Bayesian Estimation of Demand Functions
under Block-Rate Pricing

Koji Miyawaki
National Institute for Environmental Studies

Yasuhiro Omori
University of Tokyo

Akira Hibiki
National Institute for Environmental Studies

January 2010

Bayesian Estimation of Demand Functions

under Block-Rate Pricing

Koji Miyawaki[a]   Yasuhiro Omori[b] *   Akira Hibiki[c]

January 29, 2010

[a] National Institute for Environmental Studies, Ibaraki 305-8506, Japan (miyawaki.koji@nies.go.jp).

[b] Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan (omori@e.u-tokyo.ac.jp).

[c] National Institute for Environmental Studies, Ibaraki 305-8506, Japan (hibiki@nies.go.jp).

Abstract

This article proposes a Bayesian estimation of demand functions under block-rate pricing by focusing on increasing block-rate pricing. This is the first study that explicitly considers the separability condition which has been ignored in previous literature. Under this pricing structure, the price changes when consumption exceeds a certain threshold and the consumer faces a utility maximization problem subject to a piecewise-linear budget constraint. Solving this maximization problem leads to a statistical model in which model parameters are strongly restricted by the separability condition. In this article, by taking a hierarchical Bayesian approach, we implement a Markov chain Monte Carlo simulation to properly estimate the demand function. We find, however, that the convergence of the distribution of simulated samples to the posterior distribution is slow, requiring an

*Corresponding author: Tel:+81-3-5841-5516, E-mail:omori@e.u-tokyo.ac.jp

additional scale transformation step for parameters to the Gibbs sampler. These proposed methods are then applied to estimate the Japanese residential water demand function.

Key words: Discrete/continuous choice approach, Markov chain Monte Carlo method, Piecewise-linear budget constraint, Residential water demand, Separability condition.

JEL classification: C11, C24, Q25.

1   Introduction

Block-rate pricing is a nonlinear pricing system often applied to public utilities, such as water. In contrast to other goods and services offered at a single price, consumers under block-rate pricing face several prices corresponding to the level of consumption. Income tax also has this pricing structure because the marginal tax rate changes according to total income. General microeconomic theory suggests that the most efficient allocation is achieved by setting a good's unit price equal to its production cost per unit, which is called marginal cost pricing. However, in practice, several market issues exist that make marginal cost pricing inapplicable. In such cases, block-rate pricing is often employed by regulators.

To derive the demand function under block-rate pricing, we adopt a discrete/continuous choice approach, which Burtless and Hausman (1978) first used to analyze taxation's effect on the labor supply (see also Hanemann (1984); Hausman (1985); Moffitt (1986)). Model specifications of this kind are commonly used to evaluate tax policy or examine consumer behavior under block-rate pricing structures, such as labor supply (Burtless and Hausman, 1978), food stamp expenditure (Moffitt, 1989), car ownership and use (de Jong, 1990), electricity demand (Herriges and King, 1994; Reiss and White, 2005), and water demand (Hewitt and Hanemann, 1995; Olmstead, Hanemann, and Stavins, 2007).

While this approach is based on the consumer's maximization problem, a corresponding statistical model strongly restricts the model parameters. Furthermore, as Moffitt (1986)

pointed out, there is not only a computational burden but also the non-differentiability of the likelihood function. Thus, previous studies have exclusively estimated the demand function under two-block-rate pricing. The only exception seems to be Olmstead et al. (2007), who considered the multiple-block-rate pricing, where the number of blocks varies from two to four, using the maximum likelihood method. This method, however, ignored the so-called separability condition, which becomes important as the number of blocks increases. In Japan, consumers usually face more than two blocks (specifically, between five and eleven for water and between three and four for electricity), which requires the consideration of multiple-block-rate pricing.

Therefore, this article takes a hierarchical Bayesian approach and implements a Markov chain Monte Carlo (MCMC) simulation to properly estimate the demand function; see Chib (2001) and Chib and Greenberg (1996) for the MCMC methodology and its use in econometrics, respectively. The present paper allows two practical attributes that previous studies excluded. First, we allow the number of blocks to be greater than two. Then, we use the discrete/continuous approach to derive the demand function as a multinomial generalization of the Type V Tobit model; see Chapter 10 of Amemiya (1985) for Tobit classifications and Chib (1992) for a pioneering study of the Bayesian approach in Tobit modeling. Second, we explicitly consider the separability condition, which guarantees that consumer preferences are divided into disjoint blocks. This condition has been ignored in previous studies yet plays a critical role, especially in multiple-block-rate pricing.

We find, however, that the distribution of samples obtained from the Gibbs sampler converges very slowly to the posterior distribution. To improve sampling efficiency, we introduce an additional scale transformation step for parameters to the Gibbs sampler based on the generalized Gibbs step (GGS) developed by Liu and Sabatti (2000).

The rest of this article is organized as follows. In Section 2, we describe the discrete/continuous choice approach and the demand function under block-rate pricing. Section 3 explains the sta-

tistical model, derives its likelihood function and joint posterior density function, and accounts for the separability condition. With this posterior density function, this section presents the MCMC algorithm and the corresponding generalized Gibbs step. Section 4 carries out a simulation study and reveals several properties of our algorithms. After data description, Section 5 applies our proposed method to estimate the residential water demand function in Japan using microdata, and the price and income elasticities are also estimated to investigate demand sensitivity in detail. Section 6 provides the conclusion for this article.

2 Demand Function

First, we explain the model settings, following the discussion of Moffitt (1986). There are two types of goods; namely, a good under block-rate pricing and all other goods. Suppose that a consumer's demand for a good, Y, is subject to K-block-rate pricing, and that its demand is strictly positive, $Y > 0$. Let Y_a and I be the expenditure for non-Y goods and total income, respectively. The price system of Y is as follows. There are K prices, P_k ($k = 1, \ldots, K$), related to K blocks. These prices are fixed and considered as given constants throughout this article. In practice, price often changes monotonically with $P_k < P_{k+1}$ or $P_k > P_{k+1}$ for $k = 1, \ldots, K-1$. This article focuses on the price system in which price increases monotonically. That is, $P_k < P_{k+1}$ ($k = 1, \ldots, K-1$), which is known as increasing block-rate pricing. Japanese residential water demand data in Section 5 offer an example of increasing block-rate pricing. Let $\bar{Y}_k$ denote the upper limit, or threshold, of the k-th block ($k = 0, \ldots, K$), where $\bar{Y}_0 \equiv 0$ and $\bar{Y}_K \equiv \infty$ for convenience. In addition to marginal prices and thresholds, there is a fixed cost, FC, that is independent of consumption, Y. At a practical level, this fixed cost represents a minimum access charge. The threshold values, $\bar{Y}_k$, and a fixed cost, FC, are given fixed constants.

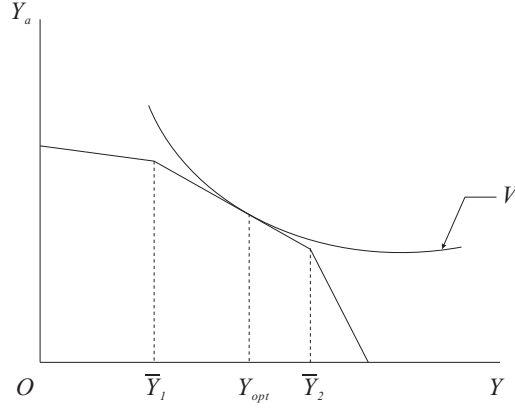Let $U(Y, Y_a)$ be the well-defined utility function. Then, the consumer's utility maximiza-

Figure 1: Utility maximization problem: three-block case.

tion problem is given by:

$$V = \max_{Y,Y_a} U(Y,Y_a) \quad \text{subject to } c(Y) + Y_a \leq I, \tag{1}$$

where $c(Y) = FC + P_k(Y - \bar{Y}_{k-1}) + \sum_{j=1}^{k-1} P_j(\bar{Y}_j - \bar{Y}_{j-1})$ if $\bar{Y}_{k-1} \leq Y < \bar{Y}_k$ for $k = 1,\ldots,K$. Figure 1 illustrates a budget constraint and indifference curve under three-block increasing block-rate pricing, where the second block is optimal with its optimal demand, Y_{opt}, and level of indirect utility, V. The budget constraint of this form is a piecewise-linear budget constraint because it becomes linear given the choice of a block.

