We are currently revising this paper. Part of my seminar presentation will be on an extension of the model in this paper and some additional empirical findings. Comments are most welcome. D.E.

A New Approach to Estimating Hedonic Equilibrium Models for Metropolitan Housing Markets*

Dennis Epple

Carnegie Mellon University and NBER

Luis Quintero

Johns Hopkins University

Holger Sieg

University of Pennsylvania and NBER

*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

We formulate and estimate an equilibrium model of metropolitan housing markets with housing differentiated by quality. Quality is a latent variable that captures all features of a dwelling and its environment. Two applications illustrate the value of our new technique. First, we develop a semi-parametric approach to estimate the model for Chicago and New York, obtaining housing prices for each quality level for each metropolitan area, stocks of each quality, and compensating variations required for a household of a given income in Chicago to be equally well off in New York. Second, we estimate rents and prices across the quality spectrum in Miami 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, Compensating Variations, Housing Market Crisis, Multiple Housing Markets.

1 Introduction

This paper develops a new, semi-parametric approach to estimate an equilibrium model of metropolitan housing markets in which housing varies in quality. Following Landvoigt, Piazzesi, & Schneider (2015), we treat housing quality as a latent unidimensional index that captures all features of a dwelling and its environment. We build on their work by incorporating this treatment of quality into an equilibrium framework for all housing in a metropolitan area. We develop a semi-parametric approach to estimate our model, providing two applications that illustrate the power of our framework. First, we estimate the model for Chicago and New York, obtaining housing prices for each quality level for each metropolitan area, stocks of each quality in each metropolitan area, and compensating variations required for a household of a given income in Chicago to be equally well off in New York. Second, we study rents and housing prices across the quality spectrum in Miami during the dramatic run-up in housing prices in the years preceding the financial crisis.

This paper makes a number of contributions to the literature on estimating hedonic equilibrium models. We develop a non-parametric matching approach for estimating user cost as a function of quality, utilizing the rent-to-value ratio of dwellings of the same latent quality. As such we provide an integrated treatment of rental rates and asset or property values. We then provide conditions such that house rental as a function of quality is identified for each of multiple metropolitan areas and/or multiple time periods in a given metropolitan area. We introduce a flexible parametrization that exploits generalized forms of log-normal distributions proposed by Vianelli (1983). This approach in turn yields a flexible closed-form solution for the equilibrium hedonic pricing function in each market in our baseline model. This baseline specification incorporates variation in income across households that have a common preference function. We then generalize our model to consider multiple household types, with the income distribution and the preference function varying across types. An innovation of our approach is that we combine a Heckman & Singer (1984) discrete factor model with the innovative use of k-means clustering to group households into types. Finally, we show how to estimate period-to-period changes in housing supply as a function of quality. As our applications demonstrate, implementation of our new approach is feasible with readily available data for metropolitan housing markets in the U.S.

To achieve our objective of formulating and estimating an equilibrium model of metropolitan-wide housing markets encompassing both owner-occupied and rental units, we require data for a representative sample of houses in the markets we study. To apply our approach to study changes over time within markets, we require repeated sampling of markets. To the best of our knowledge, the American Housing Survey (AHS) is the only data source that meets these requirements. The AHS draws a new sample each time it surveys a market. Hence, the AHS does not provide repeat observations on either dwellings or occupants of dwellings. Our model is tailored to accommodate these features of the AHS, enabling us to exploit the comprehensive multi-period coverage of metropolitan housing markets afforded by AHS. While our model does not incorporate the dynamics that would be permitted by repeat-sales data, our user-cost formulation captures expected housing appreciation in each market in each time period.

Estimation of our model does not require any assumptions regarding the extent to which households are mobile among metropolitan areas. With the further assumption that households are fully mobile across metropolitan areas, we can use our estimated model to calculate the compensating variation required for a household of a given income in a given metropolitan area to be equally well off in another metropolitan area. This calculation exploits the fact that house quality incorporates not only structural housing characteristics, but also all amenities and dis-amenities that affect the desirability of a dwelling. Thus, in addition to structural characteristics, quality incorporates the presence of a subway stop near a dwelling, an art museum in the metropolitan area in which the dwelling is located, balminess or harshness of climate, and so forth. We also compare compensating variation as a function of income across household types. We find, for example, that, for a household at the 50th income percentile in Chicago, a compensating variation of approximately 20% of income is required to make that household equally well off in New York. These measures are of interest in their own right and can also be interpreted, under additional assumptions, as measures of agglomeration economies.

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.

Our work is related to the following literature. The pioneering work of Rosen (1974) transformed modeling of markets for differentiated products and inspired an extensive literature focused on applications and associated issues of identification and estimation. A great many fruitful applications have built on Rosen's framework, including extensive research applying RosenÕs framework to the study of housing markets. An illuminating review of this literature is provided in Kuminoff, Smith, & Timmins (2013).

Recent research on hedonic identification is particularly relevant to our work. Ekeland, Heckman, & Nesheim (2004) describe limitations of prior work that uses linear systems of equations to study identification. Ekeland et al. (2004) demonstrate the payoff from fully exploiting all equilibrium implications of the hedonic framework. Investigating additive-utility models, they establish that non-parametric identification of the Rosen model is possible using data for a single market. Heckman, Matzkin, & Nesheim (2010) extend the analysis of non-parametric identification to non-additive models utilizing a unidimensional quality scale with multidimensional household types. As advocated in these papers, we exploit the full set of hedonic equilibrium conditions in our model of metropolitan equilibrium. Also, as in Heckman et al. (2010), we use a unidimensional index of housing quality with multiple household types.

Bajari & Benkard (2005) develop identification results and counterpart estimation methods for hedonic models, focusing in particular on developing methods that incorporate product characteristics observed by the consumer but not the econometrician. They develop a semi-parametric approach to estimating demand, exploiting the set of optimality conditions implied by consumer choice of product characteristics. While we employ a single index to achieve tractability for modeling supply and demand at the metropolitan level, we are mindful of the importance of unobserved characteristics demonstrated by the work of Bajari & Benkard (2005). Our latent quality approach captures both observed and unobserved characteristics.

There have also been important recent advances in study of the dynamics of housing markets. As we noted in our introduction, we build on Landvoigt et al. (2015) (LPS) who consider housing sales by owner-occupants while incorporating frictions in asset markets (e.g., collateral constraints, transaction costs, idiosyncratic shocks to housing returns). Their framework and an empirically calibrated computational model permits them to provide a rich quantitative account of the impacts of the housing boom in San Diego.

Finally, our work is related to a recent paper by Bajari, Fruehwirth, Kim, & Timmins (2012) who use repeat-sales in a rational expectations framework to estimate the marginal prices of changes in the housing bundle. As demonstrated in their application, this proves to be especially well suited to estimating the implicit prices of changes in environmental quality. Bayer, McMillan, Murphy, & Timmins (2015) formulate and estimate a dynamic model of household location choice and housing preferences. Using repeat-sales data, they estimate willingness to pay for environmental amenities while taking account of unobserved geographic heterogeneity. The dynamic modeling and use of repeat sales data in these papers mark important advances in hedonic modeling and in estimation of the value of environmental amenities.

