

Estimating Hedonic Functions for Rents and Values in the Presence of Unobserved Heterogeneity in the Quality for Housing*

Dennis Epple

Carnegie Mellon University

Luis Quintero

Carnegie Mellon University

Holger Sieg

University of Pennsylvania

July 1, 2013

*We would like to thank Pat Bayer, Markus Berliant, Moshe Bushinsky, Morris Davis, Uli Doraszelski, Fernando Ferreira, Matt Kahn, Lars Nesheim, Henry Overman, Monika Piazzesi, Martin Schneider, Paolo Somaini, Xun Tang and seminar participants at various universities and conferences for comments and discussions. Financial support for this project was provided by the NSF grant SES-0958705.

Abstract

A central challenge in estimating models with heterogeneous housing is separating quality from price. We provide a new method for estimating hedonic functions for rental rates and housing values treating housing quality as unobserved by the econometrician. Our method also deals with the problem that rental prices and housing values are partially latent showing that the rent-to-value ratio is non-parameterically identified. We, thus provide an integrated treatment of non-linear pricing rental and asset markets. Our new method can be used to gain some new insights into the causes and effects of the recent housing market crisis in the U.S. Our main application focuses on the housing markets of Miami (Fl) which experienced an average real appreciation of housing values of 65 percent during the period from 2002 to 2007. Our findings are broadly consistent with the view that changes in credit market conditions for low income houses together with changes in investor expectations about future housing prices may have accounted for the large run-up in housing values.

1 Introduction

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

Following Rosen (1974) we consider a standard hedonic equilibrium model with heterogeneous housing units that can be captured by a one dimensional (latent) index.² One of Rosen’s key insights is that the equilibrium pricing function can be characterized by the solution to a nonlinear differential equation. We show that there exists a new flexible parametrization of this model that yields tractable solutions for the equilibrium pricing function. This parametrization exploits generalized log-normal distributions (Vianelli, 1983) which provide good approximations of the observed distributions of outcomes such as income and housing values. Moreover, we

¹This has been an important agenda of hedonic theory and of associated empirical work linking house values to observed house characteristics. Ekeland, Heckman, and Nesheim (2004), Bajari and Benkard (2005), and Heckman, Matzkin, and Nesheim (2010) have established identification of various versions of hedonic models under weak functional form assumption. Closely related is the work by Berry and Pakes (2007) on the "pure characteristics" model.

²Our modeling approach is also related to recent work on nonlinear pricing in housing markets by Landvoigt, Piazzesi, and Schneider (2011) who study the impact of credit constraints on house ownership.

can derive a closed form solution to the equilibrium price function.

Having closed form solutions to the pricing function is useful, but obviously not sufficient to identify the parameters of the model. First, we consider the case in which the distribution of rental prices is fully observed by the econometrician. If we only have access to data from one market and one time period, there is an obvious identification problem since there is no inherent scale for housing quality if it is treated as latent. For every non-linear pricing model, there exists a transformation of the utility function such that this transformation is observationally equivalent to the original model and pricing is linear. We can, therefore, normalize housing quality by setting it equal to the rental price in a baseline period. This normalization then implies that we can identify preferences for housing. We need data for more than one time period to identify non-linearities in pricing of housing.³ A sufficient condition for identification of the pricing functions in all subsequent periods is that preferences are time invariant. Our proofs of identification are constructive and can be used to devise an estimator for the parameters of the model.

We then consider the empirically more relevant case in which rental rates and housing values are partially latent. In practice, some housing units are rented and others are owner-occupied. For rental units, we observe rental prices, but not housing values. For owner-occupied units, we observe housing values, but not rental prices. As a consequence, both rental rates and values are partially latent. That creates some additional identification problems. In particular, we need to identify the rent-to-value functions which characterize the equilibrium loci of feasible rent and value combinations.

We, therefore, need to provide an integrated treatment of rental and asset mar-

³Exploiting variations among multiple markets is also a useful strategy to obtain identification if characteristics are observed as discussed in Bartik (1987) and Epple (1987).

kets. We assume that the housing stock is owned at each point of time by risk neutral investors who trade real estate assets in competitive asset markets and model the equilibrium in the real estate market. Investment decisions depend on the prevailing interest rate, rental income as well as expectations about future increases in rentals. We assume that asset markets are competitive. Hence, the expected profits of risk-neutral investors must be equal to zero. Separating the rental decision from the investment-ownership decision is a simplifying assumption, but has the advantage that we can show that housing values and rents are closely linked in equilibrium. In particular, the value of the house is given by the net present value of the discounted stream of rental income. Our result basically extends the well-known asset pricing results of Poterba (1984) to the case of heterogenous housing units. At each point of time, the proportionality between rents and values can be captured by time varying rent-to-value functions which depend on housing quality. These rent-to-value functions can be non-parametrically identified by characterizing the set of households that are indifferent between owning and renting. We can estimate these rent-to-value using non-parametric matching estimators.

We then apply our methods to study the housing market in Miami (Florida) using data from the American Housing Survey. The second purpose of the paper is to show that our new flexible methods can be used to gain some new insights into the causes and effects of the dramatic run up in housing prices that occurred in places such as Miami during the period leading up to the recent recession in the U.S. As a shorthand, we will adopt the common parlance of referring to this as a "bubble" without endowing this term with any connotations as to whether investor behavior was or was not rational.

From periodic American Housing Surveys, we observe the income distributions, housing rental distributions and house value distributions for Miami for three years:

1995, 2002 and 2007. We divide this time period into the pre-bubble period (1995-2002) and the bubble period (2002-2007). Rental rates and home values were moderately higher in 2002 than in 1995. Our model accounts for these changes by changes in population and the income distribution without a change in investor expectations about rental value appreciation over this period. Real incomes and rents were relatively constant in Miami during the 2002-07 period, but housing values increased dramatically during these years. We find that the rent-to-value functions dropped by almost 50 percent from the pre-bubble levels with larger decreases at the lower end of the quality distribution. These findings are broadly consistent with the view that changes in credit market conditions for low income houses together with changes in investor expectations about future housing prices may have accounted for the large run-up in housing values.

The rest of the paper is organized as follows. Section 2 of the paper introduces the hedonic model and derives the main new theoretical results of the paper. Section 3 discusses identification of the model under various different informational requirements. Section 4 focuses on estimation and imposing supply side restrictions. Section 5 provides information about the data sources and the samples used in estimation and presents the main set of empirical results focusing on the city of Miami between 1995-2007. Finally, we offer some conclusions and discuss future work in Section 6.

2 An Equilibrium Model of Housing Markets

Our model distinguishes between housing services and housing assets. We assume that housing services can be purchased in a frictionless rental market that allows for nonlinear pricing of housing services. Housing values or prices for real estate assets depend on prevailing interest rates, rental rates for housing services and expectations

about future appreciation in rentals and are determined in asset markets. We first consider the rental markets and then discuss the asset markets. Finally, we show how to incorporate housing supply changes and population growth into the analysis.

2.1 Rental Markets

We develop a hedonic model of non-linear pricing in a rental market for housing services in which housing quality can be characterized by a one-dimensional ordinal measure denoted by h . There is a continuum of households with mass equal to one.⁴ Households differ in income denoted by y . Let $F_t(y)$ be the metropolitan income distribution at time t . Households 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 .⁵

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 .⁶ All households

⁴We allow for population changes by varying the number of households in the economy in Section 3.3.

⁵Wheaton (1982) and Henderson and Venables (2009) developed insightful models that incorporate durability. Dunz (1989) and Nechyba (1997) provide a general equilibrium treatment of economies with heterogeneous durable housing. The value of this framework is demonstrated by Nechyba (2000) in study of school choice and educational vouchers. The important role that durable housing plays in the fortunes of cities and metropolitan areas has been demonstrated empirically by Glaeser and Gyourko (2005) and Brueckner and Rosenthal (2005).

