CIRJE-F-942

A Discrete/Continuous Choice Model on a Nonconvex Budget Set

Koji Miyawaki
Kwansei Gakuin University

Yasuhiro Omori
The University of Tokyo

Akira Hibiki
Sohpia University

September 2014

A discrete/continuous choice model on a nonconvex budget set

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

September 29, 2014

[a] School of Economics, Kwansei Gakuin University, Hyogo 662-8501, Japan (miyawaki.koji@kwansei.ac.jp).

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

[c] Faculty of Economics, Sophia University, Tokyo 102-8554, Japan (hibiki@sophia.ac.jp).

Abstract

Decreasing block rate pricing is a nonlinear price system often used for public utility services. Residential gas services in Japan and the United Kingdom are provided under this price schedule. The discrete/continuous choice approach is used to analyze the demand under decreasing block rate pricing. However, the nonlinearity problem, which has not been examined in previous studies, arises because a consumer's budget set (a set of affordable consumption amounts) is nonconvex and, hence, the resulting model includes highly nonlinear functions. To address this problem, we propose a feasible, efficient method of demand estimation on the nonconvex budget. The advantages of our method are as follows: (i) the construction of an Markov chain Monte Carlo algorithm with an efficient blanket based on the Hermite-Hadamard integral inequality and the power-mean inequality, (ii) the explicit consideration of the (highly nonlinear) separability condition, which often makes numerical likelihood maximization difficult, and

*Corresponding author: Phone:+81-798-54-6204. E-mail:miyawaki.koji@kwansei.ac.jp.

(iii) the introduction of normal disturbance into the discrete/continuous choice model on the nonconvex budget set. The proposed method is applied to estimate the Japanese residential gas demand function and evaluate the effect of price schedule changes as a policy experiment.

Key words: Discrete/Continuous choice approach, Nonconvex budget set, Bayesian analysis, Residential gas demand, Hermite-Hadamard integral inequality.
JEL classification: C11, C24, D12.

1 Introduction

The decreasing block rate pricing is a nonlinear price system where the unit prices discontinuously decline with the quantity consumed. Figure 1 illustrates a typical decreasing block rate pricing. As this figure shows, there are several threshold values that divide the con-

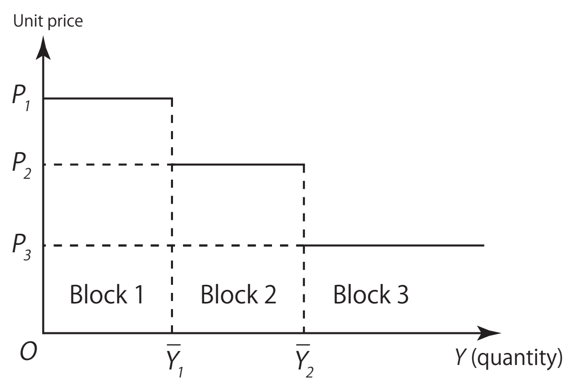


Figure 1: Price schedule of a three-block decreasing block rate pricing ($K = 3$).

sumption amount and the unit prices decrease when the quantity consumed exceeds these threshold values. The residential gas services in Japan and the United Kingdom are often provided under decreasing block rate pricing (see also Section 6 for the case of Japan). Other services, such as the mobile phone service (the personal handy-phone system) in Japan and some of the residential electricity services in the United States, also employ this price system. Such a price schedule is likely to be employed partly because the production cost is decreasing in scale and partly because this system is considered to encourage a larger amount of consumption. Formal description of this price schedule is as follows.

Let K be the number of blocks in the decreasing block rate pricing. In Figure 1, $K = 3$. The consumption amount Y is divided into these blocks. Let $\bar{Y}_k$ be the upper limit quantity for the k-th block ($k = 0, \ldots, K$). We set $\bar{Y}_0 \equiv 0$ and $\bar{Y}_K \equiv \infty$ for simplicity. Then, each block is defined as the interval $[\bar{Y}_{k-1}, \bar{Y}_k)$ for $k = 1, \ldots, K$. The unit prices are set in relation to these blocks and let P_k be the unit price for the k-th block ($k = 1, \ldots, K$). Under decreasing block rate pricing, the unit price monotonically and discontinuously declines according to the blocks, that is, $P_{k+1} < P_k$ for $k = 1, \ldots, K - 1$. When, on the other hand, $P_k < P_{k+1}$, such a price schedule is called the increasing block rate pricing. These two price schedules are special cases of block rate pricing, where the unit prices discontinuously change with the quantity consumed. Chapter 7 of Train (1991) provides a brief microeconomic analysis of block rate pricing.

