A New Approach to Estimating 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

May 7, 2018

*We would like to thank the editor of journal, James Heckman, four anonymous referees, as well as 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.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Abstract

We formulate and estimate a new 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. We estimate the model for Chicago and New York, obtaining hedonic housing price functions 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.

Keywords: Hedonic Models, Nonlinear Pricing, Housing Supply, Multiple Housing Markets, Semi-parametric Estimation.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

1 Introduction

This paper develops a new class of equilibrium models of metropolitan housing markets. Our model is comprehensive, encompassing all dwellings in a metropolitan area, whether owner-occupied or rental. By treating housing quality as a unidimensional index, we are able to capture all features of a dwelling and its environment. Our model is also comprehensive in modeling demand by all households in a metropolitan area. The model accounts for both observed and unobserved differences in preferences over housing, which proves to be essential for generating a realistic demand structure. The equilibrium in our model determines price as a function of quality for each time period in each of the metropolitan areas included in estimation of the model. These price functions coupled with the observed distributions of house values, yield an estimate of the distribution of housing stock by quality in each time period in each metropolitan area, thereby permitting comparison of housing stocks by quantity and quality across metropolitan area. The estimated price functions and type-specific preference parameters also permit calculation of the compensating variation required for a given household in one metropolitan area to be as well off in another metropolitan area. This in turn is informative about agglomeration economies. We thus obtain a flexible equilibrium model with nonlinear pricing of housing quality that provides a compelling explanation of observed patterns within and across metropolitan housing markets.

Our model also embodies a novel treatment of supply. We assume that housing supply depends on the asset price of the unit and not the current rental price, because the asset price of a dwelling is the price to a builder for constructing a dwelling. In addition, we introduce a flexible parametrization of investors expectations over future interest and tax rates as well housing price appreciation. This approach allows us to

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

express the current housing value as the product of the current rental price and a stochastic discount, interpretable as a stochastic user cost of capital.[1] Hence, supply depends on credit market conditions and expectations over future rental prices.

We then develop a new flexible parametric approach to estimate this class of hedonic equilibrium models. Our approach to identification and estimation significantly differs from the previous empirical literature. It builds on four important insights. First, we deviate from standard practice in estimating hedonic or locational equilibrium models by treating both housing quality and housing prices as latent. This approach is convenient because it circumvents the need to decompose quality into an observed and unobserved component. This decomposition typically requires the availability of suitable instruments that are orthogonal to unobserved characteristics. These types of instruments can be hard to find. It also circumvents the need to estimate prices per unit of housing services.[2]

Second, we show that this class of hedonic models can be efficiently approximated by a class of sorting models that use a convenient discretization of the quality space. These sorting models are closely related to the types of models estimated in Epple & Sieg (1999) and others. Note that computing equilibria in these sorting models is much easier than solving for the nonlinear pricing function that arises in hedonic models. Computing equilibria is essential in order to treat housing prices as latent in estimation.

Third, we follow the insights in Landvoigt, Piazzesi, & Schneider (2015) and normalize housing quality using the distribution of rents in a baseline period. This

[1]Our approach to modeling expectation over future housing values extends the work on user-costs of capital by Poterba (1992) and Poterba & Sinai (2008) to models with heterogenous housing types.

[2]See, for example, Sieg, Smith, Banzhaf, & Walsh (2002) for a discussion of the problems that are encountered in that analysis.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

normalization is feasible since rents are a monotonically increasing function of the single index of quality in equilibrium. Housing quality does not have an intrinsic cardinality.

Finally, we provide a new non-parametric method for estimating rent-to-value functions that are consistent with the requirement that these functions depend on housing quality.[3] This allows us to incorporate owner-occupied housing into the empirical analysis without losing internal consistency of the modeling approach. Moreover, it allows us to estimate our supply model.

These four insights equip us to identify and estimate the structural parameters of a broad class of flexible parameterizations of our equilibrium model. The parameters include those of the utility function, the housing supply function, and the distribution of observed and unobserved household types. We develop a two-step, sequential estimation strategy. First, we estimate the rent-to-value function and impute rents for owner-occupied housing. Second, we estimate the structural parameters of the model by matching the observed joint distributions of income and rents conditional on observed characteristics in each metropolitan housing market. In doing so, we also impose the market clearing conditions for each quality and each time period that must hold in equilibrium.

To implement our approach we require data for a representative sample of housing units in multiple 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

[3]An alternative strategy to identify and estimate the rent-to-value function or user cost locus is to use uses observations on units that were both rented and sold within a short period, as developed in Bracke (2015). Halket, Nesheim, & Oswald (2017) estimate user costs when tenure choice is endogenous.

Electronic copy available at: https://ssrn.com/abstract=3216207

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.

We then provide a compelling application that illustrates the power of our new framework. We estimate a number of models using data from Chicago and New York. We find that a model with five unobserved types provides a sufficiently flexible characterization of demand to capture all relevant features observed in the data while also providing quite plausible estimates of parameters of the demand function. In addition, our estimated model of supply is consistent with the observed changes of the housing stock. Finally, we obtain plausible estimates of the elasticity of housing supply.

Estimation of our model does not require any assumptions regarding the extent to which households are mobile among metropolitan areas. However, 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 disamenities 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,

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

as measures of agglomeration economies.

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 Rosen's framework to the study of housing markets.[4]

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 that encompasses all observed and unobserved housing characteristics.[5]

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

[4]An illuminating review of this literature is provided in Kuminoff, Smith, & Timmins (2013).

[5]See, also, Chernozhukov, Galichon, Henry, & Pass (2017) for an extension to multiple dimensions of unobserved heterogeneity.

Electronic copy available at: https://ssrn.com/abstract=3216207

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.

Our work is related to recent work 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.

There have also been important recent advances in the study of the dynamics of housing markets.[6] We build on Landvoigt, Piazzesi, & Schneider (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). An important feature of their framework, which we adopt, is treating housing as a differentiated product that varies along a single dimension. We follow LPS as well in defining housing quality using the value distribution in a baseline period. There are also significant differences between our work and theirs. As noted above, a key contribution of LPS is study of frictions in the market for owner-occupied housing. We abstract from market frictions in order to undertake a unified treatment of both owner-occupied and rental dwellings in a metropolitan area. Our unified treatment of ownership and rental housing includes development of a new approach for estimating the mapping from value to implicit rentals for owner-occupied dwellings. In modeling households, we incorporate both observed and unobserved heterogeneity in order to

[6]See also the discussion in Bayer, McMillan, Murphy, & Timmins (2016) who estimate a dynamic model of household location choice and housing preferences.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

obtain a rich demand model that can account for variation across observed household types in the correlation between income and housing expenditure. We model changes in housing supply while LPS model matching in the market for the existing housing stock. A further difference is that we undertake estimation while LPS use a methods of moments approach to calibrate their model. The above differences in modeling reflect differing objectives of the LPS analysis relative to our analysis. Our framework permits us to pursue several objectives, including comparisons of distributions of housing qualities across metropolitan areas, comparisons of prices as a function of quality across metropolitan areas, and calculation of the compensating variation that would make a household in one metropolitan area equally well off in another. The LPS framework enables them to provide a rich quantitative account of the factors, including market frictions, that drove the housing boom in San Diego while also permitting them to explain the pattern of limited price growth in the years following the boom.

Finally, our work is also related to a large class of locational sorting models. We discuss the relationship between hedonic and locational equilibrium models in more detail below.

The rest of the paper is organized as follows. Section 2 develops our model. Section 3 discusses identification and estimation. Section 4 discusses the data used in the analysis. Sections 5 and 6 provide the empirical results for the housing markets in Chicago and New York and provide a careful cross-metropolitan comparison of the two housing markets. Section 7 offers conclusions and discusses future research.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

2 Housing Markets in Metropolitan Areas

2.1 A Hedonic Model of Housing

We model housing as a differentiated product. Housing units differ by quality, which can be characterized by a one-dimensional ordinal measure denoted by h. We follow the hedonic literature allowing for non-linear pricing in markets for housing services. All households behave as renters, i.e. households who are owner-occupants make decisions about housing consumption using an implicit rental rate that equals the amount the dwelling would command on the rental market.

To simplify exposition, we consider a metropolitan area at multiple points in time, but the model applies equally well to multiple metropolitan areas, or multiple points in time for multiple metropolitan areas. There is a continuum of households with mass equal to one.[7] There are I types of households in the economy. Following Heckman & Singer (1984), we treat these types as unobserved by the econometrician in estimation. Households differ by an observed vector of characteristics x as well as income y. The population share of households with characteristics x at time t is given by $s_t(x)$. We assume that x affects the probability of being each household type i. Hence, there exists a mapping $p_{it}(x)$ that maps observed characteristics x into types i, i.e. the fraction of each type in period t is given by $p_{it}(x)$.

Households also differ in income. We interpret income as a broadly defined measure that includes not just labor income, but also income from asset holdings for wealthy households and transfer income for poor households. [8] Our approach thus

[7]It is straightforward to allow for population growth in our model. In that case, population is given by N_t and $\{N_t\}_{t=1}^{\infty}$ is treated as an exogenous process. With a suitable rescaling of all key equations, the results go through as before.

