h) \left(\frac{V_t(h)}{V_{t-1}(h)}\right)^{\zeta} \tag{16}$$

where

$$k_t = \int_0^{\infty} q_{t-1}(h) \left(\frac{V_t(h)}{V_{t-1}(h)} \right)^{\zeta} dh \tag{17}$$

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 (16) 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:

$$G_1(v_1(h)) = R_1(h) \tag{18}$$

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

$$R_t(h) = \int_0^{h} k_t \; q_{t-1}(x) \left(\frac{V_t(x)}{V_{t-1}(x)} \right)^{\zeta} dx \tag{19}$$

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:

$$G_t(v_t(h)) = R_t(h) \tag{20}$$

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

$$n_t^y(y) = N_t f_t(y) \tag{21}$$

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

$$n_t^v(v) = N_t g_t(v) \tag{22}$$

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:

$$N_t F_t(y) = N_t G_t(v) \tag{23}$$

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) Households behave optimally given the price function;
b) Housing markets clear, i.e. for each level of housing quality h, we have:

$$H_t^d(h|\, v_t(h)) = R_t(h) \tag{24}$$

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).*[8]

$$\begin{aligned} \ln(y_t) &\sim GLN4(\mu_t, \sigma_t^{r_t}, \beta_t) \\ \ln(v_t) &\sim GLN4(\omega_t, \tau_t^{m_t}, \theta_t) \end{aligned} \tag{25}$$

These functions are sufficiently flexible to fit the housing value and income distributions in the metro areas and 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

[8]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

$\theta_t - \beta_t$ is time invariant permits us to obtain a closed-form solution to the hedonic price function.[9]

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

$$y_t \quad = \quad A_t \, (v_t + \theta_t)^{b_t} - \beta_t \tag{26}$$

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

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:*

$$U = c_t(h) + \frac{1}{\alpha} \ln(y_t - v_t(h) - \kappa) \tag{27}$$

with $c_t(h) = \ln(1 - \phi(h + \eta)^{\gamma})$, where $\alpha > 0$, $\gamma < 0$, $\phi > 0$, and $\eta > 0$.[10]

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.[11] Given this parametric specification of the utility function, we have the following result:

[9]We impose both of these restrictions when estimating our model. We later estimate a discretized version without these restrictions to establish that these restrictions do not impair fit in the context of our applications.

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

[11]See Appendix C for details.

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:*

$$v_t(h) \;\; = \;\; \left(A_t \left[1 - (1 - \phi(h+\eta)^{\gamma})^{\alpha(b_t - 1)}\right]\right)^{\frac{1}{1-b_t}} - \theta_t \tag{28}$$

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 additional sources 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:

$$U_i(h, b) = \ln(1 - \phi_i(h + \eta_i)^{\gamma_i}) + \frac{1}{\alpha_i}\ln(b - \kappa_i) \tag{29}$$

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. With this generalization to multiple types, analytical solutions for the equilibrium price functions are not available, and we rely on numerical solution methods. Anticipating the need to rely on numerical solution algorithms, we develop the extension of our model using a discretized approximation of the housing stock.[12]

Given a grid of values $(h_1, ..., h_J)$, we use discrete distributions to approximate the continuous distributions characterizing housing demand and supply. We index the pricing function accordingly and let $v_{jt} = v_t(h_j)$. There will be, for each type, an income that is indifferent between each "adjacent" pair of housing qualities. These cut-off incomes $\hat{y}_{i,j,t}$ satisfy

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

As a consequence all households of type i with $\hat{y}_{i,j-1,t} \leq \hat{y}_{i,j,t}$ will consume housing quality j assuming that preferences for each type satisfy a single-crossing condition. 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:

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

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:

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

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

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

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

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

3 Identification

We consider identification of the model assuming that a) we have access to data for one market that is observed for more than one time period; 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 intrinsic unit of measurement for housing quality. The implications of the latent quality measure for identification are 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 $v(h)$. Suppose now that we have data for more than one time period in a market. A corollary of Proposition 4 is that we can normalize housing quality by setting $h = v_t(h)$ in one time 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 period preferences are invariant over time, this normalization then suffices to identify the price functions in all other time periods.

Assumption 7 *The utility function is invariant across period.*

Assumption 7 implies the following result.

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

Broadly speaking, the parameters of the utility function are identified from 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. The same argument applies to establish identification from cross-sectional data for two or more geographically distinct markets.

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.

As discussed previously, owner-occupants make their housing consumption decisions, and hence purchase decisions, based on implicit rent that corresponds to the market rent a dwelling would command. Hence households with income y consume the same quality of housing independently whether they live in a rental unit, for which we observe, $v_t(y)$, or 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.

Having identified the rent-to-value function, it is straight forward to identify 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 or two geographically distinct markets.*

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 demand depends on a vector of observed state variables.[13]

It remains 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 market clearing conditions for the J qualities in

[13] 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.

the baseline period. 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. For simplicity of exposition, we discuss estimation based on data for multiple time periods, but the logic applies equally if estimation is for multiple geographically distinct markets, or both. 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. 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:

$$L^N(\psi) = (1-W)\ (l_y^N(\psi) + l_r^N(\psi))\ +\ W\ l_h(\psi) \tag{36}$$

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.[14] We use a standard bootstrap procedure to estimate the standard errors.

This estimator can be extended to estimate the model with multiple observed types. There are two differences. First, we need to estimate 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 estimation algorithm follows the steps discussed above for the simpler model.

5 Empirical Results

We obtained data from the American Housing Survey, the most comprehensive national housing survey in the United States. There is a national and a metropolitan version, and, in selected years, also an extended metropolitan component for some

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

metropolitan areas in the national version. 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 re-surveyed. 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.[15]

Fortunately, the AHS conducted surveys in both Chicago and New York for both 1999 and 2003. We exploit data from these surveys for our first application, jointly estimating the model for those two metropolitan areas. Our second application exploits three successive surveys (1995, 2002, 2007) for Miami to study the period preceding the housing bubble and period of the housing bubble in Miami.

5.1 The Housing Markets of Chicago and New York

One of the most advantageous features of our model is its capacity to separate quality from price by identifying the prices for different levels of the quality distribution for each market at each point in time. One very interesting application is comparison across metropolitan areas over time. Fortunately, the AHS provides data for concurrent years, 1999 and 2003, for two major metropolitan areas, Chicago and New York. Hence, for our first application, we estimate our model for these two periods for these two metropolitan areas. AHS definitions of each of these metropolitan areas

[15] 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 are weighted because there is incomplete sampling lists and non response. The weights are 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.

is unchanged across these two periods.[16] We use New York metropolitan area in 1999 as the base for our normalization. As a shorthand, we will sometimes refer to the two metropolitan areas as CMA and NYMA. We present results for our continuous model with one household type and also for the discretized model with multiple household types.

