No. 778

# Estimation of the Hedonic Price Function Using Monotonicity Restrictions

by

Yasushi Kondo

June 1998

# Estimation of the Hedonic Price Function Using Monotonicity Restrictions

Yasushi Kondo*

*Doctoral Program in Policy and Planning Sciences, University of Tsukuba,*

*Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan*

May 1998

Address for correspondence:

Yasushi Kondo

Doctoral Program in Policy and Planning Sciences,

University of Tsukuba,

Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan

email: ykondo@shako.sk.tsukuba.ac.jp

**Abstract of "Estimation of the Hedonic Price Function Using Monotonicity Restrictions" by Yasushi Kondo**

In hedonic analyses, estimating imputed prices of utility-bearing characteristics is of great interest. Nevertheless, it is not so easy to estimate the imputed prices because characteristics display multicollinearity inherently. As an improved estimation method, this paper examines the constrained least squares estimator with inequality-form constraints which represent the monotonicity of the hedonic price function. The estimators of imputed prices are guaranteed to be nonnegative by definition and are expected to have smaller variances than the unconstrained least squares estimators. To compare the constrained estimator with the unconstrained one, an empirical example and Monte Carlo experiments are presented.

JEL CLASSIFICATION NUMBERS: C13, L85, R32.

# I. INTRODUCTION

This paper examines the performance of the constrained least squares method (abbreviated to CLS) for estimating the hedonic price function (abbreviated to HPF) and the imputed prices of utility-bearing characteristics. The constraints for the CLS estimator are constructed to represent monotonicity restrictions on the HPF. Compared with the (unconstrained) ordinary least squares method (abbreviated to OLS), the CLS has advantages that it necessarily provides non-negative estimates of imputed prices by definition and that its variance is expected to be smaller than that of the OLS estimator.

In hedonic analyses, estimating imputed prices of utility-bearing characteristics is of great interest. There seem to exist no other methods than the hedonic approach to estimate implicit values of characteristics which are not traded in a market explicitly but bundled into an explicitly traded product. The implicit values of characteristics are called their imputed prices. Estimated imputed prices can be used in a tax policy making and policy assessment and in constructing quality-adjusted price indices. In addition, they are used to estimate the structural parameters contained in the bid functions of demanders (Quigley, 1982; Kanemoto and Nakamura, 1986). Nevertheless, it is difficult to estimate the imputed prices "accurately" because utility-bearing characteristics display multicollinearity inherently. That is, we might unfortunately face the situation that estimated variances are large and/or that estimated imputed prices are negative.

In the economic theoretical context of the hedonic analysis, the HPF is determined by various conditions such as producers' offers, demanders' bids, market situation (competitive or not, arbitrage can be done or not), and how agents are distributed. Rosen (1974), Ohta (1975), and Arguea and Hsiao (1993) studied the properties of the HPF. They considered different market situations, and so they obtained different properties. As far as I know, Rosen's result is accepted most widely among the three studies above. He showed that the HPF is the envelope curve of producers' offers and demanders' bids so we have no a priori information about the curvature of the HPF.[1] Economic theory tells

[1] Because the functional form of an HPF is scarcely restricted by economic theory, the Box-Cox transformation (Box and Cox, 1964) is widely applied to make data themselves "speak"

us only that the HPF is monotonic in utility-bearing characteristics, which is derived from the assumption of free disposal and nonsatiation of preference.[2]

As stated above, it is difficult to estimate the imputed prices of characteristics because of inherent multicollinearity among characteristics. For example, the total number of rooms and bedrooms, and lot area do not vary independently for housing characteristics, and a heavy passenger car likely has a high horse-power. Therefore, in order to obtain "accurate" estimation results, we should cope with the inherent problem of multicollinearity using an appropriate device. I believe that a non-sample information, that is, the monotonicity restriction on the HPF, can be a device to circumvent the problem of multicollinearity; the CLS uses such a non-sample information and the OLS does not.

In general, a parameter estimate can be *irregular* even if the true parameter is hypothesized to be *regular* and the hypothesis is correct, where the word "regular" is used for the same sense as in the following cases: a conditional demand system is said to be regular if it satisfies the corresponding integrability condition; and a marginal propensity to consume is said to be regular if it belongs to the unit interval $[0, 1]$. In this paper, an HPF or an estimated HPF is said to be regular if it is monotonic. An irregular parameter estimate appears more likely as the variance of the estimator is larger. Though that is a (rigorous) analysis in statistics, econometricians sometimes not only believe the true parameters to be regular but also wish parameter estimates to be so. The reasons why we wish so are that irregular estimation results are difficult to interpret and that we have to give up continuing our analysis in the worst case. For example, since Quigley's (1982) procedure uses the logarithms of estimated imputed prices to estimate structural parameters in hedonic models, the procedure can continue only if all the estimated imputed prices are positive. Therefore, it is worthwhile to develop an estimation method which guarantees parameter estimates themselves to be regular.

---

much.

[2]Ohta (1975) and Arguea and Hsiao (1993) derived curvature conditions that the HPF should satisfy. Note that the HPF is monotonic even if we accept the market situation they assumed.

Varian (1974), Atkinson and Crocker (1987), and Gilley and Pace (1995) used Bayesian approaches from similar motivations as this paper. Their procedures to incorporate a non-sample information take the forms of prior densities. If the monotonicity restriction can be represented as inequalities only in parameters, we can use their Bayesian methods straightforwardly. A sufficient condition for the monotonicity restriction to be given by inequalities only in parameters is that the HPF is a linear combination of utility-bearing characteristics or a one-to-one transformation of it. In most existing hedonic analyses, specifications for the HPF satisfy this sufficient condition. Specification tests such as the RESET (regression specification error test), however, might reject the condition. This is the case for our empirical example presented later. In such cases, we cannot represent the monotonicity restriction on the HPF only with a prior density. In addition, we have to specify the likelihood function to conduct the Bayesian approach. The use of a misspecified likelihood might lead us to a seriously biased estimation result. In contrast, the constrained least squares estimator (abbreviated to CLSE) can be used even when Bayesian methods cannot be used straightforwardly.