⁶We abstract from borrowing and lending by assuming that each household receives and spends an exogenously determined income endowment each period. The assumption of a given income endowment that is spent each period permits us to frame our model in a way that is estimable with our data. For each metropolitan area, we have two or more periods of data comprised of observations for a sample of households and associated incomes of those households, but the sample of households differs from period to period.

are renters, and transactions cost in the rental market are zero. Hence, the household can costlessly change its housing consumption on a period-to-period basis as rental rates change. It follows that the household's optimal choice of housing to rent at each date t maximizes its period utility at date t :

$$\begin{aligned} \max_{h_t, b_t} \quad & U_t(h_t, b_t) \\ \text{s.t.} \quad & y_t = v_t(h_t) + b_t \end{aligned} \tag{1}$$

where b_t denotes expenditures on a composite good.

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

$$m_t(h_t, y_t - v_t) \equiv \frac{U_h(h_t, y_t - v_t)}{U_b(h_t, y_t - v_t)} = v'_t(h_t) \tag{2}$$

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

$$H_t^d(h|v_t(h)) = \int_0^\infty 1\{h_t(y, v_t(h)) \leq h\} dF_t(y) \tag{3}$$

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

Initially, we assume that the supply of housing is inelastic and can be characterized by a distribution of house quality $R_t(h)$.⁷ In equilibrium rental markets must clear for each value of h . We can define an equilibrium in the rental market for each point of time as follows:

Definition 1 *A hedonic housing market equilibrium is an allocation of housing consumption for each household and price function $v_t(h)$ such that*

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

$$H_t^d(h|v_t(h)) = R_t(h) \tag{4}$$

⁷We consider extension of the model to allow for changes in housing supply in Section 3.3

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

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

Assumption 1 *The utility function satisfies the following single-crossing condition:*

$$\left. \frac{\partial m_t}{\partial y} \right|_{U_t(h, y - v(h)) = \bar{U}} > 0 \quad (5)$$

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. The single-crossing condition implies the following result.

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

$$y_t(v) = F_t^{-1}(G_t(v)) \quad (6)$$

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

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

Assumption 2 *Income and housing are distributed generalized log-normal with location parameter (GLN4).⁸*

$$\begin{aligned} \ln(y_t) &\sim \text{GLN4}(\mu_t, \sigma_t^{r_t}, \beta_t) \\ \ln(v_t) &\sim \text{GLN4}(\omega_t, \tau_t^{m_t}, \theta_t) \end{aligned} \quad (7)$$

⁸The four-parameter distribution for income simplifies to the standard two-parameter lognormal when the location parameter, β_t , equals zero and the parameter $r_t = 2$. Similarly for the house value distribution. See Appendix B

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

Imposing the restriction that $r_t = m_t$ permits us to obtain a closed-form mapping from house value to income. We then establish that the further assumption that $\theta_t - \beta_t$ is time invariant permits us to obtain a closed-form solution to the hedonic price function.⁹

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

$$y_t = A_t (v_t + \theta_t)^{b_t} - \beta_t \quad (8)$$

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 3 *Let utility given by:*

$$U = u_t(h) + \frac{1}{\alpha} \ln(y_t - v_t(h) - \kappa) \quad (9)$$

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

⁹We impose both of these restrictions when estimating our model.

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

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

$$\frac{dh}{dp} \frac{p}{h} = \frac{(-\alpha\phi\gamma h + (h + \eta)((h + \eta)^{-\gamma} - \phi))}{(-\alpha\phi\gamma + (1 - \gamma)(h + \eta)^{-\gamma} - \phi)} \frac{1}{h} \quad (10)$$

and the income elasticity of demand is given by:

$$\frac{dh}{dy} \frac{y}{h} = \frac{-\alpha\phi\gamma}{p} \left[\frac{1}{-\alpha\phi\gamma + (1 - \gamma)(h + \eta)^{-\gamma} - \phi} \right] \frac{\kappa - \frac{p}{\alpha\phi\gamma} [-\alpha\phi\gamma h + (h + \eta)^{1-\gamma} - \phi(h + \eta)]}{h} \quad (11)$$

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

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

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

$$v_t(h) = \left(A_t \left[1 - (1 - \phi(h + \eta)^\gamma)^{\alpha(b_t - 1)} \right] \right)^{\frac{1}{1 - b_t}} - \theta_t \quad (12)$$

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

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

Summarizing, our analysis rental markets provides an equilibrium characterization that determines the rental price of housing, $v_t(h)$, as a function of house quality, h . The market fundamentals determining $v_t(h)$ are the quality of the housing stock and the demand for housing services arising from the distribution of income in the metropolitan population.

2.2 Asset Markets

Next we consider home ownership and asset markets for housing. For each level of housing quality h , there is an asset market in which investors can buy and sell houses at the beginning of each period. Let $V_t(h)$ denote the asset price of a house of quality h at time t .¹¹

Assumption 4 *Private investors are risk neutral.*

Investors can borrow capital in short term bond markets. The one-period interest rate is denoted by i_t . Investors (owners) 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 5 *Asset markets are competitive.*

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

$$E_t[\Pi_t(h)] = E_t \left[-V_t(h) + v_t(h) + \frac{V_{t+1}(h)(1 - \tau_{t+1}^p - \delta_{t+1})}{1 + i_t} \right] \quad (13)$$

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.

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

$$0 = E_t \left[-V_t(h) + v_t(h) + \frac{V_{t+1}(h)(1 - \tau_{t+1}^p - \delta_{t+1})}{1 + i_t} \right] \quad (14)$$

¹¹We thus treat a person that lives in an owner-occupied house as both a renter and an investor.

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

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

By successive forward substitution of the preceding, we obtain:

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

where

$$\beta_{t+j} = \prod_{k=1}^j \frac{(1 - \tau_{t+k}^p - \delta_{t+k})}{(1 + i_{t+k-1})} \quad (16)$$

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 and define $\tilde{\beta}_{t+j}$ as follows:

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

Then:

$$V_t(h) = \frac{v_t(h)}{u_t(h)} \quad (18)$$

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

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

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

$$E_t \prod_{k=1}^j \frac{(1 - \tau_{t+k}^p - \delta_{t+k})(1 + \pi_{t+k}(h))}{(1 + i_{t+k-1})} = \left[\frac{(1 - \tau^p - \delta)(1 + \pi(h))}{1 + i} \right]^j \quad (20)$$

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

Our model does not necessarily assume that investors have correct expectations about housing rental appreciation. It is possible that expectations of rental price increases prove to be greater than the actual rates of increase that are realized. Recall that changes in demand for housing services due to changes in population and the distribution of income drive changes in equilibrium rentals in our model.

2.3 Population Growth and Housing Supply

Thus far our model has treated the population and the distribution of housing quality as fixed. We can extend the model to accommodate population and housing supply change. Let N_t denote the metropolitan population at date t . Normalize the population at the initial date to be one: $N_1 = 1$ and treat $\{N_t\}_{t=1}^{\infty}$ as an exogenous process.

Let $q_t(h)$ denote the density of housing of quality h at date t . Let the housing supply function for quality h be:

$$q_t(h) = s(q_{t-1}(h), V_t(h), V_{t-1}(h)) \quad (21)$$

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

Assumption 6 *We adopt the following constant-elasticity parametric form for this*

supply function:

$$q_t(h) = k_t q_{t-1}(h) \left(\frac{V_t(h)}{V_{t-1}(h)} \right)^\zeta \quad (22)$$

where

$$k_t = \int_0^\infty q_{t-1}(h) \left(\frac{V_t(h)}{V_{t-1}(h)} \right)^\zeta dh \quad (23)$$

While this function is not explicitly derived from specification of cost function for the producer, it has attractive properties. It is parsimonious; it introduces only one additional parameter, ζ . Equation (22) also implies that the stock of housing of quality h does not change from date $t-1$ to date t if the the rental price of that quality of housing does not change. If the rental 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 . Hence, our model of unchanging supply corresponds to $\zeta = 0$.