Generally, it is often the case in consumers' demand analysis to examine how consumers respond to the unit price. However, under decreasing block rate pricing, there are several unit prices depending on the consumption amount, and makes the analysis more complicated. From the microeconomics' point of view, such a response is modeled through the demand function based on the so-called discrete/continuous choice approach, which is first proposed by Burtless and Hausman (1978).[1] As a consequence of this discrete/continuous choice approach, we can evaluate the social welfare, such as the compensating variation, under decreasing block rate pricing in comparison with that under uniform price system where there is only one fixed unit price. A formal presentation of this welfare measure is found in

[1] The discrete/continuous choice approach is named because it simultaneously considers the discrete and continuous choices. In the block rate pricing case, consumers choose both the block and the consumption amount, which are discrete and continuous, respectively. This approach has also been used to examine a wide range of topics including housing (Lee and Trost (1978); King (1980)), transportation (Mannering and Winston (1985); Hensher and Milthorpe (1987); de Jong (1990); West (2004)), labor supply (Burtless and Hausman (1978); Burtless and Moffitt (1985)), electricity demand (Herriges and King (1994)), and water demand (Hewitt and Hanemann (1995); Olmstead, Hanemann, and Stavins (2007); Miyawaki, Omori, and Hibiki (2013)). When the budget set is convex, Miyawaki et al. (2013) proposed the appropriate estimation method. However, as we will see later, the nonconvex budget case (including the decreasing block rate pricing case) is much more difficult to estimate model parameters properly.

Subsection 3.2 and the empirical analysis based on the compensating variation is given in Subsection 6.3.

This demand function, however, has not been investigated in the previous empirical studies because it requires the comparison of nonlinear functions (i.e., nonlinear indirect utility functions which will be defined in Subsection 2.1 and their functional forms will be specified in Subsection 3.1). Such a nonlinearity is caused by *Roy's identity* which is a partial differential equation based on the consumer's behavior in the microeconomic theory.[2]

To avoid this nonlinearity, Blomquist and Newey (2002) proposed a nonparametric approach. They analyzed the effect of tax reform in Sweden on working hours for married or cohabiting men from 20 to 60 years of age. For employees, the working time is influenced by the tax system, and it is interpreted as a block rate pricing. Then, the employee's decision about how much time to work can be considered as the problem under block rate pricing. Thus, Blomquist and Newey (2002) estimated the function of working time as a nonparametric function of the entire tax system. Though their approach is free of the nonlinearity caused by Roy's identity and of model misspecifications and distributional errors, it does not incorporate foundational aspects of the microeconomic theory like Roy's identity into the statistical model. Thus, this article considers a parametric model of demand to appropriately address Roy's identity.

Previous literature (e.g., Burtless and Hausman (1978); Hausman (1980); Burtless and Moffitt (1985)) has used parametric models that are based on the discrete/continuous choice approach and applied them to estimate the effect of block rate pricing involving a two-block decreasing block rate pricing using the maximum likelihood method.[3] However, two-block

[2] We face the nonlinearity under decreasing block rate pricing, though we do not under increasing block rate pricing (see, e.g., Moffitt (1986)).

[3] Recently, Szabó (2009) proposed the maximum likelihood estimation method for general block rate pricing where the linear demand function is assumed. Szabó (2009) imposed a condition that the direct utility function is quasiconcave. This condition aims to guarantee that the underlying preference relation be strictly convex, that is, the preference relation be well-behaved. However, as stated in Hurwicz and Uzawa (1971), two more conditions (the nonnegative demand condition and the separability condition) are required for the underlying preference relation to be strictly convex. These additional conditions often make it difficult to numerically maximize the likelihood function. See Miyawaki et al. (2013) for the detailed discussion on this issue.

rate pricing is too simple for use in the analysis of real data such as Japanese residential gas data, where the number of blocks is much greater than two. (Indeed, the number of blocks is three to six depending on the gas company.) If the block structure was simplified to mimic two-block rate pricing, the estimates of the demand function as used for policy-making would be biased. Thus, we consider general multiple-block decreasing block rate pricing as a type II Tobit model subject to many nonlinear constraints (see Chapter 10 of Amemiya (1985) for the Tobit classification) and propose its Bayesian estimation method using a Markov chain Monte Carlo (MCMC) simulator with an efficient blanket.