This paper is organized as follows. Section II presents the regression model of the HPF, the definition of the CLSE, and a bootstrap method to estimate the variance of the CLSE. In Section III, the CLS is applied to the hedonic office rent function in the Tokyo central business district to be compared with the OLS. Section IV conducts Monte Carlo experiments. Finally, concluding remarks are mentioned in Section V.

## II. THE MODEL AND THE ESTIMATION METHOD

Consider a purely competitive market of differentiated products and suppose that a differentiated product is distinguished completely by a finite number, say $\ell$, of observable utility-bearing characteristics, as Rosen (1974) studied. Suppose further that producers and demanders in the market have production functions and utility functions, respectively, and these functions are defined on a convex set $\mathcal{Z} \subset \mathbb{R}^{\ell}$, which coincides with the domain of the HPF. We wish to estimate the HPF for the market price and the imputed prices of characteristics using a data set $\{(p_i, z_i) : i = 1, \ldots, N\}$, where $p_i$ and $z_i$ are the price of a differentiated product and its vector of characteristics, respectively, for the $i$th observation.

Suppose that the logarithm of the HPF, $h(\cdot)$, is given by the following quadratic form

$$h(z_i) = \alpha_0 + \alpha^{\mathsf{T}} z_i + z_i^{\mathsf{T}} A z_i / 2 \quad \text{for } i = 1, \ldots, N, \tag{1}$$

where $\alpha_0$, $\alpha$, and $A$ denote parameters to be estimated, they have suitable dimensions, and the superscript $\mathsf{T}$ denotes the transpose. Note that we can assume $A$ be symmetric without any loss of generality and that this assumption is necessary for all the parameters to be identified. Adding an error term $u_i$ to the right-hand side of (1), we estimate parameters in the regression equation $\log p_i = h(z_i) + u_i$ under the assumption $Eu_i = 0$ for $i = 1, \ldots, N$. Since $p_i > 0$, the sign of imputed prices $\partial p_i / \partial z_i = p_i(\partial h / \partial z_i)$ coincides with that of the gradient $\nabla h(\cdot) = \partial h / \partial z_i$ of $h$. Thus, the monotonicity restrictions on the HPF can be written as

$$\nabla h(z) = \alpha + Az \geq 0 \quad \text{for all } z \in \mathcal{Z}. \tag{2}$$

**Remark 1.** The reasons why I choose a quadratic form to specify the HPF are: (i) While most existing empirical studies in the hedonic analysis specify the HPF as linear in characteristics, a specification test might reject the linearity of the HPF, as is the case for our empirical example presented later. The RESET is one of the most well-known and the easiest to implement specification tests, and a quadratic regression function is an exact form of the alternative hypothesis for the RESET;[3] and (ii) The optimization problem to define the CLSE takes the form of the quadratic programming problem (abbreviated to QP) if the regression function is linear in parameters. There exists a polynomial time algorithm to solve the QP, that is, the QP is one of the nonlinear programming problems which we can solve reliably and quickly.

Now I define some notations for brevity. Let $\theta$ denote the vector consisting of all the parameters $\alpha_0$, $\alpha$, and $A$. For a subset $W \subset \mathcal{Z}$, let $S(W)$ denote the system of inequalities $\alpha + Az \geq 0$ for all $z \in W$ on $\theta$. By this notation, the

[3] See Davidson and MacKinnon (1993, pp. 195–196) for detail.

system of inequalities (2) can be denoted by $S(\mathcal{Z})$. Define the function $Q(\cdot)$ representing a squared residual as $Q(\theta; p_i, z_i) = (\log p_i - \alpha_0 - \alpha^\top z_i - z_i^\top A z_i/2)^2$.

Then, using the least squares criterion, the optimization problem to define the CLSE is given by

$$\text{minimize} \quad N^{-1} \sum_{i=1}^{N} Q(\theta; p_i, z_i) \quad \text{subject to} \quad S(\mathcal{Z}). \tag{3}$$

This optimization problem, however, is generally intractable because it would have uncountably many constraints. Thus, we consider the following tractable form

$$\text{minimize} \quad N^{-1} \sum_{i=1}^{N} Q(\theta; p_i, z_i) \quad \text{subject to} \quad S(z_1, \dots, z_N), \tag{4}$$

where $S(z_1, \dots, z_N)$ stands for $S(\{z_1, \dots, z_N\})$ by a slight abuse of notation. This problem is a QP, and it has a unique global minimizer under the same condition as the least squares estimator can be identified. Note that the difference between the two problems above is only in the constraints. Since $\mathcal{Z}$ is assumed to be convex and over restrictive constraints would make the estimator inconsistent, a plausible choice of the system of inequality constraints is $S(W)$ with $W = \mathrm{conv}(z_1, \dots, z_N)$, the convex hull of all the observation points. Note also that the system $S(z_1, \dots, z_N)$ for the problem (4) is equivalent to $S(\mathrm{conv}(z_1, \dots, z_N))$ because the inequalities in the system $S(\cdot)$ are bilinear in parameters $\theta$ and characteristics $z$.

**Remark 2.** This equivalence between $S(\mathrm{conv}(z_1, \dots, z_N))$ and $S(z_1, \dots, z_N)$ can be used to reduce a computational burden as follows. First, find all the extreme points of $\mathrm{conv}(z_1, \dots, z_N)$, and let $W$ be the set of all the extreme points. Then, solve the problem (4) after replacing the constraints with $S(W)$. Finding extreme points can be done by solving linear programming problems at most $N$ times.

The CLSE defined above can easily be shown to be consistent under the assumption that the HPF is monotonic, or equivalently, the true parameters in the HPF satisfy the monotonicity restrictions. Note that, under this assumption, the CLSE coincides with the ordinary least squares estimator (abbreviated to OLSE) for a sufficiently large sample. If the regressors are nonstochastic, then the CLSE is equivalent to what Judge and Takayama (1966) and Judge and Yancey (1986) studied. Their results, however, are difficult to be applied here because their studies are made under the assumptions that error terms are independently and identically distributed as a normal distribution, regressors are nonstochastic, and the number of constraints is a fixed integer, say $K$. In addition, the distribution of the CLSE they derived is extremely complicated as a mixture of $2^K$ distributions. Therefore, I apply the bootstrap to estimate variances of CLSE.