The rest of the paper is organized as follows. Section 2 develops our model. Section 3 discusses identification. Section 4 provides our estimation strategy. Section 5 studies housing markets in NYC and Chicago, while Section 6 analyzes the markets in Miami during the recent run-up in housing prices. Section 7 offers some conclusions and discusses future research. All proofs of propositions are reported in Appendix A.

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, the two markets cannot be studied separately.

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.[1]

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:

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

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.[2]

[1] Our approach is closely related to that of Poterba (1992) and Poterba & Sinai (2008).

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

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

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

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

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

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:

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

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 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:

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

Then:

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

where $c_t(h)$ is the user cost ratio for housing capital:

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

As detailed in Gyourko & Sinai. (2003), the user cost of owner-occupied housing is affected by the deductibility of property taxes and mortgage interest from federal taxes. They show that the federal subsidy arising from deductibility varies greatly across metropolitan areas. The subsidy also varies across individuals of differing incomes. In our development in the following sections we show that our model implies an equilibrium mapping from house quality to income. Denote this mapping $y(h)$. Then let $i_t(y(h))$ and $\tau_p(y(h))$ denote respectively the equilibrium after-federal-tax mortgage interest rate and property tax rate respectively for property of quality h. The preceding derivation then applies to account for the effect of federal deductibility of mortgage interest and property taxes on the equilibrium user cost for each quality h. This in turn establishes that our non-parametric estimator of user cost as a function of h then incorporates the effects of federal deductibility of mortgage interest and property taxes.

Given that hedonic price functions of rental rates and asset value are non-linear in quality and are influenced by a variety of factors in equilibrium, the model does not imply that the user-cost is constant. Instead it is a function of housing quality. This key insight is typically overlooked in empirical work. Most empirical research that uses the hedonic framework uses a constant user-cost specification.[3] One key contribution of this paper is that we show below how to identify and estimate the

[3]An exception is the paper by Bracke (2013) who studies rent-value functions in London using a subset of housing units for which both asset prices and rental rates are observed.

user-costs as a function of latent quality using a non-parametric matching estimator.[4]

Finally, note that 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. The main advantage of this approach is that we do not have to invoke strong assumptions on the evolution of price expectations.

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. Note that we interpret income as a broadly defined measure that includes not just labor income, but also income from asset holding for wealthy households and transfer income for

[4]Note that the time-invariant case studied by Poterba (1984, 1992) implies that:

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

When τ^p, δ, π, and i are small, the preceding closely approximates the continuous time solution of Poterba (1984). Bajari, Benkard, & Krainer (2005) also develop a user cost formula within a dynamic framework.

poor households.[5] Our approach thus implicitly accounts for differences in wealth. 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 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. We assume that transactions costs in the rental market are zero. Hence, a household can 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:

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

where z_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 in-

[5] Broadly defined income measures are available in the AHS which we use in the empirical application.

come distribution yields the aggregate housing demand $H_t^d(h|v_t(h))$:

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

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 1 states that high-income households are willing to pay more for a higher quality house than low-income households – a weak restriction on preferences.

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, and 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 the following 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.

Let $G_t(v)$ and $R_t(h)$ be, respectively the metropolitan distribution of house rent and the metropolitan distribution of the stock of housing. 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 $F_t(y) = G_t(v)$.

The single-crossing condition then implies the following result. This innovative approach of equating CDF's to obtain the relationship between income and house value is due to Landvoigt et al. (2015).

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

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

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

In equilibrium rental markets must clear for each value of h, i.e. for each level of housing quality h, we have:

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

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).*[6]

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

We will show below that these functions are sufficiently flexible to fit the housing value and income distributions in 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 $\theta_t - \beta_t$ is time invariant permits us to obtain a closed-form solution to the hedonic price function.

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 $b_t = \frac{\sigma_t}{\tau_t}$.

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:*

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

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

with $m_t(h) = \ln(1 - \phi(h+\eta)^\gamma)$*, where* $\alpha > 0$*,* $\gamma < 0$*,* $\phi > 0$*, and* $\eta > 0$*.*[7]

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

Proposition 3 *If* $b_t > 1$ *and* $\kappa = \theta_t - \beta_t \ \forall t$*, the hedonic equilibrium pricing function is unique and given by:*

$$v_t \quad = \quad \left(A_t \left[1 - \frac{(1-\phi(h+\eta)^\gamma)^{\alpha(b_t-1)}}{e^{c_t}}\right]\right)^{\frac{1}{1-b_t}} - \theta_t \tag{27}$$

for all $h > (\frac{1}{\phi})^{\frac{1}{\gamma}} - \eta$*. Here* c_t *is a constant of integration that we set to zero.*[8]

With the above equations in hand, a summary of the sequence of events each period is useful. At the beginning of date t, the population of households in the metropolitan area, N_t, and the metropolitan distribution of income, $F_t(y)$, are determined by factors outside the model (migration to or from the metropolitan area,

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

[8]The constant of integration determines the intercept in the hedonic pricing function. There is not a strong a priori basis for fixing the intercept at a particular value. It is natural to expect the intercept to be close to zero (rent equal to zero for a dwelling of quality zero), though a small positive intercept might arise if there is a fixed cost of providing housing of minimum quality. With our normalization, $c_t = 0$ in all periods, we obtain intercepts close to zero in all periods (see Figure 6). Thus, setting $c_t = 0$ yields quite plausible values for the intercepts. To investigate robustness, rather than setting $c_t = 0$, we estimated the model for Chicago choosing c_t in each period to satisfy the requirement that $v_t(0) = 0$ in each period. This had minimal effect on our estimates. We thank a referee for helping us clarifying this issue.

household formation and dissolution, labor and capital markets, macroeconomic conditions). The number of households in the metropolitan area and the distribution of income across households together determine the demand for housing. The user cost ratio as a function of quality, $c_t(h)$, is also determined outside the model, by factors set forth in Section 2.1 (property tax rates, depreciation and maintenance costs of housing of each quality level, and expected interest rates and housing price appreciation rates for housing of each quality level). Next, housing markets clear. This entails the following. We take the distribution of house quality as given in the initial period. In each subsequent period, suppliers adjust the housing stock of each quality, $R_t(h)$, according to Equation (18). Flow of each quality, proportional to the stock of each quality, is thereby determined. The demand for housing of each quality is equilibrated with the supply of each quality according to equations (19) and (23), and the ratio of house value to rent for each quality level is determined by (8). Together, these determine via equation (28) the equilibrium rental rate for each quality level, $v_t(h)$, equilibrium house value for each quality level, $Vt(h)$, and the expenditure on housing for each income level, $y_t(v)$. These in turn determine the period t equilibrium CDF for rent, $G_t(v) = F_t(y_t(v))$.

