No. 194 (83-17)

# A NEW APPROACH TO THE ESTIMATION OF STRUCTURAL EQUATION IN HEDONIC MODELS

by

Yoshitsugu Kanemoto and Ryohei Nakamura

October 1983

## Introduction

Many differentiated products such as houses and automobiles can be described by a set of utility-bearing characteristics. Rosen (1974) proposed a two-step procedure to estimate structural demand and supply functions for these characteristics. His procedure has been applied to the estimations of housing demand functions by many economists such as Harrison and Rubinfeld (1978), Witte, et. al. (1979), Linneman (1980), (1981). Recently, however, Brown and Rosen (1982) pointed out major difficulties in Rosen's procedure in identifying the structural demand and supply functions.

They took an example of a quadratic equilibrium price function and linear inverse demand and supply functions, and showed that Rosen's procedure yields nonsense estimates of demand and supply functions: their estimated coefficients are simple functions of the coefficients of the equilibrium price function. Since their criticism relys on special functional forms, its general validity is not quite clear. Quigley (1982) argued that this problem can be avoided by imposing an a priori assumption on the functional form of the utility function (or the production function). He used the generalized CES form to estimate housing demand relationships.

In this paper we first point out more fundamental difficulty in the estimation of structural demand and supply equations in a hedonic model. An equilibrium allocation in a hedonic model is basically a simultaneous equilibrium for markets of many differentiated products. Observed prices in a cross-section data set are therefore one sample of a set of prices that equilibrate all markets. Obviously, we need more than one sample to estimate structural equations. In terms of a bid price function, the equilibrium data provide the bid price of the individual who bought a good with a certain set of characteristics. The data do not, however, contain the information on the bid prices for goods that the individual did not buy. It is therefore impossible to know the global shape of the bid price function from cross-section data.

Quigley's approach is to impose sufficient restrictions on the functional form of the utility function (and hence the bid price function) so that the local information completely characterizes the global preference structure. Since there is no a priori justification of a specific functional form such as a generalized CES form, the estimation method that is feasible only when such a restriction is imposed seems to have serious weakness.

In this paper, we introduce another approach that makes use of unobserved characteristics of goods. The main idea is that if there are unobserved characteristics of products, individuals with the same characteristics may buy products with different observed characteristics, since difference in unobserved characteristics can offset difference in observed characteristics. We then have more than one observation for the same type of individuals and the estimation of a bid price function becomes possible.

In section 1, the fundamental difficulty in the estimation of the structural equations is explained. Quigley's estimation method is examined in more detail in section 2. Our estimation method of the bid price function is developed in section 3 and

## 1. A Fundamental Difficulty

In hedonic models there is a fundamental difficulty in estimating structural demand and supply equations using cross-section data. A hedonic model contains many differentiated products with different amounts of characteristics, and an equilibrium allocation in the model requires that markets for all these products be in equilibrium simultaneously. Then, even though there are many prices in a cross-section data set, we essentially have only one sample of a set of prices that equilibrate all markets. It is, of course, impossible to estimate structural equations with only one sample.

In order to illustrate this problem, consider the estimation of the bid price function in the completely deterministic case where there are no observation errors, no unobserved characteristics, and no specification errors. Let $r(z,y)$ denote the bid price function of an individual with the characteristic vector $y$ for a good with the characteristic vector $z$. The equilibrium price of good $z$ then satisfies $p^*(z)=\max_{\{y\}} r(z,y)\equiv r(z,y^*(z))$, where $y^*(z)$ denotes the characteristics of the individual who has the highest bid for good $z$.

Assume, for simplicity, that characteristic vectors, $z$ and $y$, are scalars. Fig. 1 shows the relationship between the equilibrium price function $p^*(z)$ and the bid price function $r(z,y)$ in this simple case. From the above relationship, the equilibrium price function is an upper envelope of bid price functions of individuals with different characteristics.

Since the observed price is the highest bid price among all the utility function is derived from the bid price function in section 4. The estimation method is applied to Japanese housing data in section 5.

potential buyers in equilibrium, all the observations lie on the equilibrium locus, $p^*(z)$. The information one can get is the characteristics of the individual, $y$, who bought a good with a certain set of characteristics, $z$, and the price at which it was bought. If, for example, individual $y^1$ bought good $z^1$ at price $p^*(z^1)$, then we can infer that the bid price function of individual $y^1$ is tangent from below to the equilibrium price locus at $y^1$. However, it is impossible to know the shape of the bid price function at any point other than $z^1$. Even at $z^1$, the curvature of the bid price function cannot be known. Thus, the estimation of the bid price function is impossible in the completely deterministic case.

There are at least two ways of circumventing this difficulty. One is to impose a priori restrictions on the functional form of the bid price function, $r(z,y)$. If the restrictions are strong enough, all parameters of the bid price function may be estimated from the observations only along the equilibrium locus. Quigley (1982) has done this by assuming that the utility function is of the generalized CES form. The other is the approach taken in this paper. If there is an error term, such as unobserved characteristics of goods, that affect the bid price function, then individuals with the same characteristic vector may buy goods with different observed characteristics. In such a case, more than one point of the bid price function is observed and the shape of the bid price function can in principle be estimated.

The two approaches can be understood more clearly by considering the estimation of the indifference map (or the utility function) which generates the bid price function. Suppose that all individuals have the same utility function, $U(x,z)$, where x is the composite good representing all other commodities than the hedonic good. Individuals differ only in incomes and the vector y is a scalar. An individual maximizes $U(x,z)$ subject to the budget constraint, $y=x+p^*(z)$, where the composite good is taken as a numeraire. Fig. 2 illustrates this maximization problem in the case where the characteristic vector z is a scalar. Curves $J^1J^1$ and $J^2J^2$ depict budget constraints for individuals with incomes $y^1$ and $y^2$ respectively. The budget curves are horizontal displacements of each other, since all individuals face the same equilibrium price function, $p^*(z)$. Curves $I^1I^1$ and $I^2I^2$ are indifference curves and curve EE shows utility maximizing combinations of z and x.

In the completely deterministic case, all observations lie along curve EE and no other points on the (x,z) plane are observed. Since the slope of the indifference curve must equal the slope of the budget curve at the optimum, the slopes of indifference curves along curve EE can be estimated. Shapes of indifference curves at any other points, however, cannot be known. This corresponds to the difficulty in the estimation of the bid price function discussed earlier.