Two key steps set the stage for estimation of parameters of our model. One is presentation of our findings defining implicit rents. The other is presentation of our analysis of classification of households by type for the multiple-type model. We then obtain the structural parameters using an estimator that includes orthogonality conditions for predicted and observed income and rent percentiles for each metropolitan area.[17]

We start with our estimates of the user-cost mapping between rents and values.[18] We find that the estimated user cost ranges between 0.058 and 0.061 for 1999, and between 0.051 and 0.062 for 2003. In subsequent discussion, we will use the term rentals to encompass rental payments for rental units and implicit rentals for owner-occupied units.

For implementation of our multiple-type model, we reduce the dimensionality of potential household types using k-means clustering. This is a standard method in data mining.[19] The method partitions the points in a multidimensional data matrix

[16]The Chicago metropolitan area is defined by the Census in 1999 and 2003 to consist of the following counties: Cook, Du Page, Kane, Lake, McHenry, and Will. The New York metropolitan area is comprised of Bronx, Kings, Nassau, New York, Putnam, Orange, Queens, Richmond, Rockland, Westchester, and Suffolk counties.

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

[18]We estimate this equivalence by calculating the mean rents and values by income. These means give us the user cost $u(h)$. Our sample sizes within type are not sufficiently large to estimate user cost by income for different household types.

[19]Neill (2006) provides a clear treatment of this topic. Nath (2007) shows an interesting application

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. We focus on capturing the role of two important variables that influence the housing consumption decisions of households: age and number of children. Households are clustered with respect to the share of income they spend on rent.

Table 1: k-means clustering centroids.

		Cluster	# Children	Age	Share of Population
Chicago	**1999**	1	0.29	29.03	0.261
		2	1.45	45.60	0.465
		3	0.24	72.31	0.273
	2003	1	0.29	29.19	0.246
		2	1.43	45.75	0.484
		3	0.269	73.10	0.269
		Cluster	# Children	Age	Share of Population
NYC	**1999**	1	0.35	27.33	0.251
		2	1.77	46.54	0.486
		3	0.45	67.43	0.262
	2003	1	0.45	28.15	0.268
		2	1.99	48.11	0.466
		3	0.30	65.12	0.264

We obtain three clusters, which is the optimal number of clusters calculated with our sample data.[20] Table 1 shows estimated cluster shares and centroids for CMA

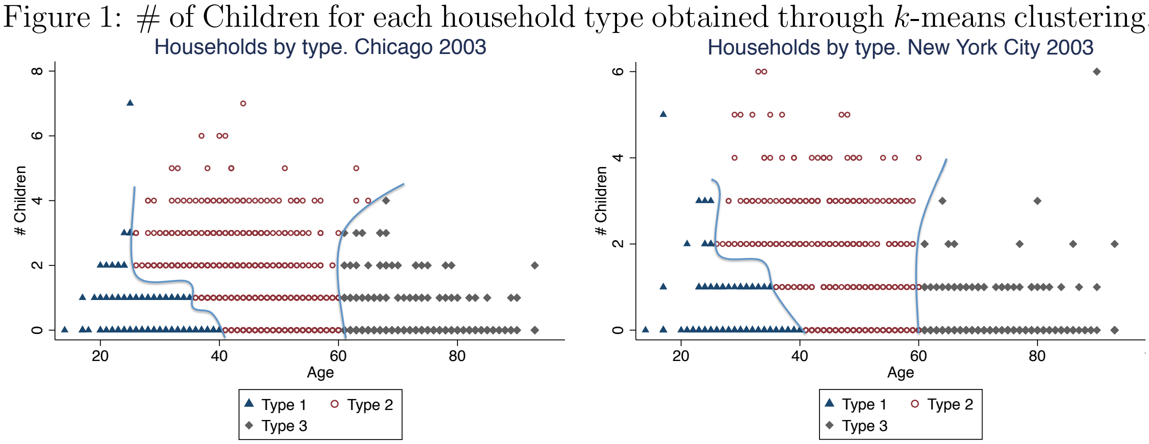
of this method to detection of crime hot spots.

[20]We used NYMA in 2003 for the cluster analysis, employing the silhouette criterion in determining the optimal number of clusters. The silhouette of a data point is a measure of its relative match to

and NYMA respectively for 1999 and 2003. As we will see, our estimation supports the claim that these types have significantly different preferences. An intuitive interpretation of the three groups is the following: Type 1 is primarily comprised of young households with few or no children, Type 2 is primarily comprised of middle aged households with more than one child, and Type 3 is comprised of older households with no children residing in the household. As can be seen in Table 1, the share of households by type is remarkably similar across the two metropolitan areas and the two time periods. The mean number of children is modestly higher for each type in NYMA relative to CMA. Mean age in the third group is somewhat higher in CMA than in NYMA. Overall, however, means by age and by number of children within each type are quite similar across metropolitan ares and across periods.

Further information about the clusters is provided in Figure 1 which shows the distribution of number of children by age of household head in the two metropolitan areas in 2003. To help in reading the figure, we have included curves that approximately demarcate the three household types.

Figure 1: # of Children for each household type obtained through k-means clustering.



Income distributions by type for 2003 in the two metropolitan areas are shown in the top panel of Figure 2; rental distributions are shown in the bottom panel. The

its cluster relative to alternative clusters.

distributions for income show that the Type 3 households tend to be poorest among the three groups in both metropolitan areas. Likewise, the rental distributions show that Type 3 households tend to have lower rental expenditures than the other two types. There is little difference in income distributions between Types 1 and 2 in NYMA whereas households Type 2 households tend to be more well-to-do in CMA. In NYMA, Type 1 households tend to have the highest rental expenditures. In CMA, Type 1 households have the highest rental expenditures over the lower half of the income range while Type 2 households have highest expenditures over the upper half of the income rang.

Figure 2: Observed distributions for each household type obtained through k-means clustering.