Before, we discuss extensions of the model we offer some additional comments. In constructing the hedonic price function in equation (27) we impose a functional form assumption on the *endogenous* equilibrium distribution of rents. We then show that there exists a unique hedonic price function that is consistent with the observed income and rental distributions. While this approach is new to the hedonic literature, parameterized equilibrium relationships are widely used in economics on topics ranging from specification of demand systems (Deaton & Muellbauer (1980)) to parameterization of expectations in rational expectations models (den Haan & Marcet (1994)). Of course, we must show that our chosen parameterization conforms to the data.

In hedonic models of housing, it is often the case that the distribution of housing supply is taken as exogenous, with assumptions then placed on that distribution as well as the distributions of household preferences and income. The objective is then to find or approximate the functional form of the endogenous equilibrium hedonic price function that clears the markets. While broadly similar, our approach differs in three respects. First, as we detail below, we impose market clearing conditions during estimation of the model. This dovetails nicely with our overall estimation approach, but does not differ in spirit from prior work. Second, the supply of housing in the initial period is identified in our model based on the observed rent distribution. This follows from the insight of LPS that housing quality can be normalized in a baseline period. We formalize this insight in the context of our model below. As a consequence we do not need to invoke a functional form assumption on the initial distribution of housing. Third, the evolution of housing supply is governed in our model by equation (19). This equation then allows us to endogenize changes in housing supply for periods after the initial period.

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. The fraction of each type i at time t is given by s_{it}. Our approach is thus in the spirit of Heckman & Singer (1984) who also use discrete types to control for observed and unobserved heterogeneity. The main difference is that we use a clustering algorithm to determine the fraction of each type, instead of treating these fractions as parameters to be estimated. Each type of household has a utility function that depends on its type $U_i(h,z)$. For example, a straight forward extension of the parametric utility function used in the previous section is the following utility function: [9]

$$U_i(h,z) = \ln(1 - \phi_i(h+\eta_i)^{\gamma_i}) + \frac{1}{\alpha_i}\ln(z-\kappa_i) \tag{28}$$

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.

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

$$U_i(h_j, \hat{y}_{i,jt} - v_{jt}) = U_i(h_{j+1}, \hat{y}_{i,j,t} - v_{j+1,t}) \tag{29}$$

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 analogous to Assumption 3. For the parametrization of the utility function discussed above, these cut-off levels are given by:

$$\hat{y}_{i,j,t} = \frac{v_{j,t} - e^{(M_{i,j+1}-M_j)\alpha_i}\ v_{j+1,t}}{1 - e^{(M_{i,j+1}-M_j)\alpha_i}} + \kappa_i \tag{30}$$

[9]Since we do not require the existence of closed form solutions for the equilibrium price function, the approach allows for other functional form assumptions for the utility function.

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:

$$H^d_{i,j,t}(v_{1t}, ..., v_{Jt}) = F_{i,t}(\hat{y}_{i,j,t}) \; - \; F_i(\hat{y}_{i,j-1,t}) \tag{31}$$

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:

$$\sum_{i=1}^{I} s_{i,t} \; H_{i,j,t}(v_{1,t}, ..., v_{J,t}) = r_{j,t} \;\; \forall j \tag{32}$$

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

$$r_{j,t} = R_t(h_j) - R_t(h_{j-1}) \tag{33}$$

Equilibrium can then be defined as above.

3 Identification

We consider identification of the model assuming that a) we have access to data for, at least, two markets; 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 periods.*

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 from the observed income and rent distributions and the hedonic price function 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 the 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 carry 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. The non-parametric matching algorithm extends to more general demand models in which demand depends on a vector of observed state variables.

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 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

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. 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. 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 $\tilde{F}^N_{t,j}$ denote the jth percentile of empirical income distribution at time t that is estimated based on a sample with size N. Similarly, let $\tilde{G}^N_{t,j}$ 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:

$$\hat{\psi}^N = \text{argmin}_{\psi \in \Psi} \ L^N(\psi) \tag{34}$$

subject to the structural constraints. The objective function is:

$$L^N(\psi) = (1 - W)\ (l^N_y(\psi) + l^N_r(\psi))\ +\ W\ l_h(\psi) \tag{35}$$

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

$$\begin{aligned}
l^N_y(\psi) &= \sum_{t=1}^{T}\sum_{j=1}^{J}([F_t(y_{t,j};\psi) - F_t(y_{t,j-1};\psi)] - [\tilde{F}^N_{t,j} - \tilde{F}^N_{t,j-1}])^2 && (36)\\
l^N_r(\psi) &= \sum_{t=1}^{T}\sum_{j=1}^{J}[G_t(v_{t,j};\psi) - G_t(v_{t,j-1};\psi)] - [\tilde{G}^N_{t,j} - \tilde{G}^N_{t,j-1}])^2 && (37)\\
l^N_h(\psi) &= \sum_{t=2}^{T}\sum_{j=1}^{J}([G_t(v_t(h_j;\psi) - R_t(h_j;\psi))^2 && (38)
\end{aligned}$$

Note that W is the weight that is assigned to the market clearing conditions.[10] We use a standard bootstrap procedure to estimate the standard errors.

[10] Our estimates are relatively insensitive to the choice of this parameter.

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 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 one-type model.

5 The Housing Markets of Chicago and New York

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. 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 a given 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.[11]

Fortunately, the AHS conducted surveys in both Chicago and New York for both

[11] 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 of 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.

1999 and 2003. We exploit data from these surveys for our first application, jointly estimating the model for those two metropolitan areas. 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.

AHS definitions of each of these metropolitan areas is unchanged across these two periods.[12] We use New York metropolitan area in 1999 as the base for our normalization. We use data from the extended metropolitan surveys conducted in the Chicago and New York metropolitan areas in 1999 and 2003. As a shorthand, we will sometimes refer to the two metropolitan areas as CHI and NYC.

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.[13] 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.[14] 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.

We obtain three clusters, which is the optimal number of clusters calculated with

[12]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.

[13]Neill (2006) provides a clear treatment of this topic.

[14]Similar results were obtained with cosine distance and are available from the authors.

our sample data.[15] Table 1 shows estimated cluster shares and centroids for CHI and NYC 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 comprised of young households with few or no children, Type 2 is 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 NYC relative to CHI. Mean age in the third group is somewhat higher in CHI than in NYC. Overall, however, means by age and by number of children within each type are quite similar across metropolitan ares and across periods.