The first way of evading this difficulty is to assume a special functional form of the utility function so that the information of the slopes of indifference curves along curve EE is sufficient to characterize the entire indifference map. The second way is to introduce unobserved characteristics of a good, v. The utility function is then $U(x,z,v)$ and the equilibrium price function must

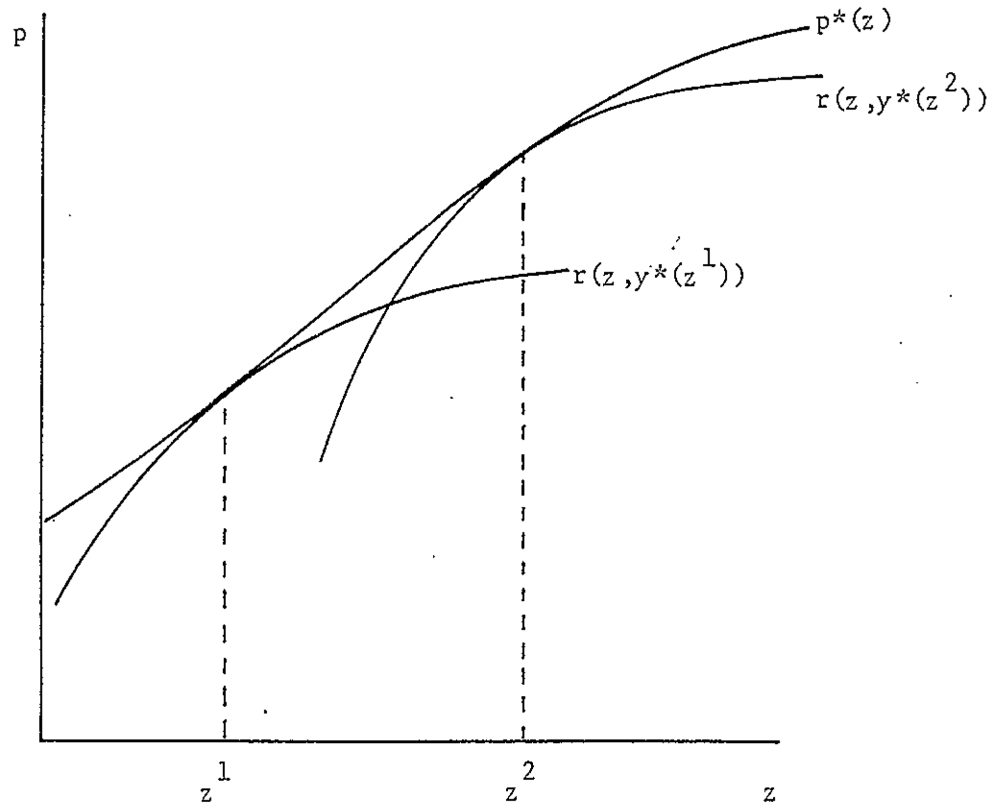


Fig. 1. The relationship between the bid price functions and the equilibrium price function

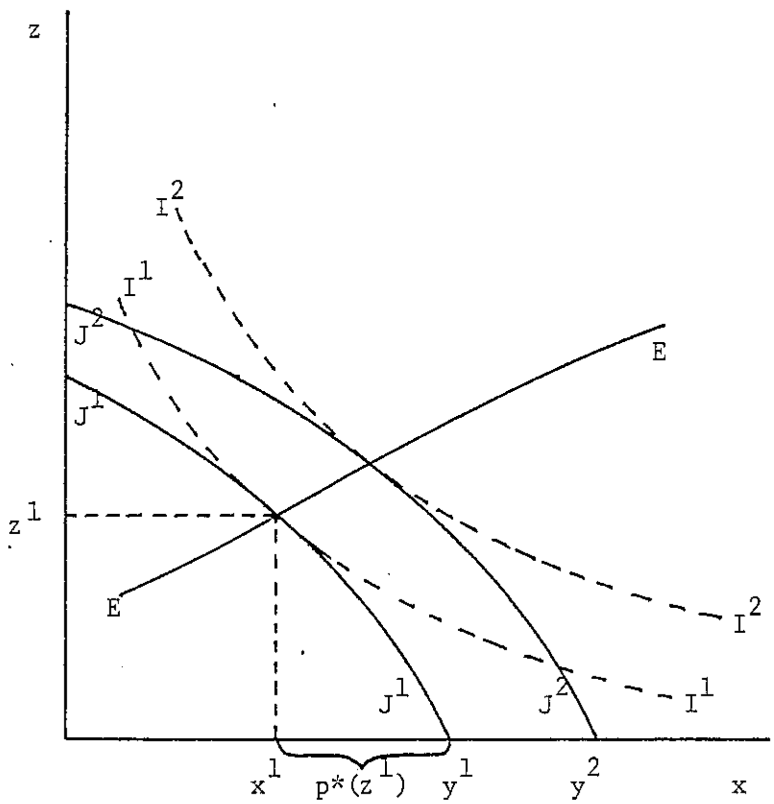


Fig. 2. The estimation of the indifference map

be written $p^*(z,v)$. Corresponding to different values of $z$, different combinations of $z$ and x are chosen by individuals with the same income and there are many curves like curve EE. Observations therefore scatter around on the $(x,z)$ plane and the global shapes of indifference curves may be estimated. The estimation is not, however, straightforward, since the indifference map on the $(x,z)$ plane does not in general remain the same when the unobserved characteristics are changed. In this paper, we develop a way of estimating the indifference maps by the appropriate use of the information on the characteristics of the good chosen by an individual with a certain characteristic vector.

In the same way as the unobserved characteristics of a commodity, the unobserved characteristics of an individual can be used to estimate the structural demand equation. In this case, individuals with different unobserved characteristics have different indifference maps and observations are scattered on the $(x,z)$ plane. If there exist unobserved characteristics of individuals but all characteristics of commodities are observed, then the multinomial-logit type estimation method developed by Ellickson (1981) and Lerman and Kern (1983) can be used to estimate the bid price function. Although it is, in principle, possible to combine our approach with theirs by assuming both types of unobserved characteristics, the computational requirement for the combined model is formidable.