The bootstrap, originated by Efron (1979), is a nonparametric method for statistical inferences based on resampling. A number of varieties of the bootstrap has been proposed in the literature (Efron and Tibshirani, 1993; Shao and Tu, 1995). The three bootstrap techniques, the naive bootstrap, the parametric bootstrap, and the wild bootstrap, are typical ones for the regression model. Given a data set $X_0 = \{(\log p_i, z_i) : i = 1, \dots, N\}$, the naive bootstrap draws bootstrap samples, which are of the same size with the original data set, from the empirical joint distribution of the dependent and the independent variables. Then a bootstrap sample for the naive bootstrap is given by $\{(\log p_{k(i)}, z_{k(i)}) : i = 1, \dots, N\}$, where $k(1), \dots, k(N)$ are integers distributed as the distribution putting mass $n^{-1}$ to each of integers $1, \dots, N$. The parametric bootstrap draws from the empirical distribution of the estimated residuals so that a bootstrap sample is $\{(\widehat{h}(z_i) + \widehat{u}_{k(i)}, z_i) : i = 1, \dots, N\}$, where $\widehat{h}$ is an estimator of the regression function $h$, and $\widehat{u}_i = \log p_i - \widehat{h}(z_i)$. The wild bootstrap draws error terms $\varepsilon_i$ from estimated distributions such that $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \widehat{u}_i^2$, and $E\varepsilon_i^3 = \widehat{u}_i^3$. Then a bootstrap sample for the wild bootstrap is $\{(\widehat{h}(z_i) + \varepsilon_i, z_i) : i = 1, \dots, N\}$.

Of the three techniques above, the wild bootstrap is the most appropriate for our study because it is consistent under a more general condition, where the others are not consistent (Shao and Tu, 1995; Mammen, 1993). That is, if the error terms are heteroskedastic, then the naive bootstrap and the wild bootstrap are consistent but the parametric bootstrap is not.[4] The dimension of the

[4]Härdle and Marron (1991) studied how to construct simultaneous error bars for a nonparametric regression estimator. They showed that the naive bootstrap is not consistent while the wild bootstrap is consistent for their purpose.

distribution that the wild bootstrap tries to mimic is smaller than that the naive bootstrap does; the former is only one and the latter is the number of regressors plus one. Results become more accurate as the bootstrap technique mimics a smaller dimensional distribution. Therefore, I will use the wild bootstrap in the next section with the following error terms $\varepsilon_i$:

$$\varepsilon_i = \widehat{u}_i((\delta_1 + v_{1i}/\sqrt{2})(\delta_2 + v_{2i}/\sqrt{2}) - \delta_1\delta_2), \tag{5}$$

where $\delta_1 = (3/4 + \sqrt{17}/12)$, $\delta_2 = (3/4 - \sqrt{17}/12)$, and $v_{1i}$ and $v_{2i}$ are independently distributed as the standard normal distribution.

Let $B$ be the number of bootstrap replications. To distinguish each of $B$ bootstrap samples, I introduce a subscript and a superscript to observations in bootstrap samples and denote the $b$th bootstrap sample by $X_b = \{(\log p_i^b, z_i^b) : i = 1, \ldots, N\} = \{(\widehat{h}(z_i) + \varepsilon_i^b, z_i) : i = 1, \ldots, N\}$, for $b = 1, \ldots, B$.

Let $T(X)$ denote the estimator of interest, which is a column vector of a certain dimension, defined as a function of a sample $X$. A bootstrap covariance matrix estimator $\widehat{V}_{\text{BOOT}}$ for $T(X)$ is given by

$$\widehat{V}_{\text{BOOT}} = \frac{1}{B}\sum_{b=1}^{B}(T_b - T_0)(T_b - T_0)^{\mathsf{T}}, \tag{6}$$

where $T_0 = T(X_0)$ and $T_b = T(X_b)$ for $b = 1, \ldots, B$.

# III. AN APPLICATION: OFFICE RENTS IN THE TOKYO CBD

This section presents an empirical application in order to show how much better the CLSE defined in the previous section is than the OLSE with the following criteria: how small estimated variances are, and how accurately out-of-sample predictions work.

### *Data*

For estimating an HPF, the OLS and the CLS are applied to the office rent data in the Tokyo central business district. The data set is a subset of the data set used by Nagai, Kondo, and Ohta (1997). They found that the office rent discount in transaction is not ignorable and that the discount rate is much higher for new offices than for used offices, and it is more so in the period of the asset deflation than in that of the bubble. So I use the data set consisting of only used offices for the time period 1985–1986 in the beginning of the bubble to avoid a high discount rate in this paper. According to their estimation result, the estimated discount rate is at most 3% for this sample period. The number of observations is 550.

The dependent variable, RENT, is the monthly office rent per square meter (yen/$m^2$). The continuous variables of characteristics are as follows: DIST (10 m) is the distance of the shortest route from the office building to the nearest station of any one of the railroads and the subways; ACCT and ACCS (minutes) are the times required to go by the shortest route from the nearest station to Tokyo Station and to Shinjuku Station, respectively, by train or subway; AGE (year) is the length of time which has passed since the building construction was completed; DENS (1000 men/$km^2$) is the employee-density of the ward where the office locates; FLOSP (1000 $m^2$) is the total floor space of the whole building in which the office is; and VOLM (absolute number) is the cubic restriction of the building volume to the lot size. The dummy variables are defined as follows: FONE is unity if the office is on the first floor, and is zero otherwise; MAJOR is unity if the lessor is one of the large companies, and is zero otherwise; and D86 is unity if the observation is of 1986, and is zero otherwise.