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:

$$G_1(v_1(h)) = R_1(h) \quad (24)$$

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

$$R_t(h) = \int_0^h k_t q_{t-1}(x) \left(\frac{V_t(x)}{V_{t-1}(x)} \right)^\zeta dx \quad (25)$$

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:

$$G_t(v_t(h)) = R_t(h) \quad (26)$$

The expressions for the remainder of the model are unchanged. The "number" of households of income y at date t is given by:

$$n_t^y(y) = N_t f_t(y) \quad (27)$$

Similarly, the number of houses at rental v is:

$$n_t^v(v) = N_t g_t(v) \quad (28)$$

Single-crossing implies that, in equilibrium, the house rental expenditure at date t by income y must satisfy:

$$N_t F_t(y) = N_t G_t(v) \quad (29)$$

or $F_t(y) = G_t(v)$.

3 Identification

We consider identification and estimation of the parameters of the model assuming we have access to data for one market and h is not observed. Housing quality is ordinal and latent. There is no well-defined unit of measurement for housing quality. This implies that we can use the values of houses in a baseline period as our measure of quality. The next result formalizes this insight.

Proposition 4 *For every model with equilibrium pricing 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

given housing market and one time period, we cannot identify $u_1(h)$ separately from $v(h)$.

A corollary of Proposition 4 is then that we can normalize housing quality by setting $h = v_t(h)$ in some baseline period t . If, in addition, we make the standard assumption that per-period preferences are not changing over time, we can establish identification of the preference parameters.¹²

Assumption 7 *The utility function is time invariant.*

Assumption 7 implies that the parameters of the utility function and the price functions in $t > 1$ are identified.

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

Moreover, it is straight forward to show that the housing supply elasticity is identified of the market clearing condition in periods $t \geq 2$. We have the following result

Proposition 6 *The parameters of housing supply function are identified if we observe the equilibrium for, at least, two periods.*

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

¹²This normalization creates a market-specific quality measure that then permits intertemporal analysis of that market. An interesting future extension would be to jointly estimate the model for multiple markets with a common normalization of housing quality. This would permit comparing house quality distributions across metropolitan areas.

for the fact that rents are not observed for owner-occupied housing and need to be imputed.

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

Proposition 7 *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 7 implies that we can easily impute rents for owner-occupied using the observed rent-value function.

In practice, our data is 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. More generally speaking, the key assumption is that the average quantity of housing consumed conditional on observed characteristics is the same for owners and renters, i.e. no sorting on unobservables into home ownership. The non-parametric matching algorithm extends to more general demand models in which demands depends on a vector of observed state variables.

4 Estimation

The identification proofs 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-costs across time. This approach thus allows to capture 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 distribution of rents and income for each time period. Third, we estimate the remaining structural parameters of the model using an extremum estimator which matches quantiles of the income and value distributions while imposing the parameter constraints in Propositions 2 and 3 and the housing market equilibrium restriction that $R_{t+j}(h) = G_{t+j}(v_{t+j}(h))$ for $j \geq 1$.

Let $\tilde{F}_{t,j}^N$ denote the j th percentile of empirical income distribution at time t that is estimated based on a sample with size N . Similarly, let $\tilde{G}_{t,j}^N$ denote the j th 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 quantiles predicted by our model. Our estimator is then defined as:

$$\hat{\psi}^N = \operatorname{argmax}_{\psi \in \Psi} L^N(\psi) \quad (30)$$

subject to the structural constraints. The objective function is:

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

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

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

We use a parametric bootstrap procedure to estimate the standard errors, i.e. we parametrically bootstrap values of $\tilde{F}_{t,j}^N$ and $\tilde{G}_{t,j}^N$ and then implement the estimator above using the bootstrap percentiles.

5 The Housing Markets of Miami (FL)

5.1 Data

Our data set is taken from the American Housing Survey which is conducted by field representatives who obtain information from occupants of homes. Interviewing occurred from May 30 through September 8. There is a national and a metropolitan version. We use the latter. The sample sizes for the metropolitan areas range from 1,300 to 3,500 addresses. The unit of observation in the survey is the dwelling together with the household. The sample is selected from the decennial census.¹³

¹³Due to incomplete sampling lists (and nonresponse), the homes in the survey do not represent all homes in the country. Therefore, the raw numbers from the survey are raised proportionally so that the published numbers match independent estimates of the total number of homes. Housing unit under-coverage and household nonresponse is about 11 percent. Compared to the level derived from the adjusted Census 2000 counts, housing unit under-coverage is about 2.2 percent.

The AHS conducts surveys each year, but the metropolitan areas surveyed change from year to year. There is no fixed interval of repetition for surveying a given metropolitan area. The number of metropolitan areas surveyed has changed over time, likely due to changes in the AHS budget.

We use data for Miami (1995, 2002, 2007). The Miami Metropolitan Area is defined in 1995 and 2002 by 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.

We use income quantiles for the corresponding metropolitan areas. We aggregate the housing data for rental units and owner occupied housing. Since the AHS does not report rent paid by households net of utilities, we use reported housing costs and calculate the fraction of rent paid for utilities for those households that do not have them included in their rent payment. We then use this fraction to deduce the net rent of households with included utilities in their rent payment. Finally, we use polynomial regression to extrapolate both data to common quantile bounds and aggregate.

To illustrate the usefulness of the methods developed in this paper, we focus on Miami (FL) between 1995 and 2007. We divide this period into two sub-periods: 1) the pre-bubble period from 1995 - 2002; 2) the bubble period from 2002 -2007. Figure 1 plot the Case-Shiller index from Miami for the time period from 2002 through 2012. From January 2002 through December 2007, the CPI increased 19 percent. During the same time period, the Case-Shiller index rose from about 140 to 275. The real increase is about $(275/(1.19*140) -1) = 65$ percent.

We can compare the movements of the Case-Shiller index with the AHS data.

Figure 1:

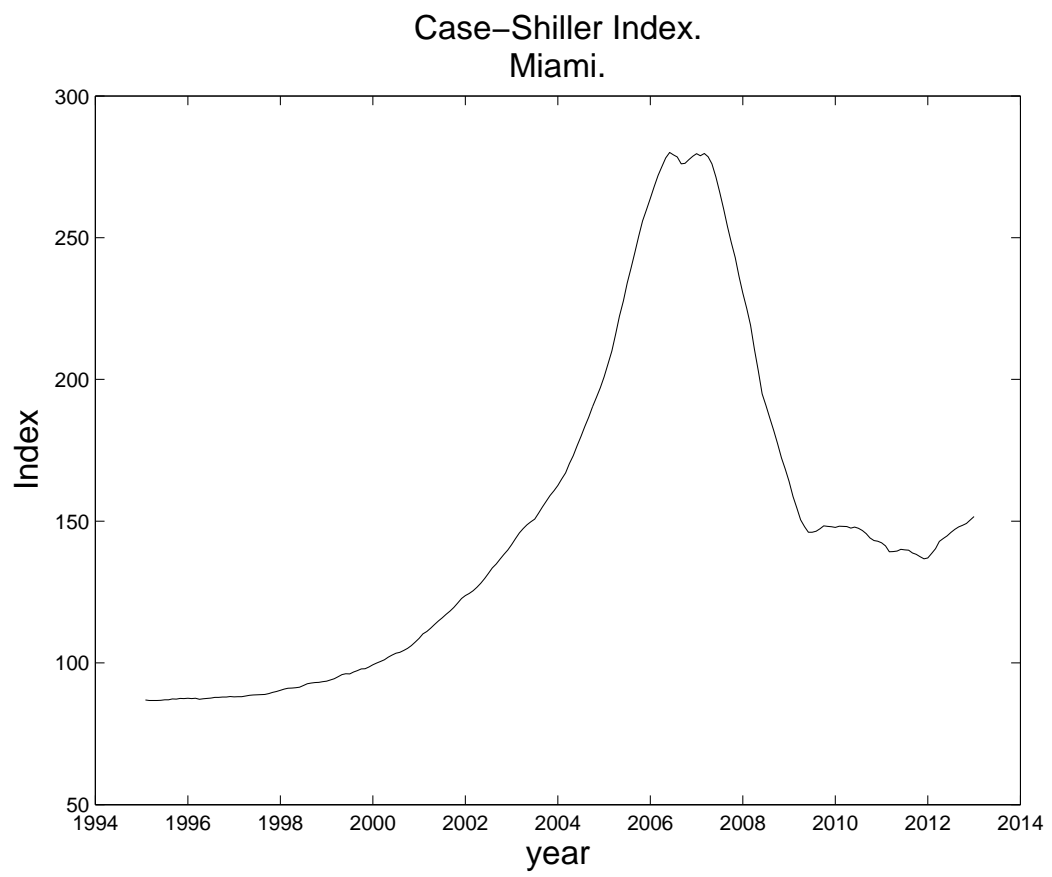
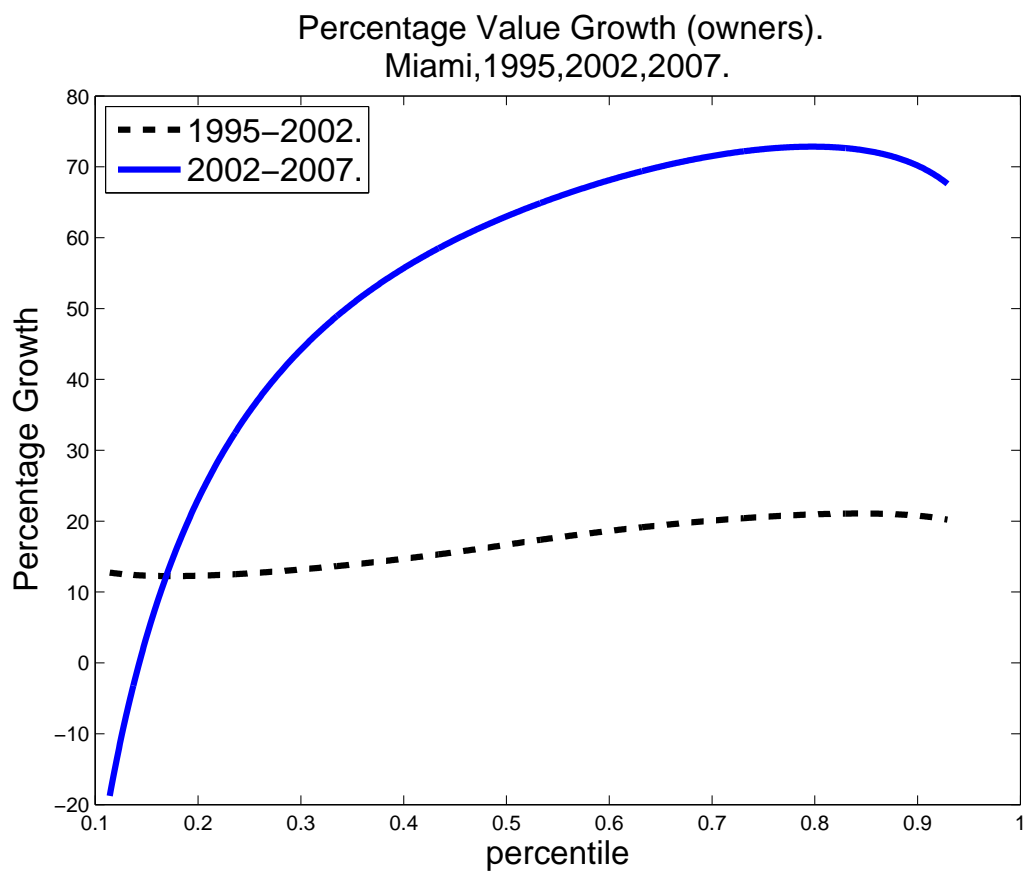


Figure 2:



First, we plot the change in housing values by quantile. The results are illustrated in Figure . We find that housing values appreciated by approximately 15 percent between 1995 and 2002. These values are consistent with the Case-Shiller index. Moreover, there are few systematic differences along the quality scale of housing. The results are different for the period between 2002-2007. Here we find that there are large differences by quality. At the low end of the quality distribution, housing prices stagnated or even declined. In contrast we find large housing price changes – ranging from approximately 30 percent at the 20th percentile to more than 70 percent for the top quantiles. We thus conclude that there is a lot of heterogeneity in housing price appreciation during the “bubble period.”

Next we plot the changes in observed rents in Figure 3. We see only moderate changes in rents during both periods. Rents increased faster during the bubble period than the pre-bubble period for the majority of rental properties. Nevertheless, rental changes were small in comparison to changes in home values. There is no evidence that rental changes were out of line with changes in fundamentals such as income or demographics.

Finally, we plot the change in the income distributions in Figure 4. We find that real income increased across the board between 1995 and 2002 while it stagnated during the later period

Summarizing the main empirical regularities observed in the AHS, we find that housing values and rents changes in the pre-bubble period are roughly in line with income changes. Rents and income were fairly flat during the bubble period. Owner occupied housing appreciated by up to 75 percent during the bubble period. Capital gains were higher for high quality than low quality homes.

Figure 3:

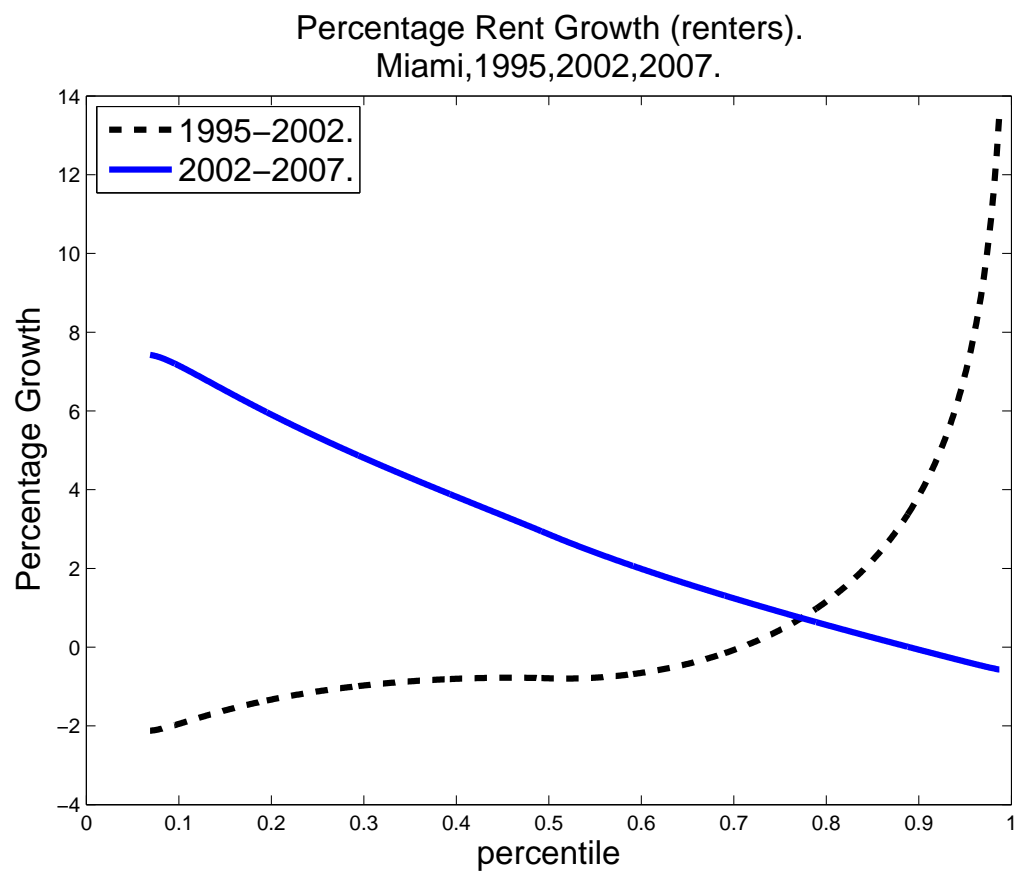
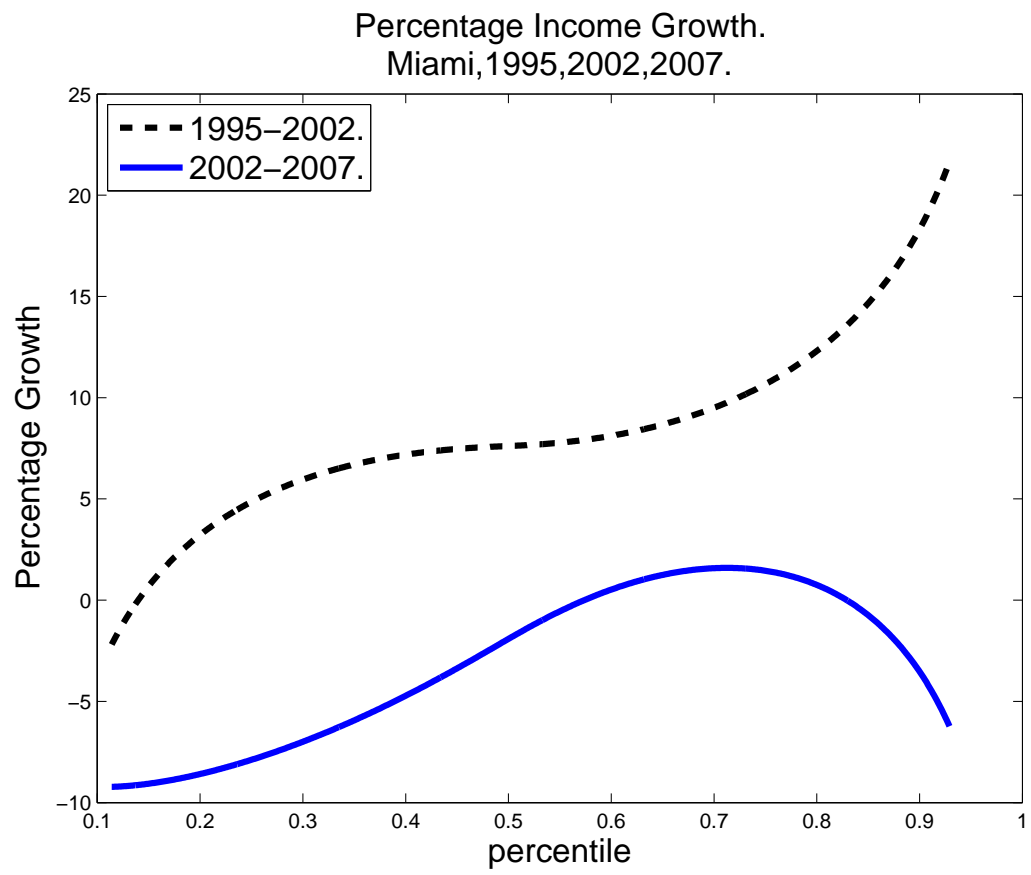