## 2. The Examination of Quigley's Estimation Method

Before presenting our estimation method, we examine Quigley's approach in more detail. In our notation, Quigley's model is as follows. All consumers have the generalized CES utility function, $U(x,z)=[\sum_{i=1}^{n} \alpha_i z_i^{\beta_i} + x^{\varepsilon}]^{\phi}$, and maximizes the utility level under the budget constraint, $y=x+p^*(z)$, where $z=(z_1,\ldots,z_n)^T$, and $\alpha_i$, $\beta_i$, and $\varepsilon$ are parameters to be estimated. Note that $\phi$ cannot be estimated since a utility function is unique only up to a monotonic transformation. The first order conditions for the maximization problem are

$$\log\frac{\partial p^*(z)}{\partial z_i} = \log\frac{\alpha_i\beta_i}{\varepsilon} + (\beta_i-1)\log z_i - (\varepsilon-1)\log x, \quad i=1,\ldots,n. \tag{2.1}$$

Quigley proposed the following two-step estimation procedure. First, estimate the equilibrium price function, $p^*(z)$. The specific form of the price function is

$$p^{(\lambda)} = a_0 + \sum_{i=1}^{n} a_i z_i + u,$$

where $u$ is an error term, $a_i$ are parameters to be estimated, $p^{(\lambda)}$ is the Box-Cox transformation of the price,

$$p^{(\lambda)} = (p^{\lambda}-1)/\lambda , \quad \lambda\neq 0$$
$$= \log p , \quad \lambda=0,$$

and the parameter $\lambda$ is also estimated. The estimated equilibrium price function can be written

$$p^*(z) = [\hat{\lambda}(\hat{a}_0+\sum_{i=1}^{n}\hat{a}_i z_i)+1]^{1/\hat{\lambda}} ,$$

where hats above $\lambda$ and $a_i$ denote their estimated values. The second step is to calculate partial derivatives of the equilibrium price function and to estimate

$$\log \frac{\partial p^*(z)}{\partial z_i} = d_0 + d_i \log z_i + d_{n+1} \log x + v_i, \quad i=1,\ldots,n,$$

where $v_i$ are error terms and $d_i$ are parameters to be estimated. The estimates of parameters in the utility function, $\alpha_i$, $\beta_i$, and $\varepsilon$ can then be obtained from estimates of $d_i$, $i=0,\ldots,n+1$:

$$\hat{\beta}_i = \hat{d}_i + 1$$

$$\hat{\varepsilon} = -\hat{d}_{n+1} + 1$$

$$\hat{\alpha}_i = e^{\hat{d}_0} \frac{1-\hat{d}_{n+1}}{\hat{d}_i+1} .$$

Even if we accept the restriction of the utility function to the generalized CES form, there is a serious difficulty in Quigley's procedure. The functional form of the equilibrium price function, $p^*(z)$, must be estimated accurately, since its partial derivatives are used in the second-step estimation. How the first partials of $p^*(z)$ change as the characteristic vector, $z$, changes is crucial in the second-step estimation, and the estimated price function must at least provide a good second order approximation of the true function. If the functional form is misspecified, therefore, the second-step estimation may be seriously biased. Usually, however, the data do not allow us to obtain an accurate estimate of the functional form. For example, in Quigley's estimation the estimated equilibrium price function explains less than 50% of the variations in monthly rents for the sampled dwelling units. In such a case, the confidence intervals are usually too large to yield an accurate estimates of the parameters in a sufficiently flexible functional form.

Furthermore, the form of the utility function has an implication on the form of the equilibrium price function. In the generalized CES case, the bid price function of an individual with income y and utility level U is $r(z,y,U)=y-[U^{1/\phi}-\Sigma\alpha_i z_i^{\beta_i}]^{1/\varepsilon}$. If the utility level of an individual with income y is U*(y) in equilibrium, then the equilibrium price function satisfies

$$\begin{aligned} p^*(z) &= \max_{\{y\}} r(z,y,U^*(y)) \\ &= \max_{\{y\}}\{y-[U^*(y)^{1/\phi}-\Sigma\alpha_i z_i^{\beta_i}]^{1/\varepsilon}\}. \end{aligned}$$

There is no guarantee that the functional form used by Quigley is a good approximation of the form obtained in this way.

## 3. The Estimation of Bid Price Functions

In this section, we consider the estimation of the bid price function in the case where there are unobserved characteristics of a good, v. The characteristic vector of an individual is denoted by $y=(y_1,y_2,\ldots,y_m)^T=(y_1,s^T)^T$, where $y_1$ denotes the income and $s=(y_2,\ldots,y_m)^T$ other characteristics such as age and the number of children. The utility function of an individual is written $U(x,z,v,s)$ and the budget constraint is $y_1=x+p^*(z,v)$. Define $x^*(z,v,s,\bar{U})$ by the identity $U(x^*(z,v,s,\bar{U}),z,v,s)\equiv\bar{U}$. Then the bid price function is $r(z,v,y,\bar{U})\equiv y_1-x^*(z,v,s,\bar{U})$. The bid price function represents the price an individual with characteristics y is willing to pay for the good with characteristics z at the utility level $\bar{U}$.

The equilibrium price is the highest bid price among those of all individuals in the market. If the utility level of an individual with characteristics y is given by $U^*(y)$ at the equilibrium allocation, then the equilibrium price function satisfies

$$\begin{aligned} p^*(z,v) &\equiv \max_{\{y\}} r(z,v,y,U^*(y)) \\ &\equiv r[z,v,y^*(z,v),U^*(y^*(z,v))], \qquad (3.1) \end{aligned}$$

where $y^*(z,v)$ denotes the characteristic vector of the individual who buys a good with characteristics (z,v).

Now, we assume that the observed price of good (z,v) and the observed characteristics of an individual who buys good (z,v) satisfy

$$p = p^*(z,v) + u, \qquad (3.2)$$

$$y = y^*(z,v), \qquad (3.3)$$

where u is the observation error of the price. Our problem is to estimate the bid price function using the data on p,z, and y. It is of course impossible to estimate the bid price function $r(z,v,y,\bar{U})$, since the utility level $\bar{U}$ is unobservable. Instead, we estimate a sort of a reduced form bid price function $\tilde{r}(z,v,y) \equiv r[z,v,y,U^*(y)]$.