A typical Bayesian approach for limited dependent variable models applies the data augmentation method (see Tanner and Wong (1987)). Pioneering works on the Bayesian approach for such models that utilize the data augmentation can be found in Chib (1992) and Albert and Chib (1993).

Because the resulting statistical model includes many nonlinear constraints on model parameters (the comparison of nonlinear functions and the separability condition, which will be explained in Subsection 3.4), the support of the full conditional distribution for some model parameters is difficult to calculate when we use a standard statistical software. One possible solution to this problem is rejection sampling. However, using a simple envelope function (or a simple blanket) for the support is extremely inefficient because the acceptance rate of the proposed samples is extremely low (see Section 4.3). Thus, this article develops an efficient blanket using two properties of convex functions: the Hermite-Hadamard integral inequality and the power-mean inequality.

Our approach also has another particular advantage. The previous studies employing maximum likelihood estimation do not explicitly consider the separability condition, though this condition is necessary for the demand model under decreasing block rate pricing with more than two blocks (see Subsection 3.4 for the detailed discussion on the separability condition under decreasing block rate pricing). Miyawaki et al. (2013) dealt with this issue in the context of increasing block rate pricing. Under increasing block rate pricing, this condition is a set of linear constraints on model parameters. In contrast, under multiple-block decreasing block rate pricing, our statistical model includes the separability condition, which is highly nonlinear, to properly estimate the model parameters. Because of this condition, the likelihood maximization requires the constrained optimization, and it is often difficult to numerically maximize the likelihood function. Thus, we need to pursue the Bayesian approach, using the MCMC simulator to estimate the model parameters.

Finally, we would like to note that our proposed method has an advantage over the other type of discrete/continuous choice analysis used in the context of the multinomial choice model, as in Dubin and McFadden (1984). The resulting statistical model is the same as that for demand under decreasing block rate pricing. Dubin and McFadden (1984) analyzed the joint choice of electric appliances and electricity demand using this approach and estimated the model parameters based on a combination of the maximum likelihood and the conditional expectation correction method. Their statistical model is simplified by introducing the logit error into the choice of electric appliance portfolios. However, such a specification implies the independence of irrelevant alternatives. The subsequent literature addresses this problem in two ways: by using the nested logit model (e.g., Goldberg (1998)) or by linearizing the nonlinear indirect utilities (e.g., Bernard, Bolduc, and Bélanger (1996)). Carpio, Wohlgenant, and Safley (2008) used a different method and applied it to the estimation of the demand for pick-your-own versus preharvested strawberries with normal error. However, their statistical model is a binary choice model: thus, they do not consider the separability condition. Therefore, this article is the first study to propose the discrete/continuous choice model on the nonconvex budget set with normal disturbance.

In presenting a parametric model for demand under decreasing block rate pricing, this article proposes the use of Bayesian analysis to make the following contributions: (i) the construction of an MCMC algorithm with an efficient blanket based on the Hermite-Hadamard integral inequality and the power-mean inequality: (ii) the explicit consideration of the (highly nonlinear) separability condition, which often makes numerical likelihood maximization difficult: and (iii) the introduction of normal disturbance into the discrete/continuous choice model on the nonconvex budget set.

Using the proposed method, we analyze the residential gas demand function and evaluate the effect of price schedule changes. The substitution between residential gas and electricity will be left for the future work because our main interest is the demand function under decreasing block rate pricing.

This article is organized as follows. Section 2 explains the consumer's choice problem under one fixed unit price, describes that under the decreasing block rate pricing, and derives the demand function under decreasing block rate pricing based on the discrete/continuous choice approach. Then, Section 3 describes the corresponding statistical model and its likelihood function, with the discussion of the separability condition. In Section 4, we discuss the Bayesian approach and its MCMC simulator with an efficient blanket. We also evaluate the adequacy of the proposed blankets. In Section 5, we explain the model without heterogeneity as an alternative model. Section 6 estimates the Japanese residential gas demand function and evaluates the effect of price schedule changes. Section 7 concludes the study.