[8]Broadly defined income measures are available in the AHS which we use in the empirical appli-

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

implicitly accounts for differences in wealth. Let $F_{it}(y)$ be the metropolitan income distribution at time t of type i. Note that the income distributions of unobserved types are linked to the income distributions of observed types by the following identity:

$$F_{it}(y) = \frac{\sum_x s_t(x)\ p_{it}(x)\ F_t(y|x)}{\sum_x s_t(x)\ p_{it}(x)} \tag{1}$$

where $F_t(y|x)$ is the (observed) distribution of income conditional on x.

Households have preferences defined over housing services h and a composite good b. Let $U_{it}(h, b)$ be the utility of a household of type i at time t. We invoke the exclusion restriction that conditional on type i, preferences do not depend on x. 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 housing unit that provides quality h. Here and subsequently the rental price $v_t(h)$ denotes both the rental price of renter-occupied units of quality h and the implicit rental price of owner-occupied units of quality h. We assume that transactions costs are zero. Hence, whether owner-occupant or renter-occupant, a household can change its housing consumption on a period-to-period basis as rental rates change. It follows that, regardless of tenure choice, a household's optimal choice of housing at each date t maximizes its period utility at date t:

$$\begin{aligned} & \max_{h_t, b_t} \ U_{it}(h_t, z_t) \\ s.t.\ y_t \ &= \ v_t(h_t) + z_t \end{aligned} \tag{2}$$

where z_t denotes expenditures on a composite good.

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

$$m_{it}(h_t, y_t - v_t) \ \equiv \ \frac{U^h_{it}(h_t, y_t - v_t)}{U^z_{it}(h_t, y_t - v_t)} \ = \ v'_t(h_t) \tag{3}$$

cation.

Electronic copy available at: https://ssrn.com/abstract=3216207

Solving this expression yields housing demand functions, $h_{it}(y_t, v_t(h))$. Integrating over the income distribution yields the aggregate housing demand of each type $H^d_{it}(h|v_t(h))$:

$$H^d_{it}(h|v_t(h)) = \int_0^\infty 1\{h_{it}(y, v_t(h)) \ \leq h\} \, dF_{it}(y) \tag{4}$$

where $1\{\cdot\}$ denotes an indicator function. Total housing demand is given by:

$$H^d_t(h|v_t(h)) = \sum_x s_t(x) \sum_{i=1}^{I} p_{it}(x) \ H^d_{it}(h|v_t(h)) \tag{5}$$

Thus $H^d_t(h|v_t(h))$ is the fraction of households 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 1 *The utility function for each type of household i satisfies the following single-crossing condition:*

$$\frac{\partial m_{it}}{\partial y}\Big|_{U_{it}(h, y - v(h)) = \bar{U}} > \ 0 \tag{6}$$

Assumption 1 states that, for each type, high-income households are willing to pay more for a higher quality house than low-income households – a weak restriction on preferences. Single-crossing implies that, in equilibrium, the house rental expenditure at date t by income y of type i must satisfy:

$$F_{it}(y) \ = \ G_{it}(v) \tag{7}$$

where $G_{it}(v)$ denoted the distribution of rents of type i. The single-crossing condition then implies the following result.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Proposition 1 *If $F_{it}(y)$ is strictly monotonic, then there exists a monotonically increasing function $y_{it}(v)$ which is defined as*

$$y_{it}(v) = F_{it}^{-1}(G_{it}(v)) \tag{8}$$

Note that $y_{it}(v)$ fully characterizes household i sorting in equilibrium. In equilibrium rental expenditures are a deterministic function of income conditional on each type. Hence the model predicts that the correlation between income and rental expenditures is almost one conditional on unobserved type. In the data we observed that income and rental expenditures are positively, but far from perfectly correlated. We use unobserved household types to explain this feature of the data. We will show below that the model is sufficiently flexible to generate realistic correlation patterns conditional on observed type.

To close the model, we need to specify the supply of housing units. We assume in the baseline model that the housing stock is owned by institutional investors who provide an exogenous supply given by $R_t(h)$. We extend the model and allow for endogenous housing supply below. In equilibrium the hedonic rental price function, $v_t(h)$, must clear the market for occupancy at each value of h, i.e. for each level of housing quality h:

$$H_t^d(h|\, v_t(h)) = R_t(h) \tag{9}$$

Define $p_t(x) = (p_{1t}(x), ..., p_{It}(x))$ and $F_t(y) = F_{1t}(y), ..., F_{It}(y)$. Summarizing the discussion above, an equilibrium for the model with exogenous housing supply can be defined as follows.

Definition 1 ***Equilibrium with Exogenous Housing Supply***
Given an exogenous process $\{s_t(x), p_t(x), F_t(y), R_t(h)\}_{t=1}^{T}$, an equilibrium of this model consists of sequence of rental price functions $v_t(h)$ and allocations for each household

Electronic copy available at: https://ssrn.com/abstract=3216207

type such that a) households behave optimally; and b) the housing market for each h at each point of time t clears.

Note the equilibrium in this model can also be computed period by period.

2.2 Equilibrium with Endogenous Housing Supply

Next we introduce two important elements that add further realism to our model. First, we assume that housing supply depends on the asset price of the unit. Previous studies typically assume that housing supply depends on the current rental price. This formulation reflects the fact that home builders produce and sell dwellings and hence are concerned about the market value of the dwelling and not implicit rent. Second, we introduce a convenient parametrization of investors expectations over future interest and tax rates as well housing price appreciation to express the current housing value as the product of the current rental price and a user cost of capital type stochastic discount.

Recall that the standard approach in the hedonic literate links the supply of housing to current rental price function $v_t(h)$. We deviate from this approach and assume instead that the supply of housing depends on the current and the previous period asset price of housing, denoted by $V_t(h)$ and $V_{t-1}(h)$, respectively. Moreover, our supply model allows for potential adjustment costs in the housing stock. In particular, we make the following assumption.

Assumption 2 *Let $q_t(h)$ denote the density of housing of quality h at date t. Taking the distribution of housing quality as given at time 0, the density of housing evolves over time according to the following law of motion:*

$$q_t(h) = f(q_{t-1}(h), V_t(h), V_{t-1}(h)) \tag{10}$$

Electronic copy available at: https://ssrn.com/abstract=3216207

Including lagged values of quantity and price serves to capture potential adjustment costs. The distribution of housing supply in period t is:

$$R_t(h) = \int_0^h f(q_{t-1}(h), V_t(h), V_{t-1}(h))\, dx \tag{11}$$

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. We thus obtain a recursive relationship governing the evolution of the supply of housing over time.

Finally, we endogenize the sequence of asset values $\{V_t(h)\}_{t=1}^T$ by explaining asset prices by the current endogenous prices of rental units and an exogenous sequence of the "user cost of capital." In particular, we assume that at each point of time, there is an asset market in which institutional 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. 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 depreciation and maintenance that occurs with rate δ_t.

Assumption 3 *Investors discount housing assets at a rate that reflects the perceived financial market risk of housing assets. The market for housing assets is competitive.*

The expected profit, 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:

$$E_t[\Pi_t(h)] = E_t\left[-V_t(h) + v_t(h) + \frac{V_{t+1}(h)(1-\tau_{t+1}^p - \delta_{t+1})}{1+i_t}\right] \tag{12}$$

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

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.[9]

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

$$0 = E_t \left[-V_t(h) + v_t(h) + \frac{V_{t+1}(h)(1 - \tau^p_{t+1} - \delta_{t+1})}{1 + i_t} \right] \tag{13}$$

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

$$V_t(h) = v_t(h) + \frac{(1 - \tau^p_{t+1} - \delta_{t+1})}{(1 + i_t)} E_t \left[V_{t+1}(h) \right] \tag{14}$$

By successive forward substitution of the preceding, we obtain:

$$V_t(h) = v_t(h) \ + \ E_t \sum_{j=1}^{\infty} \ \beta_{t+j} \ v_{t+j}(h) \tag{15}$$

where the stochastic discount factor is given by:

$$\beta_{t+j} = \prod_{k=1}^{j} \frac{(1 - \tau^p_{t+k} - \delta_{t+k})}{(1 + i_{t+k-1})} \tag{16}$$

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:

$$\tilde{\beta}_{t+j}(h) = \prod_{k=1}^{j} \frac{(1 - \tau^p_{t+k} - \delta_{t+k}) \ (1 + \pi_{t+k}(h))}{(1 + i_{t+k-1})} \tag{17}$$

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

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Then:

$$V_t(h) = \frac{v_t(h)}{c_t(h)} \tag{18}$$

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

$$c_t(h) = \frac{1}{1 + E_t \sum_{j=1}^{\infty} \tilde{\beta}_{t+j}(h)} \tag{19}$$

$c_t(h)$ is thus a summary statistic that aggregates the impact of future interest rate and tax payment as well as expectations over future rental income growth into one single number. Given $c_t(h)$, $V_t(h)$ is a function of $v_t(h)$.

Assumption 4 *The sequence of user costs $\{c_t(h)\}_{t=1}^{T}$ is exogenously determined.*[10]

Note that this idea is similar to Landvoigt et al. (2015) who also parametrize expectations over future housing prices as functions of current prices. Given an exogenous process for the user cost of capital denoted by $\{c_t(h)\}_{t=1}^{T}$ we can extend the definition of equilibrium of our model as follows.

Definition 2 ***Equilibrium with Endogenous Housing Supply and Exogenous User Costs***

Given an exogenous process $\{s_t, p_t(x), F_t(y), c_t(h)\}_{t=1}^{T}$, an equilibrium of this model consists of a sequence of rental price function $v_t(h)$ and allocations for each household type such that a) households behave optimally; b) supply satisfies the recursive formulation in equation (11); c) asset prices satisfy equation (18); and d) the market for each quality h at each point of time t clears.