The outline of our estimation procedure is as follows. The estimation of equation (3.2) does not yield the estimate of the bid price function, $\tilde{r}(z,v,y)$, since y is optimized out and $p^*(z,v) \equiv \tilde{r}(z,v,y^*(z,v))$ is the only relationship that can be estimated. Considering the unobserved characteristics, v, as error terms, however, we can estimate equation (3.3). This equation can be used to infer the vector of unobserved characteristics from the observed characteristics, y and z. That is, if the function $y^*()$ is estimated, then we can solve equation (3.3) to express v as a function of y and z, $v=\Phi(y,z)$, where $\Phi(y,z)$ satisfies the identity, $y \equiv y^*(z,\Phi(y,z))$. Then the equilibrium price function can be written $p^*(z,\Phi(y,z)) \equiv \tilde{r}(z,\Phi(y,z),y)$. Since the function $\Phi()$ is known, we can estimate $\tilde{r}(z,v,y)$ from

$$p = \tilde{r}(z,\Phi(y,z),y) + u. \tag{3.4}$$

Consider the special case of the quadratic bid price function,

$$\tilde{r}(z,v,y) = b_0+b_z^Tz+b_y^Ty+\tfrac{1}{2}z^TB_zz+\tfrac{1}{2}y^TB_yy+y^T(B_{yz}z+v), \tag{3.5}$$

where z and $b_z$ are n-dimensional vector, y, v, and $b_y$ m-dimensional vectors, $B_z$ a nxn matrix, $B_y$ a mxm matrix, $B_{yz}$ a mxn matrix, and $b_0$ a scalar. Parameters to be estimated are $b_0$, $b_z$, $b_y$, $B_z$, $B_y$, and $B_{yz}$. If $B_y$ is negative semi-definite, then the solution to the maximization problem (3.1) exists and $y^*(z,v)$ satisfies

$$B_y y^*(z,v) + b_y + B_{yz} z + v = 0. \qquad (3.6)$$

Hence, $\Phi(y,z)$ is

$$v = \Phi(y,z) = -[B_y y + B_{yz} z + b_y]. \qquad (3.7)$$

Furthermore, if $B_y$ is of full rank, then

$$y^*(z,v) = -B_y^{-1}[B_{yz} z + b_y + v]. \qquad (3.8)$$

Substituting (3.7) into (3.5) yields

$$r[z,\Phi(y,z),y] = b_0 + b_z^T z + \tfrac{1}{2} z^T B_z z - \tfrac{1}{2} y^T B_y y. \qquad (3.9)$$

The estimation of the bid price function is carried out in two steps. First, estimate

$$y = c_0 + C_z z + w, \qquad (3.10)$$

by ordinary least squares (OLS) or some other appropriate method, where w is a m-dimensional vector of error terms, $c_0$ and $C_z$ are respectively a m-dimensional vector and a mxn matrix of parameters. This estimation yields the estimates of parameters, $\hat{c}_0$ and $\hat{C}_z$. The relationship between these parameters and those in the bid price function is

$$b_y = -B_y c_0 \qquad (3.11)$$

$$B_{yz} = -B_y C_z. \qquad (3.12)$$

In the second step, estimate

$$p = b_0 + b_z^T z + \tfrac{1}{2} z^T B_z z - \tfrac{1}{2} y^T B_y y + u \qquad (3.13)$$

to obtain the estimates $\hat{b}_0$, $\hat{b}_z$, $\hat{B}_z$, and $\hat{B}_y$. Substituting $\hat{B}_y$ into (3.11) and (3.12) yields $\hat{b}_y$ and $\hat{B}_{yz}$. Thus all parameters in the bid price function (3.5) are estimated. Note that in the estimation of (3.13) matrix $B_y$ is restricted to be negative semi-definite.

The identification of structural equations depends crucially on the existence of unobserved characteristics of a good. If there are no such characteristics, i.e., v=0, then there is multicollinearity bytween y and z and y can be expressed as a deterministic function of z. In such a case, it is impossible to identify all parameters in (3.13). Essentially, we can only estimate

$$p = a_0 + a_z^T z + \tfrac{1}{2} z^T A_z z + u. \qquad (3.14)$$

The bid price functions compatible with (3.10) and (3.14) are

$$\tilde{r}(z,0,y) = a_0 + a_z^T z + \tfrac{1}{2} z^T A_z z - \tfrac{1}{2}(y-c_0-C_z z)^T B_y (y-c_0-C_z z), \quad (3.15)$$

where $B_y$ is <u>any</u> negative semi-definite matrix with dimension mxm. Thus, if there are no unobserved characteristics, parameters $B_y$ are always left undetermined. It is ironical that the absence of error terms causes a serious identification problem.

In the bid price function (3.5), unobserved characteristics are multiplied by the characteristic vector of an individual. We can obtain a more general case by including those, $v_z$, multiplied by the characteristic vector of a good and that, $v_0$, which appears alone. The bid price function then contains the terms, $z^T v_z + v_0$, in addition to equation (3.5). This generalization does not affect (3.6), but changes (3.9) into

$$\tilde{r}(z,\Phi(y,z),y) = b_0 + b_z^T z + \tfrac{1}{2} z^T B_z z - \tfrac{1}{2} y^T B_y y + z^T v_z + v_0. \qquad (3.15)$$

Therefore, the first step estimation remains the same, but the error term in (3.13) is now $z^T v_z + v_0 + u$. Since the error term contains z, there is a heteroschedasticity problem in the second step estimation: The variance of the error term is $E(z^T v_z + v_0 + u)^2$ $= g_0 + 2g_z^T z + z^T G_z z$, where $g_0 = E(v_0+u)^2$, $g_z = E[(v_0+u)v_z]$, and $G_z = E(v_z v_z^T)$. In this case, the OLS estimators are inefficient. Efficient estimates can be obtained by the following two-step procedure. Compute the

OLS estimators, get the estimated residuals, and regress the squares of these residuals on z and $zz^T$ to get estimates of $g_0$, $g_z$, and $G_z$. Then, use the weighted-least-squares method: minimize

$$\sum_j \{[p^j - b_0 - b_z^T z^j - \tfrac{1}{2}(z^j)^T B_z z^j + \tfrac{1}{2}(y^j)^T B_y y^j]/[g_0 + 2g_z^T z^j + (z^j)^T G_z z^j]^{\frac{1}{2}}\}^2$$

with respect to $b_0$, $b_z$, $B_z$, and $B_y$, where superscripts j indicate the j-th sample. The estimated residuals are used to obtain the new estimates of $g_0$, $g_z$, and $G_z$, and the same procedure is iterated till convergence is attained.