2 Demand function under decreasing block rate pricing

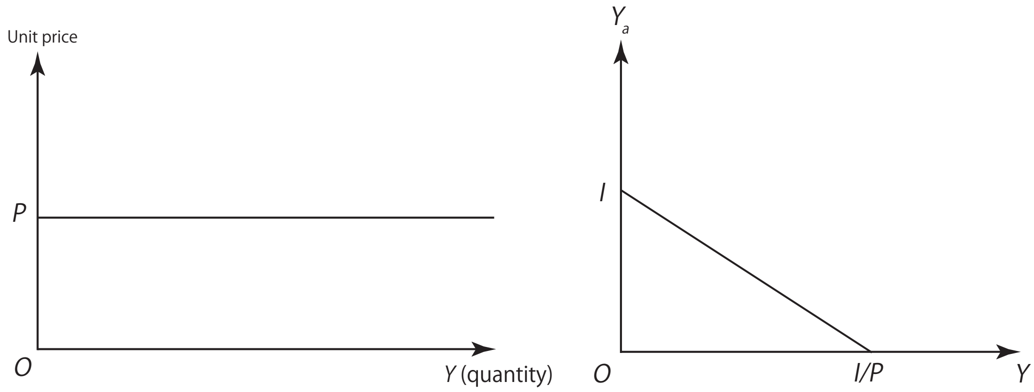
2.1 Consumer's problem under uniform price system

This subsection describes the consumer's choice problem under uniform price system and introduces several terminology that is common in microeconomics.

Consider a consumer's choice problem between two goods, that is, the consumer needs to decide the consumption amount for each good. In this subsection, both goods are supplied with one fixed unit prices. Because there are only two goods, it is sufficient to consider a (relative) unit price P for one good. The unit price for the other good is normalized to one. Given the total income I, we define the budget set for the consumer. The budget set is a set of consumption amounts which the consumer can afford. More precisely,

$$\{(Y, Y_a) \mid PY + Y_a \leq I, Y \geq 0, Y_a \geq 0\}, \tag{1}$$

where Y and Y_a are the consumption amounts for goods with the unit price P and one, respectively. Figure 2(a) and Figure 2(b) shows the price schedule for Y and the budget set given above, respectively.



(a) Price schedule. (b) Budget set.

Figure 2: A uniform price system.

In the classical demand theory, the consumer's optimal consumption amount is determined by maximizing its utility subject to the budget set. The utility is a measure that compares possible choices and is usually represented by a real-valued function of these choices. Let $U(Y, Y_a)$ be the well-defined utility function of the consumption amounts. Then, the utility maximization problem is defined as

$$\max_{Y, Y_a} U(Y, Y_a) \quad \text{subject to } PY + Y_a \leq I. \tag{2}$$

The solution is called the demand functions for these goods denoted by $Y(P, I)$ and $Y_a(P, I)$. The maximum is termed as the indirect utility function represented by $V(P, I)$. These two functions are related by the so-called Roy's identity, which is given by

$$Y(P,I) = -\frac{\partial V(P,I)/\partial P}{\partial V(P,I)/\partial I}. \tag{3}$$

See, e.g., Proposition 3.4.G of Mas-Colell, Whinston, and Green (1995) for the derivation of this identity.

2.2 Consumer's problem under decreasing block rate pricing

This subsection examines the consumer's choice problem under decreasing block rate pricing. Suppose there are two goods: a good that is provided under decreasing block rate pricing and the numeraire good. The numeraire good represents all the other good except the good under decreasing block rate pricing and its price is normalized to one. Let Y be the demand for a good under decreasing block rate pricing. Then, because the unit price declines as the consumption amount grows, the budget set becomes nonconvex (see Figure 3 for the three-block case).

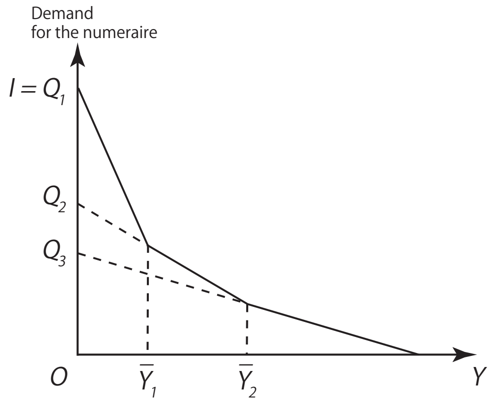


Figure 3: Budget set of a three-block decreasing block rate pricing ($K = 3$).