The main data source is *Survey Tables of the Circumstances of Buildings* (*Biru Jittai Chôsa Hyô,* in Japanese), the 1985 to the 1987 fiscal year editions, edited and published by the corporate juridical person named Tokyo Building Owners and Managers Association (Shadan Hôjin Tôkyô Birujingu Kyôkai, in Japanese). This data source gives us data of RENT, AGE, FLOSP, FONE, MAJOR, and the time dummy D86. DIST is measured on the map in *Housing Map* (*Jûtaku Chizu,* in Japanese), edited and published by Zenrin Co., Ltd., 1990. The data sources for ACCT and ACCS are *Attached Line Maps of Weekly Housing Information* (*Shûkan Jûtaku Jôhô Huzoku Rosenzu,* in Japanese), edited and published by Recruit Co., Ltd., July 1, 1992, and *Time Tables* of railroads and subways. The data sources of DENS and VOLM are, respectively, *Establishment Census of Japan 1986* (*Jigyôsho Tôkei Chôsa Hôkoku 1986,* in Japanese), edited by Statistics Bureau of Management and Coordination Agency (Sômuchô Tôkeikyoku, in Japanese), 1993, and *Map of City*

*Planning of Tokyo Metropolis* (*Tôkyô Toshi Keikakuzu*, in Japanese), edited by Tokyo Metropolis and published by Kokusai Chigaku Kyôkai Co., Ltd., 1994.

Among the above nine characteristics, DIST, ACCT, and ACCS indicate transportational conveniences, and AGE, DENS, FLOSP, VOLM, FONE, and MAJOR indicate the benefit of agglomeration or the performance of the office building. DIST, ACCT, ACCS, and AGE are *disutility*-bearing characteristics while DENS, FLOSP, VOLM, FONE, and MAJOR are *utility*-bearing characteristics. AGE, DENS, FLOSP, and VOLM are positive for all the observations by their definition so that we can take the logarithms of them. Summary statistics of the variables are shown in Table 1. For a more detailed explanation of the data set, see Nagai, Kondo, and Ohta (1997, section 2 and appendix).

*Preliminary Estimations and Tests*

For selecting a specification of the HPF, the four linear-in-parameter models are considered: linear in characteristics' levels (LV), linear in the logarithms of characteristics(LL), quadratic in characteristics' levels (QV), and quadratic in the logarithms of characteristics (QL). In the models LL and QL, characteristics that can take nonpositive values are included in levels. When the disutility-bearing characteristics, DIST, ACCT, ACCS, and AGE, are included as regressors; their signs are negated so that all the imputed prices should be positive. For AGE in the models LL and QL, I will negate its sign after taking its logarithm.

All the four models have the intercept term. The dummy characteristics FONE, MAJOR, and D86 are included as intercept-term type and their quadratic terms are not constructed. All the calculations reported in this subsection are done with TSP version 4.4.

The testing results are presented in Table 2. All the models are estimated by the OLS. First, to check whether there are higher-order nonlinearity in characteristics or not, two RESETs were performed for each model. One, called RESET2, has the squared fitted value as an additional regressor, and the $t$-value for its coefficient is used as a test statistic. The other, called RESET3, has the squared and the cubed fitted values as regressors, and the usual $F$-test is used. According to these criteria, higher-order nonlinearity was detected for the model LL at the 1% significance level.

Second, to check whether the continuous characteristics should be included in level or in logarithm, I performed a non-nested test called $J$-test (Davidson and MacKinnon, 1993, 381–383). The testing procedure for the null hypothesis LL against the alternative LV is as follows: first, estimate the model LV and obtain fitted values; second, estimate the model LL with including the fitted values as an additional regressor; and finally, perform the standard $t$-test for the coefficient of the fitted values. According to this criterion, the models LV and QV were strongly rejected against LL and QL, respectively. In contrast, the model LL was not rejected against LV and QL was not rejected against QV at the 1% significance level.

As a result, the model QL was selected because a higher-order nonlinearity was not detected and it was not rejected against QV.

*Estimation Performance for Imputed Prices of Characteristics*

For notational simplicity, let $\pi_{\mathsf{CHAR}}$ denote the imputed price of the characteristic called CHAR: for example, $\pi_{\mathsf{DIST}}$ is the imputed prices of DIST.

The model QL, which were selected by specification tests in the last subsection, were estimated both by the OLS and by the CLS. The number of bootstrap replications for estimating the variances of the CLSE is 2500. All the calculations reported in this subsection and the next section are done with MATLAB version 5.1. To save space, only the selected results are reported in Table 3. Both the OLS and the CLS provide quite good fits, judging from the R-squares and considering that cross-section data are used.

Since the imputed prices of characteristics depend on levels of characteristics, I first report the estimates of the imputed prices evaluated at the sample mean of characteristics; those evaluated at all the observation points are reported later. Each of the standard errors estimated by the CLS is smaller than the corresponding standard error estimated by the OLS. Among the seven continuous characteristics, the OLSEs of $\pi_{\mathsf{ACCT}}$ and $\pi_{\mathsf{ACCS}}$ at the sample mean are negative so that they violate the monotonicity of the HPF, while all the CLSEs of the imputed prices of characteristics are nonnegative by their definition. Based on the results of the OLS, the imputed prices are not significantly different from zeros for the four characteristics DIST, ACCT, ACCS, and DENS. In contrast, the CLSEs of $\pi_{\text{DIST}}$ and $\pi_{\text{DENS}}$ are larger than the OLSEs of them and each of their standard errors by the CLS is smaller than the corresponding standard error by the OLS so that $\pi_{\text{DIST}}$ and $\pi_{\text{DENS}}$ are significantly different from zeros.

For the estimated results of the OLS, the monotonicity restrictions are violated at almost all the observation points; they are satisfied at only 3.82% of the observation points. Table 4 shows how many observation points satisfy the monotonicity restrictions. Looking at the results for each characteristics separately, the partial monotonicities for DIST, ACCT, ACCS, and DENS are violated at more than a half of the observation points. In contrast, the partial monotonicities for AGE, FLOSP, and VOLM are satisfied at almost all the observation points. Figure 1 shows how the estimates of the imputed prices evaluated at all the observation points are distributed. From the Figure 1 and Table 3, it is observed that the constrained estimators have smaller estimated standard errors than the OLS estimators and that, most CLSEs of the imputed prices of the transportational conveniences by train or subway (ACCT and ACCS) are nearly equal to zeros.[5]

*Out-of-Sample Prediction Performance*

By construction, the in-sample R-square of the OLS must be greater or equal to that of the CLS. This relation, however, does not hold between their out-of-sample prediction errors. The CLS can outperform the OLS by the following reasons: (i) the monotonicity restrictions can limit the influence of potentially erroneous or unrepresentative observations; and (ii) the estimated residuals within sample possibly forecast the out-of-sample errors *with* bias. Excessively optimistic results of error estimates would be obtained by using the observations in a single data set both for fitting a model and for evaluating the goodness of fit. A solution to this problem is dividing a sample into two parts (the sample observations and the out-of-sample observations). As a method of the cross validation, one part is used to fit a model and the other is used to evaluate the goodness of fit.