Finally, the demand function is derived. Before its derivation, we must define K conditional utility maximization problems. For $k = 1,\ldots,K$, the k-th conditional problem is given by:

$$\max_{Y,Y_a} U(Y,Y_a) \quad \text{subject to } P_k Y + Y_a \leq Q_k, \quad \text{where } Q_k = I - FC - \sum_{j=1}^{k-1}(P_j - P_{j+1})\bar{Y}_j, \tag{2}$$

and Q_k is an augmented income (also referred to as virtual income). Under Problem (2), the consumer can maximize utility as if facing a single price, P_k, and virtual income, Q_k. Let

5

conditional demand Y_k be the solution to this conditional utility maximization problem, which results in the demand function under increasing block-rate pricing:

$$Y = \begin{cases} Y_k, & \text{if } \bar{Y}_{k-1} < Y_k < \bar{Y}_k \text{ and } k = 1,\ldots,K, \\ \bar{Y}_k, & \text{if } Y_{k+1} \le \bar{Y}_k \le Y_k \text{ and } k = 1,\ldots,K-1. \end{cases} \tag{3}$$

In studying the demand function under block-rate pricing, there are several functional forms, such as linear, quadratic, and log-linear functions, for the conditional demand, Y_k, in Equation (3); the log-linear conditional demand model is one of the most popular models used in previous studies (Hewitt and Hanemann, 1995). Thus, this article focuses on the log-linear model for conditional demand, but our proposed estimation method would apply to other models in a similar manner. The log-linear model is given by $\ln Y_k = \beta_1 \ln P_k + \beta_2 \ln Q_k$, where the parameters β_1 and β_2 represent the price and income elasticities conditional on block choice, respectively. For simplicity, let y, y_k, $\bar{y}_k$, p_k, and q_k denote the logarithm of demand (Y), k-th conditional demand (Y_k), k-th threshold ($\bar{Y}_k$), k-th marginal price (P_k), and k-th virtual income (Q_k), respectively. Then, Equation (3) under the log-linear model can be written as follows:

$$y = \begin{cases} y_k, & \text{if } \bar{y}_{k-1} < y_k < \bar{y}_k \text{ and } k = 1,\ldots,K, \\ \bar{y}_k, & \text{if } y_{k+1} \le \bar{y}_k \le y_k \text{ and } k = 1,\ldots,K-1, \end{cases} \tag{4}$$

$$y_k = \beta_1 p_k + \beta_2 q_k \equiv \boldsymbol{x}_k' \boldsymbol{\beta}, \tag{5}$$

where $\boldsymbol{x}_k = (p_k, q_k)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, $\bar{y}_0 \equiv -\infty$, and $\bar{y}_K \equiv \infty$.

6

3 Bayesian Analysis of Demand Functions under Block-Rate Pricing

3.1 Statistical Model

From this section, we append the subscript i to the i-th consumer's variables ($i = 1, \ldots, n$) and the superscript $*$ to latent variables. For example, $y_i, \bar{y}_{ik}, p_{ik}, q_{ik}, K_i$ are observed variables, whereas w_i^*, s_i^* are unobserved that will be explained in the following paragraph. We note that y_{ik}, the k-th log conditional demand, is unobserved, as it has no superscript $*$ to avoid tedious notation. Because we construct a statistical model that assumes different block-rate pricing for different observation, variables for block-rate pricing are also marked by the subscript i.

Our statistical model, which is a multinomial extension of Moffitt (1986), is described as follows; see also Hewitt and Hanemann (1995). First, we introduce two unobserved random variables into the demand function of the i-th consumer, namely, heterogeneity, w_i^*, and state variable, s_i^*. Heterogeneity is a stochastic term that models consumer characteristics. The w_i^* is assumed to follow the linear model:

$$w_i^* = \mathbf{z}_i' \boldsymbol{\delta} + v_i, \quad v_i \sim \text{i.i.d. } N(0, \sigma_v^2), \tag{6}$$

where $\mathbf{z}_i$ and $\boldsymbol{\delta}$ are $d \times 1$ vectors of explanatory variables for heterogeneity and corresponding parameters, respectively, and v_i is an independently and identically distributed disturbance term with a normal distribution of mean 0 and variance σ_v^2. We assume that heterogeneity for the i-th observation, w_i^*, is additive to log conditional demand, y_{ik}. Then, the log conditional

demand with heterogeneity, y_i^*, for the i-th consumer is given by:

$$y_i^* = \begin{cases} y_{ik} + w_i^*, & \text{if } \bar{y}_{i,k-1} < y_{ik} + w_i^* < \bar{y}_{ik} \text{ and } k = 1,\ldots,K_i, \\ \bar{y}_{ik}, & \text{if } y_{i,k+1} + w_i^* \le \bar{y}_{ik} \le y_{ik} + w_i^* \text{ and } k = 1,\ldots,K_i-1, \end{cases} \tag{7}$$

where $y_{ik} = \boldsymbol{x}_{ik}'\boldsymbol{\beta}$ and $\boldsymbol{x}_{ik} = (p_{ik},q_{ik})'$.

Another latent variable is the state variable, s_i^*. There are $2K_i - 1$ potential outcomes in the demand function Equation (7), including K_i conditional demands with heterogeneity $(y_{ik} + w_i^*)$ and $K_i - 1$ threshold demands $(\bar{y}_{ik})$. The state variable, s_i^*, is an unobserved discrete random variable taking values from 1 to $2K_i - 1$ and indicates which outcome the i-th observation selects; if s_i^* is odd, observation i chooses conditional demand with heterogeneity, and if s_i^* is even, it selects the kink point as its demand. More precisely:

$$s_i^* = \begin{cases} 2k - 1, & \text{if } y_i^* = y_{ik} + w_i^* \text{ and } k = 1,\ldots,K_i, \\ 2k, & \text{if } y_i^* = \bar{y}_{ik} \text{ and } k = 1,\ldots,K_i-1. \end{cases} \tag{8}$$

It is straightforward from Equation (7) that the condition regarding whether y_i^* equals $y_{ik} + w_i^*$ or $\bar{y}_{ik}$ is equivalent to the interval condition for heterogeneity.

$$\bar{y}_{i,k-1} < y_{ik} + w_i^* < \bar{y}_{ik} \iff w_i^* \in R_{i,2k-1} = (\bar{y}_{i,k-1} - y_{ik}, \bar{y}_{ik} - y_{ik}) \iff s_i^* = 2k - 1, \tag{9}$$

$$y_{i,k+1} + w_i^* \le \bar{y}_{ik} \le y_{ik} + w_i^* \iff w_i^* \in R_{i,2k} = (\bar{y}_{ik} - y_{ik}, \bar{y}_{ik} - y_{i,k+1}) \iff s_i^* = 2k. \tag{10}$$

Furthermore, we assume that conditional demand y_i for the i-th consumer is observed with a disturbance:

$$y_i = y_i^* + u_i, \quad u_i \sim \text{i.i.d. } N(0,\sigma_u^2), \quad i = 1,\ldots,n, \tag{11}$$

where u_i may imply optimization, specification, and measurement error (Hausman, 1985).

Finally, the statistical model for the demand function under increasing block-rate pricing is given by following equations:

$$y_{ik} = x'_{ik}\boldsymbol{\beta}, \quad x_{ik} = (p_{ik}, q_{ik})', \quad k = 1, \ldots, K_i, \tag{12}$$

$$w_i^* = z_i'\boldsymbol{\delta} + v_i, \quad v_i \sim \text{i.i.d. } N(0, \sigma_v^2), \tag{13}$$

$$s_i^* = \begin{cases} 2k-1, & \text{if } w_i^* \in R_{i,2k-1} \text{ and } k = 1, \ldots, K_i, \\ 2k, & \text{if } w_i^* \in R_{i,2k} \quad \text{and } k = 1, \ldots, K_i - 1, \end{cases} \tag{14}$$

$$y_i^* = \begin{cases} y_{ik} + w_i^*, & \text{if } s_i^* = 2k-1 \text{ and } k = 1, \ldots, K_i, \\ \bar{y}_{ik}, & \text{if } s_i^* = 2k \text{ and } k = 1, \ldots, K_i - 1, \end{cases} \tag{15}$$

$$y_i = y_i^* + u_i, \quad u_i \sim \text{i.i.d. } N(0, \sigma_u^2). \tag{16}$$

Error terms, including measurement error, u_i, and the error for heterogeneity, v_i, are assumed to be mutually independent conditional on the block choice s_i^* because they represent different sources of error. This model is a multinomial extension of the Type V Tobit model; see Section 10.10 in Amemiya (1985) for more on the Type V Tobit model.

Remark 1. Suppose $K_i = 2$. Then, Equations (14) and (15) reduce to:

$$s_i^* = \begin{cases} 1, & \text{if } w_i^* \in R_{i1} = (-\infty, \bar{y}_{i1} - y_{i1}), \\ 2, & \text{if } w_i^* \in R_{i2} = (\bar{y}_{i1} - y_{i1}, \bar{y}_{i1} - y_{i2}), \\ 3, & \text{if } w_i^* \in R_{i3} = (\bar{y}_{i1} - y_{i2}, \infty), \end{cases} \quad y_i^* = \begin{cases} y_{i1} + w_i^*, & \text{if } s_i^* = 1, \\ \bar{y}_{i1}, & \text{if } s_i^* = 2, \\ y_{i2} + w_i^*, & \text{if } s_i^* = 3. \end{cases} \tag{17}$$

Remark 2. There may be consumers whose first block is a zero marginal price; that is, $P_{i1} = 0$. They are assumed to consume more than or equal to the first threshold, $\bar{y}_{i1}$, as suggested by economic theory, which implies that $s_i^* = 2, \ldots, K_i$ and $R_{i2} = (-\infty, \bar{y}_{i1} - y_{i2})$.