To derive the demand function under decreasing block rate pricing, it is popular to use the so-called discrete/continuous choice approach (see, e.g., Moffitt (1986)). This approach is a two-step procedure used to solve the utility maximization problem under block rate pricing. For each k-th block, let $Q_k = I - FC - \sum_{j=1}^{k-1}(P_j - P_{j+1})\bar{Y}_j$, where FC is the minimum access charge as the fixed cost. This variable is called the virtual income for the k-th block. We note that $Q_{k+1} < Q_k$ for $k = 1, \dots, K-1$ (see also Figure 3).

Then, under decreasing block rate pricing, the discrete/continuous choice approach is described as follows.

Step 1. For each k-th block ($k = 1, \dots, K$), maximize the utility under the uniform price system, where a consumer faces the single unit price P_k and its corresponding virtual income Q_k. As the solution and maximum, we obtain the demand function Y_k and the indirect utility function V_k, respectively.

Step 2. Find the block k such that $V_k = \max_j V_j$. Then, Y_k is the optimal demand.

In Step 1, both the price and the virtual income are given as constants. Thus, this step can be interpreted as the consumer's choice problem under uniform price system with the unit price P_k and the virtual income Q_k, which has been described in the previous subsection. The obtained solution and maximum in this step are called the conditional demand and the conditional indirect utility, respectively, because they are derived by fixing the block choice k.

Finally, by following the above two steps, we obtain the demand function under decreasing block rate pricing:

$$Y = Y_k, \quad V_k = \max_j V_j. \tag{4}$$

3 Type II Tobit model with nonlinear indirect utility comparisons

3.1 Log-linear demand specification

To derive the statistical form to be used for the empirical analysis, we need to specify the functional form of the conditional demand or the conditional indirect utility (see equation

(3)). Following the discussion by Hausman (1985), this article assumes the conditional demand to be linear in logarithm.[4]That is,

$$\ln Y_k = \beta_1 \ln P_k + \beta_2 \ln Q_k. \tag{5}$$

The log-linear function is popular in the analysis of demand under block rate pricing, because β_1 and β_2 can be directly interpreted as price and (virtual) income elasticities, respectively, conditional on block choice (see, e.g., Hewitt and Hanemann (1995); Olmstead et al. (2007)). The price elasticity, for example, is the percentage change in demand with respect to a percentage change in price. These elasticity parameters play an important role in the microeconomic theory and in the policy-making.

After specifying the conditional demand function, Roy's identity implies

$$V_k = -\frac{P_k^{1+\beta_1}}{1+\beta_1} + \frac{Q_k^{1-\beta_2}}{1-\beta_2}, \tag{6}$$

where $\beta_1 \neq -1$ and $\beta_2 \neq 1$, as derived in Burtless and Hausman (1978). Plugging equations (5) and (6) into equation (4), we have the demand function under decreasing block rate pricing based on the discrete/continuous choice approach.[5]

We note that this theoretical framework does not exclude cases in which multiple blocks are simultaneously optimal. Such a case is excluded by introducing a continuous random disturbance into the consumer's heterogeneity in preferences. Subsection 3.3 describes its specification.

Remark 1. Hanemann (1984) proposed two other demand functions that are less popular in the literature: the linear expenditure system (LES) model and the price independent general-

[4]This functional form implicitly assumes that $P_k > 0$ and $Q_k > 0$ for all k. Under decreasing block rate pricing, these assumptions are equivalent to $P_K > 0$ and $Q_K > 0$.

[5]As pointed out in Hausman (1985), our approach that involves deciding the demand function first and deriving its corresponding indirect utility function has two advantages: (i) we can flexibly choose the functional form of the demand function based on the empirical dataset, and (ii) the stochastic specification becomes convenient.

ized log-linear (PIGLOG) model.

3.2 Compensating variation

Because the demand function includes the (conditional) indirect utility, we can evaluate the effect of the price schedule changes on welfare using the compensating variation. The compensating variation is a quantitative measure of welfare changes due to the price schedule changes and is defined as the difference between the current income and the income required to attain the current utility level under the new price schedule. The amount of positive (negative) difference can be interpreted as the degree of improvement (decline) in consumer welfare under the new price schedule (see Chapter 3 of Mas-Colell et al. (1995) for a general discussion of the compensating variation).