Figure 4:



5.2 First Stage: The Rent-to-Value Ratios

Next we report the estimates of the rent-value ratio. As we discussed above in detail we can estimate these functions for each time period using our non-parametric matching estimator. Note that data limitations make it difficult to estimate the locus outside a range of 50 and 350 thousand dollars for the 1995 and 2002, and 50 and 500 dollars in 2007. We find that the rent-to-value ratio was ranging between 0.07 and 0.06 in 1995. The function changed little between 1995 and 2002 with user costs slightly decreasing at the low and high end of the quality distribution. In contrast, we see large changes in the rent-value ratio during the bubble period between 2002 and 2007. The rent-value ratio drops to values between 0.046 and 0.035.

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. Of course, there is widespread evidence that credit became more available during the the bubble period especially for first time home owners at the lower end of the distribution. We also find that the largest changes in the rent-value ratio are for lower quality houses which is consistent with the notion that demand for these asset may have increased more strongly due to changes in credit markets.¹⁴

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

¹⁴Property tax rates as well as maintenance and depreciation rates did not change much either during the time period.

Figure 5:

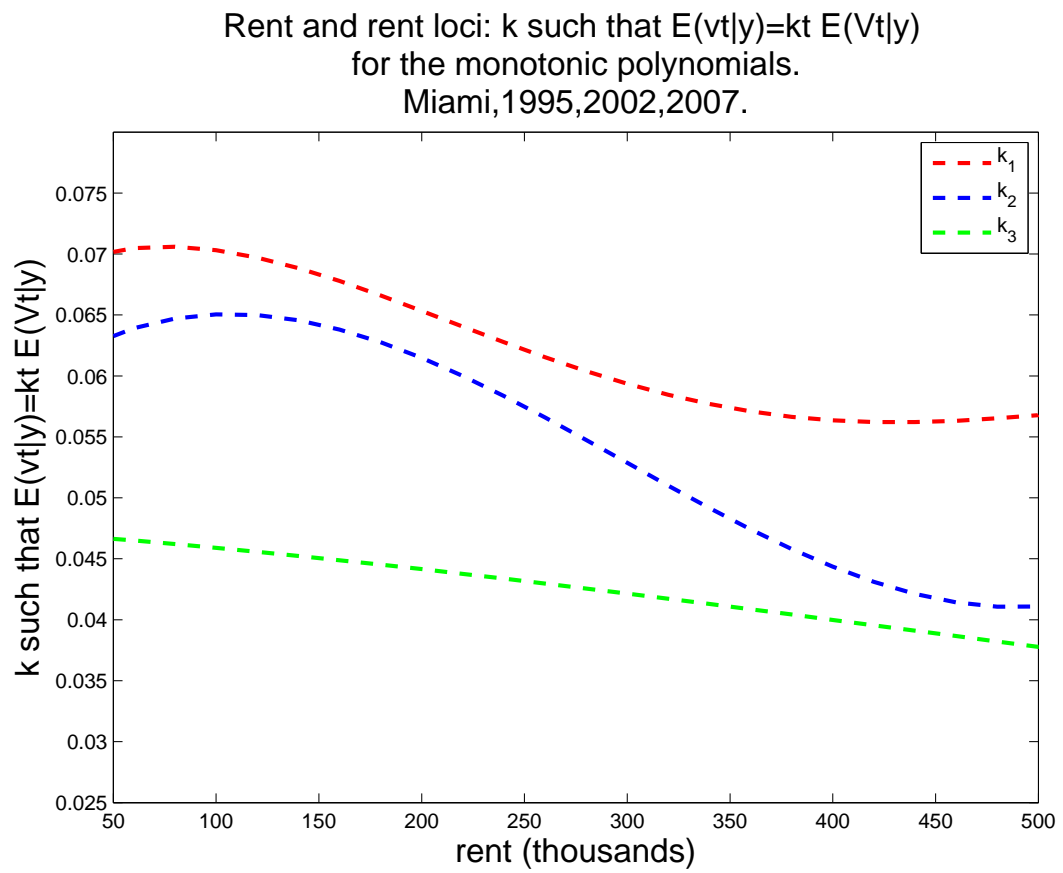
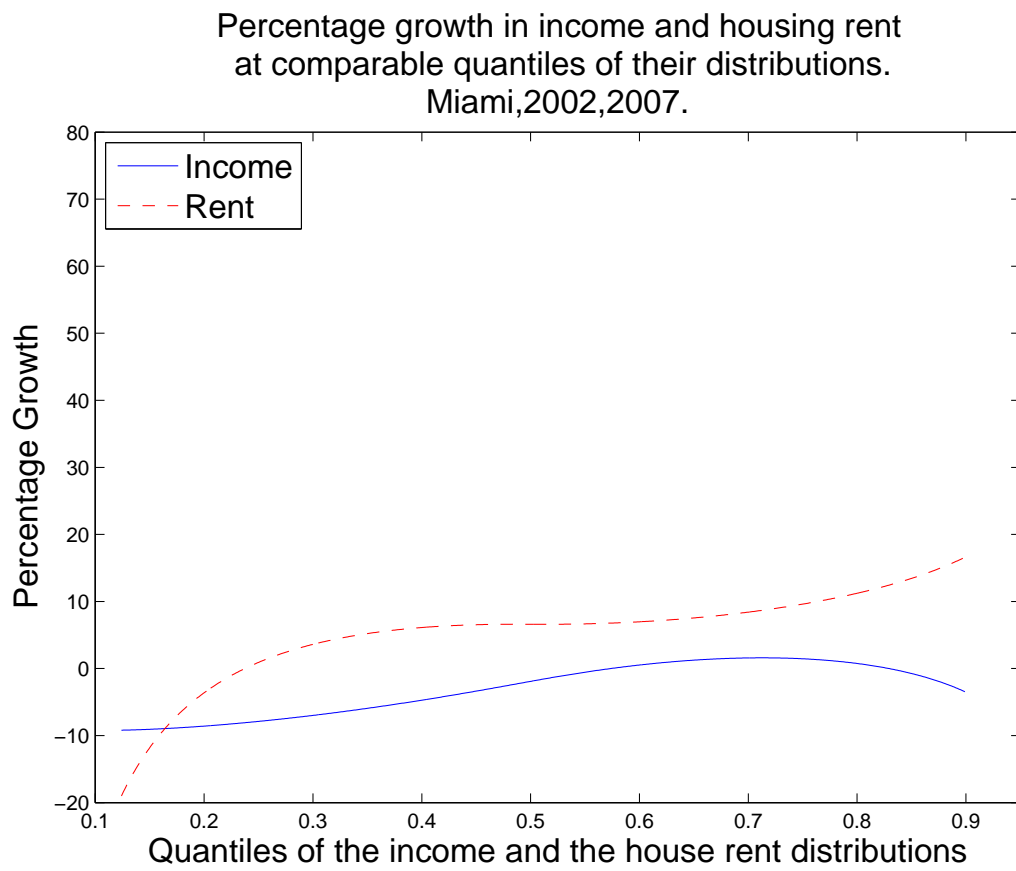