[10] 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.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Note that the equilibrium can still be computed period by period in each market since we treat the expectations over the future asset prices as given.

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 asset price expectations.

2.3 The Relationship between Hedonic and Sorting Models

At this stage of the analysis it useful to compare the hedonic model discussed above with a sorting model along the lines discussed Epple & Sieg (1999). On the demand side, a key difference between the two approaches is the characterization of the choice set. In the hedonic model, we treat choices as continuous while in the sorting model we treat the choice set as discrete and finite. To show the close link between hedonic and sorting models, let us define a grid for housing quality variable, $h_1, ..., h_J$. As the grid gets finer and finer, the discretized sorting model will more closely approximate the continuum of choices available in a hedonic model.

Similarly, we can discretize the supply of housing. Let r_{jt} be the fraction of units in each quality bin h_{jt} available at time t:

$$r_{jt} = R_t(h_j) - R_t(h_{j-1}) \tag{20}$$

Finally, we define the pricing function for the discrete model as follows:

$$v_{jt} = v(h_{t,j}) \tag{21}$$

Define the vector of prices $v_t = (v_{1t}, .., v_{Jt})$.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Sorting implies that there exists type specific cut-off points $\hat{y}_{ijt}$ such that the demand by type i for quality h_j in period t as a function of rental prices is given by

$$H_{ijt}(v_{1,t},...,v_{jt}) \;\; = \;\; F_i(\hat{y}_{ijt}) \; - \; F_i(\hat{y}_{i,j-1,t}) \tag{22}$$

where $\hat{y}_{ijt}$ is defined as the level of income such that household type i is indifferent between consuming quality h_j and quality h_{j+1} at their corresponding prices in period t, i.e. the level that satisfies the following for each type i:

$$U_i(h_j, \hat{y}_{ijt} - v_{jt}) \;\; = \;\; U_i(h_{j+1}, \hat{y}_{ijt} - v_{j+1t}) \tag{23}$$

Equilibrium requires that demand equals supply for each housing type at each point of time. Hence, housing prices are the solution to the following system of nonlinear equations:

$$r_{jt}(v_t) \;\; = \;\; \sum_x \; s_t(x) \sum_i p_{it}(x) \; H_{ijt}(v_t) \quad \forall j, t \tag{24}$$

Note that there are four important differences between this model and the model considered in Epple and Sieg (1999) (ES). First, the model in ES was designed to investigate household sorting in a system of local jurisdictions and the provision of local expenditures on public goods via majority rule within each community, including estimation of household preferences and the variation in housing prices across jurisdictions. The objectives of the current model are to estimate the equilibrium distribution of house qualities at each point in time within each metropolitan area, to estimate the associated variation of price with quality, to compare the distributions of prices and qualities over time both within and across metropolitan areas, and to calculate the compensating variation required for households of a given income and type to be equally well off in different metropolitan areas.

Second, ES explicitly differentiated between housing services and local neighborhood characteristics, with local neighborhood characteristics treated as observed. The

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

ES approach is well suited to comparisons across jurisdictions within a metropolitan area. Here we bundle neighborhood and housing characteristics together in one variable, h, which we treat as latent. This approach is well suited to comparisons across metropolitan areas. It permits a comprehensive measure of the services provided by a dwelling and its environment while circumventing the potentially intractable task of cataloguing, measuring, and comparing the factors that make a location in one metropolitan equivalent to a location in another metropolitan area.

Third, most previous hedonic or locational choice models treat the housing stock as fixed over time or, as in ES, assume that aggregate housing supply in a community is a function of the current housing price prevailing in the community.[11] Our current approach seeks to to capture durability and heterogeneity of housing while also capturing dynamics of adjustment of housing stock. Finally, we use discrete types to capture preference heterogeneity in this model, while ES used continuous distributions to capture preference heterogeneity in preferences for neighborhood amenities and local public goods.

Next we discuss how to identify and estimate this discrete approximation of the hedonic model. As we will see, our approach here is also significantly different from the previous literature including ES.

3 Identification and Estimation

In this section, we develop a new approach for estimating this class of models that differs significantly from previous research. Before, we formalize our approach, we offer four observations to further clarify our contributions.

[11]See, for example, Rosen (1974) or Epple & Romer (1991) who assume that supply changes with current rental prices or Nechyba (1997) who uses models with a fixed housing stock.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

First, our approach to identification and estimation differs significantly from the existing literature on sorting models. Almost all papers that have estimated equilibrium models of location choice models have followed ES in assuming that per unit housing prices and, as a consequence, housing services are observed or can be estimated by the econometrician.[12] The approach outlined in this paper allows us to treat both housing prices and housing qualities as latent.[13] Identification of the model requires the observation of equilibria in multiple markets. As we will illustrate below, we can either observe a) equilibria for the same metropolitan area over time, b) equilibria in a cross section of metropolitan areas, or c) both types of equilibria. In the use of data for multiple markets, our identification strategy is related to the early literature on identification of hedonic models, but, in contrast to that literature, we invoke all equilibrium implications of our model in estimating the model.

Second, locational sorting models as well as hedonic models typically assume that all relevant neighborhood characteristics are observed up to an idiosyncratic shock that is well behaved, i.e. there exists a set of instruments that are independent of the unobserved neighborhood characteristics. Our approach does not distinguish between housing services and neighborhood amenities and, therefore, does not require such a strong assumption on unobserved neighborhood characteristics. Hence, we do not need instruments to differentiate between a house and associated neighborhood amenities. As we demonstrate below, there are important applications where such a decomposition is not essential.

Finally, we provide a new method for imputing rents for owner-occupied housing. We implement a simple non-parametric method that is consistent with our model and

[12]The most popular approach for estimating cross-sectional housing prices is outline in Sieg et al. (2002).

[13]This idea is also used in Landvoigt et al. (2015).

Electronic copy available at: https://ssrn.com/abstract=3216207

allows user-costs to depend on latent housing quality.

3.1 A Parametrization of the Model

To identify and estimate the model, it is necessary to introduce flexible parameterizations of the key functions of interest.

Assumption 5 *Let the utility provided by housing quality h for household type i at each period t be:*

$$U = u_i(h) + \frac{1}{\alpha_i} \ln(y_t - v_t(h)) \tag{25}$$

with $u(h) = \ln(1 - \phi_i(h + \eta_i)^{\gamma_i})$, where $\alpha_i > 0$, $\gamma_i < 0$, $\phi_i > 0$, and $\eta_i > 0$.

This utility function requires the following two assumptions be satisfied $1 - \phi_i(h + \eta_i)^{\gamma_i} > 0$ and $y_t - v_t > 0$. We impose these structural constraints in estimation. Note that we also assume that the parameters of the utility function are time-invariant.

This utility function gives rise to flexible forms for the price and income elasticity equations. It also allows us to obtain a closed-form solution for the income cutoffs in (22):

$$\hat{y}_{ijt} = \frac{v_{jt} - e^{(M_{i,j+1} - M_j)\alpha_i}\; v_{j+1,t}}{1 - e^{(M_{i,j+1} - M_j)\alpha_i}} \tag{26}$$

where $M_{i,j} = \ln(1 - \phi_i(h_j + \eta_i)^{\gamma_i})$.

Note that our approach to estimation is flexible and does not require this specific specification of utility. It can be applied for many other utility functions as well. We find the specification we have fits our data well and also provides quite reasonable estimates of income and price elasticities.

With respect to the unobserved household types we assume that x is discrete and hence $p_{it}(x)$ can be treated as a discrete distribution. Hence we do not need any

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

functional form assumptions.[14]

Finally, we need to parametrize the law of motion for the housing supply.

Assumption 6 *The law of motion for housing supply is given by:*

$$r_{jt} \quad = \quad A_t \, r_{jt-1} \left(\frac{V_{jt}}{V_{jt-1}} \right)^{\zeta} \quad \forall j, t \tag{27}$$

where A_t is a scalar that guarantees the fractions of housing types sum to one.

This function has some attractive properties. It is parsimonious; it introduces only one additional parameter, ζ. If the value of housing type j rises, the quantity rises as a constant elasticity function of the proportion by which the value increases. If the value of housing type j 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.

3.2 Rent-to-value Functions

In the data, we observe rents for rental units and values for owner-occupied units. We never observe rents and values for the same unit. As a consequence, we have to impose some assumptions to identify the rent-to-value functions.

Assumption 7 *All households are indifferent between renting and owning.*

Consider the simple case with just one preference type. In that case, households with income y consume the same housing quality, independently of whether they live in a rental unit, for which we observe, $v_t(y)$, or live in an owner-occupied unit, for

[14] It is straightforward to extend our approach treating x as a continuous random variable.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

which we observe $V_t(y)$. In order to identify the rent-to-value function, $v_t(V)$, we vary income and trace the equilibrium locus between rents and values for each income level.