3.2 Likelihood Function

The augmented likelihood function for observation i is derived by multiplying two probability density functions. First, we derive the joint probability density function of unobserved variables, s_i^* and w_i^*. These variables are modeled using Equations (13) and (14). Thus:

$$f\left(s_i^*, w_i^* \mid \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_v^2\right) = f\left(w_i^* \mid \boldsymbol{\delta}, \sigma_v^2\right) f\left(s_i^* \mid w_i^*, \boldsymbol{\beta}\right)$$

$$\propto \sigma_v^{-1} \exp\left[-\frac{1}{2\sigma_v^2}\left(w_i^* - \boldsymbol{z}_i'\boldsymbol{\delta}\right)^2\right] I\left(w_i^* \in R_{is_i^*}\right) \prod_{k=1}^{K_i-1} I\left(\boldsymbol{x}_{i,k+1}'\boldsymbol{\beta} \leq \boldsymbol{x}_{ik}'\boldsymbol{\beta}\right), \quad (18)$$

where $I(A)$ is an indicator function taking a value of 1 if A is true and 0 otherwise. The last truncation term, $\prod_{k=1}^{K_i-1} I(\boldsymbol{x}_{i,k+1}'\boldsymbol{\beta} \leq \boldsymbol{x}_{ik}'\boldsymbol{\beta})$, is the separability condition that is first explicitly considered in this article. The role of the separability condition is explained in the next subsection.

After the unobserved variables are determined by Equation (18), the conditional probability density function of y_i is derived using Equations (15) and (16) and given by:

$$f\left(y_i \mid s_i^*, w_i^*, \boldsymbol{\beta}, \sigma_u^2\right) \propto \begin{cases} \sigma_u^{-1} \exp\left[-\frac{1}{2\sigma_u^2}\left(y_i - \boldsymbol{x}_{ik}'\boldsymbol{\beta} - w_i^*\right)^2\right], & \text{if } s_i^* = 2k-1 \text{ and } k = 1,\ldots,K_i, \\ \sigma_u^{-1} \exp\left[-\frac{1}{2\sigma_u^2}(y_i - \bar{y}_{ik})^2\right], & \text{if } s_i^* = 2k \text{ and } k = 1,\ldots,K_i-1, \end{cases}$$

$$\tag{19}$$

$$= \sigma_u^{-1} \exp\left[-\frac{1}{2\sigma_u^2}\left(y_i - y_i^*\right)^2\right]. \tag{20}$$

Finally, multiplying these two probability density functions (18) and (20), we obtain the augmented likelihood function for observation i:

$$f\left(y_i, s_i^*, w_i^* \mid \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2\right) = f\left(y_i \mid s_i^*, w_i^*, \boldsymbol{\beta}, \sigma_u^2\right) f\left(s_i^*, w_i^* \mid \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_v^2\right)$$

$$\propto \sigma_u^{-1} \sigma_v^{-1} \exp\left[-\frac{1}{2}\left\{\sigma_u^{-2}\left(y_i - y_i^*\right)^2 + \sigma_v^{-2}\left(w_i^* - \boldsymbol{z}_i'\boldsymbol{\delta}\right)^2\right\}\right]$$

$$\times I\left(w_i^* \in R_{is_i^*}\right) \prod_{k=1}^{K_i-1} I\left(\boldsymbol{x}_{i,k+1}'\boldsymbol{\beta} \leq \boldsymbol{x}_{ik}'\boldsymbol{\beta}\right).$$

$$(21)$$

3.3 The Separability Condition

We briefly describe the role of the separability condition in our model. The separability condition is a condition that creates disjointed heterogeneity intervals. It guarantees that the upper limit is greater than or equal to the lower limit for intervals in Equation (10). Under the multiple-block-rate pricing, this condition is given by:

$$y_{i,k+1} \leq y_{ik}\left(\Longleftrightarrow \boldsymbol{x}_{i,k+1}'\boldsymbol{\beta} \leq \boldsymbol{x}_{ik}'\boldsymbol{\beta}\right) \quad \text{for } k = 1, \ldots, K_i - 1 \text{ and } i = 1, \ldots, n. \qquad (22)$$

Because $\boldsymbol{\beta}$ is two-dimensional in our case, the separability condition reduces to two conditions. Let $\bar{r} = \max_{i,k} -(p_{i,k+1} - p_{ik})/(q_{i,k+1} - q_{ik})$, and let $\underline{r} = \min_{i,k} -(p_{i,k+1} - p_{ik})/(q_{i,k+1} - q_{ik})$. Then, Equation (22) is equivalent to

$$\beta_2 \leq \bar{r}\beta_1 \text{ and } \beta_2 \leq \underline{r}\beta_1, \qquad (23)$$

which is the shadowed region found in Figure 2. When the conditional demand includes more than two variables, the separability condition is given by Equation (22). In this case, we must address as many inequality conditions as the number of observations and blocks to numerically maximize the likelihood function.

Let us illustrate the role of the separability condition. Consider the maximization of augmented likelihood under two-block increasing block-rate pricing, assuming that there is only one observation. Condition (22) then reduces to $y_{i2} \leq y_{i1}$ as the only condition. Without this condition, the upper and lower limits for kink point demand are allowed to be reversed, and
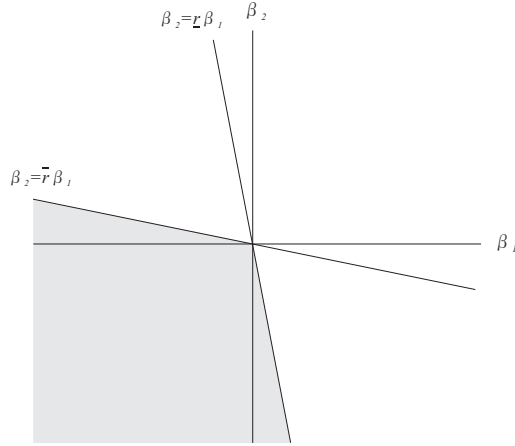
Figure 2: Region suggested by the separability condition.

hence, there could be a case in which $\bar{y}_1 - y_{i2} < w_i^* < \bar{y}_1 - y_{i1}$ (see R_{i2} of Equatioin (17)). Such a situation leads to ambiguity in the state variable if $s_i^* = 1$ or 3. Therefore, any estimation without the separability condition causes disagreement within the model.

3.4 Posterior Density Function and Gibbs Sampler

First, we assume proper prior distributions for the model parameters $(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2)$. For these parameters, we assume normal distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, conditional on σ_u^2 and σ_v^2, and inverse gamma distributions for σ_u^2 and σ_v^2.

$$\boldsymbol{\beta} \mid \sigma_u^2 \sim N_2\left(\boldsymbol{\mu}_{\boldsymbol{\beta},0}, \sigma_u^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta},0}\right), \boldsymbol{\delta} \mid \sigma_v^2 \sim N_d\left(\boldsymbol{\mu}_{\boldsymbol{\delta},0}, \sigma_v^2 \boldsymbol{\Sigma}_{\boldsymbol{\delta},0}\right), \sigma_u^2 \sim IG\left(\frac{n_{u,0}}{2}, \frac{S_{u,0}}{2}\right), \sigma_v^2 \sim IG\left(\frac{n_{v,0}}{2}, \frac{S_{v,0}}{2}\right),$$

(24)

where $\boldsymbol{\mu}_{\boldsymbol{\beta},0}$, is a 2×1 known vector; $\boldsymbol{\Sigma}_{\boldsymbol{\beta},0} = \text{diag}(\sigma_{\beta_1,0}^2, \sigma_{\beta_2,0}^2)$ is a 2×2 known diagonal matrix with diagonal elements $(\sigma_{\beta_1,0}^2, \sigma_{\beta_2,0}^2)$; $\boldsymbol{\mu}_{\boldsymbol{\delta},0}$ is a $d \times 1$ known vector, $\boldsymbol{\Sigma}_{\boldsymbol{\delta},0}$ is a known $d \times d$ covariance matrix; and $n_{u,0} > 0$, $S_{u,0} > 0$, $n_{v,0} > 0$, $S_{v,0} > 0$ are some known constants. In this article, the subscript on the normal distribution indicates its dimension.

12

The posterior density function for the statistical model in Equations (12) through (16) is obtained by multiplying the augmented likelihood function in Equation (21) over all observations with the prior density function, $\pi\left(\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2\right)$:

$$
\begin{aligned}
\pi\left(\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2,\boldsymbol{s}^*,\boldsymbol{w}^* \mid \boldsymbol{y}\right) &\propto \pi\left(\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2\right) \\
&\times \sigma_u^{-n}\sigma_v^{-n}\exp\left[-\frac{1}{2}\left\{\sigma_u^{-2}(\boldsymbol{y}-\boldsymbol{y}^*)'(\boldsymbol{y}-\boldsymbol{y}^*)+\sigma_v^{-2}(\boldsymbol{w}^*-\boldsymbol{Z}\boldsymbol{\delta})'(\boldsymbol{w}^*-\boldsymbol{Z}\boldsymbol{\delta})\right\}\right] \\
&\times \prod_{i=1}^{n}\left\{I(w_i^*\in R_{is_i^*})\prod_{k=1}^{K_i-1}I(y_{i,k+1}\leq y_{ik})\right\}, \quad (25)
\end{aligned}
$$

where $\boldsymbol{y}=(y_1,y_2,\ldots,y_n)'$, $\boldsymbol{y}^*=(y_1^*,y_2^*,\ldots,y_n^*)'$, $\boldsymbol{s}^*=(s_1^*,s_2^*,\ldots s_n^*)'$, $\boldsymbol{w}^*=(w_1^*,w_2^*,\ldots w_n^*)'$, and $\boldsymbol{Z}=(\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_n)'$.