## 4. The Derivation of the Utility Function from the Bid Price Function

Once the bid price function is estimated, it is easy to obtain the utility function. The bid price function satisfies $\tilde{r}(z,v,y) = y_1 - x$, and solving this equation with respect to $y_1$ yields $y_1 = \tilde{y}_1(x,z,v,s)$, where $\tilde{y}_1()$ is defined implicitly by the identity

$$\tilde{r}(z,v,\tilde{y}_1(x,z,v,s),s) \equiv \tilde{y}_1(x,z,v,s)-x. \tag{4.1}$$

The utility function is then $U(x,z,v,s) \equiv U^*(\tilde{y}_1(x,z,v,s),s)$. Since we do not know the shape of $U^*(y)$, the utility function $U(x,z,v,s)$ cannot be identified. However, $\tilde{y}_1(x,z,v,s)$ provides preference relations among x, z, and v for any individual s and can be used as a utility function.

In the case of the quadratic bid price function (3.5), $\tilde{y}(x,z,v,s)$ is the solution to

$$b_0+b_z^Tz+b_y^Ty+\tfrac{1}{2}z^TB_zz+\tfrac{1}{2}y^TB_yy+y^T(B_{yz}z+v)-y_1+x=0. \tag{4.2}$$

Decompose $b_y$, $B_y$, $B_{yz}$, and v as

$$b_y = (b_{y_1}, b_s^T)^T, \tag{4.3a}$$

$$B_y = \begin{bmatrix} B_{y_1y_1}, & B_{sy}^T \\ B_{sy_1}, & B_{ss} \end{bmatrix}, \tag{4.3b}$$

$$B_{yz} = (B_{y_1z}^T, B_{sz}^T)^T, \tag{4.3c}$$

$$v = (v_{y_1}, v_s^T)^T, \tag{4.3d}$$

where the dimensions of $b_{y_1}$, $b_s$, $B_{y_1y_1}$, $B_{sy_1}$, $B_{ss}$, $B_{y_1z}$, $B_{sz}$, $v_{y_1}$, and $v_s$ are respectively, 1x1, (m-1)x1, 1x1, (m-1)x1, (m-1)x(m-1), 1xn, (m-1)xn, 1x1, (m-1)x1. Then, equation (4.2) becomes

$$\tfrac{1}{2}B_{y_1y_1}(y_1)^2+(b_{y_1}+s^TB_{sy_1}+B_{y_1z}z+v_{y_1}-1)y_1+b_0+b_z^Tz+b_s^Ts+\tfrac{1}{2}z^TB_zz$$
$$+\tfrac{1}{2}s^TB_{ss}s+s^TB_{sz}z+s^Tv_s+x=0 \tag{4.4}$$

This equation has two solutions but $B_{y_1y_1} \leqq 0$ from the negative semi-definiteness of $B_y$ and $\partial\tilde{y}/\partial x > 0$ from

$$\frac{\partial\tilde{y}}{\partial x} = \frac{1}{1-\partial\tilde{r}/\partial y_1} = \frac{1}{1-\partial r/\partial y-(\partial r/\partial\bar{U})(\partial U^*/\partial y_1)} = \frac{1}{-(\partial r/\partial\bar{U})(\partial U^*/\partial y_1)},$$

$$> 0 \tag{4.5}$$

where $\partial r/\partial\bar{U} < 0$ and $\partial U^*/\partial y_1 > 0$ since a rise in the utility level reduces the bid price and an individual with a higher income has a higher utility level. The only solution compatible with these two conditions is

$$y_1 = \tilde{y}_1(x,z,v,s)$$

$$= \frac{1}{B_{y_1y_1}}\{-(b_{y_1}+s^TB_{sy_1}+B_{y_1z}z+v_{y_1}-1)-[(b_{y_1}+s^TB_{sy_1}+B_{y_1z}z+v_{y_1}-1)^2$$

$$-B_{y_1y_1}(z^TB_zz+s^TB_{ss}s+2(b_0+b_z^Tz+b_s^Ts+s^TB_{sz}z+s^Tv_s+x))]^{\frac{1}{2}}\}. \tag{4.6}$$

This function serves as the utility function of an individual with characteristics s.

## 5. Empirical Examples

Before presenting estimation results, we briefly describe our data set. The data have two sources: the Tokyo Metropolitan High-Rise Housing Survey and the household survey data collected by us.[1/] The former data contain descriptions of almost all high-rise houses (1339 buildings with 50412 apartments) which were sold between April 1980 and May 1981. Although this data set has detailed information on housing characteristics such as living area and accessibility to the CBD, it has no information on characteristics of households that bought the apartments. We therefore conducted a survey of those households. Because of the lack of resources, the household survey is conducted only in the two prefectures (Chiba and Saitama) in the suburb of Tokyo.[2/] Household data obtained in this way have household annual income, age, education, the number of children, etc. [3/] Finally, we linked these two types of data and obtained 282 samples of apartments and households who bought them.

In the estimations, the following variables are used as the characteristics of housing and household.

$z_1$: living area (square meters),

$z_2$: balcony area (square meters),

$z_3$: time distance to the CBD from the nearest railway station (minutes),

$z_4$: time distance to the nearest railway station (minutes),

$y_1$: annual income of each household (ten thousand yens),

$y_2$: age of the head of a household.

## 5-1. Quigley's Estimation Method

First, we apply Quigley's method to our housing data. The first step of his method is to estimate the equilibrium hedonic price function with Box-Cox transformation of the dependent variable,

$$\frac{p^{\lambda}-1}{\lambda} = a_0 + a_1 z_1 + a_3 z_3 + a_4 z_4 + u, \tag{5.1}$$

where notations are the same as those used in section 2. The estimation result of this function is shown in Table 1. Table 1 also presents the mean marginal prices for housing characteristics.

In the second step the first order conditions for household optimization (2.1) are estimated. In the estimation the dependent variables are marginal prices calculated from the estimate of equation (5.1). Table 2 reports the estimated parameters of the utility function, which yields

$$U(x,z) = [13.284\ z_1^{1.077} + 15.884\ z_2^{1.042} - 8.543\ z_3^{0.973} - 21.45\ z_4^{0.977} + x^{0.978}]^{\phi} . \tag{5.2}$$

## 5-2. An Estimation with Unobserved Attributes of Housing -- One Household Characteristic

Next, the bid price function is estimated by using the estimation method developed in section 3. In this subsection, we assume there is only one household characteristic, i.e., household income ($y_1$).