With multiple types, the expected rent for a household of type (x, y) is given by:

$$v_t(x, y) = \sum_i p_{it}(x)\ v_t(h_i(y)) \tag{28}$$

and the expected housing value is

$$V_t(x, y) = \sum_i p_{it}(x)\ V_t(h_i(y)) \tag{29}$$

We can estimate these expectations using a non-parametric estimator. By varying y we can trace out the indifference mapping, $v_t^x(V)$, for each observed household type x. Equilibrium requires that these mappings be the same for all x. Hence we define the rent-to-value function, $v_t(V)$, as the following projection:

$$\min_{v_t(V)} \ \sum_x \ s_t(x) \ \int \|v_t(V) - v_t^x(V)\| \ dv \tag{30}$$

For the discretized version of the model, this equation can be written as

$$\min_{v_{1t},..,v_{Jt}} \ \sum_x \ s_t(x) \ \sum_j (v_{jt} - v_{jt}^x)^2 \tag{31}$$

Summarizing the discussion above, we have the following result:

Proposition 2 *The rent-to-value function $v_t(V)$ and its inverse function, $V_t(v)$ are non-parametrically identified.*

Having identified the rent-to-value functions, we can convert values for owner-occupied houses into rents and hence construct the complete rental distribution for each market and each time period.

Electronic copy available at: https://ssrn.com/abstract=3216207

3.3 Latent Housing Quality and Prices

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 3 *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 an arbitrary monotonic transformation of h and redefine the utility function accordingly. The proposition then implies that if we only observe data in one housing market and one time period, we cannot identify the utility function separately from the pricing function. A corollary of this proposition is that we can normalize housing quality by setting $h = v_t(h)$ in one time period t. Recall that Landvoigt et al. (2015) use the same approach for normalizing housing quality in their analysis.

Given our normalization of housing quality in the baseline period, the housing price function is linear in the baseline period in one metropolitan area, and hence observed. For all other periods and all other metropolitan areas, equilibrium prices can be computed as a function of the structural parameters using equation (24).

For some simple models, we can actually compute a closed form solution for the nonlinear pricing functions. In that case we can provide an analytic proof of identification of the structural parameters of both the utility and the supply function.[15] For more realistic models we have to rely on numerical algorithms to compute equilibrium housing prices. As a consequence, we do not have a constructive proof of identification for these models.

[15]See Appendix A for a constructive proof of identification for a simplified version of our model that allows for a closed form solution of the equilibrium pricing function.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Nevertheless, key elements of the logic captured by the proof for the simpler models carry over to more general model specifications. In particular, the predicted demand for each period depends on the latent equilibrium prices, which in turn depend on the supply elasticity. There is then a natural exclusion restriction that helps with identification. Given our normalization of housing quality and prices in the first period, the sorting in the first period only depends on the preference parameters. Moreover, supply is observed in the first period in one metropolitan area. In contrast, sorting in all subsequent periods and all other metropolitan areas also depends on the parameters of the housing supply function. This exclusion restriction that follows from our normalization of quality in one market in one time period allows us to disentangle demand from supply parameters, which is the most fundamental problem encountered in the analysis.

Of course, another main source of identification of the supply function comes from exogenous changes in the user costs of capital that are captured by the rent-to-value functions that we observe. Changes in the user cost of capital drive a wedge between rents and values. Demand in our model only depends on rents, while supply depends on values and as a consequence on the user cost of capital. A decrease in the user cost of capital then leads to an increase in values holding rents constant. A decrease in the user cost, therefore, affects supply, but not demand. This additional exclusion restriction also helps us distinguish between demand and supply side parameters.

The main difference between the model with and without unobserved types is the following. In the model with only observed types, housing demand is a deterministic function of income for each observed type. As a consequence, these models have difficulties explaining the correlations of incomes and rents observed in the data. The main purpose of adding unobserved types to the model specification is to generate more realistic correlation patterns between income and rent. In that sense unobserved

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

heterogeneity in housing plays the same role as the unobserved heterogeneity for public goods in ES, which is needed to generate realistic household sorting patterns by income among neighborhoods and communities.

3.4 A Method of Moments Estimator

The intuition behind our estimator is then the following. First, we estimate the rent-to-value function and impute rents for owner occupied housing. We then estimate the remaining structural parameters by matching the observed joint distributions of income and rents conditional on observed characteristics. More formally, we estimate the structural parameters of the model using the following estimation algorithm:

1. Estimate for each t the rent-to-value function, $v_t(V_t)$, and its inverse, using a non-parametric matching estimator.

2. Use the rent-to-value function to impute rents for owner-occupied units, obtain the market distribution for rents, and compute the joint distributions for income and rent conditional on x, denoted by $F_t^N(y, v|x)$.

3. Discretize the rent distribution in period 1 into J intervals indexed by j, and normalize housing quality to obtain h_j and r_{j1} for all j.

4. Choose a vector of structural parameters, denoted by θ, which includes the parameters of the type distribution and the parameters of the income distributions for each unobserved type.

5. Solve for the implied equilibrium prices in all periods $t > 1$.

    (a) Guess values of v_{jt}.

Electronic copy available at: https://ssrn.com/abstract=3216207

(b) Calculate the implied income cutoffs $\hat{y}_{ijt}$ such that

$$U_i(h_j, \hat{y}_{ijt} - v_{t,j}) = U_i(h_{j+1}, \hat{y}_{i,j} - v_{t,j+1}) \;\; \forall j, it \tag{32}$$

Note that these cutoff points depend on $F_{it}(y|\theta)$.

(c) Calculate the implied demands:

$$H^d_{ijt}(v_t) = F_{it}(\hat{y}_{ijt}) \; - \; F_{it}(\hat{y}_{i,j-1,t}) \;\; \forall j, it \tag{33}$$

(d) Calculate the supplies:

$$r_{jt}(v_t) = A_t \; r_{jt-1} \; \left(\frac{V_{jt}(v_{jt})}{V_{jt-1}(v_{jt-1})} \right)^{\zeta} \quad \forall j, t > 1 \tag{34}$$

(e) Check whether equilibrium holds

$$\sum_x \; s_t(x) \; \sum_{i=1}^{I} p_{it}(x) \; H^d_{ijt}(v_t) = r_{jt}(v_t) \quad \forall j, t > 1 \tag{35}$$

(f) Repeat until solution of equilibrium prices has been found for each time period.

6. For each unobserved type i compute the predicted joint $F^{\theta}_{it}(y, v)$ as well as:

$$F^{\theta}_t(y, v|x) = \sum_i p_{it}(x) \; F^{\theta}_{it}(y, v) \tag{36}$$

7. Form orthogonality conditions that are based on the difference between the observed joint distributions of income and rents, denoted by $F^N_t(y, v|x)$, and their predicted counterparts, $F^{\theta}_t(y, v|x)$. Here we use quantiles of the marginal distributions and correlations between y and v conditional on x. Evaluate the objective function $Q^N_t(\theta)$.

8. Update θ until $Q^N_t(\theta)$ is minimized.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

We use a standard bootstrap procedure to estimate the standard errors. We use clustering algorithms to reduce the dimensionality of x. By combining clustering strategies with unobserved heterogeneity along the lines suggested by Heckman and Singer (1984) we obtain a fairly parsimonious model that is sufficiently flexible to capture all relevant dimensions of sorting observed in the data.

4 Data

We use 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.[16]

Fortunately, the AHS conducted surveys in both Chicago and New York for both 1999 and 2003. We exploit data from Chicago for our first application to illustrate our new method. We then use both surveys jointly estimating the model for those two metropolitan areas. One of the most advantageous features of our model is its

[16]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.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

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.

Households also differ in income. We interpret income as a broadly defined measure that includes not just labor income. Income in the AHS is based on the respondent's reply to a number of detailed questions about different income categories for the 12 months before the interview. Income is the sum of the amounts reported for wage and salary income, net self-employment income, Social Security, public assistance or welfare payments, and all other money income, which includes income from assets. The figures represent the amount of income received before deductions for personal income and payroll taxes as we as deductions. Accuracy of income data has been studied for this survey. Although individual over- and under-reporting has been found, the accuracy is similar to information included in other surveys such as the Current Population Survey.

AHS definitions of each of these metropolitan areas is unchanged across these two periods.[17] We use the Chicago 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 model, we reduce the dimensionality of potential household types using k-means clustering. This is a standard method in data mining initially used by MacQueen (1967) as a method to group observations into clusters based on some similarity measurement criterion. In a two-dimensional space, the method

[17] 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.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

is intuitive. First, K points are selected randomly and then each observation is assigned to its nearest point. Next, centroids are computed for the K and observations are reassigned to the nearest centroid, which will cause some observations to move from their original clusters to a new cluster. The algorithm iterates until no more reallocation of observations into clusters happens.

It is well-known that k-means clustering results are sensitive to the choice of initial points, i.e. the local minimum determined by the iterative procedure is not necessarily a global minimum. The usual practice to address this problem is to perform the algorithm multiple times with different randomly selected starting points and select the solution with the smallest squared error (Bernhardt and Robinson, 2007). We follow this procedure and use 10,000 repetitions of the algorithm. This iterative partitioning minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances.