To construct the rental distributions we need to convert values into imputed rents. We estimate the user-cost ratio for each household type. We find that the estimates are similar across types. This suggests constructing a single user-cost function by pooling the data. The results are illustrated in Figure 1. We find that the estimated user costs are very similar in NYC and CHI in 1999 and fairly constant across quality. The average user cost is approximately 0.06 in 1999 in both cities.[16] Not much changes in Chicago between 1999 and 2003. In contrast, we find that the user cost declined markedly in NYC between 1999 and 2003, perhaps reflecting expectations of greater price appreciation in NYC relative to CHI in 2003.

[15] We used NYC 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 its cluster relative to alternative clusters. Computations were done using Matlab.

[16] We normalize housing quality using the observed rents and values in NYC in 1999. A more detailed analysis of the estimation of these rent-value functions by household type is available upon request from the authors.

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

Figure 1: User Cost in NY and CHI in 1999 and 2003

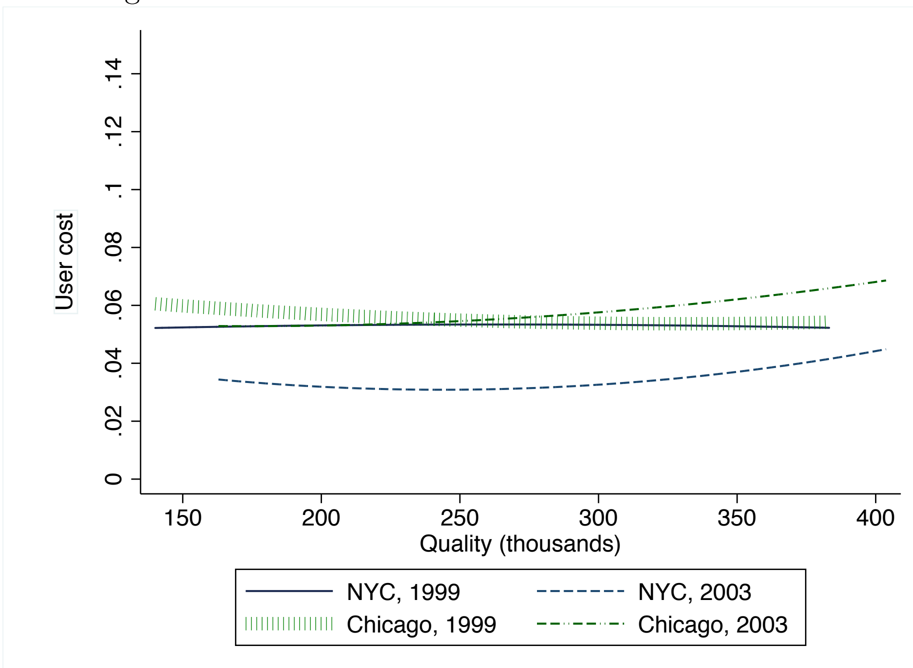
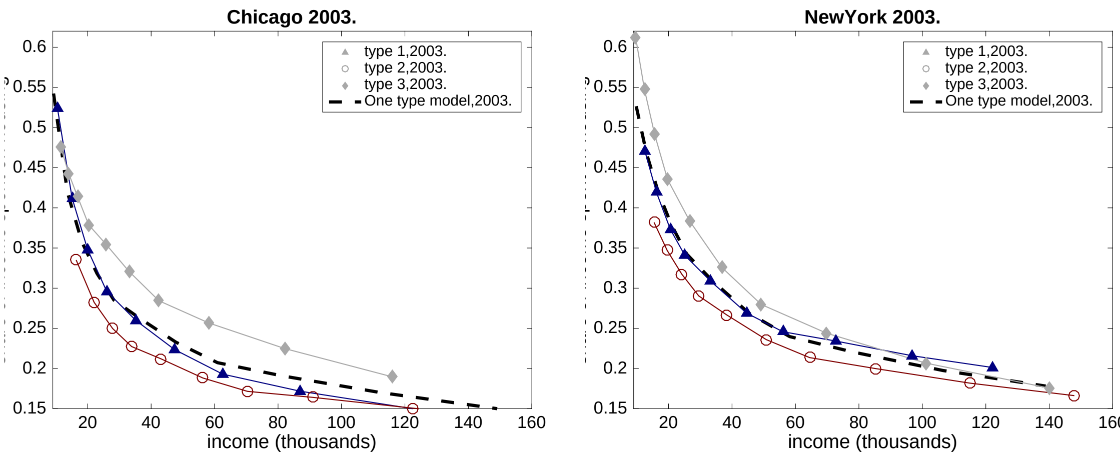


Figure 2 provides rental expenditure shares as a function of income in 2003 for CHI and NYC. The ordering of expenditure shares across types is the same in both metro areas. At each income level, the ordering of types from highest to lowest expenditure is Type 3, Type 1, and Type 2. Type 3 households spend a higher share of income on rents at each level income than the other two household types. 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. We conclude that the clustering algorithm has identified types with differing income distributions and differing housing expenditure patterns.

The parameter estimates and estimated standard errors for the preference and supply functions are reported in Table 2. The first row of Table 2 shows results for our baseline one-type model. The remaining three rows show results from the

Figure 2: Share of income spent on housing for each household type.



three-type model. All parameters have the correct algebraic signs. The bootstrapped standard errors account for the sequential nature of our estimator.

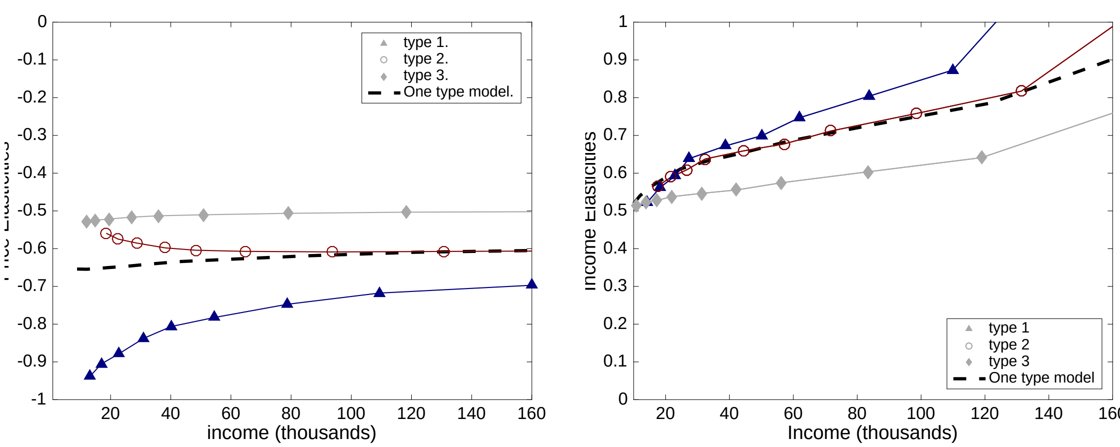
Price and income elasticities implied by our preference parameters estimates are plotted as a function of income in Figure 3. The income elasticities have the same order across the income range with Type 1 having the highest elasticity at each income level and and Type 3 the lowest. Type 3 households 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 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

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

Type	α	ϕ	η	γ	ζ
One Type	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)	

Figure 3: Income and Price Elasticities



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 (16). 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. Our supply elasticity estimates are consistent with the estimates of supply elasticities summarized in Glaeser (2004).