Under log-linear conditional demand and above priors, the full conditional posterior distributions are all standard distributions, as provided in Appendix A.1. We implement a standard Gibbs sampler to draw samples from the posterior density function (25), which is summarized in the following seven steps:

Algorithm 1.1: MCMC algorithm for the model in Equations (12)-(16)

Step 1. Initialize $\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{s}^*,\boldsymbol{w}^*,\sigma_u^2$ and σ_v^2.

Step 2. Generate β_1 given $\beta_2,\boldsymbol{s}^*,\boldsymbol{w}^*,\sigma_u^2$.

Step 3. Generate β_2 given $\beta_1,\boldsymbol{s}^*,\boldsymbol{w}^*,\sigma_u^2$.

Step 4. Generate $\left(\sigma_v^2,\boldsymbol{\delta}\right)$ given $\boldsymbol{w}^*$.

Step 5. Generate $\left(s_i^*,w_i^*\right)$ given $\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2$ for $i=1,\ldots,n$.

Step 6. Generate σ_u^2 given $\boldsymbol{\beta},\boldsymbol{s}^*,\boldsymbol{w}^*$.

Step 7. Go to Step 2.

A blocking technique is used to sample (s_i^*,w_i^*) in order to isolate the relationship in which w_i^* determines s_i^*, while blocking in $(\sigma_v^2,\boldsymbol{\delta})$ is used to accelerate the convergence of MCMC draws.

3.5 Convergence Acceleration

As we shall see in Sections 4 and 5, the obtained samples of parameters are sometimes highly autocorrelated so that their convergence to the posterior distribution is slow. This subsection introduces a generalized Gibbs step proposed by Liu and Sabatti (2000) to improve sampling efficiency. While its implementation is simple, the GGS improves sampling efficiency to some extent in the estimation of the discrete/continuous choice model.

The main idea of the GGS is to add one more sampling step for a transformation group while keeping the transition kernel of MCMC invariant, so that we can obtain acceleration effects similar to those of re-parametrization or blocking; see Section 2 of Liu and Sabatti (2000) and Section 8.3 of Liu (2001) for a general definition of the GGS.

In our case, we apply the GGS to all parameters $\zeta = (\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{w}^*, \sigma_u, \sigma_v)$ so as to implement the one-step Metropolis-Hastings (MH) algorithm described below, and take a scale transformation group, that is, $\Gamma = \{g > 0 : g(\zeta) = g\zeta\}$. Then, the full conditional probability density function of $\tilde{g} \equiv g^{-1}$ is derived as Equation (A.37) in Appendix A.2. This full conditional probability density function is a nonstandard density function, so that the MH algorithm is adopted to draw a sample of $\tilde{g}$. Starting from the initial value $\tilde{g} = 1$, we draw a candidate $\tilde{g}'$, which follows the truncated normal distribution with mean $\mu_{\tilde{g}}$, variance $\sigma_{\tilde{g}}^2$, and truncation interval $R_{\tilde{g}}$:

$$TN_{R_{\tilde{g}}}\left(\mu_{\tilde{g}}, \sigma_{\tilde{g}}^2\right), \tag{26}$$

where $\mu_{\tilde{g}} = a_2/a_1$ and $\sigma_{\tilde{g}}^2 = a_1^{-1}$. See Equations (A.38), (A.39), and (A.41) in Appendix A.2 for definitions of a_1, a_2, and $R_{\tilde{g}}$. The candidate is accepted with probability $\alpha(\tilde{g}, \tilde{g}') = \min[1, (\tilde{g}'/\tilde{g})^{a_0-1}]$, where $a_0 = n + n_{u,0} + n_{v,0}$.

The MH algorithm typically must be repeated in order to obtain a sample from the conditional posterior density function of $\tilde{g}$. As we have proved in Appendix A.3, however, it is sufficient to draw a sample only once using the initial value $\tilde{g} = 1$. Therefore, the GGS is

14

implemented by replacing Step 7 of Algorithm 1.1 described in the previous subsection as follows:

Algorithm 1.2: Generalized Gibbs step for the model in Equations (12)-(16)

Step 7. Generate $\tilde{g}$ given $\boldsymbol{\beta}, \boldsymbol{s}^*, \boldsymbol{w}^*, \sigma_u, \sigma_v$.

(a) Generate $\tilde{g}' \sim TN_{R_{\tilde{g}}}(\mu_{\tilde{g}}, \sigma_{\tilde{g}}^2)$ and $u \sim U(0,1)$, where $U(0,1)$ denotes a uniform distribution on interval $(0,1)$.

(b) Accept a candidate $\tilde{g}'$ if $u \leq \alpha(1, \tilde{g}')$. If rejected, let $\tilde{g} = 1$.

Step 8. Transform parameters $(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{w}^*, \sigma_u, \sigma_v)$ by multiplying parameters by $g = \tilde{g}^{-1}$. The state variable $\boldsymbol{s}^*$ is also updated by this new $\boldsymbol{w}^*$.

Step 9. Go to Step 2.

4 Illustration Using Simulated Data

This section illustrates our Bayesian estimation of the statistical model based on Equations (12) to (16), using simulated data. We consider two-block increasing block-rate pricing with 1,000 observations. The marginal price for the first block is generated using $|N(1, 0.5^2)|$, which is the absolute value of a random number following a normal distribution with mean 1 and variance 0.5^2. The absolute value is taken to guarantee a positive marginal price. The second block's marginal price is similarly generated by adding $|N(1, 0.5^2)|$ to the first block's price. There is one threshold in this price system, which is set equal to 0.5. As for other variables, income is generated by $|N(2, 0.5^2)|$, and fixed cost is equal to 0. We consider only one explanatory variable other than the constant term for heterogeneity following $N(0,1)$. Thus, $\boldsymbol{\delta} = (\delta_0, \delta_1)'$.

The true parameter values are $(\beta_1, \beta_2, \delta_0, \delta_1, \sigma_u, \sigma_v) = (-0.6, 0.3, 0.1, 0.1, 0.3, 0.1)$. The regression parameter for price is set to be more elastic than for income based on evidence pro-

Table 1: Estimation summary with simulated data by GGS

Parameter	True	Mean	SD	95% interval		INEF GGS / GS *	CD **
β_1	−.6	−.62	.067	[−.76	−.49]	17.2 / 41.7	.116
β_2	.3	.35	.10	[.14	.54]	96.4 / 385.4	.108
δ_0 (constant)	.1	.090	.11	[−.11	.30]	134.9 / 498.8	.331
δ_1	.1	.084	.010	[.064	.10]	1.7 / 2.7	.205
σ_u (measurement error)	.3	.29	.019	[.24	.32]	35.5 / 69.2	.468
σ_v (heterogeneity)	.1	.12	.040	[.054	.20]	49.4 / 90.9	.413

* "INEF GGS / GS" denotes the estimated inefficiency factors using the Gibbs sampler with Generalized Gibbs step and the standard Gibbs sampler

** "CD" denotes the convergence diagnostics.

vided in previous studies on the water demand function; see Table 2 of Hewitt and Hanemann (1995). The prior distributions are:

$$\boldsymbol{\beta} \mid \sigma_u^2 \sim N_2\left(\mathbf{0}, 10^2\sigma_u^2 \boldsymbol{I}\right), \boldsymbol{\delta} \mid \sigma_v^2 \sim N_2\left(\mathbf{0}, 10^2\sigma_v^2 \boldsymbol{I}\right), \sigma_u^2 \sim IG\left(10^{-2}, 10^{-2}\right), \sigma_v^2 \sim IG\left(10^{-2}, 10^{-2}\right).$$
(27)

The mean and variance for the precision parameters σ_u^{-2} and σ_v^{-2} are 1 and 10^2, respectively. In hierarchical modeling, it is often pointed out that flat or improper prior distributions for variance parameters may lead to (almost) improper posterior distributions (see, for example, Section 5.3 of Gelman, Carlin, Stern, and Rubin (1995)), which makes Bayesian inference unreliable. Thus, we use relatively tight proper prior distributions for σ_u^2 and σ_v^2 to avoid (almost) improper posterior distributions.

We draw MCMC samples via the Gibbs sampler under Algorithm 1.1 and find its sample autocorrelations are very high. Thus, we apply the generalized Gibbs step under Algorithm 1.2, to accelerate convergence of the samples to their posterior distribution. After deleting 3×10^4 samples, we draw 10^5 samples by applying the acceleration step. The obtained samples are reduced to 10^4 by selecting every 10-th value. The GGS results are reported in Table 1.

Because the other results obtained by the Gibbs sampler are very similar to those obtained by the GGS, we omit them.

Table 1 reports true values, posterior means, posterior standard deviations, 95% credible intervals, estimated inefficiency factors, and the convergence diagnostics (i.e., the two-sided p-values of the test for convergence). The inefficiency factor is defined as $1 + 2\sum_{j=1}^{\infty}\rho(j)$, where $\rho(j)$ is the sample autocorrelation at lag j. It is estimated by using the spectral density; see Section 3.2 of Chib (2001) for details. It is interpreted as the ratio of the variance of the sample mean from the Markov chain to the variance of uncorrelated draws. If the inefficiency factor is close to one, the sampling method is almost as efficient as an independent draw. The greater the inefficiency factor becomes, the more samples we need to take to reach convergence. The convergence diagnostic is the test statistic with a null hypothesis of convergence, as proposed in Section 3.2 of Geweke (1992). We use the first 10% and the last 50% of samples to calculate this test statistic, as suggested by Geweke (1992).