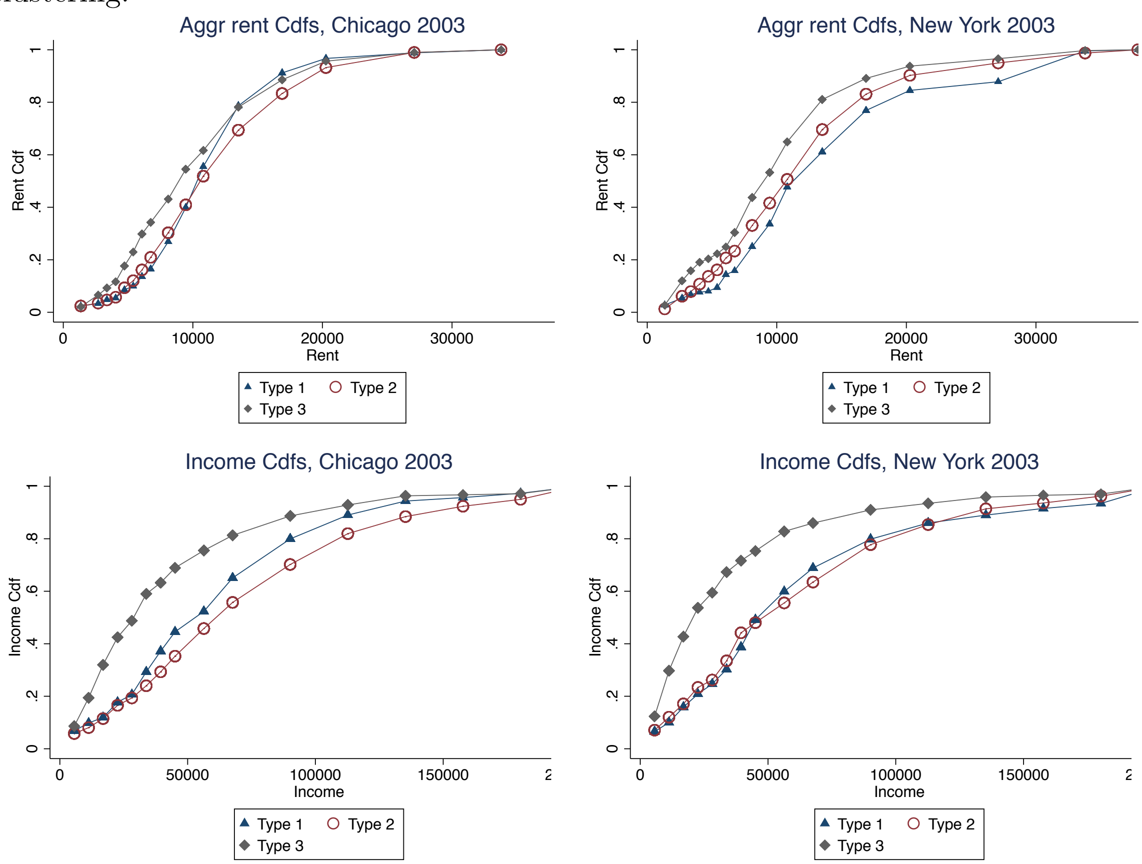
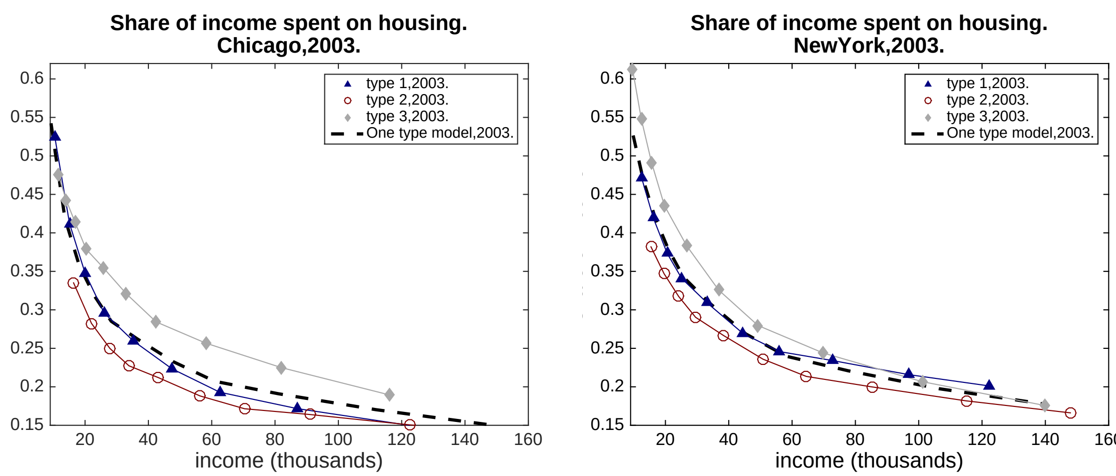


Figure 3 provides rental expenditure shares as a function of income in 2003 for CMA and NYMA. The ordering of expenditure shares across types is the same in both metro areas. At each income level, the ordering of types from highest to lowest expenditure is Type 3, Type 1, and Type 2. Type 3 households spend a higher share of income on rents at each level income than the other two household types. [21] At high income levels, Type 1 and Type 2 households spend similar shares on housing. Type 2 households spend a somewhat smaller share than Type 1 at each income level, perhaps sacrificing housing quality in order to provide for other needs of their children.

Overall Figures 2 and 3 demonstrate that the clustering algorithm has identified types with differing income distributions and differing housing expenditure patterns.

Figure 3: Share of income spent on housing for each household type obtained through k-means clustering.



The estimates for the preference and supply functions are reported in Table 1. The

[21]The AHS income measure is designed to include all wage and non-wage income. Research by Susin (2003) shows that the AHS estimate of non-wage income is lower than the CPS measure. Hence, the higher expenditure share of older (i.e., Type 3) households may, in part, reflect underestimate of non-wage income.

first row of Table 1 shows results for our baseline one-type model. The remaining three rows show results from the three-type model. All parameters have the correct algebraic signs and bootstrapped standard errors show the estimates to be highly significant. The implications of our preference parameter estimates are best illustrated by a few select figures.

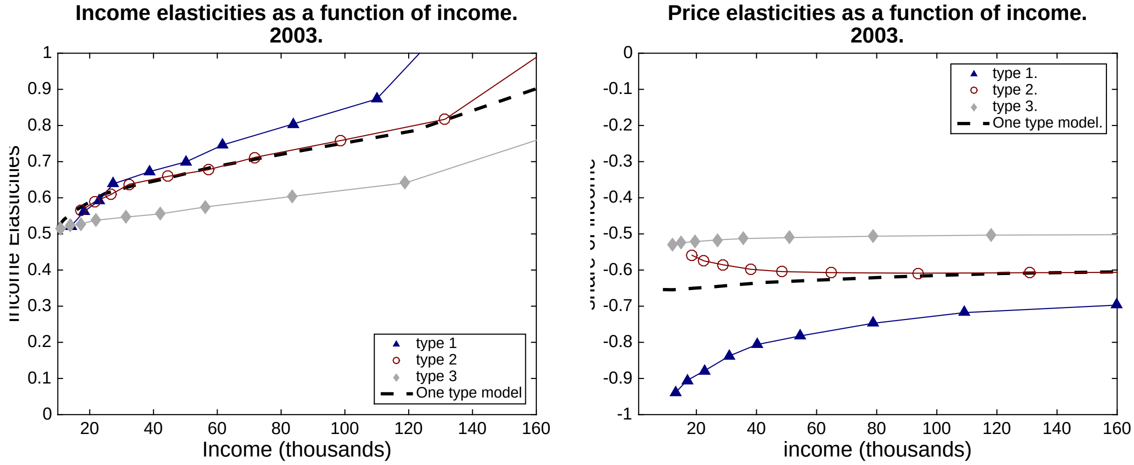
Table 2: Joint estimates 1999-2003 for New York City and Chicago.