For the case of decreasing block rate pricing, the compensating variation is derived as follows. Let $P = \{\{P_k, \bar{Y}_k\}_{k=1}^{K-1}, P_K, FC\}$ and $P' = \{\{P'_k, \bar{Y}'_k\}_{k=1}^{K'-1}, P'_{K'}, FC'\}$ denote the current and the suppositional price schedule, respectively. Then, by solving

$$V = \text{(the right hand side of equation (6) evaluated with } P'), \tag{7}$$

for I, where V is a certain utility level, we obtain the expenditure at the certain utility level under the suppositional price schedule P', which is given by

$$E_{k'}(P', V) = \left[(1-\beta_2) \left\{ V + \frac{(P'_{k'})^{1+\beta_1}}{1+\beta_1} \right\} \right]^{1/(1-\beta_2)} + FC' + \sum_{j=1}^{k'-1} \left(P'_j - P'_{j+1} \right) \bar{Y}'_j, \tag{8}$$

where $k' = \operatorname{argmax}_j V'_j$ and V'_j is the (suppositional) indirect utility conditional on the j-th block under P' (see Hausman (1981) for the case in which there is a single unit price). With equation (8), the compensating variation is defined as

$$CV = I - E_{k'}(P', V_k), \tag{9}$$

where $k = \text{argmax}_j V_j$ and V_j is the j-th (current) conditional indirect utility under P.

When we assume P' to be the uniform price system, that is, $P' = \{P^*, FC^*\}$, we have

$$E_{k'}(P', V) = \left[(1-\beta_2) \left\{ V + \frac{(P^*)^{1+\beta_1}}{1+\beta_1} \right\} \right]^{1/(1-\beta_2)} + FC^*. \tag{10}$$

The conditional indirect utility under P is given by equation (6). Therefore, the compensating variation is calculated as

$$CV = I - \left[(1-\beta_2) \left\{ \frac{(P^*)^{1+\beta_1} - P_k^{1+\beta_1}}{1+\beta_1} + \frac{Q_k^{1-\beta_2}}{1-\beta_2} \right\} \right]^{1/(1-\beta_2)} - FC^*. \tag{11}$$

In subsection 6.3, we will conduct the welfare analysis based on the compensating variation using the empirical data.

Remark 2. Another welfare measure is the equivalent variation, which is given by

$$EV = E_k\left(P, V'_{k'}\right) - I.$$

Because both EV and CV show similar patterns with our empirical dataset, the discussion and the results of EV are suppressed.

3.3 Statistical model

This subsection describes a statistical model that is a nonlinear type II Tobit model based on the theoretical framework with equations (4)-(6). There are n consumers. Let subscript i denote the consumer i ($i = 1, \dots, n$) and let $(y_i, p_{ik}, q_{ik}) = (\log Y_i, \log P_{ik}, \log Q_{ik})$.

Then, the statistical model for the demand function under decreasing block rate pricing is given by

$$y_i = \boldsymbol{x}'_{is_i^*}\boldsymbol{\beta} + w_i^* + u_i, \quad u_i \sim \text{i.i.d. } N\left(0, \sigma_u^2\right), \tag{12}$$

where $\boldsymbol{x}_{is_i^*} = (p_{is_i^*}, q_{is_i^*})'$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$,

$$s_i^* = k, \quad \text{if } w_i^* \in R_{ik} = \left\{w_i^* \mid V_{ik} > V_{ij} \text{ for } k \neq j\right\} \text{ and } k = 1, \ldots, K_i, \tag{13}$$

$$w_i^* = \boldsymbol{z}_i'\boldsymbol{\delta} + v_i, \quad v_i \sim \text{i.i.d. } N\left(0, \sigma_v^2\right), \tag{14}$$

$$V_{ik} = -\exp\left(w_i^*\right) \frac{P_{ik}^{1+\beta_1}}{1+\beta_1} + \frac{Q_{ik}^{1-\beta_2}}{1-\beta_2}, \tag{15}$$

$\beta_1 \neq -1$, and $\beta_2 \neq 1$. Because of the log-linear demand specification, we require $P_{iK_i} > 0$ and $Q_{iK_i} > 0$ for all i. In our empirical dataset, there are no households whose $Q_{iK_i} \leq 0$.

In this statistical model, there are three components in addition to the theoretical framework with equations (4)-(6). The first component is w_i^*, which represents the consumer's heterogeneity in preferences. We introduce a hierarchical structure into the heterogeneity and assume it to be linear in the d-dimensional covariate vector $\boldsymbol{z}_i$ with its corresponding coefficient vector $\boldsymbol{\delta}$. The disturbance v_i of the heterogeneity is normally distributed with a mean of 0 and a variance of σ_v^2.