In Table 1, we found smaller inefficiency factors under GGS than GS. We compare these two samplers in terms of their sample autocorrelation functions. Figure 3 shows the sample autocorrelation function of β_2 for GS and GGS results; autocorrelation decays more quickly in GGS than in GS. Thus, we conclude that the GGS is effective for β_2 in improving sample convergence.

There are two findings regarding this simulation. The first is the role of the kink point. The state change helps us to separately estimate measurement error σ_u and heterogeneity error σ_v. When the kink point is chosen, however, other information, such as regression coefficients $\boldsymbol{\beta}$ and coefficients for heterogeneity $\boldsymbol{\delta}$ is lost, which causes the sampling of our model to become inefficient.

Second, initial values are important for efficient sampling, which is also a difficult task in the maximum likelihood method; see Section 4 of Moffitt (1986). Because the full conditional distributions, especially for $\boldsymbol{\beta}$, are restricted to the tight parameter space, samples cannot
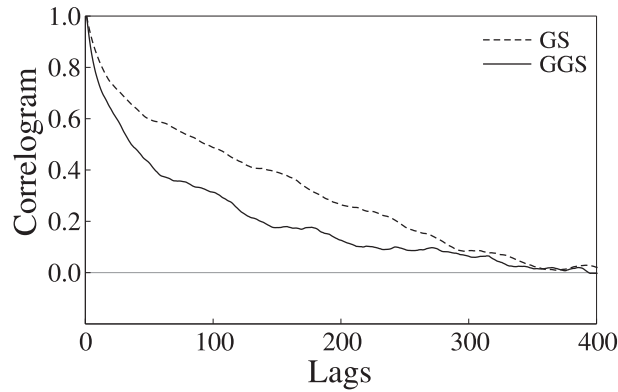
Figure 3: Sample autocorrelation functions for β_2.

move freely in their state space. Thus, when initial values are far from true values, it takes a considerable amount of time for the distribution of the MCMC samples to converge to the posterior distribution. Because we do not know much about the true parameter values in the empirical analysis, testing several initial values might be effective when sampling appears to be inefficient.

5 Estimation of the Japanese Residential Water Demand Function

5.1 Data Description

We use household-level dataset collected by INTAGE Inc. (www.intage.co.jp/english) through an online questionnaire. The population consists of households between the age of 20 and 79 years in the Tokyo and Chiba prefectures. In June 2006, there were over 8 million such households. Among these, INTAGE Inc. collected data from $47,239$ households, from which we have randomly chosen $1,678$ households. Overall, $1,250$ households out of $1,678$ participated in our research.

Table 2: Variables used in the water demand function

Variable	Coefficient	Description
year		June 2006
num. of obs.		365
price	β_1	water+sewer (log $\yen 10^3/m^3$)
virtual income	β_2	two-month income augmented by price (log $\yen 10^3$)
variables for w_i^*	δ_0	the constant
	δ_1	the number of members in a household (person)
	δ_2	the number of rooms in a house/apartment (room)
	δ_3	the total floor space of a house/apartment (50m^2)

Table 3: Summary statistics of variables used in the water demand function

Variable	Unit	Mean	SD	Min.	Max.
the amount of water	log m^3	3.53	.51	.098	4.87
the two-month income	$\yen 10^3$	1145.00	566.11	166.67	4666.70
the number of members in a household	person	3.05	1.22	1	7
the number of rooms in a house/apartment	room	4.29	1.10	1	8
the total floor space of a house/apartment	50m^2	1.66	.70	.24	4.60

All households face increasing block-rate pricing, where the number of blocks varies from five to eleven (see Figure 4(a)). The dependent variable is the amount of water calculated from each payment using corresponding price tables. This amount is the water amount that has been used in the last two months because water is charged every two month in Japan. The explanatory variables to be used for empirical analysis are summarized in Table 2, and their summary statistics are found in Table 3 and Figure 4. The number of observations was reduced to 365 because of the following reasons.

1. Missing data concerning household attributes.

2. Missing data concerning water usage.

3. Difficulty in identifying the corresponding rate table.

4. Consumption within the zero marginal price block was observed.

19

(a) Number of blocks (number).　(b) Price where the consumption is actually made ($¥10^3/m^3$).　(c) Fixed cost ($¥10^3$).
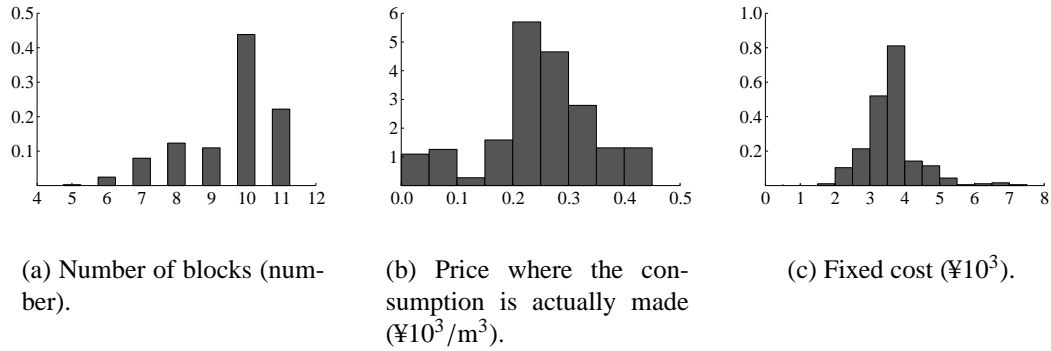
Figure 4: Histograms of the number of blocks, price, and fixed cost.

Observations linked to any of the above issues were omitted.

Regarding the income variable, asking households for their exact income level is a sensitive task. In our research, the household annual income is recorded in one of eight intervals in million yen, including 0-2, 2-4, 4-6, 6-8, 8-10, 10-12, 12-15, and over 15 million yen. Then, we use the median of the interval divided by 6 as a proxy for two-month income for most households. Households with annual income over 15 million yen were asked to approximate their annual income, and we divided it by 6 to estimate two-month income for those households.

5.2 Empirical Result

Initially, the following prior distributions are assumed for the parameters of the demand function:

$$\boldsymbol{\beta} \mid \sigma_u^2 \sim N_2\left(\mathbf{0}, 10\sigma_u^2\boldsymbol{I}\right), \boldsymbol{\delta} \mid \sigma_v^2 \sim N_4\left(\mathbf{0}, 10\sigma_v^2\boldsymbol{I}\right), \sigma_u^2 \sim IG\left(10^{-1}, 10^{-1}\right), \sigma_v^2 \sim IG\left(10^{-1}, 10^{-1}\right).$$

(28)

Because the Gibbs sampler (Algorithm 1.1) in Subsection 3.4 is very slow to converge

20

Table 4: Water demand function

Parameter	Mean	SD	95% interval		INEF GGS / GS*	CD**
β_1 (price)	-1.09	.22	$[-1.52$	$-$ .67]	242.55 / 370.44	.201
β_2 (income)	.067	.044	$[-$ .028	.14]	272.18 / 583.76	.025
δ_0 (constant)	.23	.51	$[-$ .89	1.12]	312.61 / 500.76	.055
δ_1 (num. of members)	.23	.039	[.16	.31]	54.90 / 63.16	.983
δ_2 (num. of rooms)	.14	.049	[.038	.23]	7.65 / 9.31	.198
δ_3 (floor space)	.041	.077	$[-$ .11	.20]	7.07 / 7.47	.987
σ_u (measurement error)	.42	.018	[.38	.45]	8.91 / 19.41	.531
σ_v (heterogeneity)	.20	.038	[.14	.28]	17.39 / 10.60	.021

* "INEF GGS / GS" denotes the estimated inefficiency factors using the Gibbs sampler with Generalized Gibbs step and the standard Gibbs sampler
** "CD" denotes the convergence diagnostics.

to the posterior distribution, we accelerate the convergence of the MCMC samples using the GGS described in Algorithm 1.2 of Subsection 3.5. The initial 16×10^5 samples are discarded, and the subsequent 4×10^6 samples are recorded. The recorded samples are reduced to 10^4 samples by picking up every 400-th value. These estimation results are shown in Table 4.