Table 1: k-means clustering centroids

Chicago			
Cluster	# Children	Age	$s_t(x)$
1	0.287	29.13	0.260
2	1.390	45.00	0.433
3	0.299	71.00	0.307
Chicago & New York			
Cluster	# Children	Age	$s_t(x)$
1	0.345	21.50	0.256
2	1.660	46.50	0.475
3	0.314	69.49	0.268

The results presented here are for squared Euclidean distances and a multidimen-

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

sional data matrix with variables that influence housing consumption.[18] This method allows us to reduce dimensionality based on observable characteristics that are relevant to the preferences for housing without imposing ad hoc thresholds, using an algorithm that is efficient in large datasets and easy to implement. When we implement the cluster analysis for our two sample, we obtain three clusters, which is the optimal number of clusters calculated with our sample data employing the silhouette criterion.[19]

The top panel of Table 1 shows estimated cluster shares and centroids for Chicago pooling the data for 1999 and 2003. 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. To estimate the model jointly, we also conducted a joint cluster analysis using data for both cities. The results are shown in the bottom panel of Table 1 . The results for the two samples are very similar.

5 Empirical Results

Here we report the empirical results when we estimate the model using data for Chicago in 1999 and 2003. We then turn to joint model of Chicago and New York in the next section.

[18]Christopher (2016) provides more details of this method. We used the Matlab package kmeans.

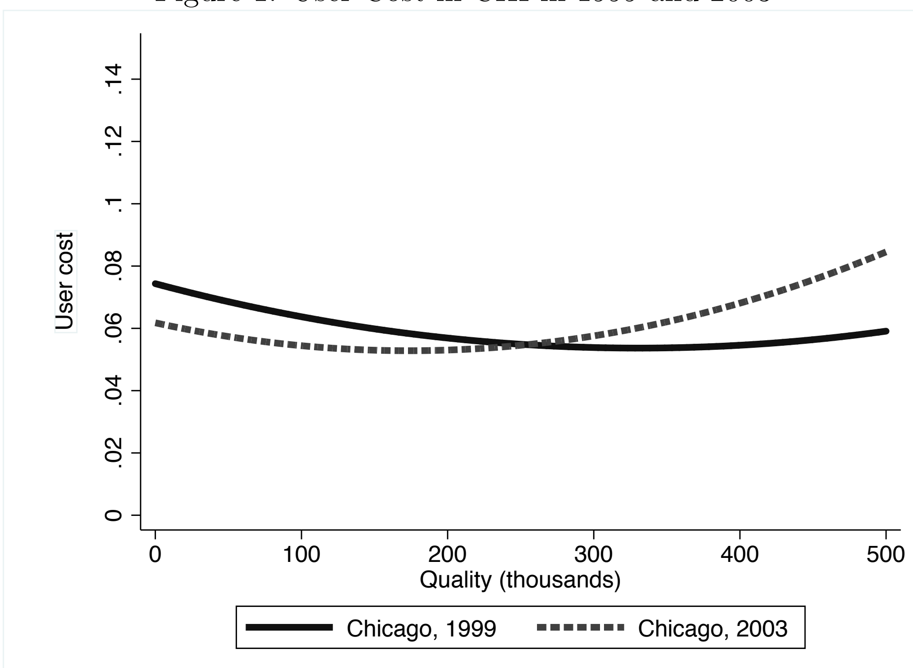
[19]The silhouette of a data point is a measure of the average distance to other points in its cluster less the average distance to points in the alternative closest cluster, normalized by the maximum of both distances (Kaufman and Rousseeuw, 2009).

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

5.1 User Costs

To incorporate owner occupied housing into our analysis we need to convert values into imputed rents. We estimate the user-cost ratios as a function of quality for each household type. We find that the estimates are similar across types. We then construct a single user-cost function as discussed in the previous section. (Note that we also need to these user cost functions to estimate our supply function as discussed above.) The results are illustrated in Figure 1.

Figure 1: User Cost in CHI in 1999 and 2003



We find that the average user cost is approximately 0.06 in 1999 in Chicago. Between 1999 and 2003 the user costs moderately decreased for low quality units and increased for higher quality units.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

5.2 Parameter Estimates

We then estimate the equilibrium model using data for the Chicago metropolitan area. We estimated different versions with up to five unobserved types of households. As we explain in detail below, our preferred model specification has five unobserved types. Table 2 shows the estimated probabilities of each type and the estimated standard errors.

Table 2: Unobserved Type Probabilities $p_{it}(x)$.

$p_{it}(x)$	x_1	x_2	x_3
i_1	0.41	0.08	0.04
	(0.01)	(0.05)	(0.02)
i_2	0.07	0.50	0.17
	(0.02)	(0.07)	(0.02)
i_3	0.03	0.03	0.65
	(0.00)	(0.05)	(0.01)
i_4	0.40	0.09	0.10
	(0.01)	(0.04)	(0.04)
i_5	0.09	0.30	0.04
	(0.03)	(0.05)	(0.02)

Table 3 shows the implied probabilities of each observed-unobserved pair in the population as well as the marginal probability for each unobserved type.

Our results imply that observed type 1 is primarily matched into unobserved types 1 and 4 while observed type 2 is primarily matched into types 2 and 5. Observed type 3 is matched primarily into unobserved types 2 and 3.

The parameter estimates and estimated standard errors for the preference of the

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Table 3: Implied Probability of Type (i, k)

π_{ikt}	x_1	x_2	x_3	all
i_1	0.107	0.035	0.012	0.154
i_2	0.018	0.217	0.052	0.287
i_3	0.008	0.013	0.200	0.220
i_4	0.104	0.039	0.031	0.174
i_5	0.023	0.130	0.012	0.166

Table 4: Preference Parameter Estimates

	α	ϕ	η	γ
i_1	1.22	9.29	0.42	-1.77
	(0.32)	(3.32)	(0.10)	(0.34)
i_2	1.61	4.32	1.19	-1.22
	(0.23)	(3.10)	(0.19)	(0.25)
i_3	2.12	1.15	5.71	-1.618
	(0.30)	(1.61)	(0.22)	(0.39)
i_4	1.11	9.11	0.99	-1.51
	(0.29)	(5.21)	(0.18)	(0.33)
i_5	1.33	7.77	1.98	-1.99
	(0.42)	(5.51)	(0.33)	(0.23)

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

different types are reported in Table 4. The bootstrapped standard errors account for the sequential nature of our estimator. We find that all parameter estimates have the correct algebraic signs and have reasonable magnitudes.

We find that the unobserved types have different preference parameters and hence different housing demand. Since each observed type matches into more than one unobserved type, our model can generate realistic correlations between income and housing demand as explained in detail below.

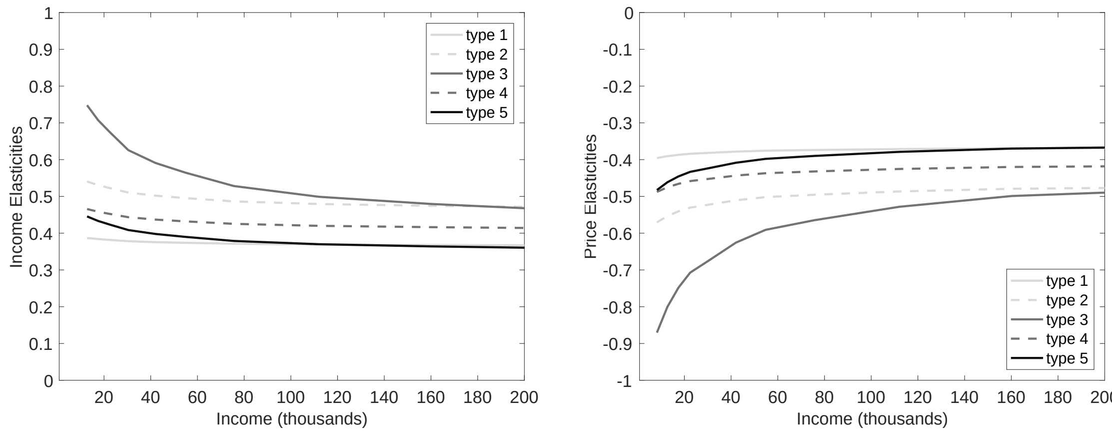
Price and income elasticities implied by our preference parameters estimates are plotted as a function of income in Figure 2. We find that income elasticities vary across type, with Types 1 and 5 having the lowest income elasticities and Type 3 the highest. With the exception of Type 3, income elasticities do not vary much across income within type. For Type 3, the income elasticity declines from 0.8 to 0.5 as income rises. We also find variation across type in price elasticities and modest variation across income within type, except for Type 3. Types 1 and 5 exhibit the lowest price sensitivity. Type 3 exhibits the greatest sensitivity to price and has declining price sensitivity as income rises.

Recall that the changes over time in the supply of stock of a given quality depend on changes in values and on the estimated supply elasticity. Our point estimate for the annual supply elasticity from the model with five household types is 0.072 with a standard error of 0.018.[20] The implied supply growth rates of quality from our parameter and value estimates are between 4.05% and 6.70% for the 4 year period, which correspond to average annualized changes of approximately 1.1% and 1.6% respectively. The estimated annual growth in total number of units was 5.48% 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

[20]We obtain a similar estimate when we estimate the joint for Chicago and New York.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Figure 2: Income and Price Elasticities



estimates are broadly consistent with the estimates of supply elasticities summarized in Glaeser (2004).