Figure 4 shows the fit of model to the income and rent distributions. The fitted lines are from structural estimation that imposes all equilibrium conditions.The left two graphs show the fit to income distributions in NYC in 1999 and 2003 respectively, while the right two graphs show fit to 1999 and 2003 NYC rental distributions. The corresponding graphs for CHI are shown below those for NYC.

These graphs illustrate that the fit to the income and rent data is quite good in both metropolitan areas in both time periods. We thus conclude that the generalized log-normal distributions are sufficiently flexible to fit the observed income and rent distributions.

The graphs in Figure 5 illustrate the resulting equilibrium in the housing markets. We plot supply and demand for each quality level in the two time periods in the two metropolitan areas. The upper pair of graphs are for CHI in 1999 and 2003 and the lower pair for NYC 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.

Figure 4: Distribution Fit

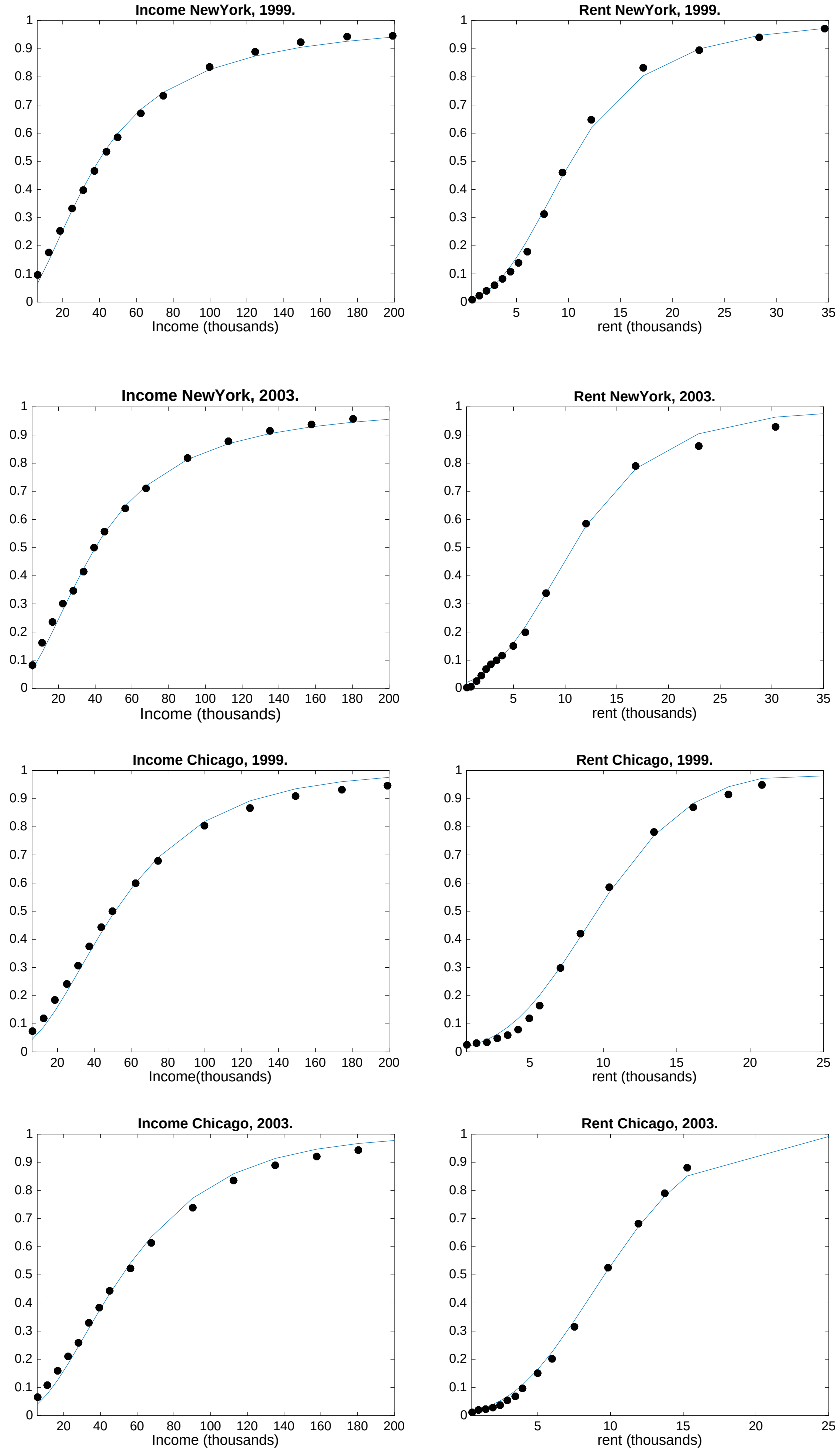


Figure 5: Supply and Demand Equilibrium

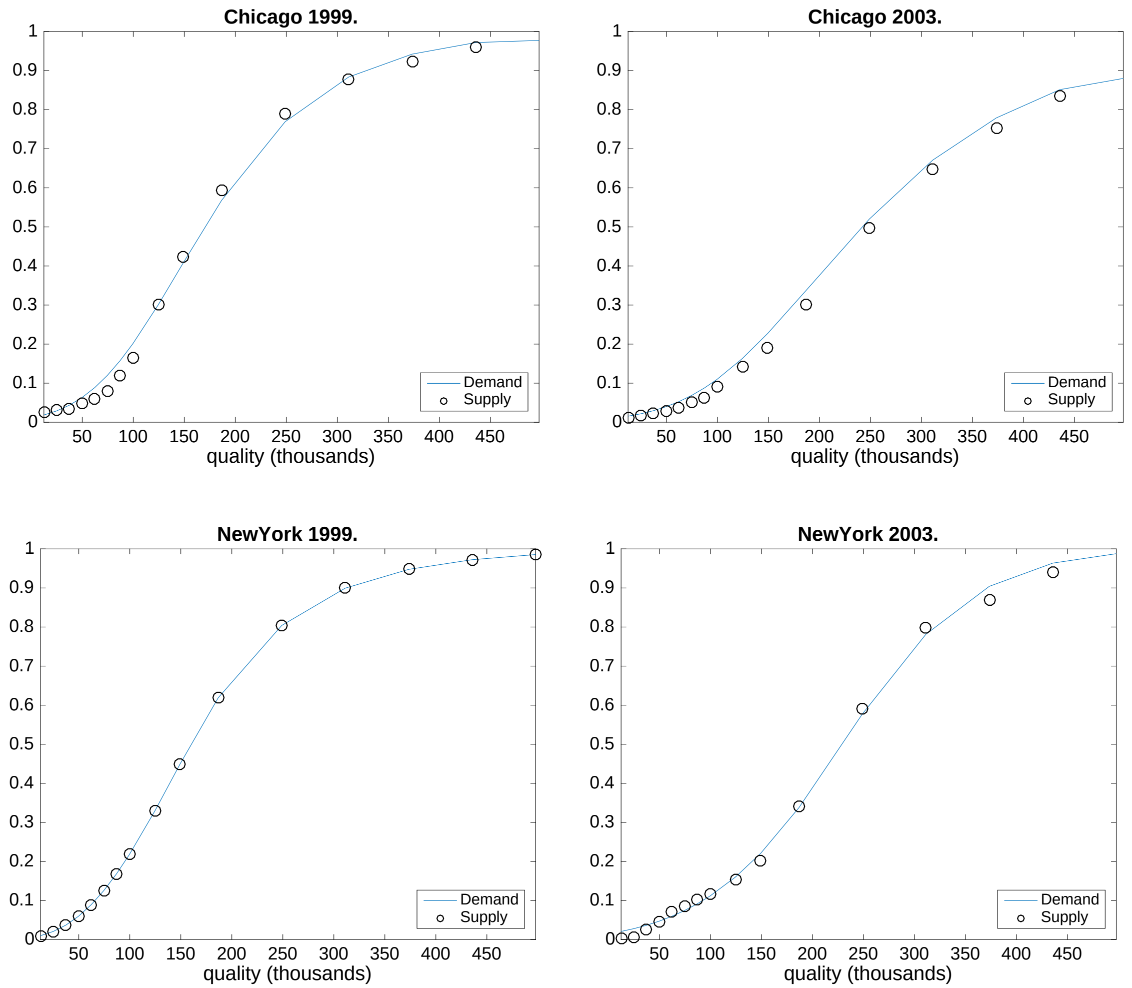


Figure 6: Equilibrium Price Functions in 1999 and 2003

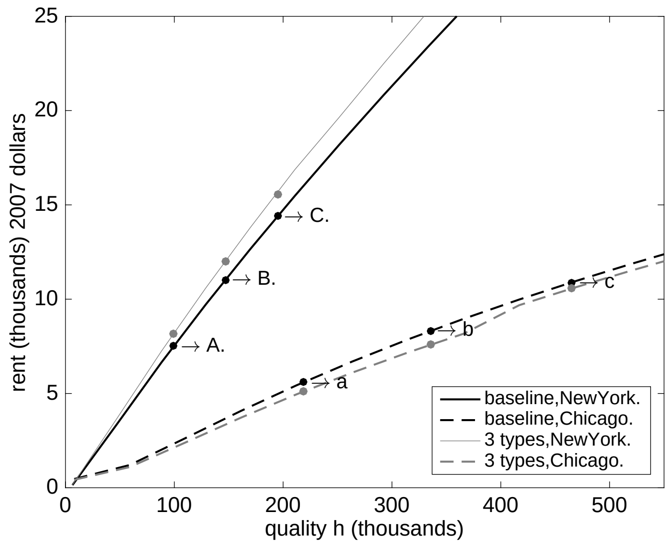
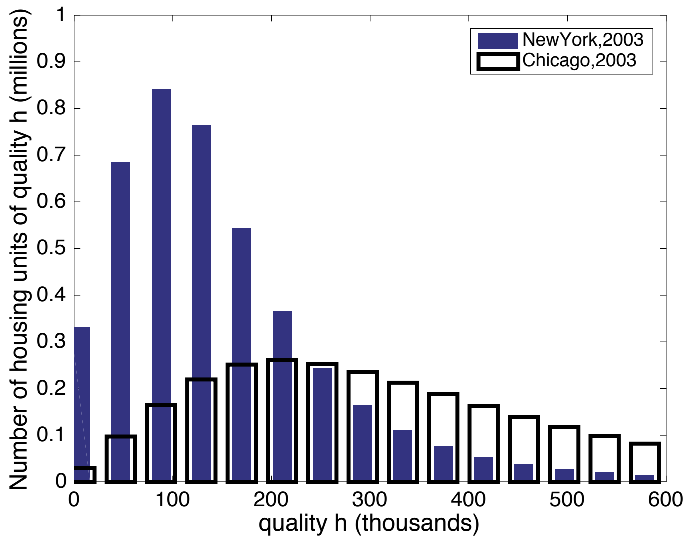