Type	α	ϕ	η	γ	ζ
Baseline	1.399	2.921	9.111	-0.880	0.055
	(0.031)	(0.578)	(2.104)	(0.199)	(0.019)
Type 1	3.123	2.733	4.554	-0.749	
	(0.012)	(0.451)	(0.023)	(0.832)	
Type 2	1.332	4.576	14.675	-0.799	0.079
	(0.013)	(0.231)	(2.235)	(0.140)	(0.009)
Type 3	0.104	2.018	0.1779	-1.030	
	(0.098)	(0.143)	(0.234)	(0.056)	

Price and income elasticities implied by our preference parameters estimates are plotted as a function of income in Figure 4.[22] The income elasticities have the same order across the income range with Type 1 having the highest elasticity at each income level and and Type 3 the lowest. Type 3 households have the lowest price elasticity, with the elasticity being approximately -.5 throughout the range of income. The elasticity for Type 2 households is also relatively constant across the income range at approximately -.6. Type 1 households exhibit the greatest sensitivity to price, especially at the lower range of incomes. The price elasticity for Type 1 households increases from approximately -1 at low income levels to -.8 at high income levels. As

[22]Recall from our discussion of equations 66 and 67 the assumptions implicit in the derivation of elasticities.

Figure 4: Income and Price Elasticities.



one would expect, for each income level, the elasticities from the one-type model lie near the middle of the elasticities of the three-type model.

The right-most column of Table 2 shows the supply elasticity estimates. Our estimate for the annual supply elasticity from the model with three household types is 0.079. The estimated elasticity from the one-type model is somewhat lower, .055. Recall that the changes in supply stock of a certain quality over time depend on changes in values and the estimated elasticity through the supply equation (17). The implied supply growth of quality from our estimates are between 4.30% and 6.82% for the 4 year period, which correspond to average annualized changes of approximately 1.1% and 1.7% respectively. The estimated annual growth in total number of units was 5.2% per year over the 4 year period, with the largest number of additional units being created in the qualities located around the middle of the distribution.[23] Our supply elasticity estimates are along the lines of the estimates of supply elasticities summarized in Glaeser (2004).

We next illustrate fit of our model to the data. In the interest of space, we

[23]The reported numbers correspond to Chicago

illustrate for our one-type model. Figure 5 shows the fit of the one-type model to the income and rent distributions. The left two graphs in Panel A show the fit to income distributions in NYMA in 1999 and 2003 respectively, while the right two graphs show fit to 1999 and 2003 NYMA rental distributions. The corresponding graphs for Chicago are shown in Panel B. These graphs illustrate that the fit to the income and rent data is quite good in both metropolitan areas in both time periods.

Figure 6 illustrate the resulting equilibration of supply and demand for each quality level in the two time periods in the two metropolitan areas. The upper pair of graphs are for CMA in 1999 and 2003 and the lower pair for NYMA in 1999 and 2003. As these graphs illustrate, our approach results in close correspondence of supply and demand over the quality range in both metro areas in both time periods.

We next turn to a presentation of the implications of our model. In particular, we compare hedonic price functions for the two metropolitan areas for the two time periods; we compare estimated housing stocks for each housing quality level in the two metropolitan areas, and we provide an analysis of differential agglomeration economies across the two metropolitan areas.

Figure 7 shows the price functions in 2003 for CMA and NYMA from both the one-type model and the three-type model. The steeper curves show the hedonic price functions for 2003 for NYMA from the two alternative models while the shallower curves show the estimates for CMA. In each metropolitan area, the estimates from the two alternative models are strikingly close to each other.