5.3 Goodness of Fit

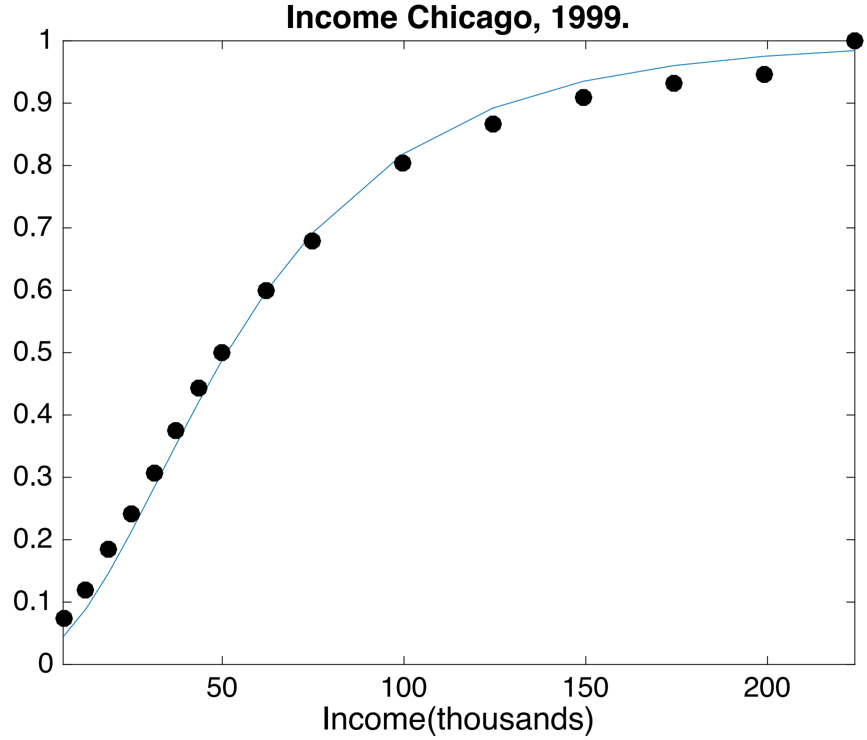
Figures 3 and 4 show the fit of model to the income and rent distributions. The fitted lines are from structural estimation that imposes all equilibrium conditions. Figure 3 shows the fit to income distributions in CHI in 1999 and 2003 respectively, while Figure 4 shows the rental distributions in 1999 and 2003. These graphs illustrate that the model fit of the income and rent data is quite good in both time periods.

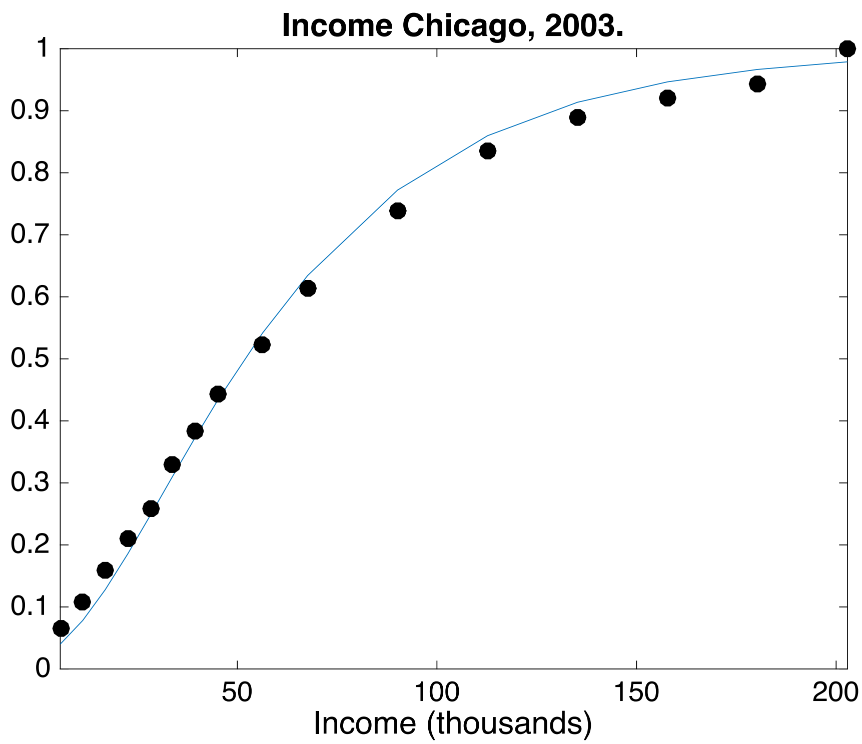
Table 5 provides a detailed analysis of the observed and the predicated correlations between income and rent. We report both unconditional correlations for each year as well as correlations conditional on observed type for each year. Standard errors are reported in parentheses.

Recall that the correlation between income and rent conditional on unobserved type is high. It typically exceeds 0.92. However, each observed type is mapped in

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

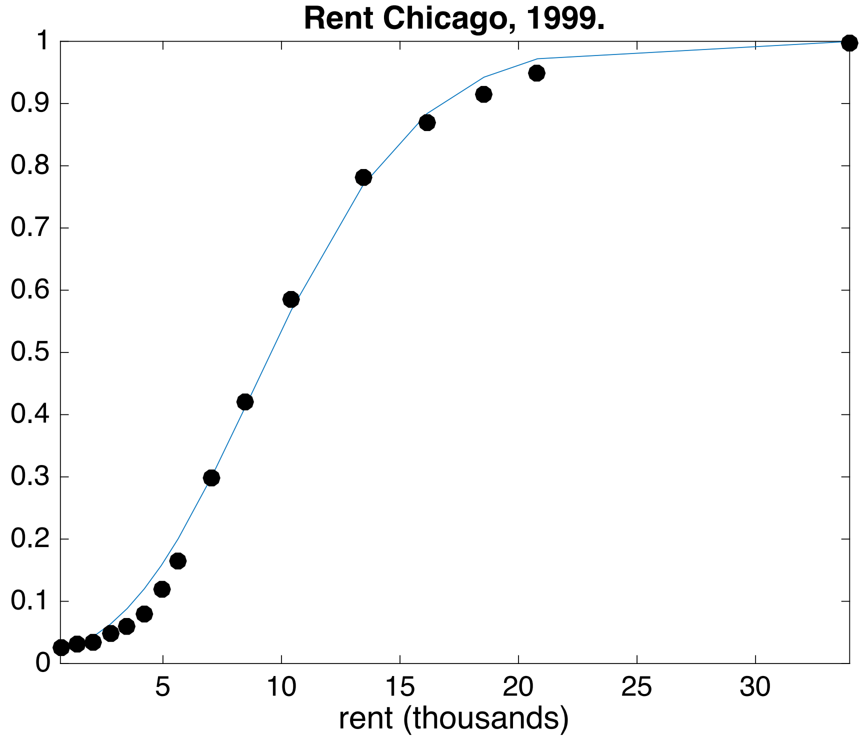
Figure 3:

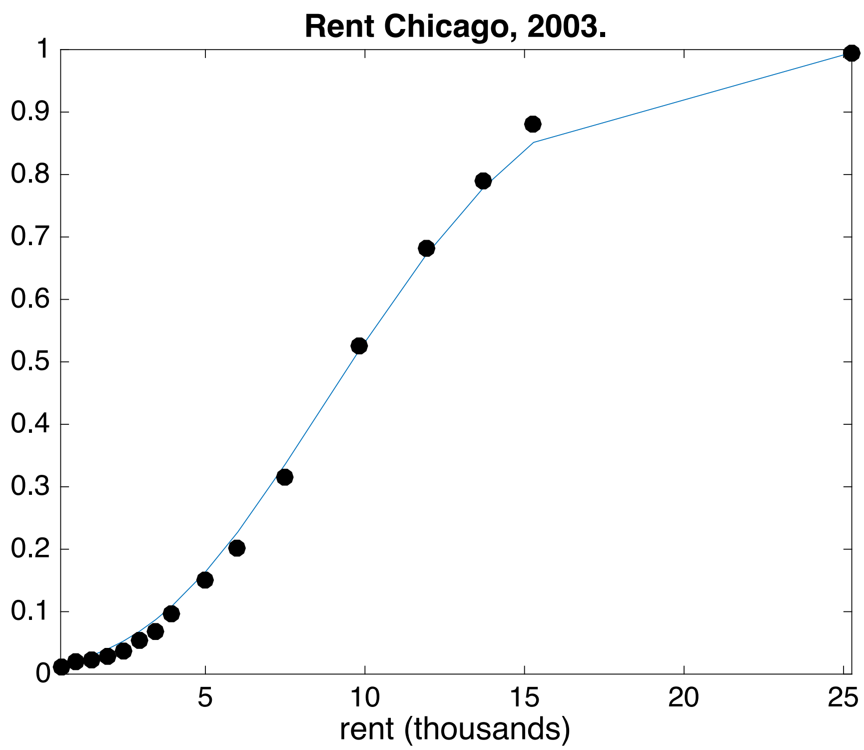




Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Figure 4:





Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Table 5: Rent and Income Correlations