There are two following motivations for the introduction of this term. At first, as thoroughly discussed in Moffitt (1986), this term is introduced to explain unobserved tastes included in the utility function. It is natural to assume that the utility function may vary across consumers due to their unmeasured individual attributes. Then, the solutions to the utility maximization problem (i.e., the optimal demands) will differ among consumers even if they face the same price structure and earn the same income level.

Next motivation is to impose zero probability on the multiple optima to the utility maximization problem. To see this, let us solve the comparison of conditional indirect utilities with respect to heterogeneity. This is because the indirect utility conditional on the block choice is derived from the sum of y_{ik} and w_i^* using Roy's identity. The resulting interval is called the heterogeneity interval and is denoted by R_{ik}. The explicit formula for the heterogeneity interval is given in Appendix A.1.

To be rigorous, this interval must be $\bar{R}_{ik} = \{w_i^* \mid V_{ik} = \max_j V_{ij}\}$, where a tie among the conditional indirect utilities is allowed. Clearly, $R_{ik} \subseteq \bar{R}_{ik}$. However, the set where V_{ik} is equal to V_{ij} ($j \neq k$) has a probability of zero in our statistical model. The reason is as follows. Conditional on β_1 and β_2, the condition $V_{ik} = V_{ij}$ leads to the condition that w_i^* must equal to a certain real value, $\ln E_{kj}$, which is derived in Appendix A.1. Because w_i^* is a continuous random variable, this condition has a zero probability. Thus, we are allowed to replace $\bar{R}_{ik}$ with R_{ik}. This zero probability implies that the statistical model excludes the multiple optima.

The second component is the state variable, s_i^*, and we can use the data augmentation method to estimate the model parameters (see Tanner and Wong (1987) for more information on this method). The s_i^* is a discrete latent variable that takes one of the values from 1 to K_i and indicates the optimal block for the i-th consumer. Thus, in the proposed statistical model, the observed block where the consumption is actually made differs from the optimal block s_i^* due to u_i, which will be explained in the next paragraph.

The third component is the measurement error u_i for demand that follows a normal distribution with a mean of 0 and a variance of σ_u^2. This term is assumed to be independent of v_i. As discussed in Hausman (1985), u_i also represents an optimization error by the consumer and a misspecification error by the statistician. Furthermore, Moffitt (1986) pointed out that it is expected to give nonzero probability on the consumption amount between heterogeneity intervals (see also Figure 6 of Moffitt (1986)). When the measurement error is excluded from the model, such a consumption amount will not be observed because the upper limit of the k-th heterogeneity interval is less than the lower limit of the $k+1$-st interval. This situation is alleviated by the introduction of the error term.

We refer to the identification problem of two errors: u_i for the observed demand and v_i for heterogeneity. They cannot be fully identified unless there is additional information through the prior distribution about these errors because there is only one equation for them: $y_i = \boldsymbol{x}'_{is_i^*}\boldsymbol{\beta} + \boldsymbol{z}'_i\boldsymbol{\delta} + v_i + u_i$. However, due to the utility maximization condition, they are weakly separated depending on the dataset.[6]

Alternatively, it is possible to consider the model without heterogeneity. Section 5 describes such a model. Subsection 6.4 gives its estimation result with the gas demand data, and the model comparison between models with and without heterogeneity in terms of the log of the marginal likelihood. Although there is the identification issue, the two-error component model is preferred because of its features described above as well as its fit to the empirical dataset.

3.4 Likelihood function subject to many nonlinear constraints

The likelihood function augmented by the latent variables is given by

$$f\left(y_i, s_i^*, w_i^* \mid \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2\right) \propto (\sigma_u \sigma_v)^{-1} \exp\left[-\frac{1}{2}\left\{\sigma_u^{-2}\left(y_i - \boldsymbol{x}_{is_i^*}'\boldsymbol{\beta} - w_i^*\right)^2 + \sigma_v^{-2}\left(w_i^* - \boldsymbol{z}_i'\boldsymbol{\delta}\right)^2\right\}\right] I\left(w_i^* \in R_{is_i^*}\right) \times \prod_{k=2}^{K_i-1} I(RL_{ik} \leq RU_{ik}), \quad (16)$$

where $I(A)$ is the indicator function: $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. Because we take a Bayesian approach as described later and treat $\boldsymbol{\beta}$ as a continuous random vector, the conditions $\beta_1 \neq -1$ and $\beta_2 \neq 1$ are omitted hereafter.