Let $I$ and $O$ denote subsets of the index set $\{1, \ldots, N\}$ such that $I \cup O =$

---

[5]This weakness of the characteristics related to transportational convenience for explaining price variation is similar to that Nagai, Kondo, and Ohta (1997) found.

$\{1, \ldots, N\}$ and $I \cap O = \emptyset$. Given a pair of $I$ and $O$, the corresponding sample observations and out-of-sample observations can be written as $\{(\log p_i, z_i) : i \in I\}$ and $\{(\log p_i, z_i) : i \in O\}$, respectively. The mean squared error of the out-of-sample prediction is given by

$$\frac{1}{|O|} \sum_{i \in O} (\log p_i - \widehat{h}_I(z_i))^2, \tag{7}$$

where $|O|$ is the number of elements of the set $O$, and $\widehat{h}_I(\cdot)$ is an estimate of the regression function estimated with the sample observations. Only with a single partition, the mean squared error would be large or small by chance. The mean squared error should be calculated with all possible partitions conceptually. A number of random partitions, however, is used because the number of all possible partitions is quite huge. Then the root mean squared error (abbreviated to RMSE) is defined as

$$\text{RMSE} = \left( \frac{1}{R|O|} \sum_{r=1}^{R} \sum_{i \in O_r} (\log p_i - \widehat{h}_{I_r}(z_i))^2 \right)^{1/2}, \tag{8}$$

where $R$ is the number of random partitions, $I_r$ and $O_r$ are the index sets for the sample observations and the out-of-sample observations, respectively, for the $r$th random partition, $r = 1, \ldots, R$.

To evaluate not the absolute performances of the OLS and CLS but the relative performance of them for out-of-sample prediction, Table 5 presents the ratio of RMSE by OLS to that by CLS. Five sizes of sample observations, 20%, 35%, 50%, 65%, and 80% of all the observations, are examined. For each case, the number of random partitions is 1000.

As these results show, the CLS outperforms the OLS much for the sample observations of 20% and 35% of all the observations. For 50%, 65%, and 80% cases, the OLS and the CLS produce similar levels of out-of-sample prediction errors. Thus, we can say that the CLS outperforms the OLS in the out-of-sample prediction (the OLS displays 11.86% higher RMSE than the CLS, on average).

## IV. A MONTE CARLO EXAMINATION OF PERFORMANCE FOR ESTIMATING IMPUTED PRICES

Empirical applications in the last section show the possibility that the CLSEs provide smaller variances than the OLSEs. From the viewpoint of the statistical decision theory, however, the mean squared error (abbreviated to MSE) is a more appropriate criterion than the variance because the CLSE is a *biased* estimator. Calculating the MSE requires true parameters to be known. Therefore, I examine the performance of OLS and CLS for estimating imputed prices of characteristics by Monte Carlo experiments.

A key point to determine the relative performance of the OLSE and the CLSE is the probability that the OLSE violates the monotonicity restrictions; the CLSE, by definition, provides estimation results almost identical to the OLSE if such a probability is nearly equal to zero. To control the probability of violation, I generate 75 cases. These consist of all the combinations of the five different vectors of parameters, the three levels of R-square, and the five matrices of regressors that display different degrees of multicollinearity. Of course, the relative performance of the OLSE and the CLSE partly depends on other factors such as the number of observations and the stringentness of monotonicity restrictions.[6] Nevertheless, I hold these factors constant in order to concentrate on the influence of the changes of R-square and that of multicollinearity on the relative performance.

*Generation of Data*

To make the Monte Carlo experiments realistic, artificial data are generated so that various features of the observed data are, if possible, kept intact except for the factors to be controlled. The utility-bearing characteristics used for the experiments are the six continuous characteristics, DIST, ACCT, AGE, DENS, FLOSP, and VOLM, while ACCS, FONE, MAJOR, and D86 are discarded. The discarded variables are dummies or characteristics having the imputed price nearly equal to zero in the empirical application of the last section. Therefore, the original data matrix $X$ of characteristics has the $i$th row $x_i$ such that $x_i = [-\mathsf{DIST}_i, -\mathsf{ACCT}_i, -\log(\mathsf{AGE}_i), \log(\mathsf{DENS}_i), \log(\mathsf{FLOSP}_i), \log(\mathsf{VOLM}_i)]$. Let $N$ and $\ell$ denote the number of observations ($N = 550$), and that of characteristics

[6]When the true parameters locate close to the boundary of the vector of parameters satisfying monotonicity restrictions, the probability of violation rises so that the CLSE has a much smaller variance and a much larger bias than the OLSE. If, on the other hand, the true parameters lie far from any boundary point, the bias of the CLSE will be close to zero.

($\ell = 6$), respectively. After discarding three characteristics and the time dummy, the predicted values of log(RENT) are constructed with the CLSEs of parameters in the last section. These predicted values are used to set up the true parameters in Monte Carlo experiments by regressing them on artificial regressors.