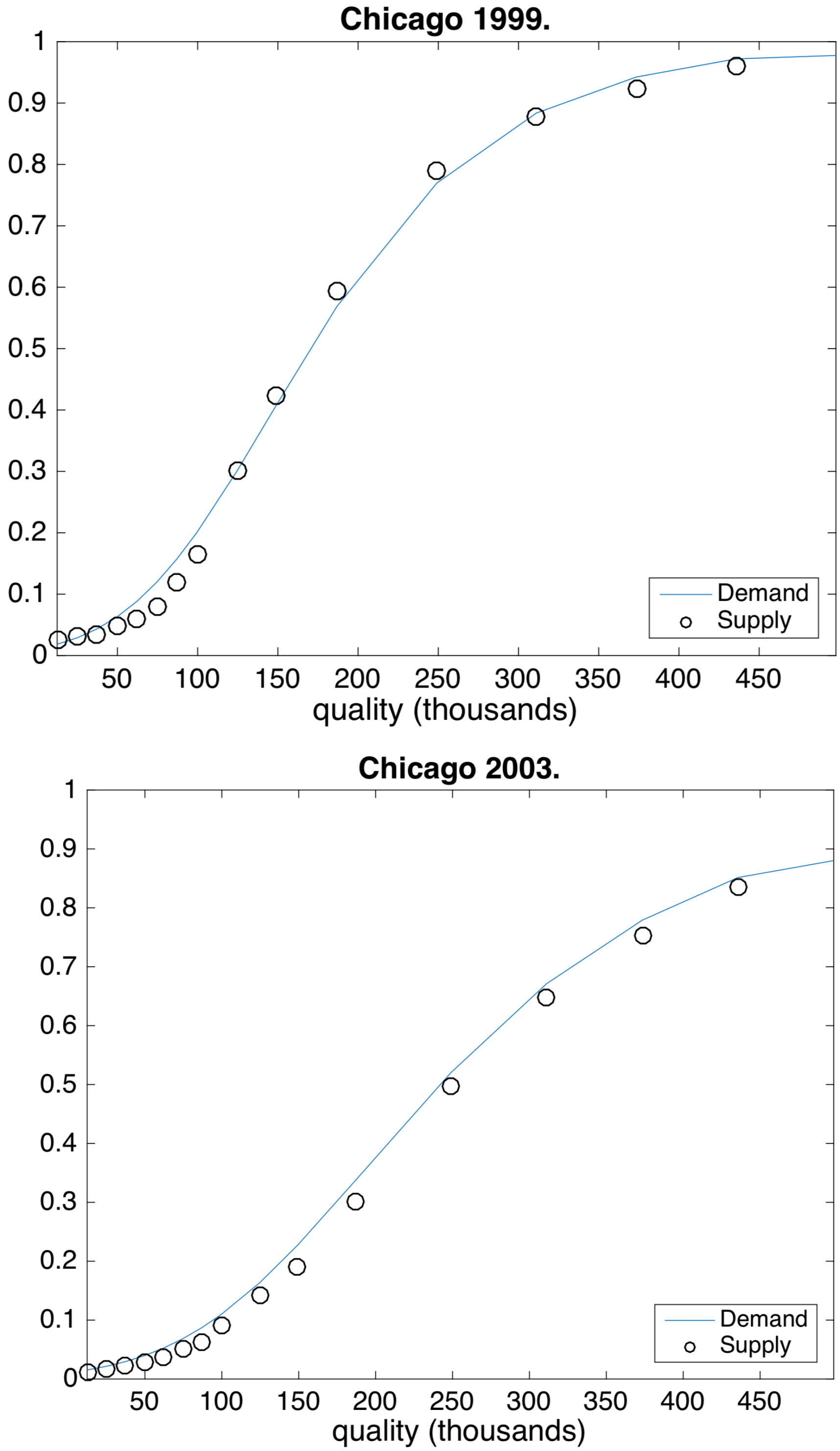
	1999			
Observed Type (x_k)	1	2	3	All
Data	0.27	0.40	0.54	0.40
Model	0.27	0.43	0.41	0.38
	(0.01)	(0.08)	(0.01)	(0.04)
	2003			
Observed Type (x_k)	1	2	3	All
Data	0.50	0.32	0.32	0.42
Model	0.31	0.41	0.39	0.37
	(0.02)	(0.07)	(0.01)	(0.04)

more than one unobserved type as shown in Table 3. As a consequence, the model can generate much lower correlations between income and rents conditional on observed types.

Overall, we find that our model explains the unconditional correlation between income and rental expenditures in both periods reasonably well. We also explain the average correlation conditional on observed type. To obtain reasonable correlation structures we need at least four types. Going from four to five types improves the fit significantly. Allowing for more than five types does not significantly improve the overall fit. Our model does not perform as well period by period. There are some variation in correlation conditional on type between 1999 and 2003. In our model, we assume that preferences and the fractions of each unobserved type are time invariant. As a consequence, our model faces more difficulties explaining variation in correlations across time.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Figure 5: Supply and Demand Equilibrium



Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

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 graph is for 1999 and the lower graph is for are for 2003. As these graphs illustrate, our approach results in close correspondence of supply and demand over the quality range in Chicago in both time periods.

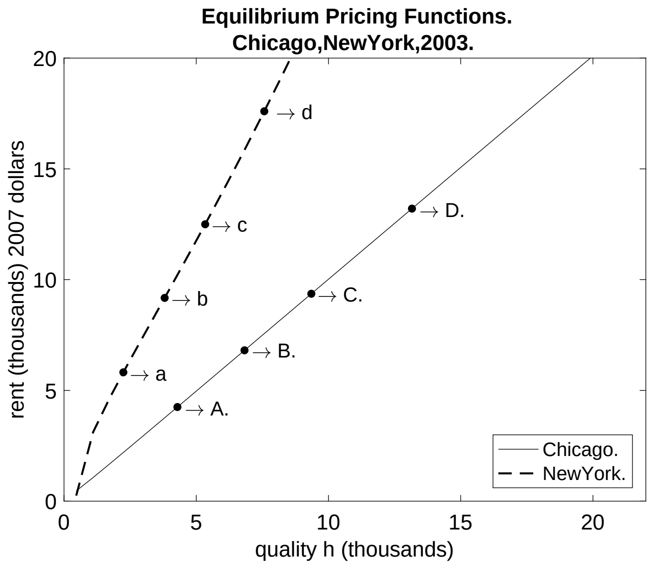
6 Cross-Metropolitan Comparison

In this section we highlight the useful of our approach by comparing the rental markets in New York with those in Chicago. To start the analysis, we estimate our model using data from both metropolitan areas. Tables 6 and 7 in the Appendix B report the parameter estimates and estimated standard errors for the joint model. Overall, we find that our model fits the data in both periods and both metropolitan areas well. If anything joint estimation improves the accuracy of the estimation procedure and yields smaller estimated standard errors. Moreover, we obtain a good fit of the correlations between income and rent by types.

Figure 6 shows the hedonic price functions in 2003 for CHI and NYC. The steeper curve shows the hedonic price functions for NYC while the shallower curve shows the estimates for CHI. 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 New York 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 Chicago. 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

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Figure 6: Equilibrium Price Functions in 2003

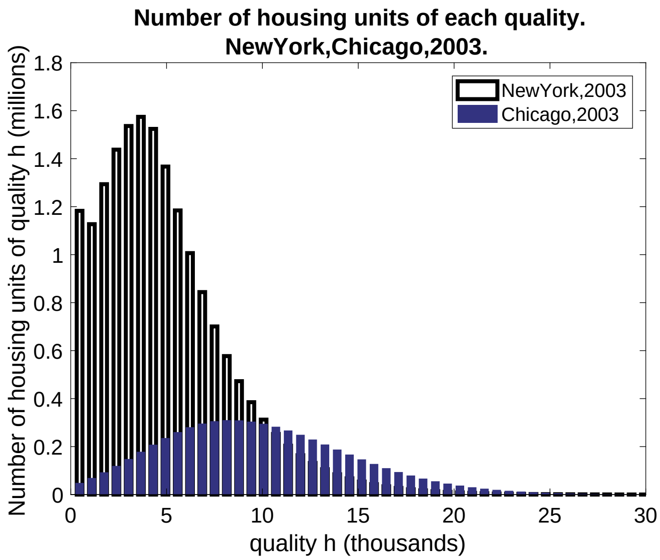


in addition to the housing structure itself.

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 metropolitan areas. The CDF for income for the Chicago metro area is slightly shifted to the right relative to the New York metro area. Median income was slightly higher in Chicago in 1999 and 2003 than in New York. In addition, New York also had more inequality than Chicago. 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.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Figure 7: Housing Stock in NYC and CHI



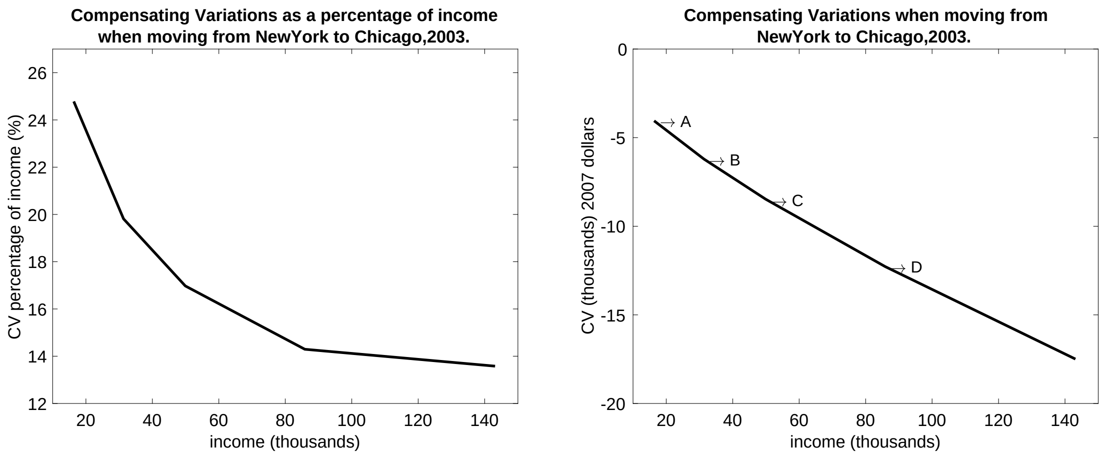
We can compute compensating variations required for a household of a given income in New York to be equally well off in Chicago. Housing at each quality level in the New York metro area is more expensive than in Chicago. To be equally well off in the two metro areas, a given household must then earn more in New York metro area than in the Chicago metro area. Hence, the compensation required in Chicago metro area for a household to be as well off as its counterpart in New York metro area is a measure of the reduction in earnings required in the Chicago metro area, i.e. the compensating variation is negative.