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

Figure 5: Goodness of Fit

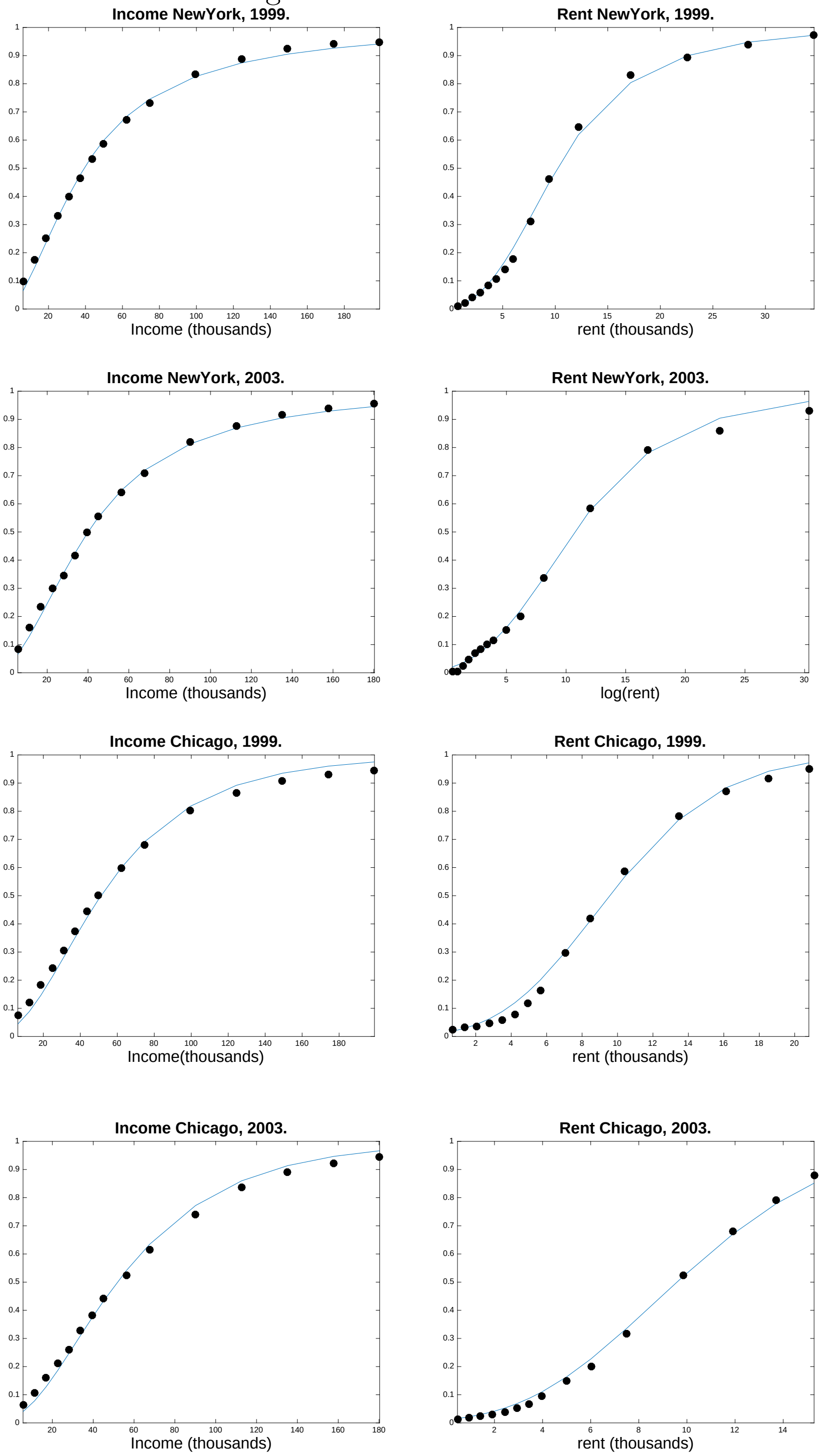


Figure 6: Supply and Demand Equilibrium

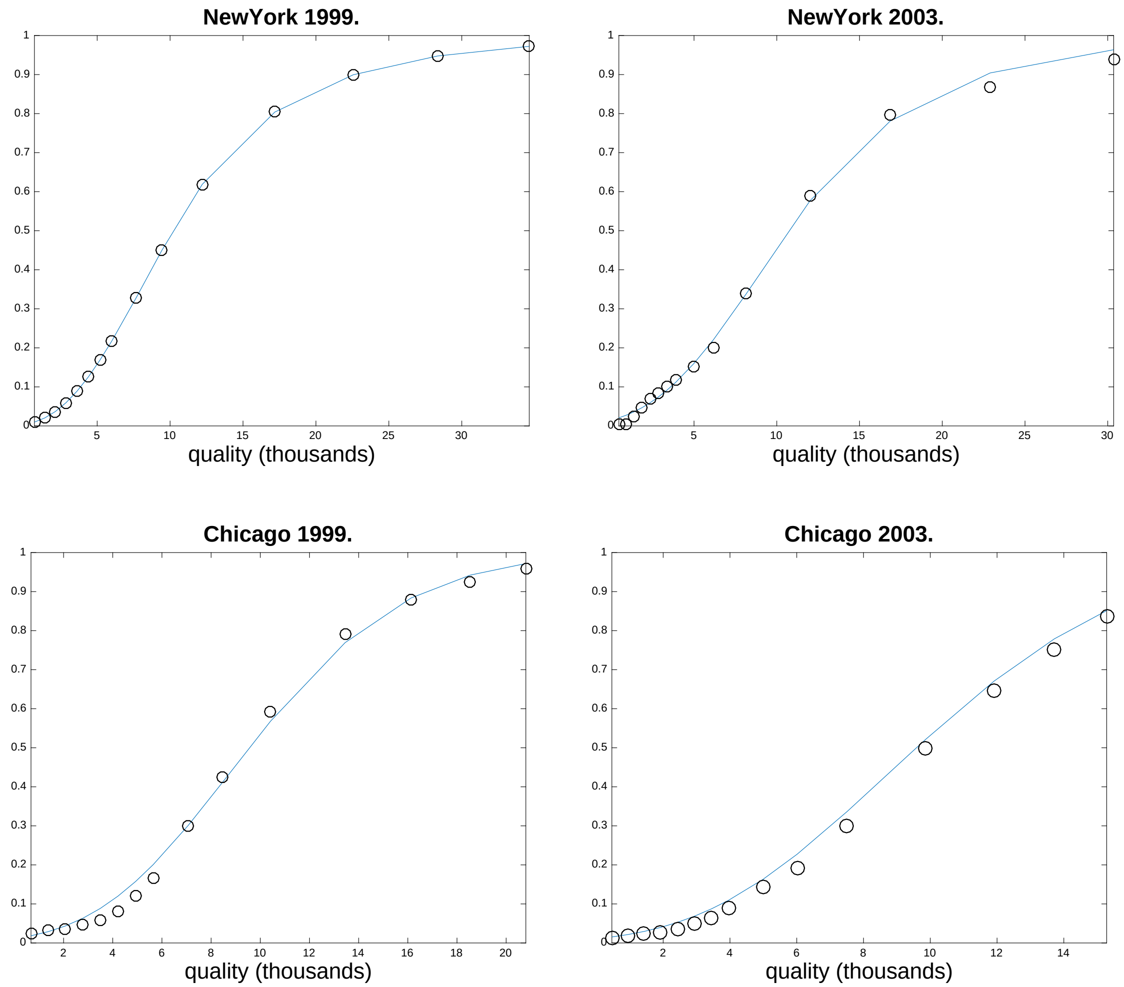
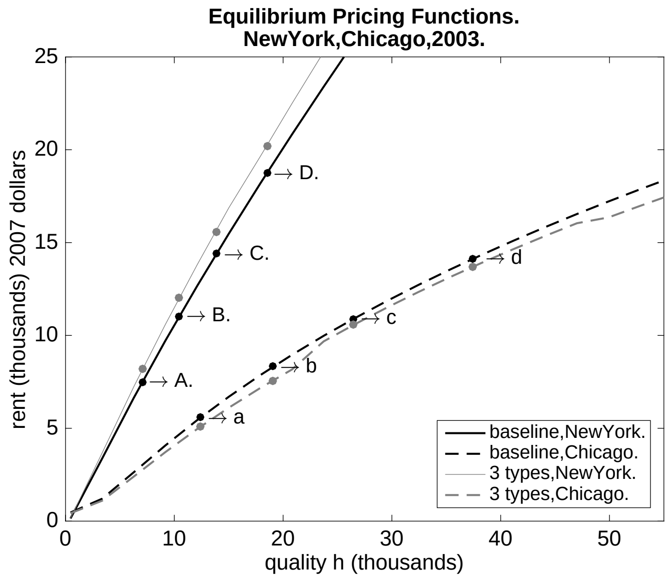