To first check the plausibility of our proposed model, we carried out the numerical posterior predictive checks (PPCs) based on these results; see Chapter 6 of Gelman, Carlin, Stern, and Rubin (2003). Seven test quantities (including the first and third quartile, mean, median, standard deviation, minimum, and maximum) were chosen to conduct PPCs, and the results are found in Figure 5. The density plots represent those of test quantities based on the replicated data from the predictive distribution, and the vertical lines denote the values of test quantities based on the observed data. We also calculated the posterior predictive p-values, which are shown in parentheses. All density plots and p-values, except those for the minimum, indicate that the discrete/continuous choice model would be plausible to represent our Japanese residential water demand data. The small p-value for the minimum indicates that we may need to improve our model for low levels of consumption; the five lowest levels of
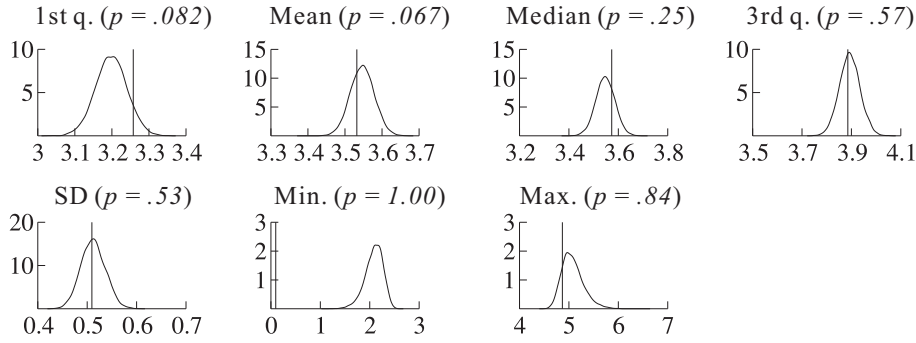
Figure 5: Posterior predictive checks.

consumption are 0.098, 1.79, 1.86, 1.92, and 1.95 $\log \mathrm{m}^3$. As stated in Remark 2 of Subsection 3.1, for simplicity, we excluded households that consume within the zero marginal price block so that all consumptions are above this block. The PPC result for the minimum is improved when we include in the model these households that consume little water, which could be a future exercise in follow-up studies.

Next, we analyze the GGS results. The posterior mean of the price elasticity, β_1, is estimated to be negative at -1.09. Because the 95% credible interval does not include zero, the probability that $\beta_1 < 0$ is greater than 0.95. This is consistent with economic theory. However, the posterior mean of the income elasticity, β_2, may equal to zero, because its 95% credible interval includes zero.

Among the independent variables that are expected to explain individual heterogeneity, the number of members in a household and the number of rooms in a house or apartment exhibit a positive relationship with residential water demand because $\Pr(\delta_j > 0 \mid \mathbf{y}) > .95$ ($j = 1, 2$). Furthermore, the former has a larger marginal effect on demand than does the latter. That is, a one-person increase in a household has greater effect on water demand than does a one-room extension to a house/apartment does. In contrast, the total floor space in a house/apartment (δ_3) has no effect on water demand in terms of its 95% credible interval.

We compare these parameter estimates with those of previous studies on water demand.

22

Hewitt and Hanemann (1995) used microdata from Denton, Texas and employed the discrete/continuous choice model as the underlying statistical model. Hewitt and Hanemann (1995) then estimated the water demand function under block-rate pricing using the maximum likelihood method. Because of the complex form of the likelihood function, that analysis simply focused on households under two-block increasing block-rate pricing. They reported that the price and income parameters are -1.8989 and 0.1782, respectively. While these parameters are larger in terms of absolute value than ours, Hewitt and Hanemann (1995)'s estimates show a similar pattern to ours (i.e., the larger price and smaller income elasticities in their absolute values).

Olmstead et al. (2007) also applied the discrete/continuous choice model to estimate the water demand function. They used microdata from the United States and Canada that incorporates different price schedules, such as two-block and four-block increasing block-rate pricings and the uniform price system. They showed that the price and income parameters for households under block-rate pricing are estimated at -0.6411 and 0.1959, respectively.

Dalhuisen, Florax, de Groot, and Nijkamp (2003) analyzed 64 studies on water demand and presented the meta-analysis on price and income elasticities. They showed that the price and income elasticities are dispersed with means -0.41 and 0.43, respectively, and standard deviations 0.86 and 0.79, respectively. Their estimates are somewhat similar to ours when we take their large standard deviations into consideration.

5.3 Predictive Analysis

At the end of this section, we conduct a posterior predictive analysis on water demand when the block-rate price schedule is changed to uniform pricing. We consider two types of uniform pricing; namely, the same uniform price for all households and different uniform prices for each households. For the former, let the unit price be ¥100/m^3, ¥250/m^3, or ¥500/m^3 and set

the fixed cost to ¥3,500 for every price scenario. These unit prices are inexpensive, almost as high as, or expensive ones for the majority of households as compared with the present increasing block-rate pricing, and the fixed cost is set close to the present fixed cost for most households (see Figures 4(b) and 4(c)). Other variables remain the same. For the latter, we use the price of the block where water is actually consumed and use it as the single price for the suppositional uniform price system. The fixed cost remains the same as with the present schedule. Furthermore, the virtual income for the block where water is actually consumed is used for total income. Other variables remain the same. Under this hypothetical change, the budget line becomes the straight line that matches the current piecewise-linear budget line at the block where water is actually consumed. To analyze the effect of these price schedule changes, we generate samples of predictive demand using the Gibbs with GGS samples, and draw boxplots of predictive distributions for each household found in Figure 6.
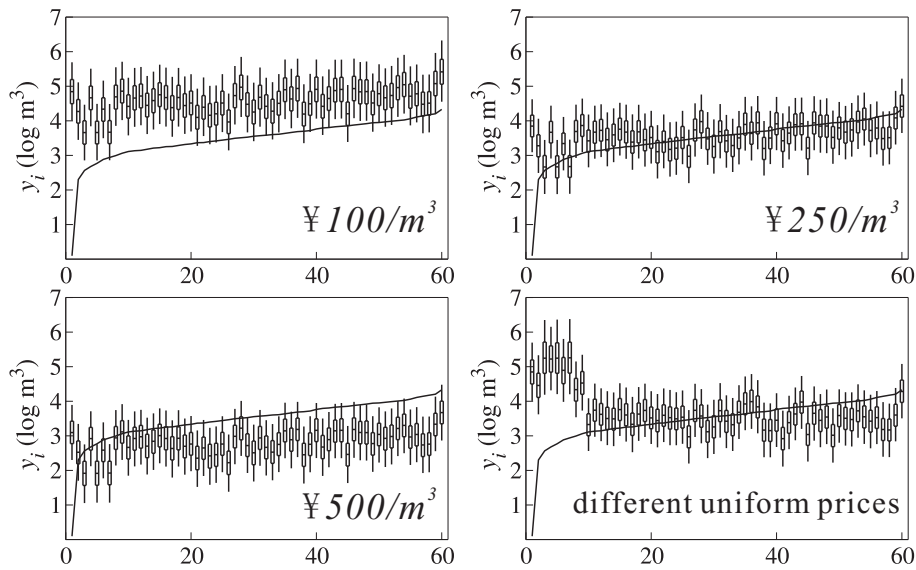

Figure 6: Effects of uniform pricings.

In this figure, solid lines and boxplots represent plots of actual log demands and boxplots of predictive distribution for each households, respectively. Water consumption of each

household is arranged in ascending order, and the number of households is reduced to 60 by selecting every 6-th household. Each box represents the range between the first and third quartiles. The upper and lower whiskers denote the 95-th and 5-th percentiles, respectively.

The first three panels in Figure 6 reveal that most households consume more water as price decreases, which is consistent with the negatively-estimated price elasticity. The bottom right panel in Figure 6, however, shows another aspect of our data. From a microeconomic view-point, a change of this kind has no effect on consumption when an underlying preference satisfies regular assumptions. However, this result is different from what we would expect based on the microeconomic theory, especially for households that consume less than approximately 3 log m^3. When we assume that our statistical modeling is correct, the actual consumption block for these households is suboptimal partly due to heterogeneity and/or partly due to measurement error. Thus, this result implies that we may need to take special care with these households in terms of heterogeneity and/or measurement error.

6 Concluding Remarks

This article proposes a Bayesian estimation method for demand functions under block-rate pricing and reports empirical analysis using Japanese residential water demand data. Furthermore, the separability condition is explicitly considered to obtain appropriate estimates. Our method is useful for analyzing demand for water services, as well as for other goods or services involving block-rate pricing, including taxes. Furthermore, it would be possible to apply our method to examine a consumer's choices over multiple product categories and brands (Song and Chintagunta, 2007) and a consumer's selection of wireless calling plans (Iyengar, 2004).

Future research can be conducted on several related issues. First, the supply structure must be considered when applying our method to other goods under block-rate pricing. Water

companies are regional monopolists and are obliged to supply as much water as consumers require. Thus, we excluded firm competition. Other suppliers, such as telecommunication services and deregulated electricity services, face no such obligation and compete fiercely with each other. To analyze demand of these services, it is necessary to explicitly consider the supply structure in our model. Disequilibrium models represent a framework that address such market structures. See Kunitomo and Sato (1996) and Maddala (1983) for a discussion of disequilibrium models.

Second, as pointed out in the previous section, there are households that consume less than the zero marginal price block. The discrete/continuous choice model proposed by this article excludes such behavior. Developing a structural approach to these consumers is a necessary aspect of future studies.

Third, substitution among electricity, gas, and other fuels must be considered. It is possible for the block-rate pricing model proposed here to be extended to a multivariate setting in a natural way. Furthermore, Japanese gas companies provide natural gas under a decreasing block-rate pricing system. Thus, a subsequent study should examine the energy demand function under a mixture of increasing and decreasing block-rate pricing.

Finally, an improved convergence acceleration method must be developed. Although the generalized Gibbs step improved sampling efficiency, the regression coefficients β still show high sample autocorrelation. Further improvement of convergence acceleration should be a subject of future research.