To control the degrees of multicollinearity, the five matrices of characteristics with the condition numbers 10, 20, 30, 50, and 70 are generated as follows:[7] (i) Scale each column of $X$ to unit length by letting $Y = XD^{-1}$ with the diagonal matrix $D$ whose $j$th diagonal element is the square root of the sum of squared elements in the $j$th column of $X$; (ii) Performing the singular value decomposition of $Y$, obtain the matrices $U$, $V$, and $S$ such that they are $N \times N$, $\ell \times \ell$, and $N \times \ell$, respectively, $Y = USV^{\mathsf{T}}$, $U$ and $V$ are unitary matrices, the matrix consisting of the first $\ell$ rows of $S$ has the singular values $s_1, \ldots, s_\ell$ of $Y$ as the main diagonal elements, and the other elements of $S$ are all zeros. The singular values $s_1, \ldots, s_\ell$ are descendingly ordered. Note that $s_1/s_\ell$ is the condition number of $Y$ and that $\sum_{j=1}^{\ell} s_j^2 = \ell$; (iii) Construct $\lambda_j$'s representing the relative position of $s_j$ between $s_1$ and $s_\ell$ so that $s_j = s_\ell + \lambda_j(s_1 - s_\ell)$ for $j = 1, \ldots, \ell$; (iv) Given a condition number, say $c$, find $t_1, \ldots, t_\ell$ satisfying $t_1/t_\ell = c$, $t_j = t_\ell + \lambda_j(t_1 - t_\ell)$ for $j = 1, \ldots, \ell$, and $\sum_{j=1}^{\ell} t_j^2 = \ell$. Then, construct an $N \times \ell$ matrix $T$ such that the first $\ell$ rows of $T$ is $\mathrm{diag}\{t_1, \ldots, t_\ell\}$ and the other elements are all zeros; and (v) Perform the singular value decomposition and scaling in reverse so that the matrix of artificial characteristics $Z$ with a condition number $c$ is given by $Z = (UTV^{\mathsf{T}})D$.

By the above procedure, the five matrices of artificial characteristics are generated. Then, adding their cross-products and a constant term to them, the five different matrices of artificial regressors are constructed. The fitted values of log(RENT) are regressed on the generated regressors by CLS in order to set up the five different vectors of parameters. The error terms are independently and identically distributed as a normal distribution with zero mean. The variance of the error terms is set so that the average R-square for OLS regressions is equal to a controlled level for each of 75 cases. The three levels of R-square to be examined are 0.5, 0.7, and 0.9, which covers the typical range for hedonic

[7] The procedure conducted here to generate not all regressors but characteristics (a part of regressors) is very similar to what Gilley and Pace (1995) used to generate all regressors.

studies. The number of Monte Carlo replications is 1000.

For all the CLS estimations to construct parameters above and to estimate parameters in Monte Carlo replications, the same monotonicity restrictions are used. The restrictions are constructed so that the domain of the HPF is given by a hyper rectangular whose extreme points have the maximum or the minimum of the artificial characteristics as their components. This construction of the monotonicity restrictions has advantages that the computational burden will be reduced because the number of constraints is small ($2^{\ell} = 64$), and that we can concentrate on examining the influences of two factors to be controlled (the degrees of multicollinearity and R-square) by keeping other conditions fixed.

*Results of Simulations*

To summarize the estimation performances of the OLS and the CLS, the average relative RMSEs about the imputed prices of the six characteristics are reported below. The imputed prices are evaluated at all the artificial data points and then averaged.

Table 6 presents the average relative RMSEs about the imputed prices of the six characteristics for each degree of multicollinearity, where the averages are taken across the 15 cases, all the combinations of the three levels of R-square and the five vectors of the true parameters. As the table shows, all the average relative RMSEs (OLS/CLS) are larger than unity, that is, the CLSEs of imputed prices have smaller RMSEs than the OLSEs. The average relative RMSEs are larger as the condition number becomes larger, except for a reverse relation for $\pi_{\text{DIST}}$ between the condition numbers 10 and 20.

Table 7 presents the average relative RMSEs about the imputed prices of the six characteristics for each level of R-square, where the averages are taken across the 25 cases, all the combinations of the five degrees of R-square and the five vectors of the true parameters. As the table shows, the CLSEs of imputed prices have smaller RMSEs than the OLSEs. The average relative RMSEs are smaller as the R-square becomes larger.

As a whole, improvements by the CLS about RMSEs are considerable. In addition, the influence of the condition number and the R-square appears as expected: there is a large margin of improvements by the CLS about RMSEs if the probability that the OLSE violates monotonicity restrictions is a high level.

## V. CONCLUDING REMARKS

This paper has studied the performance of the CLS for estimating the HPF. The constraints are inequalities which represent monotonicity restrictions on the HPF. Throughout our empirical application and Monte Carlo experiments, the CLSE outperforms the OLSE in the following sense: (i) Estimated variances of the CLSE are smaller than those of the OLSE; (ii) out-of-sample prediction errors by the CLSE are much smaller than those by the OLSE when a sample used for estimation is a half or smaller sample from the original data set, while out-of-sample prediction errors by the CLSE are only a little larger than those by the OLSE when a sample used for estimation is larger than a half sample from the original data set; and (iii) the CLSE has a much smaller mean square error than the OLSE.

Not only the CLS provides more accurate estimation results than the OLS but also the CLSEs of imputed prices of utility-bearing characteristics are guaranteed to be nonnegative by definition. Even if its computational burden is taken into consideration, therefore, the CLS should be a useful estimation method for hedonic price models.

## Acknowledgement

* I am grateful to my supervisor, Professor Makoto Ohta, University of Tsukuba, for his valuable comments and encouragement. Needless to say, I am solely responsible for any possibly remaining errors or shortcomings.

## References

Arguea, N. M., and Hsiao, C. (1993) Econometric Issues of Estimating Hedonic Price Functions, *Journal of Econometrics*, **56**, 243–267.

Atkinson, S. E., and Crocker, T. (1987) A Bayesian Approach to Assessing the Robustness of Hedonic Property Value Studies, *Journal of Applied Econometrics*, **2**, 27–45.

Box, G. E. P., and Cox, D. R. (1964) An Analysis of Transformations, *Journal of Royal Statistical Society, Series B*, **26**, 211–243.

Davidson, R., and MacKinnon, J. G. (1993) *Estimation and Inference in Econometrics.* Oxford University Press, New York.

Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife, *Annals of Statistics*, **7**, 1–26.

Efron, B., and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Gilley, O. W., and Pace, R. K. (1995) Improving Hedonic Estimation with an Inequality Restricted Estimator, *Review of Economics and Statistics*, **77**, 609–621.