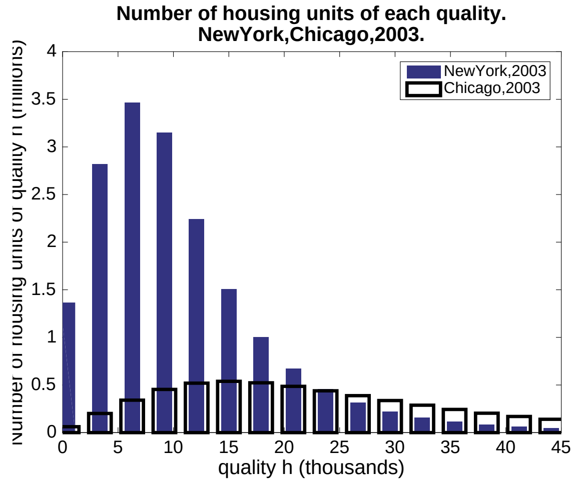
Figure 7: Price Functions



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. This is an important feature of our modeling approach.

We next turn to a comparison of the distributions of stocks of housing by quality in the two metropolitan areas. The distributions of housing by quality are shown in Figure 8. As we have just seen, at every income level, a household in NYMA consumes lower quality than the corresponding household in CMA. The effect of this difference in consumption levels is to shift the distribution of quality in NYMA to the left relative to that in CMA. This effect is augmented to some extent by differences in the income distributions in the two metro areas. The CDF for income for the Chicago metro area is shifted to the right relative to the New York metro area, i.e., CMA incomes tend to be higher. The relatively higher concentration of low-income

Figure 8: Quality Distributions



households in New York accentuates the leftward shift in the quality distribution in New York relative to Chicago. CMA has *relatively* more high quality housing. Given its much higher population, however, NYMA has a larger number of housing units at almost at all quality levels than CMA.

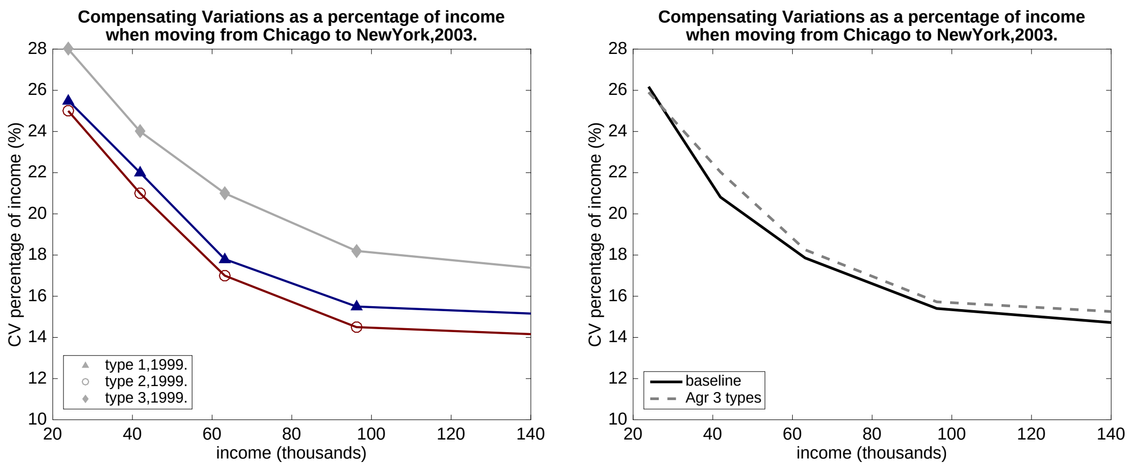
Our framework provides a new approach for measuring agglomeration economies.[24] The logic of our approach is the following. Housing at each quality level in the New York metro area is more expensive than in Chicago. To be equally well off in the two metro areas, a given household must then earn more in New York metro area than in the Chicago metro area. Hence, the compensation required in New York metro area for a household to be as well off as its counterpart in Chicago metro area is a measure of the additional earnings required in New York metro area.

In the top panel of Figure 9, we plot for each household type for each income the compensating variation (CV) that would be required for that type and income

[24] Rosenthal and Strange (2003) provide an in-depth analysis of the spatial and organizational features of agglomeration economies and a discussion of alternative approaches to measuring agglomeration economies.

in Chicago to be equally well off in New York. For all three types, CV declines with income, reflecting the declining share of income spent on housing as income rises for all three types. We see that, across the income range, the highest compensation would be required for Type 3 households. This reflects the higher propensity of Type 3 to consume housing and the relatively inelastic response of Type 3 to price. Types 1 and 2 have similar CV values at each income level, with Type 1 being slightly higher than Type 2 at all incomes.

Figure 9: Compensating Variations



In the right panel of Figure 9, we plot the CV calculated from the one-type model and the CV aggregated across types from the three-type model. The estimates are remarkably close over the entire income range. For a household earning \$24,000 in CMA (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. Productivity, and hence earnings, in NYMA would need to be higher by these amounts to compensate a household for the differences in housing price functions between the two metropolitan areas.

5.2 The Housing Market Bubble: The Case of Miami

As a further application of the one type model, we study changes in prices and rents across the quality dimension in Miami during the period leading up to the financial crisis. We take advantage of three successive AHS surveys of the Miami (FL) metropolitan area in 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.[25]

First, we estimate the rent-to-value functions using our non-parametric matching estimator for the three periods in our data set.[26] Figure 10 plots the estimated functions for the three time periods.

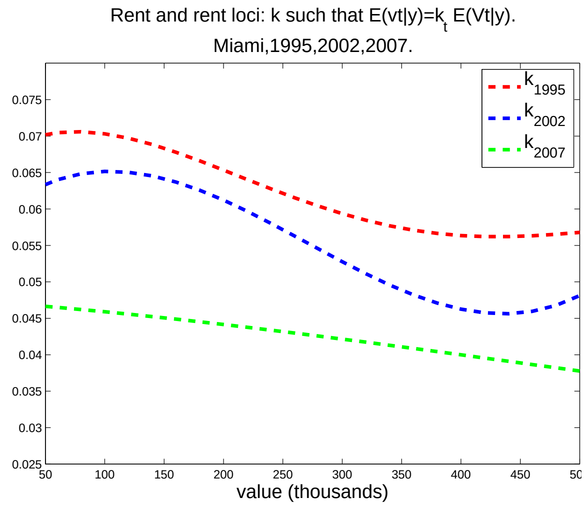
We find that the rent-to-value ratio ranged between 0.07 and 0.06 in 1995. The function showed a relatively modest decline between 1995 and 2002, with somewhat larger decreases at the low and high ends of the quality distribution. In contrast, we see large changes in the rent-value function during the bubble period between 2002 and 2007. The range of the function is from 0.035 to 0.046. We clearly reject the null hypothesis that the user costs do not depend on quality.