The bid price function which we estimate is

$$\tilde{r}(z,v,y_1) = b_0 + \sum_{i=1}^{4} b_{z_i} z_i + b_{y_1} y_1 + \tfrac{1}{2} B_{y_1 y_1} y_1^2 + \left(\sum_{i=1}^{4} b_{y_1 z_i} z_i + v_1\right) y_1 + v_0 , \tag{5.3}$$

Table 1

Parameter Estimates of Hedonic Price Function
and
Mean Marginal Prices

| Parameter | Estimated Value | t-value | Mean Marginal Price |
|---|---|---|---|
| $a_0$ | 102.07 | 25.1 | |
| $a_1$ | 0.967 | 16.9 | 22.65 |
| $a_2$ | 0.831 | 7.8 | 19.45 |
| $a_3$ | -0.408 | 10.0 | -9.52 |
| $a_4$ | -1.066 | 10.7 | -24.85 |
| $\lambda$ | 0.59 | | |
| $R^2$ | 0.650 | | |
| log likelihood | -1925.8 | | |

Table 2

Estimates
of
Generalized CES Utility Function Parameters

| Parameter | Estimated Value | t-value |
|---|---|---|
| $\log \frac{\alpha_1 \beta_1}{\varepsilon}$ | 2.683 | 57.73 |
| $\log \frac{\alpha_2 \beta_2}{\varepsilon}$ | 2.829 | 71.45 |
| $\log \frac{\alpha_3 \beta_3}{\varepsilon}$ | 2.138 | 53.03 |
| $\log \frac{\alpha_4 \beta_4}{\varepsilon}$ | 3.085 | 75.96 |
| $\beta_1 - 1$ | 0.077 | 42.37 |
| $\beta_2 - 1$ | 0.042 | 17.79 |
| $\beta_3 - 1$ | -0.027 | 21.57 |
| $\beta_4 - 1$ | -0.003 | 24.02 |
| $-(\varepsilon - 1)$ | 0.022 | 3.09 |

where for simplicity $B_z$ is assumed to be a null matrix. The estimation of (3.10) and (3.13) with the use of (3.11) and (3.12) yields the estimates of the parameters in the bid price function (5.3). The estimates are shown in Table 3. The implied bid price function and the utility function are

$$\tilde{r}(z,v,y_1) = 934.256 + 20.971z_1 + 19.383z_2 - 9.035z_3 - 23.196z_4 + 0.0841y_1 - 0.000206y_1^2 + (0.00269z_1 + 0.000604z_2 - 0.000818z_3 - 0.00256z_4 + v_1)y_1 + v_0, \quad (5.4)$$

$$\tilde{y}_1(x,z,v) = 2427.2\{f_1(z,v_1) + [(f_1(z,v_1))^2 + 0.000824f_2(x,z,v_0)]^{\frac{1}{2}}\} \quad (5.5)$$

where

$$f_1(z,x) = 0.00269z_1 + 0.000604z_2 - 0.000818z_3 - 0.00256z_4 + v_1 - 0.916,$$

$$f_2(x,z,v_0) = 934.256 + 20.971z_1 + 19.383z_2 - 9.035z_3 - 23.196z_4 + x + v_0.$$

## 5-3. Two Household Characteristics

Finally, the estimation in the preceding subsection is extended to the case of two household characteristics, i.e., household income ($y_1$) and the age of the head of the household ($y_2$). The estimation result is reported in Table 4. The estimated bid price function and the utility function are

$$\tilde{r}(z,v,y) = 921.10 + 20.605z_1 + 19.340z_2 - 8.979z_3 - 23.054z_4 + 0.0725y_1 + 1.500y_2 - 0.000355y_1^2 - 0.0656y_2^2 + (0.00232z_1 + 0.000521z_2 - 0.000705z_3 - 0.00221z_4 + v_1)y_1 + (0.0150z_1 + 0.00380z_2 - 0.00308z_3 - 0.00884z_4 + v_1)y_2, \quad (5.6)$$

Table 3

Parameter Estimates of the New Approach
(One household characteristic case)

| Parameter | Estimated Value | t-value | Parameter | Estimated Value | t-value |
|---|---|---|---|---|---|
| $b_0$ | 934.256 | 10.13 | $B_{y_1y_1}$ | -0.000412 | 4.97 |
| $b_{z_1}$ | 20.791 | 15.60 | $B_{y_1z_1}$ | 0.00269 | 3.71 |
| $b_{z_2}$ | 19.383 | 8.05 | $B_{y_1z_2}$ | 0.000604 | 0.69 |
| $b_{z_3}$ | -9.035 | 9.73 | $B_{y_1z_3}$ | -0.000818 | 2.15 |
| $b_{z_4}$ | -23.195 | 10.15 | $B_{y_1z_4}$ | -0.00256 | 2.61 |
| $b_{y_1}$ | 0.0841 | 2.20 | | | |
| $B_{y_1y_1}$ | -0.000412 | 4.97 | | | |

Table 4

Parameter Estimates of the New Approach
(Two household characteristics)

| Parameter | Estimated Value | t-value | Parameter | Estimated Value | t-value |
|---|---|---|---|---|---|
| $b_0$ | 921.100 | 9.94 | $B_{y_1z_1}$ | 0.00232 | 3.20 |
| $b_{z_1}$ | 20.606 | 15.16 | $B_{y_1z_2}$ | 0.000521 | 0.67 |
| $b_{z_2}$ | 19.339 | 8.04 | $B_{y_1z_3}$ | 0.000705 | 2.04 |
| $b_{z_3}$ | -8.979 | 9.67 | $B_{y_1z_4}$ | 0.00221 | 2.42 |
| $b_{z_4}$ | -23.055 | 10.10 | $B_{y_2z_1}$ | 0.0150 | 1.28 |
| $b_{y_1}$ | 0.0725 | 2.09 | $B_{y_2z_2}$ | 0.00380 | 0.66 |
| $b_{y_2}$ | 1.500 | 1.30 | $B_{y_2z_3}$ | 0.00308 | 1.18 |
| $B_{y_1y_2}$ | -0.000355 | 3.78 | $B_{y_2z_4}$ | 0.00884 | 1.08 |
| $B_{y_2y_2}$ | -0.0654 | 1.29 | | | |

$$\tilde{y}_1(x,z,v,y_2) = 2816.9\{g_1(z,v_1)+[(g_1(z,v_1))^2+0.00142g_2(x,z,v_0,y_2)]^{\frac{1}{2}}\} \tag{5.7}$$

where

$$g_1(z,v_1) = 0.00232z_1+0.00521z_2-0.000705z_3-0.000221z_4+v_1-0.928$$

$$g_2(x,z,v_0,y_2) = 921.1+20.605z_1+19.34z_2-8.979z_3-23.054z_4+x$$
$$+(0.015z_1+0.00380z_2-0.00308z_3-0.00884z_4+v_1)y_2+v_0.$$