In the left panel of Figure 8, we plot the compensating variation in percent as a function of income. We find that CV as a share of income declines with income, reflecting the declining share of income spent on housing as income rises.[21] In the right

[21]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

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Figure 8: Compensating Variations



panel of Figure 8, we plot the CV aggregated across types. For a household earning \$24,000 in CHI (the 20th income percentile in Chicago), compensating variation of approximately 25% of the household's income (\$4,000) would be required. For a household at the 80th percentile, CV of approximately 14% of income (\$14,000) would be required.

These measures are of interest in their own right and can also be interpreted, under additional assumptions, as measures of agglomeration economies. 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.[22]

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. Extending our approach to address the impact of price stabilization is an important issue for future research.

[22]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 agglom-

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Finally, we can also compute the compensating variation for price changes within cities. We observe, for example, some significant price increases in New York between 1999 and 2003 for most low and medium quality housing units. As a consequence, compensating variations in NYC are positive for most low- and moderate income households and are as high as $800. Hence we conclude that low- and moderate-income households were significantly affected by these rental price increases.

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 our model for Chicago, we find that a model with five unobserved types captures the observed heterogeneity in rental market conditions in 1999 and 2003. Not surprisingly, some differences in housing demand are driven by income. However, there is also much heterogeneity in price and income elasticities for housing

eration economies.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

among unobserved household types. The joint analysis for New York and Chicago provides a contrast of hedonic price functions and house quality distributions across the two metropolitan areas. We also calculate compensating variations that leave households indifferent between the two areas. These estimates can be interpreted as measures of agglomeration externalities.

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 Amsterdam. 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.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

References

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

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

Bayer, P., Ferreira, F., & McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, *115 (4)*, 588–638.

Bayer, P., McMillan, D., Murphy, A., & Timmins, C. (2016). A Dynamic Model of Demand for Houses and Neighborhoods. *Econometrica*, *84*, 893–942.

Bernhardt, D. & Robinson, P. (2007). Detecting affect from non-stylised body motions. In *International conference on affective computing and intelligent interaction*, pp. 59–70. Springer.

Bracke, P. (2015). House Prices and Rents: Micro Evidence from a Matched Dataset in Central London.. *Real Estate Economics*, *43 (2)*, 403–420.

Chernozhukov, V., Galichon, A., Henry, M., & Pass, B. (2017). Single Market Nonparametric Identification of Multi-Attribute Hedonic Equilibrium Models. Working Paper.

Christopher, M. B. (2016). *Pattern Recognition and Machine Learning.* Springer-Verlag New York.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

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

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

Epple, D. & Romer, T. (1991). Mobility and Redistribution. *Journal of Political Economy*, *99(4)*, 828–858.

Epple, D. & Sieg, H. (1999). Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy*, *107 (4)*, 645–681.

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.

Halket, J., Nesheim, L., & Oswald, F. (2017). The Housing Stock, Housing Prices, and Usr Costs: The Roles of Location, Structure and Unobserved Quality. Working Paper.

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.

Kaufman, L. & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, Vol. 344. John Wiley & Sons.

Electronic copy available at: https://ssrn.com/abstract=3216207

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.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, No. 14, pp. 281–297. Oakland, CA, USA.

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

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.

Rezankova, H., Loster, T., & Husek, D. (2011). Evaluation of categorical data clustering. In *Advances in Intelligent Web Mastering–3*, pp. 173–182. Springer.

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).

Electronic copy available at: https://ssrn.com/abstract=3216207

Sieg, H., Smith, V. K., Banzhaf, S., & Walsh, R. (2002). Interjurisdictional Housing Prices in Locational Equilibrium. *Journal of Urban Economics*, *52*, 131–153.

Sieg, H., Smith, V. K., Banzhaf, S., & Walsh, R. (2004). Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods. *International Economic Review*, *45 (4)*, 1047–1077.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

A A Constructive Proof of Identification Based on an Example With Closed Form Solutions

Here we consider a simplified version of our model with continuous quality measures and no unobserved heterogeneity among households. To obtain a closed form solution for the equilibrium pricing function, we impose additional functional form assumptions.

Assumption 8 *Income and housing are distributed generalized log-normal with location parameter (GLN4).*[23]

$$\begin{aligned} \ln(y_t) &\sim GLN4(\mu_t, \sigma_t^{r_t}, \beta_t) \\ \ln(v_t) &\sim GLN4(\omega_t, \tau_t^{m_t}, \theta_t) \end{aligned} \tag{37}$$

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

$$y_t = A_t \left(v_t + \theta_t\right)^{b_t} - \beta_t \tag{38}$$

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

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

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

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 9 *Let utility given by:*

$$U = m_t(h) + \frac{1}{\alpha}\ln(y_t - v_t(h) - \kappa) \tag{39}$$

with $m_t(h) = \ln(1 - \phi(h+\eta)^\gamma)$*, where* $\alpha > 0$*,* $\gamma < 0$*,* $\phi > 0$*, and* $\eta > 0$.[24]

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

$$v_t = \left(A_t \left[1 - \frac{(1 - \phi(h+\eta)^\gamma)^{\alpha(b_t - 1)}}{e^{c_t}} \right] \right)^{\frac{1}{1-b_t}} - \theta_t \tag{40}$$

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

$$v_t = \left(A_t \left[1 - \frac{(1 - \phi(h+\eta)^\gamma)^{\alpha(b_t - 1)}}{e^{c_t}} \right] \right)^{\frac{1}{1-b_t}} - \theta_t \tag{41}$$

Note 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

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

Electronic copy available at: https://ssrn.com/abstract=3216207

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

$$v_t = \left(A_t \left[1 - [1 - \phi(h + \eta)^\gamma]^{\alpha(b_t - 1)} \right] \right)^{\frac{1}{1-b_t}} - \theta_t \tag{42}$$

Setting

$$\alpha = \frac{1}{b_t - 1} \tag{43}$$

implies

$$v_t = (A_t [1 - [1 - \phi(h + \eta)^\gamma]])^{\frac{1}{1-b_t}} - \theta_t \; = \; (A_t \phi(h + \eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{44}$$

Setting

$$\phi = \frac{1}{A_t} \tag{45}$$

implies

$$v_t = ((h + \eta)^\gamma)^{\frac{1}{1-b_t}} - \theta_t \tag{46}$$

Setting

$$\gamma = 1 - b_t \tag{47}$$

implies

$$v_t = (h + \eta) - \theta_t \tag{48}$$

Finally, setting

$$\eta = \theta_t \tag{49}$$

implies.

$$v_t = h \tag{50}$$

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.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

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

$$v_{t+s} = \left(A_{t+s} \left[1 - \frac{(1 - \phi(h + \eta)^{\gamma})^{\alpha(b_{t+s}-1)}}{e^{c_{t+s}}} \right] \right)^{\frac{1}{1-b_{t+s}}} - \theta_{t+s} \tag{51}$$

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}.

B Preference Parameter Estimates for the Joint NYC - CHI Model

The estimates of the parameters of the utility functions of the different household types for the joint NYC and Chicago model are summarized in Table 6.[25] Overall we find that the results are similar to the one we obtained when we just used the Chicago subsample. We hence conclude that preferences for housing are similar in both metropolitan areas. The main difference is that we obtain slightly smaller standard errors for the structural parameters of the utility function which suggests that there are some efficiency gains from pooling across metropolitan areas.

Tables 7 report the estimates for the unobserved types and the implied probabilities for the observed types. We find that a model with five unobserved types fits the data well.

[25] A more detailed summary of the joint estimation results is available upon request from the authors.

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207

Table 6: Preference Parameter Estimates: CHI-NYC

	α	ϕ	η	γ
i_1	1.33	9.31	0.45	-1.91
	(0.42)	(2.00)	(0.08)	(0.14)
i_2	2.34	3.23	1.56	-1.12
	(0.35)	(1.51)	(0.17)	(0.08)
i_3	2.77	2.13	4.78	-1.13
	(0.45)	(0.81)	(0.15)	(0.13)
i_4	1.01	7.32	1.11	-1.43
	(0.50)	(0.51)	(0.15)	(0.11)
i_5	1.87	5.65	2.38	-1.23
	(0.60)	(2.01)	(0.21)	(0.10)

Table 7: Implied probability of types: CHI-NYC

π_{ikt}	x_1	x_2	x_3	agr
i_1	0.096	0.024	0.013	0.134
i_2	0.011	0.299	0.027	0.337
i_3	0.014	0.017	0.147	0.178
i_4	0.100	0.057	0.030	0.187
i_5	0.035	0.078	0.050	0.164

Electronic copy available at: https://ssrn.com/abstract=3216207
Electronic copy available at: https://ssrn.com/abstract=3216207