Figure 6:



Figure 7:



Based on our estimates of the rent-value ratio, we can convert housing values into (imputed) rents. Figures 6 and 7 plot the changes in rents and the changes in income for the two key periods that we analyze in this paper.

5.3 The Second Stage: Nonlinear Pricing in Rental and Asset Markets

Table 1 reports the parameter estimates and estimated standard errors for our model. The parameter estimates of the utility function are reasonable. Our estimates imply income elasticities that range between 0.60 and 0.72. Similarly, the price elasticities range between -1.1 for low income households to -0.67 for high income households.

Next we study the overall within sample fit of our model Figures 8 and 9 plot the estimated and the observed income and rent distributions for the three time periods. Overall, we find that our model fits the data very well.

Figure 10 shows that the resulting demand and supply functions in all three periods as predicted by our model.

Next we plot the estimated rental price functions to illustrate the importance of non-linear pricing in the rental market. The pricing function in the base period (1995) is, by construction, linear quality. Figure 11 then shows that Rents were fairly stagnant during the pre-bubble period between 1995 and 2002. Higher quality houses had steeper rental price increase between 2002 and 2007.

Note that our model of rental prices accounts for changes in real income, housing supply and population growth that occurred during the period. We have seen that our model can explain the rental price distributions. This finding is consistent with research by Sommer et al (2011) who also find that there was no “bubble” in rental

Table 1: Parameter Estimates: 1995-2007

	μ	σ	r	β
$t = 95$	10.79	0.66	1.73	7.42
std errors	0.12	0.01	0.03	0.08
$t = 02$	10.77	0.72	2.49	6.09
std errors	0.03	0.02	0.02	0.33
$t = 07$	10.66	0.81	2.18	-0.01
std errors	0.01	0.01	0.02	0.01
	ω	τ	m	θ
t=95	10.04	0.19	1.73	13.54
std errors	0.02	0.005	0.03	0.89
t=02	10.00	0.22	2.49	12.21
std errors	0.04	0.01	0.02	0.91
t=07	9.71	0.29	2.18	6.11
std errors	0.013	0.01	0.02	0.24
	α	ϕ	η	γ
	0.43	5993539	13.54	-2.31
std errors	0.05	0.16	0.89	0.02

Figure 8:

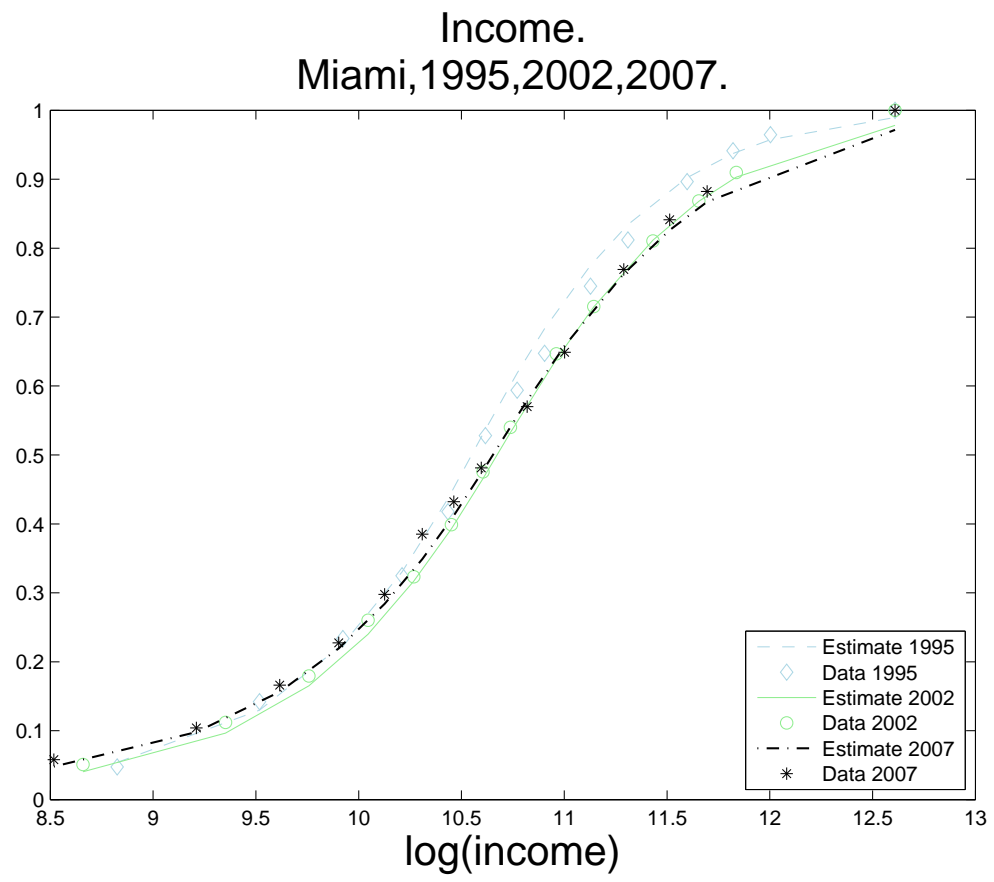


Figure 9:

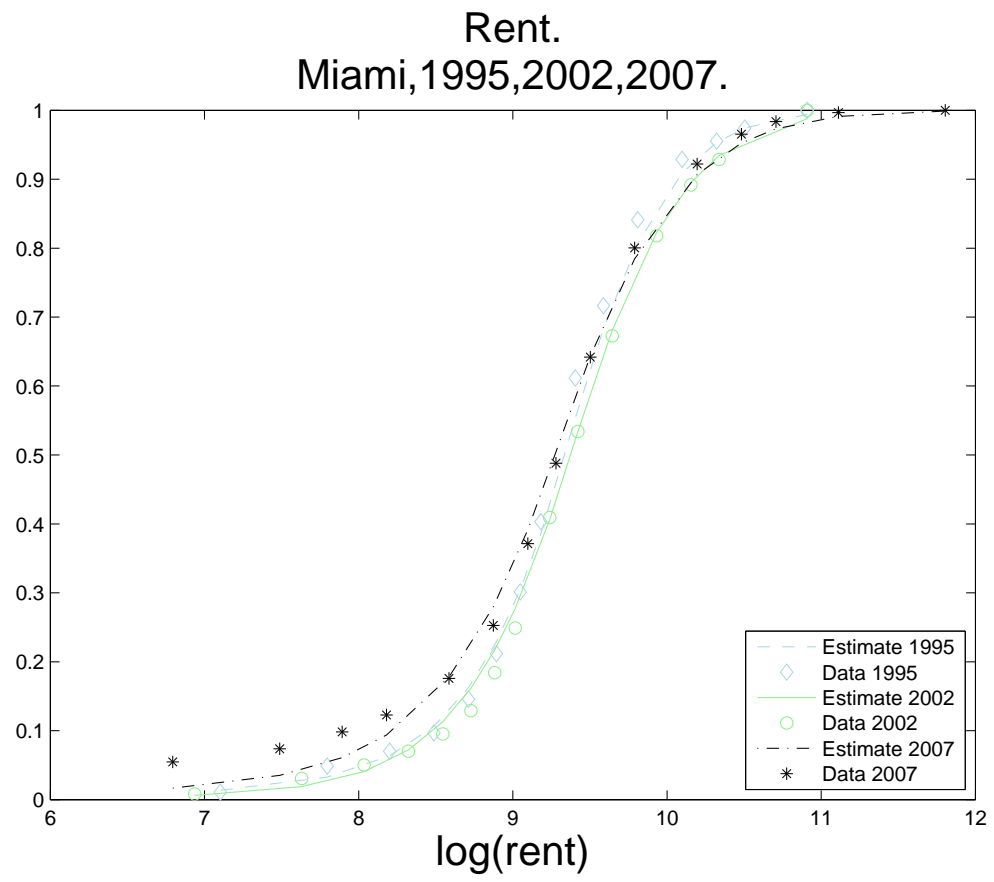


Figure 10:

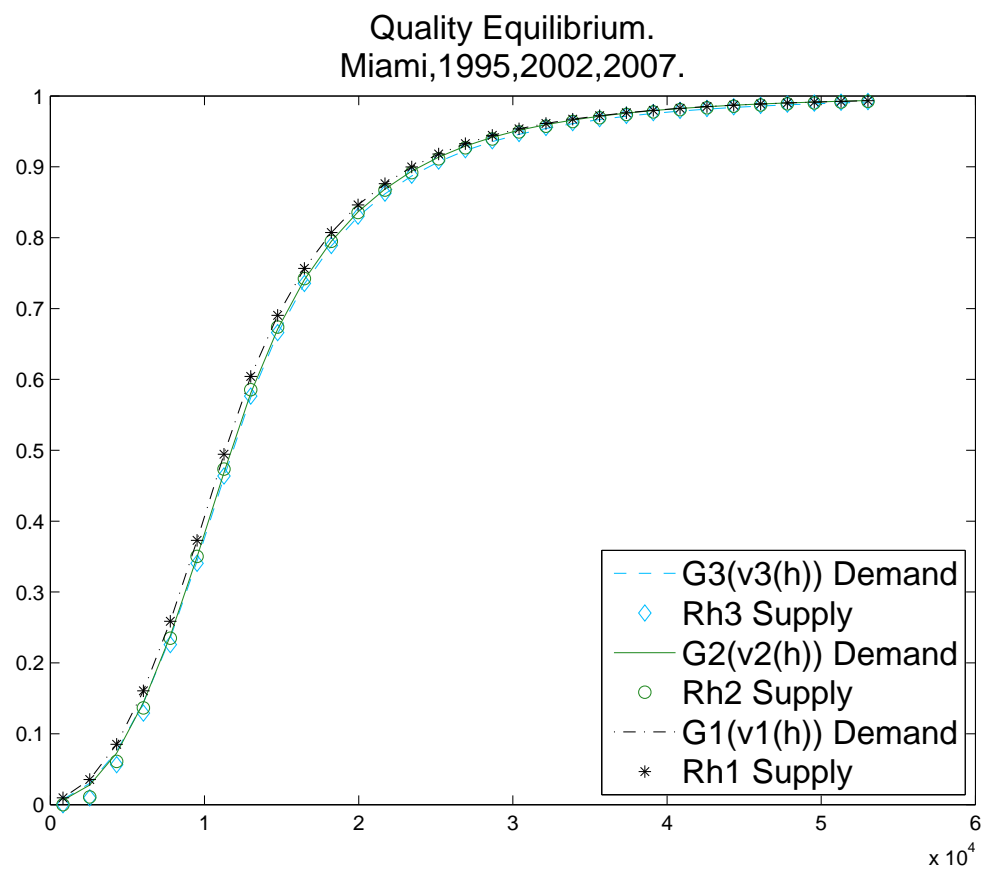
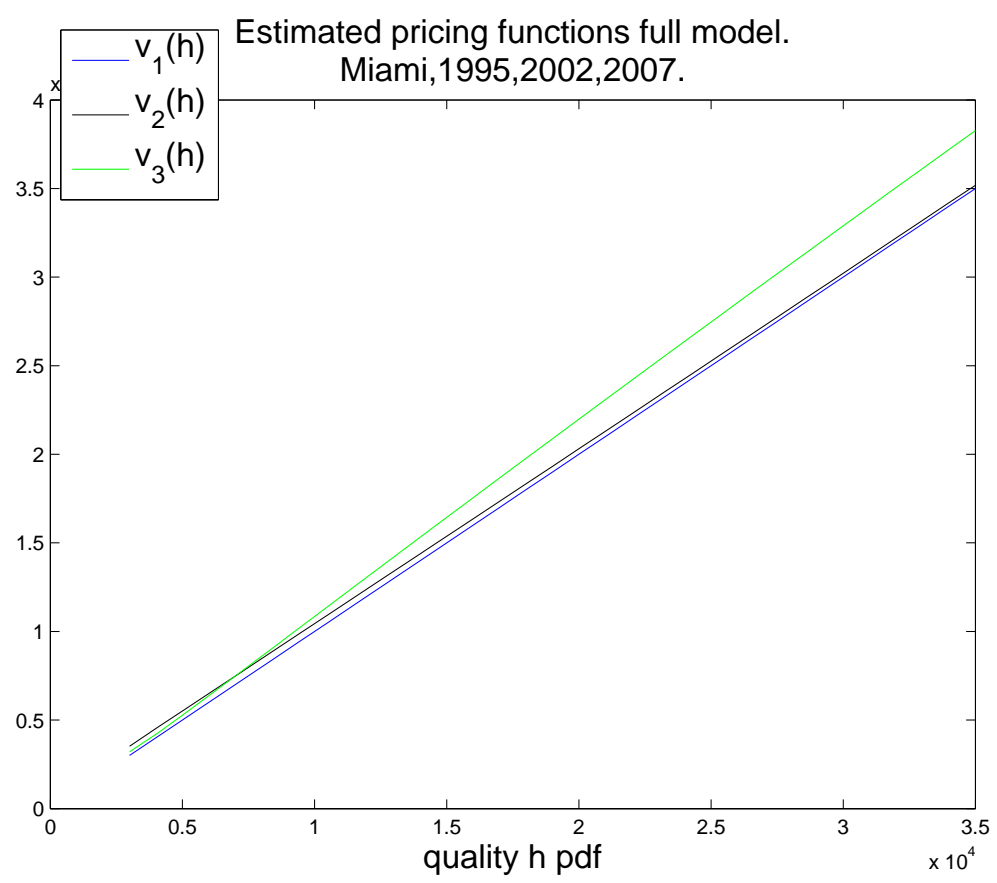


Figure 11:



rates for housing. Our findings confirm this result.

6 Conclusions

We have developed a new approach for estimating the price-quality frontier in markets for durable housing. 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 the distribution of characteristics of households. Third, it provides a straightforward summary of the changes in prices across the house quality distribution. In particular, we do not need to collapse the change in the distribution of prices into one number, as, for example, with the Case-Shiller index. Fourth, it gives insights into the mechanism that generates those price changes. Finally, we have shown that it provides some new insights into recent developments in local real estate markets.