## 5-4. A Comparison of Estimation Results

In this subsection, we compare the estimation results in subsections 5-1 and 5-2. Since the functional forms of the estimated utility functions are quite different, the results are not easy to compare. Here, we compare marginal rates of substitution (MRS's) between the composite non-housing good and housing characteristics calculated from the estimated utility functions.

After straightforward manipulations, the MRS's are

(i) Quigley's Method:

$$\frac{\partial U}{\partial z_i} \Big/ \frac{\partial U}{\partial x} = (\alpha_i\beta_i/\varepsilon)x^{1-\varepsilon}z_i^{\beta_i-1}, \qquad i=1,\ldots,4,$$

(ii) Our Approach:

$$\frac{\partial U}{\partial z_i} \Big/ \frac{\partial U}{\partial x} = F(x,z,v)^{-1}\{2427.2(\partial f_1/\partial z_i)+1213.6F(x,z,v)[2f_1(z,v_1)(\partial f_1/\partial z_i)$$
$$+0.000824(\partial f_2/\partial z_i)]\}, \quad i=1,\ldots,4,$$

where $F(x,z,v)=[f_1(x,v_1)^2+0.000824f_2(x,z,v_0)]^{-\frac{1}{2}}$.

Table 5

Marginal Rate of Substitution between x (nonhousing good) and $z_1$ (living area)

| x | $z_1$ | MRS | | x | $z_1$ | MRS | | x | $z_1$ | MRS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | Q | | | N | Q | | | N | Q |
| 205.9 | 40.0 | 24.06 | 21.85 | 388.3 | 40.0 | 24.36 | 22.16 | 200.0 | 75.4 | 25.48 | 22.93 |
| 231.6 | 45.0 | 24.31 | 22.11 | 388.3 | 45.0 | 24.56 | 22.36 | 250.0 | 75.4 | 25.55 | 23.04 |
| 257.3 | 50.0 | 24.54 | 22.34 | 388.3 | 50.0 | 24.76 | 22.54 | 300.0 | 75.4 | 25.63 | 23.13 |
| 283.1 | 55.0 | 24.79 | 22.55 | 388.3 | 55.0 | 24.95 | 22.71 | 350.0 | 75.4 | 25.71 | 23.21 |
| 308.8 | 60.0 | 25.03 | 22.75 | 388.3 | 60.0 | 25.15 | 22.86 | 400.0 | 75.4 | 25.78 | 23.28 |
| 334.6 | 65.0 | 25.26 | 22.93 | 388.3 | 65.0 | 25.35 | 23.00 | 450.0 | 75.4 | 25.86 | 23.34 |
| 360.3 | 70.0 | 25.50 | 23.10 | 388.3 | 70.0 | 25.54 | 23.13 | 500.0 | 75.4 | 25.93 | 23.40 |
| 386.1 | 75.0 | 25.73 | 23.25 | 388.3 | 75.0 | 25.74 | 23.26 | 600.0 | 75.4 | 26.08 | 23.49 |
| 411.8 | 80.0 | 25.97 | 23.40 | 388.3 | 80.0 | 25.93 | 23.37 | 700.0 | 75.4 | 26.23 | 23.57 |
| 437.5 | 85.0 | 26.20 | 23.54 | 388.3 | 85.0 | 26.13 | 23.48 | 800.0 | 75.4 | 26.37 | 23.64 |
| 463.3 | 90.0 | 26.43 | 23.68 | 388.3 | 90.0 | 26.32 | 23.59 | 900.0 | 75.4 | 26.51 | 23.70 |
| 489.0 | 95.0 | 26.66 | 23.80 | 388.3 | 95.0 | 26.51 | 23.68 | 1000.0 | 75.4 | 26.66 | 23.76 |
| 514.8 | 100.0 | 26.89 | 23.93 | 388.3 | 100.0 | 26.71 | 23.78 | 1100.0 | 75.4 | 26.80 | 23.81 |
| 566.3 | 110.0 | 27.35 | 24.15 | 388.3 | 110.0 | 27.09 | 23.95 | 1200.0 | 75.4 | 27.07 | 23.89 |

N.B. N:Our New Approach, Q:Quigley's Method.

Table 6

Marginal Rate of Substitution between x (nonhousing good) and $z_3$ (accessibility to the CBD)

| x | $z_3$ | MRS | | x | $z_3$ | MRS | | x | $z_3$ | MRS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | Q | | | N | Q | | | N | Q |
| 205.9 | 70.6 | 10.08 | 8.49 | 388.3 | 70.6 | 10.08 | 8.61 | 200.0 | 37.5 | 10.21 | 8.66 |
| 231.6 | 62.8 | 10.17 | 8.53 | 388.3 | 62.8 | 10.13 | 8.62 | 250.0 | 37.5 | 10.26 | 8.70 |
| 257.3 | 56.5 | 10.23 | 8.56 | 388.3 | 56.5 | 10.17 | 8.64 | 300.0 | 37.5 | 10.31 | 8.74 |
| 283.1 | 51.3 | 10.30 | 8.60 | 388.3 | 51.3 | 10.21 | 8.66 | 350.0 | 37.5 | 10.35 | 8.77 |
| 308.8 | 47.1 | 10.35 | 8.63 | 388.3 | 47.1 | 10.26 | 8.69 | 400.0 | 37.5 | 10.40 | 8.79 |
| 334.6 | 43.4 | 10.40 | 8.67 | 388.3 | 43.4 | 10.31 | 8.70 | 450.0 | 37.5 | 10.44 | 8.82 |
| 360.3 | 40.3 | 10.45 | 8.70 | 388.3 | 40.3 | 10.36 | 8.72 | 500.0 | 37.5 | 10.49 | 8.84 |
| 386.1 | 37.6 | 10.49 | 8.74 | 388.3 | 37.6 | 10.41 | 8.74 | 600.0 | 37.5 | 10.55 | 8.87 |
| 411.8 | 35.5 | 10.52 | 8.78 | 388.3 | 35.5 | 10.46 | 8.77 | 700.0 | 37.5 | 10.62 | 8.90 |
| 437.5 | 33.2 | 10.56 | 8.83 | 388.3 | 33.2 | 10.52 | 8.80 | 800.0 | 37.5 | 10.68 | 8.93 |
| 463.3 | 31.4 | 10.59 | 8.87 | 388.3 | 31.4 | 10.57 | 8.84 | 900.0 | 37.5 | 10.75 | 8.95 |
| 489.0 | 29.7 | 10.62 | 8.93 | 388.3 | 29.7 | 10.63 | 8.88 | 1000.0 | 37.5 | 10.81 | 8.97 |
| 514.8 | 28.5 | 10.65 | 8.99 | 388.3 | 28.5 | 10.69 | 8.94 | 1100.0 | 37.5 | 10.87 | 8.99 |
| 566.3 | 25.7 | 10.70 | 9.18 | 388.3 | 25.7 | 10.81 | 9.11 | 1200.0 | 37.5 | 10.99 | 9.03 |