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 somewhat cheaper between 1995 and 2002, consistent with our finding of a decline in user cost shown in Figure 10.

[25]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.

[26]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 10: Rent-to-Value Functions



There was little change in mortgage rates between 2002 and 2007. There is, however, widespread evidence that credit became more available during the the period leading up to the bubble, especially for applicants with low credit ratings (Keys, Mukherjee, Seru, and Vig (2010)). Figure 10 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.

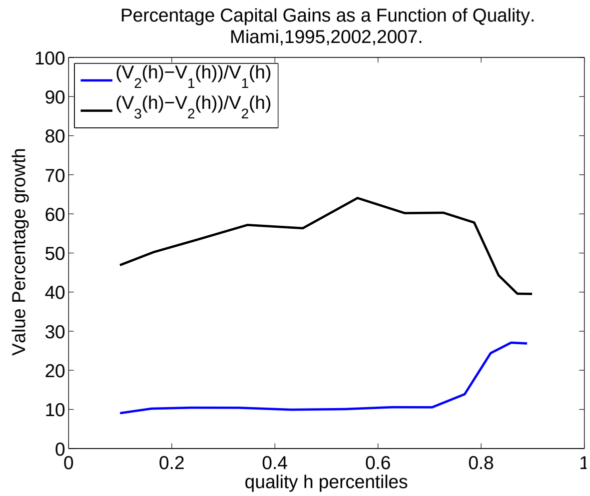
Another possible explanation for the change in the user-cost function is that investor expectations about future appreciation in rental rates and, thus, housing values changed during that time period. Our estimates indicate the average expectations of annual real rental appreciation must have been on 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. Brueckner, Calem, and Nakamura (2012) and Brueckner, Calem, and Nakamura (2015) present a theoretical framework and empirical evidence that these phenomena reinforced each other–expected housing

Figure 11: Growth in income and rent

price inflation encouraged relaxation of lending standards that in turn fed housing price inflation.

We combine data for owner-occupied and rental dwellings for each period, as in the preceding section. Our model accounts for changes in the distribution of real income, housing supply, and also population growth of more than 1% per year that occurred in Miami during the period from 1995-2007. The left panel in figure 11 shows that rents were relatively stable across the quality distribution during the pre-bubble period between 1995 and 2002; at each quantile rent increased somewhat less than the increase in income. (Recall that rents reflect both rental properties and implicit rent on owner-occupied properties.) The right panel of Figure 11 shows that rents increased at a somewhat faster rate during the bubble period with 5 year increases ranging up to 10 percent at the upper end of the quality spectrum. These 5-year rental increases were, however, quite modest relative to the run-up of housing prices discussed next. Thus, 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.

Figure 12: Capital Gains



Next, we compute the capital gains across the quality spectrum during the pre-bubble and bubble periods. The predicted capital gains combine our estimates of the rent-to-value functions with the predicted equilibrium hedonic rent functions. The results are illustrated in Figure 12. First consider the pre-bubble period, shown by the lower curve in the graph. Our estimates imply that 7-year capital gains were approximately 10 percent for most houses. Houses in the upper two deciles of the quality distribution had larger gains of up to 20 precent. During the bubble period, our model yields 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. This pattern of gains is consistent with loosening of credit market constraints playing a substantial role in explaining the run-up in housing markets. In particular, we would expect that relaxation of lending standards would increase access to credit by buyers of low- and middle-quality housing units, thereby bidding up prices of those units. These results are also consistent with the model and results reported in Landvoigt et al. (2015).

6 Conclusions

We have developed a new approach for estimating hedonic price functions for rents and values. 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, complementing single-index measures such as the Case-Shiller index. Fourth, it is comprehensive in incorporating all location-specific amenities in addition to services provided by the dwelling. Fifth, it provides a new, comprehensive approach to measuring agglomeration economies. Sixth, it gives new insights into the mechanism that generates housing price changes.

Estimating the model jointly for New York and Chicago, we provide a contrast of hedonic price functions and house quality distributions across the two metropolitan areas. We also calculate differences in agglomeration economies between the two areas, including calculations, specific to household type, of differences in earnings required to make a household in New York as well off as its counterpart in Chicago. Applying our framework to Miami, 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. We thus provide an accounting of how rentals and prices changed across the quality distribution during the bubble period as contrasted to the pre-bubble period.

There are variety of other potential applications of our approach. Our framework permits investigation of how changes in the real interest rate affects prices, rentals, and quantities across the quality spectrum in a metropolitan area–via the impact of the real interest rate on user cost of capital. By incorporating multiple household types, our framework also permits analysis of how changes in demographic composition and the income distribution affect housing prices and rents across the quality spectrum in a metropolitan area, and the associated impact on supply across the quality distribution. Similarly, the model can be used to study how housing price changes from growth in size or income distribution of one demographic group impact welfare of other demographic groups. Data are available that permit applying the model to make comparisons across other metropolitan housing markets, such as London and New York. More challenging generalizations are also of interest. For example, it may be feasible to extend the model to incorporate tenure choice. This wold permit investigation of how demographic composition, income distributions, and population size, via impacts on equilibrium prices and rents, affect tenure composition across the house quality spectrum in a metropolitan area.

References

Bajari, P. and 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., and 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. and 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.

Brueckner, J., Calem, P., and Nakamura, L. (2012). Subprime Mortgages and the Housing Bubble. *Journal of Urban Economics*, *71*, 230–243.

Brueckner, J., Calem, P., and Nakamura, L. (2015). House-Price Expectations, Alternative Mortgage Products and Default. *Journal of Money, Credit, and Banking.*

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

Keys, B., Mukherjee, T., Seru, A., and Vig, V. (2010). Did Securitization Lead to Lax Screening? Evidence from Subprime Loans. *The Quarterly Journal of Economics*, 307–362.

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

Landvoigt, D., Piazzesi, M., and Schneider, M. (2015). The Housing Market(s) of San Diego. *American Economic Review*, *105*(4), 1371–1407.

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 (4)*, 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.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics*, *85*(2).

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

Susin, S. (2003). Discrepancies Between Measured Income in the American Housing Survey (AHS) and the Current Population Survey (CPS)..