Appendices

A.1 Full Conditional Distributions for the Statistical Model in Equations (12) to (16)

This section provides the full conditional distributions for the statistical model in Equations (12) to (16) following the standard Gibbs sampler's steps (see Algorithm 1.1 in Subsection 3.4). Let $\mathcal{A}$ denote a set of observations that do not select the threshold or kink point as their demand. That is, $\mathcal{A} = \{i | s_i^* \text{ is odd and equal to } 2k_i - 1\}$. Furthermore, without loss of generality, we assume that $p_{i1}, q_{i1}, \bar{y}_{i1}$ are strictly positive; i.e., $p_{i1}, q_{i1}, \bar{y}_{i1} > 0$. This can be accomplished by adjusting the unit of measurement for price and income. When $P_{i1} = 0$ ($p_{i1} = -\infty$), we assume $s_i^* \geq 2$ and let $p_{i2} > 0$.

Step 2. Generate β_1 given $\beta_2, s^*, w^*, \sigma_u^2$. The full conditional distribution for β_1 is the truncated normal distribution $TN_{R_1}(\mu_1, \sigma_u^2 \sigma_1^2)$, where:

$$
\begin{aligned}
\mu_1 &= \sigma_1^2 \left[\sigma_{\beta_1,0}^{-2} \mu_{\beta_1,0} + \sum_{i \in \mathcal{A}} p_{ik_i} \left(y_i - \beta_2 q_{ik_i} - w_i^* \right) \right], \quad \sigma_1^{-2} = \sigma_{\beta_1,0}^{-2} + \sum_{i \in \mathcal{A}} (p_{ik_i})^2, \\
R_1 &= \left(\max_i (-\infty, BL_i), \min_{i,k} \left(BU_i, -\beta_2 \frac{q_{i,k+1} - q_{ik}}{p_{i,k+1} - p_{ik}} \right) \right),
\end{aligned}
\tag{A.29}
$$

and $\mu_{\beta_1,0}$ is the prior mean of β_1. The BL_i and BU_i are the respective lower and upper bounds of the interval B_i such that:

$$
B_i = \begin{cases} \left(\frac{\bar{y}_{i,k-1} - \beta_2 q_{ik} - w_i^*}{p_{ik}}, \frac{\bar{y}_{i,k} - \beta_2 q_{ik} - w_i^*}{p_{ik}} \right), & \text{if } s_i^* = 2k - 1, \\ \left(\frac{\bar{y}_{i,k} - \beta_2 q_{ik} - w_i^*}{p_{ik}}, \frac{\bar{y}_{i,k} - \beta_2 q_{i,k+1} - w_i^*}{p_{i,k+1}} \right), & \text{if } s_i^* = 2k. \end{cases}
\tag{A.30}
$$

These B_is are constructed from intervals $R_{is_i^*}$ and defined in Equations (9) and (10) of Subsection 3.1. To sample from the truncated normal distributions, we use the inverse cumulative

distribution function sampling method; see Section 1.3 of Gamerman (1997).

Step 3. Generate β_2 given $\beta_1, \boldsymbol{s}^*, \boldsymbol{w}^*, \sigma_u^2$. As in Step 2, the full conditional distribution for β_2 is the truncated normal distribution $TN_{R_2}(\mu_2, \sigma_u^2\sigma_2^2)$, where:

$$
\begin{aligned}
&\mu_2 = \sigma_2^2\left[\sigma_{\beta_2,0}^{-2}\mu_{\beta_2,0} + \sum_{i\in\mathscr{A}} q_{ik_i}\left(y_i - \beta_1 p_{ik_i} - w_i^*\right)\right], \quad \sigma_2^{-2} = \sigma_{\beta_2,0}^{-2} + \sum_{i\in\mathscr{A}}(q_{ik_i})^2, \\
&R_2 = \left(\max_i\left(-\infty, BL_i^{\dagger}\right), \min_{i,k}\left(BU_i^{\dagger}, -\beta_1\frac{p_{i,k+1}-p_{ik}}{q_{i,k+1}-q_{ik}}\right)\right),
\end{aligned}
\tag{A.31}
$$

and $\mu_{\beta_2,0}$ is the prior mean of β_2. The $BL_i^{\dagger}$ and $BU_i^{\dagger}$ are the respective lower and upper bounds of interval $B_i^{\dagger}$ such that:

$$
B_i^{\dagger} = \begin{cases} \left(\frac{\bar{y}_{i,k-1}-\beta_1 p_{ik}-w_i^*}{q_{ik}}, \frac{\bar{y}_{i,k}-\beta_1 p_{ik}-w_i^*}{q_{ik}}\right), & \text{if } s_i^* = 2k-1, \\ \left(\frac{\bar{y}_{i,k}-\beta_1 p_{ik}-w_i^*}{q_{ik}}, \frac{\bar{y}_{i,k}-\beta_1 p_{i,k+1}-w_i^*}{q_{i,k+1}}\right), & \text{if } s_i^* = 2k. \end{cases}
\tag{A.32}
$$

Step 4. Generate $(\sigma_v^2, \boldsymbol{\delta})$ given $\boldsymbol{w}^*$. Because a blocking technique is applied in this step, $\boldsymbol{\delta}$ is integrated over the full conditional distribution of $(\boldsymbol{\delta}, \sigma_v^2)$ to obtain the full conditional distribution of σ_v^2. Thus, generate the sample of σ_v^2 from the inverse gamma distribution $IG(\frac{n_{v,1}}{2}, \frac{S_{v,1}}{2})$, and the sample of $\boldsymbol{\delta}$ from the multivariate normal distribution $N_d(\boldsymbol{\mu}_{\boldsymbol{\delta},1}, \sigma_v^2\boldsymbol{\Sigma}_{\boldsymbol{\delta},1})$, where $n_{v,1} = n_{v,0} + n$,

$$
\begin{aligned}
&S_{v,1} = S_{v,0} + \boldsymbol{\mu}_{\boldsymbol{\delta},0}'\boldsymbol{\Sigma}_{\boldsymbol{\delta},0}^{-1}\boldsymbol{\mu}_{\boldsymbol{\delta},0} + \boldsymbol{w}^{*'}\boldsymbol{w}^* - \boldsymbol{\mu}_{\boldsymbol{\delta},1}'\boldsymbol{\Sigma}_{\boldsymbol{\delta},1}^{-1}\boldsymbol{\mu}_{\boldsymbol{\delta},1}, \\
&\boldsymbol{\mu}_{\boldsymbol{\delta},1} = \boldsymbol{\Sigma}_{\boldsymbol{\delta},1}\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta},0}^{-1}\boldsymbol{\mu}_{\boldsymbol{\delta},0} + \boldsymbol{Z}'\boldsymbol{w}^*\right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\delta},1}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\delta},0}^{-1} + \boldsymbol{Z}'\boldsymbol{Z}.
\end{aligned}
\tag{A.33}
$$

Step 5. Generate (s_i^*, w_i^*) given $\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2$ for $i = 1, \ldots, n$. We again apply a blocking technique in drawing samples of (s_i^*, w_i^*). Then, the conditional posterior distribution of s_i^* is discrete

such that:

$$Pr\left(s_i^* = s \mid \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2\right) \propto \tau_s \left[\Phi\left\{\tau_s^{-1}(RU_{is} - \theta_{is})\right\} - \Phi\left\{\tau_s^{-1}(RL_{is} - \theta_{is})\right\}\right] \exp\left(-\frac{m_{is}}{2}\right), \quad \text{(A.34)}$$

for $s = 1, \ldots, 2K_i - 1$, where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution; RU_{is} and RL_{is} denote the respective upper and lower limit of the interval for heterogeneity; R_{is} is given by Equations (9) and (10); and $(m_{is}, \theta_{is}, \tau_s^2)$ is defined as follows:

$$\left(m_{is}, \theta_{is}, \tau_s^2\right) = \begin{cases} \left(\dfrac{\sigma_u^{-2}\sigma_v^{-2}\left(y_i - y_{ik} - \mathbf{z}_i'\boldsymbol{\delta}\right)^2}{\sigma_u^{-2} + \sigma_v^{-2}}, \ \dfrac{\sigma_u^{-2}(y_i - y_{ik}) + \sigma_v^{-2}\mathbf{z}_i'\boldsymbol{\delta}}{\sigma_u^{-2} + \sigma_v^{-2}}, \ \{\sigma_u^{-2} + \sigma_v^{-2}\}^{-1}\right), \\ \qquad\qquad\qquad\qquad\qquad \text{if } s = 2k - 1, \\ \left(\sigma_u^{-2}(y_i - \bar{y}_{ik})^2, \ \mathbf{z}_i'\boldsymbol{\delta}, \ \sigma_v^2\right), \quad \text{if } s = 2k. \end{cases} \quad \text{(A.35)}$$

Given $s_i^* = s$, we generate w_i^* from $TN_{R_{is}}(\theta_{is}, \tau_s^2)$.