Figure 6 shows the hedonic price functions in 2003 for CHI and NYC from both the one-type model and the three-type model. The steeper curves show the hedonic price functions for 2003 for NYC from the two alternative models while the shallower curves show the estimates for CHI. We also find that in each metropolitan area the estimates from the one-type and multiple-type 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 their 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. It is important to keep in mind that our house quality measure is comprehensive and includes all locational amenities in addition to

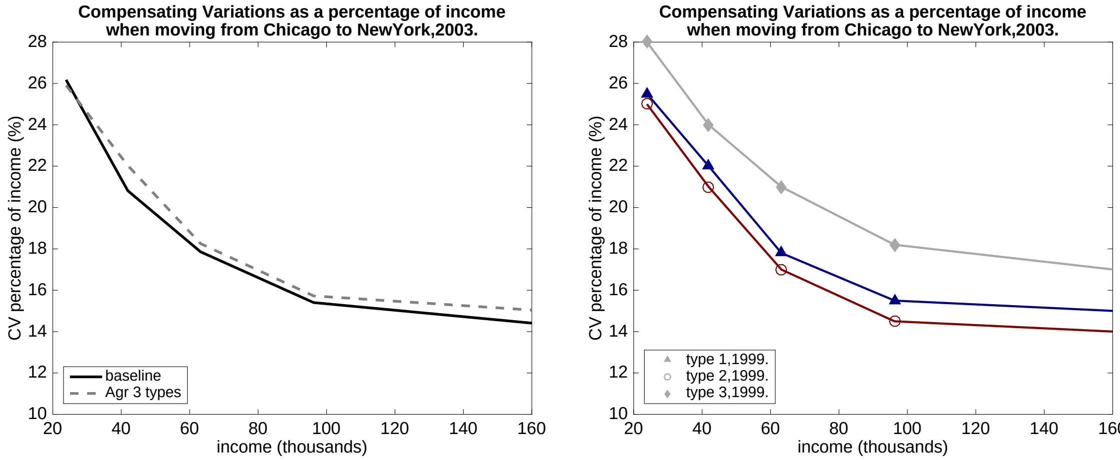
Figure 7: Housing Stock in NYC and CHI



the housing structure itself.

As we have just seen, at every income level, a household in NYC consumes lower quality than the corresponding household in CHI. The effect of this difference in consumption levels is to shift the distribution of quality in NYC to the left relative to that in CHI. 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., CHI incomes tend to be higher. The 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 predicted numbers of housing units by quality are shown in Figure 7 for each city. CHI has *relatively* more high quality housing. Given its larger population, however, NYC has a larger number of housing units overall than CHI.