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

References

- Bajari, P. and Benkard, L. (2005). Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, 113, 1239–77.
- Bartik, T. (1987). The Estimation of Demand Parameters in Hedonic Price Models. *Journal of Political Economy*, 95, 81–88.
- Berry, S. and Pakes, A. (2007). The Pure Characteristics Demand Model. *International Economic Review*, 48 (4), 1193–1225.
- Brueckner, J. and Rosenthal, S. (2005). Gentrification and Neighborhood Housing Cycles: Will America’s Future Downtowns Be Rich?. Working Paper.
- Dunz, K. (1989). Some Comments on Majority Rule equilibria in Local Public Goods Economies. *Journal of Economic Theory*, 47, 228–234.
- Ekeland, I., Heckman, J., and Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, 112(1), S60–S101.
- Epple, D. (1987). Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products. *Journal of Political Economy*, 95, 59–80.
- Glaeser, E. and Gyourko, J. (2005). Urban Decline and Durable Housing. *Journal of Political Economy*, 113, 345–75.
- Heckman, J., Matzkin, R., and Nesheim, L. (2010). Nonparametric Identification and Estimation of Nonadditive Hedonic Models. *Econometrica*, 78(5), 1569–1591.

- Henderson, J. V. and Venables, A. (2009). The Dynamics of City Formation. *Review of Economic Dynamics*, 12(233-54).
- Landvoigt, D., Piazzesi, M., and Schneider, M. (2011). The Housing Market(s) of San Diego. Working Paper.
- Nechyba, T. (1997). Local Property and State Income Taxes: the Role of Interjurisdictional Competition and Collusion. *Journal of Political Economy*, 105 (2), 351–84.
- Nechyba, T. (2000). Mobility, Targeting and Private School Vouchers. *American Economic Review*, 90, 130–146.
- Poterba, J. (1984). Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach. *Quarterly Journal of Economics*, 99 (4), 729–52.
- Poterba, J. (1992). Taxation and Housing: Old Questions, New Answers. *American Economic Review*, 82 (2), 237–242.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82, 34–55.
- Sommer, K., Sullivan, P., and Verbrugge, R. (2011). Run-up in the House Price-Rent Ratio: How Much Can Be Explained by Fundamentals?. BLS Working Paper.
- Vianelli, S. (1983). The family of normal and lognormal distributions of order r . *Metron*, 41, 3–10.
- Wheaton, W. (1982). Urban Spatial Growth under Perfect Foresight. *Journal of Urban Economics*, 12, 1–21.

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:*

$$F_t(y) = G_t(v) \quad (31)$$

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:*

$$\frac{\int_0^{[(\ln(y_t + \beta_t) - \mu_t)/\sigma_t]^{r_t}/r_t} e^{-t} t^{1/r_t - 1} dt}{2r\Gamma(1 + 1/r_t)} = \frac{\int_0^{[(\ln(v_t + \theta_t) - \omega_t)/\tau_t]^{m_t}/m_t} e^{-t} t^{1/m_t - 1} dt}{2m\Gamma(1 + 1/m_t)} \quad (32)$$

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

$$\frac{\ln(y_t + \beta_t) - \mu_t}{\sigma_t} = \frac{\ln(v_t + \theta_t) - \omega_t}{\tau_t} \quad (33)$$

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

Solving (33) yields:

$$y_t = e^{(\mu_t - \frac{\sigma_t}{\tau_t} \omega_t)} (v_t + \theta_t)^{\frac{\sigma_t}{\tau_t}} - \beta_t \quad (34)$$

Q.E.D.

Proof 3 *The household's FOC is:*

$$\alpha u'(h) \cdot dh = \frac{dv}{(y_t - v_t - \kappa)} \quad (35)$$

Substituting the income loci (8):

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

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

$$\alpha_i u'_i(h) dh = \frac{dv}{A_t(v_t + \theta_t)^{b_t} - (v_t + \theta_t)} \quad (36)$$

Integrating the right hand side yields:

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

which implies:

$$\alpha u(h) = \frac{1}{b_t - 1} \ln \left(1 - \frac{(v_t + \theta_t)^{1-b_t}}{A_t} \right) + c_t \quad (37)$$

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

$$\alpha \ln(1 - \phi(h + \eta)^\gamma) = \frac{1}{b_t - 1} \ln \left(1 - \frac{(v_t + \theta_t)^{1-b_t}}{A_t} \right) + c_t \quad (38)$$

Solving for v_t

$$(1 - \phi(h + \eta)^\gamma)^{\alpha(b_t-1)} = \left(1 - \frac{(v_t + \theta_t)^{1-b_t}}{A_t} \right) e^{c_t} \quad (39)$$

and hence

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

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

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

$$\max_h u_1(h) + u_2(y - v(h)) \quad (41)$$

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

$$u'_1(h) - u'_2(y - v(h)) v'(h) = 0 \quad (42)$$

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

$$\max_{h^*} u_1(v^{-1}(h^*)) + u_2(y - h^*) \quad (43)$$

and the FOC with respect to h^* is

$$u'_1(v^{-1}(h^*)) v^{-1'}(h^*) - u'_2(y - h^*) = 0 \quad (44)$$

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

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

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

Setting

$$\alpha = \frac{1}{b_t - 1} \quad (46)$$

implies

$$v_t = (A_t [1 - [1 - \phi(h + \eta)^\gamma]]^{\frac{1}{1 - b_t}}) - \theta_t = (A_t \phi(h + \eta)^\gamma)^{\frac{1}{1 - b_t}} - \theta_t \quad (47)$$

Setting

$$\phi = \frac{1}{A_t} \quad (48)$$

implies

$$v_t = ((h + \eta)^\gamma)^{\frac{1}{1 - b_t}} - \theta_t \quad (49)$$

Setting

$$\gamma = 1 - b_t \quad (50)$$

implies

$$v_t = (h + \eta) - \theta_t \quad (51)$$

Finally, setting

$$\eta = \theta_t \quad (52)$$

implies.

$$v_t = h \quad (53)$$

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

$$v_{t+s}(h) = \left(A_{t+s} \left[1 - [1 - \phi(h + \eta)^\gamma]^{\alpha(b_{t+s}-1)} \right] \right)^{\frac{1}{1-b_{t+s}}} - \theta_{t+s} \quad (54)$$

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

Proof 6 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. Moreover, $V_1(h)$ and $V_2(h)$ are observed by the econometrician. As a consequence ζ is identified of the market clearing condition:

$$R_2(h) = k_2 \int_0^h q_1(x) \left(\frac{V_2(x)}{V_1(x)} \right)^\zeta dx \quad (55)$$

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:

$$f(y) = \frac{1}{2(x + \beta)r^{\frac{1}{r}}\sigma\Gamma\left(1 + \frac{1}{r}\right)} e^{-\frac{1}{r\sigma^r}|\ln(x+\beta)-\mu|^r} \quad (56)$$

The CDF of the GNL4 distribution is given by:

$$F_t(y) = \begin{cases} \frac{\Gamma\left(\frac{1}{r}, B(y + \beta)\right)}{2\Gamma\left(\frac{1}{r}\right)} & \text{for } y < \exp(\mu) - \beta, \\ \frac{1}{2} & \text{for } y = \exp(\mu) - \beta, \\ \frac{1}{2} + \frac{\gamma\left(\frac{1}{r}, M(y + \beta)\right)}{2\Gamma\left(\frac{1}{r}\right)} & \text{for } y > \exp(\mu) - \beta. \end{cases} \quad (57)$$

where

- $B(y) = \frac{\left[\frac{\mu - \log(y + \beta)}{\sigma}\right]^r}{r}, M(y) = \frac{\left[\frac{\log(y + \beta) - \mu}{\sigma}\right]^r}{r},$

and

- $\Gamma(s, z) = \int_z^\infty e^{-t} t^{s-1} dt, \gamma(v, z) = \int_0^z e^{-t} t^{v-1} dt$

are the incomplete gamma functions.