N.B. N:Our New Approach, Q:Quigley's Method.

Table 5 shows MRS's between the nonhousing good, x, and the living area, $z_1$, along three lines on $(x,z_1)$ plane: (i) a ray from the origin passing through the point of the sample averages of x and $z_1$, i.e., $(x,z_1)=(388.3, 75.4)$; (ii) a horizontal line with the value of x fixed at 388.3; (iii) a vertical line with the value of $z_1$ fixed at 75.4. All other characteristics ($z_2$, $z_3$, and $z_4$) are also fixed at the sample averages.

The MRS at the average point is about 25.7 (in ten thousand yen per square meter) in our estimation method and 23.3 in Quigley's method. It can be seen from the Table that our estimates are about 10 to 15 percent higher than those obtained by Quigley's method. The estimated MRS's in case (i) indicates that the preference structure is not homothetic: the MRS's are higher at points farther away from the origin. The degree of nonhomotheticity measured by the ratio between MRS's at $z_1$=40 and $z_1$=110 is higher in our estimation method than in Quigley's. The result in case (ii) shows that the MRS becomes higher as living area is increased with the amount of nonhousing good held fixed in both estimation methods. This result is somewhat surprising although it does not contradict the standard quasi-concavity assumption on the utility function. In case (iii) the MRS becomes higher as the amount of nonhousing good is increased with living area fixed. In both case (ii) and case (iii), our estimation method yields larger increases in the MRS.

In Table 6, the absolute values of the MRS's between the nonhousing good, x, and time distance to the CBD from the nearest railway station, $z_3$, are reported in a similar way to Table 5. Because the time distance to the CBD is a bad, the MRS is negative and a ray from the origin on $(x, z_3)$ plane has a very different meaning from that on $(x, z_1)$ plane. Instead of a ray from the origin on $(x, z_3)$ plane, therefore, we consider a ray from the origin on $(x, 1/z_3)$ plane in case (iv). Cases (v) and (vi) are similar to cases (ii) and (iii) and would require no explanation. The MRS at the average point $(x, z_3)$ =(388.3, 37.5) is about 10.4 (in ten thousand yen per minute) in our estimation method and 8.74 in Quigley's method (in absolute value).

From the Table it can be seen that the MRS's are 15 to 20 percent higher in our estimates than Quigley's. When x and $1/z_3$ are increased proportionately in case (iv), the MRS becomes higher in both estimates, but the rate of increase is faster in Quigley's method. If only $z_3$ is decreased of if only x is increased, then the MRS becomes higher, and the rate of increase is faster in our estimation method than in Quigley's.

The above comparison shows significant difference in the estimated MRS's between Quigley's and our estimation methods. The choice of an estimation method is therefore very important and search for a better method must be continued. Also, it is not possible, at least at this stage, to test which method yields better estimates. It would be desirable to construct a general framework in which such a test can be conducted.

## References


Brown, J.N. and H.S. Rosen, (1982), "On the Estimation of Structural Hedonic Price Models," Econometrica 50, 765-768.

Ellickson, B., (1981), "An Alternative Test of the Hedonic Theory of Housing Markets," Journal of Urban Economics 9, 56-79.

Harrison, D. and D.L. Rubinfeld, (1978), "Hedonic Housing Prices and the Demand for Clean Air," Journal of Environmental Economics and Management 5, 81-102.

Lerman, S.R. and C.R. Kern, (1983), "Hedonic Theory, Bid Rents, and Willingness-to-Pay: Some Extensions of Ellickson's Results," Journal of Urban Economics 13, 358-363.

Linneman, P., (1980), "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market," Journal of Urban Economics 8, 47-68.

Linneman, P., (1981), "The Demand for Residence Site Characteristics," Journal of Urban Economics 9, 129-148.

Nihon Kohsoh Jutaku Kyokai (Japan High-Rise Housing Association), (1980), Tokyo Metropolitan High-Rise Housing Survey.

Quigley, J.M., (1982), "Nonlinear Budget Constraints and Consumer Demand: An Application to Public Programs for Residential Housing," Journal of Urban Economics 12, 177-201.

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

Witte, A.D., H.J. Sumka and H. Erekson, (1979), "An Estimate of a

Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets," Econometrica 47, 1151-1173.

# Footnotes

1. The Tokyo Metropolitan High-Rise Housing Survey has information on the following housing characteristics: the purchase price of an apartment, living area, balcony space, lot size of a building, accessibility to the CBD, accessibility to the nearest railway station, the number of elevators, parking space, the availability of heating and air conditioning systems, and so on. In order to make our estimation comparable to Quigley's, the purchase price of an apartment is converted into an annual rental by multiplying it by the discount rate of 0.08.
2. In the Tokyo metropolitan area there are four prefectures: Tokyo, Chiba, Saitama, and Kanagawa. It is usually considered Chiba and Saitama to be similar and to constitute the same housing market.
3. Household data were obtained by interviews with individual households in March 1982. The data contain annual family income, wife's income, family size, age of each family member, the number of children, job, education, and sources of funds used to finance the condominium.