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:*

$$F_t(y) = G_t(v) \tag{37}$$

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:*

$$\frac{\int_0^{[(\ln(y_t+\beta_t)-\mu_t)/\sigma_t]^{r_t}/r_t} e^{-t} t^{1/r_t-1} dt}{2r\Gamma(1+1/r_t)} = \frac{\int_0^{[(\ln(v_t+\theta_t)-\omega_t)/\tau_t]^{m_t}/m_t} e^{-t} t^{1/m_t-1} dt}{2m\Gamma(1+1/m_t)} \tag{38}$$

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

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

Similar steps lead to the same conclusion when $y_t < \exp(\mu_t) - \beta_t$, and $v_t < \exp(\omega_t) - \theta_t$. Solving (39) yields:

$$y_t \quad = \quad e^{(\mu_t - \frac{\sigma_t}{\tau_t}\omega_t)} (v_t + \theta_t)^{\frac{\sigma_t}{\tau_t}} - \beta_t \tag{40}$$

Q.E.D.

Proof 3 *The household's FOC is:*

$$\alpha u'(h) \cdot dh = \frac{dv}{(y_t - v_t - \kappa)} \tag{41}$$

Substituting the income loci (26):

$$\alpha u'(h) dh = \frac{dv}{A_t(v_t+\theta_t)^{b_t} - \beta_t - v_t - \kappa} \tag{42}$$

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

$$\alpha_i u_i'(h)dh = \frac{dv}{A_t(v_t+\theta_t)^{b_t} - (v_t+\theta_t)} \tag{43}$$

Integrating the right hand side yields:

$$\int \frac{dv}{A_t(v+\theta_t)^{b_t} - (v+\theta_t)} = \frac{1}{b_t-1}\left(\ln\left(-\frac{1}{A_t}\left(v_t+\theta_t - A_t\left(v+\theta_t\right)^{b_t}\right)\right) - b_t \ln v + \theta_t\right) + c_t \tag{44}$$

which implies:

$$\alpha u(h) = \frac{1}{b_t-1}\ln\left(1 - \frac{\left(v_t+\theta_t\right)^{1-b_t}}{A_t}\right) + c_t \tag{45}$$

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

$$\alpha \ln(1-\phi(h+\eta)^{\gamma}) = \frac{1}{b_t-1}\ln\left(1 - \frac{\left(v+\theta_t\right)^{1-b_i}}{A_t}\right) + c_t \tag{46}$$

Solving for v_t

$$(1-\phi(h+\eta)^{\gamma})^{\alpha(b_t-1)} \quad = \quad \left(1 - \frac{\left(v_t+\theta_t\right)^{1-b_t}}{A_t}\right)e^{c_t} \tag{47}$$

and hence

$$v_t \quad = \quad \left(A_t\left[1 - \frac{(1-\phi(h+\eta)^{\gamma})^{\alpha(b_t-1)}}{e^{c_t}}\right]\right)^{\frac{1}{1-b_t}} - \theta_t \tag{48}$$

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:*

$$\max_{h} \quad u_1(h) + u_2(y - v(h)) \tag{49}$$

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

$$u_1'(h) - u_2'(y-v(h))\ v'(h) = 0 \tag{50}$$

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

$$\max_{h^*} \;\; u_1(v^{-1}(h^*)) + u_2(y - h^*) \tag{51}$$

and the FOC with respect to h^* is

$$u_1'(v^{-1}(h^*))\; {v^{-1}}'(h^*) - u_2'(y - h^*) = 0 \tag{52}$$

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:*

$$v_t = \left(A_t \left[1 - [1 - \phi(h + \eta)^\gamma]^{\alpha(b_t - 1)}\right]\right)^{\frac{1}{1-b_t}} - \theta_t \tag{53}$$

Setting

$$\alpha = \frac{1}{b_t - 1} \tag{54}$$

implies

$$v_t = (A_t \left[1 - [1 - \phi(h + \eta)^\gamma]\right])^{\frac{1}{1-b_t}} - \theta_t \;=\; (A_t \phi(h + \eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{55}$$

Setting

$$\phi = \frac{1}{A_t} \tag{56}$$

implies

$$v_t = ((h + \eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{57}$$

Setting

$$\gamma = 1 - b_t \tag{58}$$

implies

$$v_t = (h + \eta) - \theta_t \tag{59}$$

Finally, setting

$$\eta = \theta_t \tag{60}$$

implies.

$$v_t = h \tag{61}$$

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

$$v_{t+s}(h) = \left(A_{t+s} \left[1 - [1 - \phi(h+\eta)^{\gamma}]^{\alpha(b_{t+s}-1)} \right] \right)^{\frac{1}{1-b_{t+s}}} - \theta_{t+s} \tag{62}$$

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:

$$R_2(h) = k_2 \int_0^h q_1(x) \left(\frac{V_2(x)}{V_1(x)} \right)^{\zeta} dx \tag{63}$$

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:

$$f(y) = \frac{1}{2(x+\beta)r^{\frac{1}{r}}\sigma\Gamma\left(1+\frac{1}{r}\right)} e^{-\frac{1}{r\sigma^r}|\ln(x+\beta)-\mu|^r} \tag{64}$$

The CDF of the GNL4 distribution is given by:

$$F_t(y) = \begin{cases} \dfrac{\Gamma\left(\frac{1}{r}, B(y+\beta)\right)}{2\Gamma(\frac{1}{r})} & \text{for } y < \exp(\mu) - \beta, \\ \frac{1}{2} & \text{for } y = \exp(\mu) - \beta, \\ \frac{1}{2} + \dfrac{\gamma\left(\frac{1}{r}, M(y+\beta)\right)}{2\Gamma(\frac{1}{r})} & \text{for } y > \exp(\mu) - \beta. \end{cases} \tag{65}$$

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.

C 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:

$$\frac{dh}{dp}\frac{p}{h} = \frac{(-\alpha\phi\gamma h + (h+\eta)((h+\eta)^{-\gamma} - \phi))}{(-\alpha\phi\gamma + (1-\gamma)(h+\eta)^{-\gamma} - \phi)}\frac{1}{h} \tag{66}$$

and the income elasticity of demand is given by:

$$\frac{dh}{dy}\frac{y}{h} = \frac{-\alpha\phi\gamma}{p}\left[\frac{1}{-\alpha\phi\gamma + (1-\gamma)(h+\eta)^{-\gamma} - \phi}\right]\frac{\kappa - \frac{p}{\alpha\phi\gamma}\left[-\alpha\phi\gamma h + (h+\eta)^{1-\gamma} - \phi(h+\eta)\right]}{h} \tag{67}$$

We have seen that this specification of household preferences yields plausible price and income elasticities.