Härdle, W., and Marron, J. S. (1991) Bootstrap Simultaneous Error Bars for Nonparametric Regression, *Annals of Statistics*, **13**, 1465–1481.

Judge, G. G., and Takayama, T. (1966) Inequality Restrictions in Regression Analysis, *Journal of the American Statistical Association*, **61**, 166–181.

Judge, G. G., and Yancey, T. A. (1986) *Improved Methods of Inference in Econometrics.* Elsevier Science Publishers, New York.

Kanemoto, Y., and Nakamura, R. (1986) A New Approach to the Estimation of Structural Equations in Hedonic Models, *Journal of Urban Economics*, **19**, 218–233.

Mammen, E. (1993) Bootstrap and Wild Bootstrap for High Dimensional Linear Models, *Annals of Statistics*, **21**, 255–285.

Nagai, K., Kondo, Y., and Ohta, M. (1997) An Hedonic Analysis of the Rental Office Market in the Tokyo CBD: 1985–1994 Fiscal Years, forthcoming in *Japanese Economic Review.*

Ohta, M. (1975) Production Technologies of the U.S. Boiler and Turbogenerator Industries and Hedonic Price Indexes for Their Products: A Cost-Function Approach, *Journal of Political Economy*, **83**, 1–26.

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.

Shao, J., and Tu, D. (1995) *The Jackknife and Bootstrap.* Springer-Verlag New York, Inc., New York.

Varian, H. R. (1974) A Bayesian Approach to Real Estate Assesment, in S. E. Fienberg, and A. Zellner (Eds) *Studies in Bayesian Econometrics and Statistics*, North-Holland, New York.

Table 1: Data Summary

| Variable | | Mean | Std.Dev. | Minimum | Maximum |
|---|---|---|---|---|---|
| RENT | (yen/m$^2$) | 7334.908 | 3251.940 | 1815.808 | 22330.822 |
| DIST | (10 m) | 21.583 | 23.844 | 0.000 | 168.000 |
| ACCT | (minutes) | 7.956 | 6.146 | 0.000 | 26.000 |
| ACCS | (minutes) | 15.085 | 7.185 | 0.000 | 43.000 |
| AGE | (year) | 16.602 | 10.132 | 1.500 | 62.500 |
| DENS | (1000 men/km$^2$) | 50.551 | 22.858 | 6.433 | 73.590 |
| FLOSP | (1000 m$^2$) | 24.575 | 37.830 | 0.308 | 190.595 |
| VOLM | (abs. num.) | 7.351 | 1.648 | 2.000 | 10.000 |
| FONE | (dummy) | 0.069 | 0.254 | 0.000 | 1.000 |
| MAJOR | (dummy) | 0.338 | 0.474 | 0.000 | 1.000 |
| D86 | (dummy) | 0.482 | 0.500 | 0.000 | 1.000 |

*Note:* Mean of the dummy (FONE, MAJOR, or D86) is equal to the ratio of the number of cases where the corresponding dummy takes the value unity to the total number of observations, 550.

*Data Sources:* See Section III.

Table 2: Results of Model Selection: RESET and $J$-test

| Model | A.$R^2$ | RESET2 | RESET3 | $J$-test |
|---|---|---|---|---|
| Linear in Level | .574 | 1.509 [.132] | 2.731 [.066] | 9.130 [.000] |
| Linear in Log | .626 | 3.739 [.000] | 7.098 [.001] | 1.155 [.249] |
| Quadratic in Level | .697 | −1.807 [.071] | 2.435 [.089] | 5.323 [.000] |
| Quadratic in Log | .708 | 1.318 [.188] | 2.734 [.066] | 2.511 [.012] |

*Note:* The column labeled "A.$R^2$" is adjusted R-squared. The columns labeled "RESET2" and "RESET3" are the test statistics for the RESETs, respectively, with including the squared fitted value and with including the squared and cubed fitted values. The column labeled "$J$-test" is the test statistic for the $J$-test. The figures in square brackets are P-values. All the calculations are done with TSP version 4.4.

Table 3: Estimates of Imputed Prices of Continuous Characteristics

| | OLS | | | CLS | | |
|---|---|---|---|---|---|---|
| | Estimate | Std. Err. | $t$-value | Estimate | Std. Err. | $t$-value |
| DIST | 2.235 | 8.327 | 0.268 | 22.855 | 5.765 | 3.964 |
| ACCT | −85.366 | 43.555 | −1.960 | 36.439 | 24.075 | 1.514 |
| ACCS | −27.129 | 24.765 | −1.095 | 3.792 | 12.100 | 0.313 |
| AGE | 39.003 | 13.570 | 2.874 | 40.317 | 9.581 | 4.208 |
| DENS | 10.021 | 9.824 | 1.020 | 14.569 | 5.814 | 2.506 |
| FLOSP | 49.163 | 5.442 | 9.033 | 40.976 | 4.784 | 8.565 |
| VOLM | 1137.600 | 120.081 | 9.474 | 680.112 | 72.279 | 9.410 |
| R-square | | | 0.7279 | | | 0.6768 |

*Note:* The estimates of imputed prices are obtained by evaluating the estimated functions at the sample mean. Disutility-bearing characteristics, DIST, ACCT, ACCS, and AGE, are negated their signs before estimations, and so all the imputed prices are to be nonnegative.

Table 4: Local Monotonicity of HPF: The Percentage of Observation Points

| Char. | DIST | ACCT | ACCS | AGE | DENS | FLOSP | VOLM | All |
|---|---|---|---|---|---|---|---|---|
| Ratio | 46.91 | 48.73 | 47.27 | 97.45 | 43.64 | 99.09 | 95.45 | 3.82 |

*Note:* Each of the reported figure is the percentage of the observation points where the HPF is locally monotonic within all the observation points. The column labeled with "All" is of monotonicity in all the characteristics, and the other columns are of partial monotonicity in each of characteristics.