Step 6. Generate σ_u^2 *given* $\boldsymbol{\beta}, \mathbf{s}^*, \mathbf{w}^*$. It is straightforward to show that the full conditional posterior distribution of σ_u^2 is the inverse gamma distribution $IG(\frac{n_{u,1}}{2}, \frac{S_{u,1}}{2})$, where $n_{u,1} = n_{u,0} + 2 + n$ and

$$S_{u,1} = S_{u,0} + \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta},0}\right)' \Sigma_{\boldsymbol{\beta},0}^{-1} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta},0}\right) + (\mathbf{y} - \mathbf{y}^*)'(\mathbf{y} - \mathbf{y}^*). \quad \text{(A.36)}$$

A.2 Full Conditional Density of $\tilde{g}$

We assume that $\bar{y}_{i1}$ is strictly positive. The full conditional probability density function of $\tilde{g}$ ($= g^{-1}$) is derived as follows. First, we substitute parameters multiplied by g ($= \tilde{g}^{-1}$) into the posterior density (25). Because the number of parameters to be accelerated is $4 + d + n$, the Jacobian of this transformation is $g^{-(4+d+n)}$. Transforming g to $\tilde{g}$, the conditional probability

density function of $\tilde{g}$ is given by:

$$\pi(\tilde{g} \mid \boldsymbol{\beta}, \boldsymbol{s}^*, \boldsymbol{w}^*, \sigma_v, \sigma_u) \propto \tilde{g}^{a_0} \exp\left[-\frac{1}{2}\left\{a_1\tilde{g}^2 - 2a_2\tilde{g}\right\}\right] I\left(\tilde{g} \in R_{\tilde{g}}\right) L(d\tilde{g}), \qquad \text{(A.37)}$$

where $L(d\tilde{g}) = \tilde{g}^{-1}d\tilde{g}$ is the left-Haar measure; $a_0 = n + n_{u,0} + n_{v,0}$; and

$$a_1 = \sigma_u^{-2}\left(S_{u,0} + \boldsymbol{\mu}'_{\boldsymbol{\beta},0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta},0}\boldsymbol{\mu}_{\boldsymbol{\beta},0} + \sum_{i=1}^{n} a_{4i}^2\right) + \sigma_v^{-2}\left(S_{v,0} + \boldsymbol{\mu}'_{\boldsymbol{\delta},0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\delta},0}\boldsymbol{\mu}_{\boldsymbol{\delta},0}\right), \qquad \text{(A.38)}$$

$$a_2 = \sigma_u^{-2}\left(\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta},0}\boldsymbol{\mu}_{\boldsymbol{\beta},0} + \sum_{i=1}^{n} a_{3i}a_{4i}\right) + \sigma_v^{-2}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\delta},0}\boldsymbol{\mu}_{\boldsymbol{\delta},0}, \qquad \text{(A.39)}$$

$$(a_{3i}, a_{4i}) = \begin{cases} \left(y_{ik} + w_i^*,\ y_i\right), & \text{if } s_i^* = 2k-1, \\ (0,\ y_i - \bar{y}_{ik}), & \text{if } s_i^* = 2k, \end{cases} \qquad \text{(A.40)}$$

$$R_{\tilde{g}} = \left(\max_i\left(0, BL_i^*\right), \min_i\left(BU_i^*, \infty\right)\right). \qquad \text{(A.41)}$$

The BL_i^* and BU_i^* are the respective lower and upper bounds of the interval B_i^*, which is given by:

$$B_i^* = \begin{cases} \left(\frac{\beta_1 p_{ik} + \beta_2 q_{ik} + w_i^*}{\bar{y}_{ik}}, \frac{\beta_1 p_{ik} + \beta_2 q_{ik} + w_i^*}{\bar{y}_{i,k-1}}\right), & \text{if } s_i^* = 2k-1, \\ \left(\frac{\beta_1 p_{i,k+1} + \beta_2 q_{i,k+1} + w_i^*}{\bar{y}_{ik}}, \frac{\beta_1 p_{ik} + \beta_2 q_{ik} + w_i^*}{\bar{y}_{ik}}\right), & \text{if } s_i^* = 2k. \end{cases} \qquad \text{(A.42)}$$

A.3 Proof for the One-step MH Algorithm

We prove that it suffices to implement a one-step MH algorithm using the initial value $\tilde{g} = 1$ in our GGS. By Theorem 2 of Liu and Sabatti (2000), it is sufficient to prove that for all $\tilde{g}, \tilde{g}', \tilde{g}_0 \in \tilde{\Gamma} = \{\tilde{g} > 0 : \tilde{g}(x) = \tilde{g}^{-1}x\}$,

$$T_{\boldsymbol{\zeta}}(\tilde{g}, \tilde{g}') L(d\tilde{g}') = T_{\tilde{g}_0^{-1}\boldsymbol{\zeta}}(\tilde{g}\tilde{g}_0, \tilde{g}'\tilde{g}_0) L(d\tilde{g}'), \qquad \text{(A.43)}$$

where $T_{\boldsymbol{\zeta}}(\tilde{g}, \tilde{g}') L(dg')$ is the transition kernel of our Markov chain.

Let $q_\zeta(\tilde{g}')$ denote our proposal density function. Then, the transition kernel becomes $T_\zeta(\tilde{g}, \tilde{g}') = q_\zeta(\tilde{g}')\alpha(\tilde{g}, \tilde{g}')\tilde{g}'$, where the last $\tilde{g}'$ is the adjustment term for the left-Haar measure. It is obvious that the acceptance probability $\alpha(\tilde{g}, \tilde{g}')$ is invariant to the scale transformation of $\tilde{g}_0$. Moreover, we find that:

$$q_\zeta(\tilde{g}')\tilde{g}' = \sigma_{\tilde{g}'}^{-1}\phi\left(\frac{\tilde{g}' - \mu_{\tilde{g}'}}{\sigma_{\tilde{g}'}}\right)I\left(\tilde{g}' \in R_{\tilde{g}}\right)\tilde{g}', \tag{A.44}$$

$$q_{\tilde{g}_0^{-1}\zeta}(\tilde{g}'\tilde{g}_0)\tilde{g}'\tilde{g}_0 = \left(\tilde{g}_0\sigma_{\tilde{g}'}\right)^{-1}\phi\left(\frac{\tilde{g}'\tilde{g}_0 - \tilde{g}_0\mu_{\tilde{g}'}}{\tilde{g}_0\sigma_{\tilde{g}'}}\right)I\left(\tilde{g}'\tilde{g}_0 \in \tilde{g}_0 R_{\tilde{g}}\right)\tilde{g}'\tilde{g}_0 = q_\zeta(\tilde{g}')\tilde{g}', \tag{A.45}$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution. Thus, the transition kernel of the Markov chain is invariant to the transformation $\tilde{g}$, which completes the proof.

Acknowledgements

References

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.

Burtless, G. and J. A. Hausman (1978). The effect of taxation on labor supply: Evaluating the Gary negative income tax experiment. *Journal of Political Economy 86*(6), 1103–1130.

Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics 51*, 79–99.

Chib, S. (2001). Markov chain Monte Carlo methods: Computation and inference. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter 57, pp. 3569–3649. Amsterdam: North-Holland.

Chib, S. and E. Greenberg (1996). Markov chain Monte Carlo simulation method in econometrics. *Econometric Theory 12*, 409–431.

Dalhuisen, J. M., R. J. G. M. Florax, H. L. F. de Groot, and P. Nijkamp (2003). Price and income elasticities of residential water demand: A meta-analysis. *Land Economics 79*(2), 292–308.

de Jong, G. C. (1990). An indirect utility model of car ownership and private car use. *European Economic Review* (34), 971–985.

Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox* (3rd ed.). London: Timberlake Consultants Press and Oxford.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Texts in Statistical Science. London: Chapman and Hall.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis* (2nd ed.). London: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 169–193. Oxford: Oxford University Press.

Hanemann, W. M. (1984). Discrete/continuous models of consumer demand. *Econometrica 52*(3), 541–562.

Hausman, J. A. (1985). The econometrics of nonlinear budget sets. *Econometrica 53*(6), 1255–1282.

Herriges, J. A. and K. K. King (1994). Residential demand for electricity under inverted block rates: Evidence from a controlled experiment. *Journal of Business and Economic Statistics 12*(4), 419–430.

Hewitt, J. A. and W. M. Hanemann (1995). A discrete/continuous choice approach to residential water demand under block rate pricing. *Land Economics 71*, 173–192.

Iyengar, R. (2004). A structural demand analysis for wireless services under nonlinear pricing schemes. unpublished working paper, Marketing Department, Wharton School, University of Pennsylvania.

Kunitomo, N. and S. Sato (1996). Asymmetry in economic time series and the simultaneous switching autoregressive model. *Structural Change and Economic Dynamics 7*, 1–34.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer-Verlag New York.

Liu, J. S. and C. Sabatti (2000). Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika 87*(2), 353–369.

Maddala, G. S. (1983). *Limited-dependent and Qualitative Variables in Econometrics*. Number 3 in Econometric Society Monographs. Cambridge, New York: Cambridge University Press.

Moffitt, R. (1986). The econometrics of piecewise-linear budget constraint. *Journal of Business and Economic Statistics 4*(3), 317–328.

Moffitt, R. (1989). Estimating the value of an in-kind transfer: the case of food stamps. *Econometrica 57*(2), 385–409.

Olmstead, S. M., W. M. Hanemann, and R. N. Stavins (2007). Water demand under alternative price structures. *Journal of Environmental Economics and Management 54*(2), 181–198.

Reiss, P. C. and M. W. White (2005). Household electricity demand, revisited. *Review of Economic Studies 72*, 853–883.

Song, I. and P. K. Chintagunta (2007). A discrete-continuous model for multicategory purchase behavior of households. *Journal of Marketing Research 44*(4).