Figure 8: Compensating Variations



Finally, we compute compensating variations required for a household of a given income in Chicago to be equally well off in New York. 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 left panel of Figure 8, we plot for each household type for each income the compensating variation that would be required for that type and income in Chicago to be equally well off in New York. For all three types, CV as a share of income 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.[17]

In the right panel of Figure 8, we plot 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 CHI (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 NYC, arising from greater agglomeration economies in NYC, would need to be higher by these amounts to compensate a household for the differences in housing price functions between the two metropolitan areas.[18]

6 Housing Prices in Miami: 1995-2007

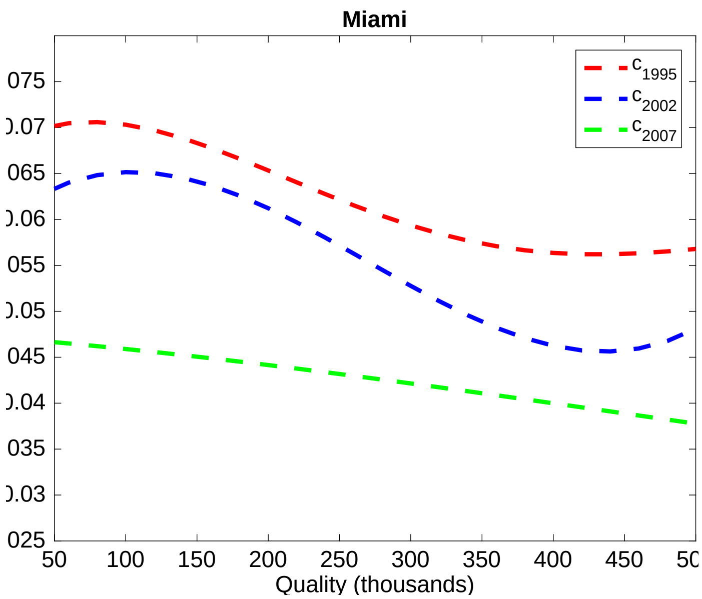
Our second application takes advantage of three successive AHS surveys of the Miami (FL) metropolitan area in 1995, 2002 and 2007. We divide this period into two sub-periods: a) the pre-bubble period from 1995 - 2002; b) the bubble period from

[17] Rent stabilization in New York City may result in an estimated hedonic price function that to some degree understates the price conditional on quality in New York City. Our model is estimated using data for the metropolitan area, and not just New York City. Nonetheless, rent stabilization in New York City may impact our estimates, resulting in some underestimation of the compensating variation required for a household in Chicago to be as well off in the New York metropolitan area. Because of the lock-in effect of rent stabilization, the impact on our estimates is likely to be greatest for older (Type 3) households. Extending our approach to address the impact of price stabilization is an important issue for future research.

[18] Rosenthal & 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.

2002 -2007.[19] First, we estimate the rent-to-value functions using our non-parametric matching estimator for the three periods in our data set. To simplify the analysis, we focus on the model with one type. Figure 9 plots the estimated functions for the three time periods.

Figure 9: User Cost Functions



We find that the user cost 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

[19]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.

large changes in the user cost 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 somewhat cheaper between 1995 and 2002, consistent with our finding of a decline in user cost shown in Figure 9. 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, & Vig (2010)). Figure 9 shows that the largest changes in user cost are for lower quality houses which is consistent with the notion that demand for these assets 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 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, & Nakamura (2012) and Brueckner, Calem, & Nakamura (2015) present a theoretical framework and empirical evidence that these phenomena reinforced each other–expected housing price inflation encouraged relaxation of lending standards that in turn fed housing price inflation.

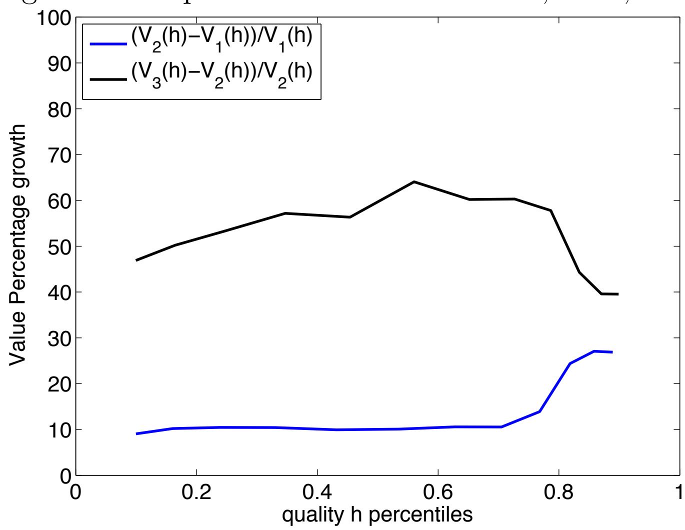
Our model accounts for changes in the distribution of real income, housing supply, and also population growth of more than 1% per year that occurred in Miami during the period from 1995-2007. The left panel in Figure 10 shows that rents were relatively stable across the quality distribution during the pre-bubble period between 1995 and 2002; at each quantile rent increased somewhat less than the increase in income. The right panel of Figure 10 shows that rents increased at a somewhat faster rate

Figure 10: Growth in income and rent across periods

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, & Verbrugge (2011) who also report that there was no “bubble ”in rental rates for housing.

Next, we compute the annualized capital gains across the quality spectrum during the pre-bubble and bubble periods. The predicted capital gains combine our estimates of the user cost functions with the predicted equilibrium hedonic rent functions. The results are illustrated in Figure 11. 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 as discussed in detail

Figure 11: Capital Gains in Miami 1995, 2002, 2007



in LPS. 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.

7 Conclusions

We have developed a new approach for estimating hedonic equilibrium models in metropolitan housing markets. 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 compensating variations, which in turn provides insights into 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. In doing so, we extend Heckman & Singer (1984) discrete type models using innovative k-means clustering algorithms to classify households. We also calculate compensating variations that leave households indifferent between the two areas, including calculations, specific to household type. Applying our framework to Miami, we find that user cost functions are highly non-linear in quality. Moreover, user costs increased by up to 50 percent from the pre-bubble levels during the bubble period. Rents only increased moderately as we would have expected given the relatively modest increase in population and income. 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. & Benkard, L. (2005). Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, *113*, 1239–77.

Bajari, P., Benkard, L., & Krainer, J. (2005). House prices and consumer welfare. *Journal of Urban Economics*, *58*(3), 474–487.

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.

Bayer, P., McMillan, D., Murphy, A., & Timmins, C. (2015). A Dynamic Model of Demand for Houses and Neighborhoods. Working Paper.