Table 5: Relative Performance for Out-of-Sample Prediction Errors

| In-Sample Ratio | 20.0% | 35.0% | 50.0% | 65.0% | 80.0% | Average |
|---|---|---|---|---|---|---|
| RMSE(OLS/CLS) | 1.4687 | 1.1256 | 1.0307 | 0.9919 | 0.9759 | 1.1186 |

*Note:* "In-Sample Ratio" is the number of sample observations divided by the total number of observations, 550, in percentage. "RMSE(OLS/CLS)" is the ratio of RMSE by OLS to that by CLS. The column labeled with "Average" is the simple arithmetic average of the five relative RMSE.

Table 6: The Relative RMSEs for Each Condition Number

| Cond.Num. | 10 | 20 | 30 | 50 | 70 |
|---|---|---|---|---|---|
| DIST | 2.0664 | 2.0532 | 2.0729 | 2.1144 | 2.1879 |
| ACCT | 2.7815 | 2.8309 | 2.9214 | 3.1684 | 3.4030 |
| AGE | 1.5057 | 1.5445 | 1.6993 | 1.8819 | 1.9803 |
| DENS | 1.8919 | 2.6212 | 2.9919 | 3.6597 | 4.1856 |
| FLOSP | 1.3762 | 1.5114 | 1.5900 | 1.6611 | 1.7438 |
| VOLM | 1.3358 | 1.6754 | 1.9859 | 2.3964 | 2.8057 |

*Note:* Reported figures are the average relative RMSEs (OLS/CLS) about the imputed prices of characteristics. The label "Cond.Num." means the condition numbers.

Table 7: The Relative RMSEs for Each R-square

| R-square | 0.5 | 0.7 | 0.9 |
|---|---|---|---|
| DIST | 2.3033 | 1.9304 | 1.5497 |
| ACCT | 3.3096 | 2.8143 | 2.2064 |
| AGE | 1.6795 | 1.4770 | 1.2949 |
| DENS | 3.1255 | 2.6144 | 1.9392 |
| FLOSP | 1.6185 | 1.4010 | 1.1619 |
| VOLM | 2.3170 | 1.9519 | 1.4770 |

*Note:* Reported figures are the average relative RMSEs (OLS/CLS) about the imputed prices of characteristics.

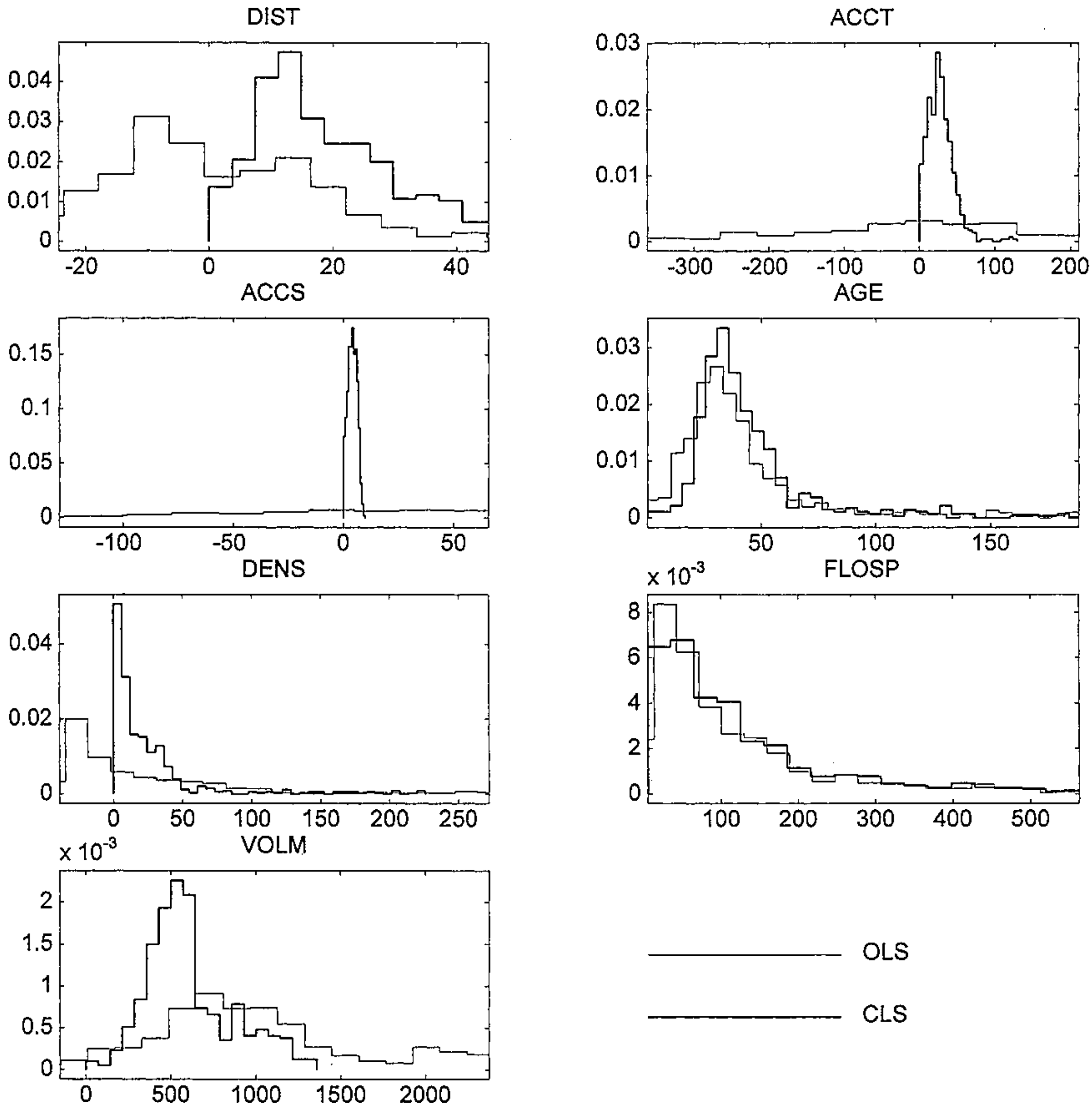


Figure 1: Distributions of Imputed Prices at All the Observations
The probability densities of the imputed prices of characteristics are estimated by the density histogram. Thin lines are of the OLSEs of the imputed prices, and thick lines are of the CLSEs of them.