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

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

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

Deaton, A. & Muellbauer, J. (1980). An Almost Ideal Demand System. *The American Economic Review*, *70*(3), 312–326.

den Haan, W. & Marcet, A. (1994). Accuracy in Simulations. *Review of Economic Studies*, 3–17.

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

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

Gyourko, J. & Sinai., T. (2003). The Spatial Distribution of Housing-Related Ordinary Income Tax Benefits. *Real Estate Economics*, *31* (1), 527–76.

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

Heckman, J. & Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data. *Econometrica*, *52 (2)*, 271–320.

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

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. (2015). The Housing Market(s) of San Diego. *American Economic Review*, *105* (4), 1371–1407.

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.

Poterba, J. & Sinai, T. (2008). Tax Expenditures for Owner-Occupied Housing: Deductions for Property Taxes and Mortgage Interest and the Exclusion of Imputed Rental Income.. *American Economic Review*, *98*(2), 84–89.

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

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

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:*

$$F_t(y) = G_t(v) \tag{39}$$

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:*

$$\frac{\int_0^{[(\ln(y_t+\beta_t)-\mu_t)/\sigma_t]^{r_t}/r_t} e^{-t} t^{1/r_t-1} dt}{2r\Gamma(1+1/r_t)} = \frac{\int_0^{[(\ln(v_t+\theta_t)-\omega_t)/\tau_t]^{m_t}/m_t} e^{-t} t^{1/m_t-1} dt}{2m\Gamma(1+1/m_t)} \tag{40}$$

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

$$\frac{\ln(y_t+\beta_t)-\mu_t}{\sigma_t} = \frac{\ln(v_t+\theta_t)-\omega_t}{\tau_t} \tag{41}$$

Similar steps lead to the same conclusion when $y_t < \exp(\mu_t) - \beta_t$, and $v_t < \exp(\omega_t) - \theta_t$. Solving (41) yields:

$$y_t = e^{(\mu_t - \frac{\sigma_t}{\tau_t}\omega_t)} (v_t + \theta_t)^{\frac{\sigma_t}{\tau_t}} - \beta_t \tag{42}$$

Q.E.D.

Proof 3 *The household's FOC is:*

$$\alpha u'(h) \cdot dh = \frac{dv}{(y_t - v_t - \kappa)} \tag{43}$$

Substituting the income loci (25):

$$\alpha u'(h) dh = \frac{dv}{A_t (v_t + \theta_t)^{b_t} - \beta_t - v_t - \kappa} \tag{44}$$

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

$$\alpha_i u_i'(h) dh = \frac{dv}{A_t (v_t + \theta_t)^{b_t} - (v_t + \theta_t)} \tag{45}$$

Integrating the right hand side yields:

$$\int \frac{dv}{A_t (v + \theta_t)^{b_t} - (v + \theta_t)} = \frac{1}{b_t - 1} \left(\ln \left(-\frac{1}{A_t} \left(v_t + \theta_t - A_t \left(v + \theta_t \right)^{b_t} \right) \right) - b_t \ln v + \theta_t \right) + c_t \tag{46}$$

which implies:

$$\alpha u(h) = \frac{1}{b_t - 1} \ln \left(1 - \frac{\left(v_t + \theta_t \right)^{1 - b_t}}{A_t} \right) + c_t \tag{47}$$

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

$$\alpha \ln(1 - \phi(h + \eta)^\gamma) = \frac{1}{b_t - 1} \ln \left(1 - \frac{\left(v + \theta_t \right)^{1 - b_i}}{A_t} \right) + c_t \tag{48}$$

Solving for v_t

$$(1 - \phi(h + \eta)^\gamma)^{\alpha(b_t - 1)} \quad = \quad \left(1 - \frac{\left(v_t + \theta_t \right)^{1 - b_t}}{A_t} \right) e^{c_t} \tag{49}$$

and hence

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

Q.E.D.

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

$$\max_h \quad u_1(h) + u_2(y - v(h)) \tag{51}$$

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

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

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

$$\max_{h^*} \; u_1(v^{-1}(h^*)) + u_2(y - h^*) \tag{53}$$

and the FOC with respect to h^* is

$$u_1'(v^{-1}(h^*))\, v^{-1'}(h^*) - u_2'(y - h^*) = 0 \tag{54}$$

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 $v_t(h) = h$. Assume that $c_t = 0$. Hence, the equilibrium hedonic pricing function is given by:*

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

Setting

$$\alpha = \frac{1}{b_t - 1} \tag{56}$$

implies

$$v_t = (A_t\, [1 - [1 - \phi(h+\eta)^\gamma]])^{\frac{1}{1-b_t}} - \theta_t \; = \; (A_t \phi (h+\eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{57}$$

Setting

$$\phi = \frac{1}{A_t} \tag{58}$$

implies

$$v_t = ((h+\eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{59}$$

Setting

$$\gamma = 1 - b_t \tag{60}$$

implies

$$v_t = (h + \eta) - \theta_t \tag{61}$$

Finally, setting

$$\eta = \theta_t \tag{62}$$

implies.

$$v_t = h \tag{63}$$

That establishes identification of the parameters of the utility function. Hence the normalizations that $c_t = 0$ and that $v_t(h) = h$ are sufficient to identify the parameters of the utility function.

The price equation in any other period $t + s$ is then given by:

$$v_{t+s} \;=\; \left(A_{t+s}\left[1 - \frac{(1 - \phi(h+\eta)^{\gamma})^{\alpha(b_{t+s}-1)}}{e^{c_{t+s}}}\right]\right)^{\frac{1}{1-b_{t+s}}} - \theta_{t+s} \tag{64}$$

The assumption of constant utility across time then implies that $v_{t+s}(h)$ is identified by the parameters b_{t+s}, A_{t+s}, and θ_{t+s} and the normalization that $v_{t+s}(0) = 0$. Note that we need the last normalization to determine the constant of integration c_{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:

$$R_2(h) = k_2 \int_0^h q_1(x) \left(\frac{V_2(x)}{V_1(x)}\right)^{\zeta} dx \tag{65}$$

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:

$$f(y) = \frac{1}{2(x+\beta)r^{\frac{1}{r}}\sigma\Gamma\left(1+\frac{1}{r}\right)} e^{-\frac{1}{r\sigma^r}|\ln(x+\beta)-\mu|^r} \tag{66}$$

The CDF of the GNL4 distribution is given by:

$$F_t(y) = \begin{cases} \dfrac{\Gamma\left(\frac{1}{r}, B(y+\beta)\right)}{2\Gamma(\frac{1}{r})} & \text{for } y < \exp(\mu) - \beta, \\ \frac{1}{2} & \text{for } y = \exp(\mu) - \beta, \\ \frac{1}{2} + \dfrac{\gamma\left(\frac{1}{r}, M(y+\beta)\right)}{2\Gamma(\frac{1}{r})} & \text{for } y > \exp(\mu) - \beta. \end{cases} \tag{67}$$

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.