RL_{ik} and RU_{ik} are the respective lower and upper limits of the heterogeneity interval R_{ik}, and their definitions are given in equation (39) in Appendix A.1. The heterogeneity intervals cover the real line, that is, $\cup_{k=1}^{K_i} R_{ik} \subseteq (-\infty, \infty)$. Further, as noted in the appendix $R_{ik} \cap R_{ij} =$ $(k \neq j)$ for all i. However, depending on values of β_1 and β_2, the upper limit of the interval can be less than the lower limit. To eliminate such a situation, we restrict the parameter space by the last term of the likelihood function.

[6] With the empirical dataset that will be used in Section 6, we conducted the estimation of the gas demand function normalizing the variance of heterogeneity to one. The results are affected by this normalization. In particular, the posterior mean of β_1 is estimated to be -0.094 and its 95% credible interval is $(-0.31, -0.003)$.

The last term, the product of the $K_i - 2$ indicator functions, is the condition that the heterogeneity intervals are separable, that is, $R_{ik} \neq \emptyset$ (for all k). We call this condition the separability condition. This condition is a set of nonlinear constraints on β_1 and β_2, and the number of nonlinear constraints increases as the number of observations and blocks grows. Because of this condition, it is often difficult to numerically maximize the likelihood.

Figure 4 is included to show how the separability condition restricts (β_1, β_2) by using the empirical dataset. Because the separability condition is difficult to calculate, each point is

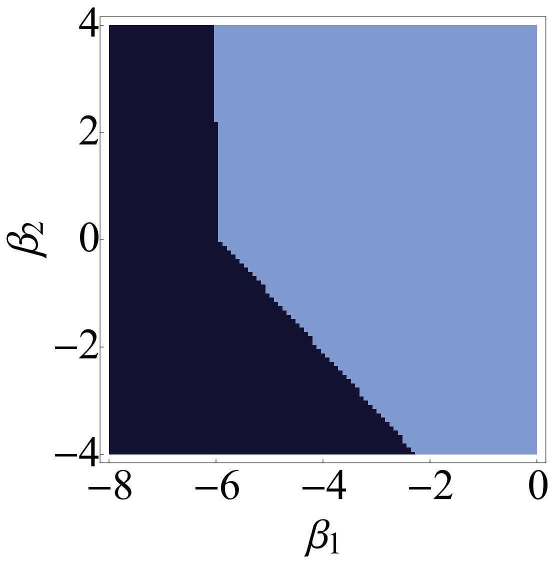


Figure 4: Region implied by the separability condition.

checked whether it satisfies the condition to draw this figure. The light blue area is the area in which the separability condition holds, whereas the deep blue area is the area in which it does not. We can see that the separability condition simulated by the empirical dataset imposes nonlinear (piecewise-linear) constraints on (β_1, β_2).

In general, when we analyze the multinomial choice model, such a condition is always required so that every choice is separable. Similarly, Miyawaki et al. (2013) analyzed the demand model under increasing block rate pricing, which is another multinomial choice model, and explicitly considered the requirement that the choice intervals be separable. In this case, the separability condition is a set of linear constraints on elasticity parameters.

Furthermore, the separability condition is one of the sufficient conditions to make the underlying preference relation strictly convex (see Hurwicz and Uzawa (1971) for the sufficient conditions).

With the likelihood function (16), the data generating process is as follows. First, true model parameters and a dataset ($\{\boldsymbol{x}_{ik}\}_{k=1}^{K_i}$ and $\boldsymbol{z}_i$ for $i = 1,\ldots,n$) are given. We check if these values satisfy the separability condition. Then, given them, the heterogeneity decides the optimal block s_i^*. Finally, the gas demand is generated given this optimal block.[7]

4 Efficient MCMC simulator based on two inequalities

4.1 Prior-Posterior analysis

This article assumes the following proper prior distributions.

$$\begin{aligned} \beta_j \mid \sigma_u^2 &\sim TN_{B_j}\left(\mu_{\beta_j,0}, \sigma_u^2\sigma_{\beta_j,0}^2\right),\ (j=1,2), & \sigma_u^2 &\sim IG\left(\frac{n_{u,0}}{2}, \frac{S_{u,0}}{2}\right), \\ \boldsymbol{\delta} \mid \sigma_v^2 &\sim N_d\left(\boldsymbol{\mu}_{\boldsymbol{\delta},0}, \sigma_v^2\boldsymbol{\Sigma}_{\boldsymbol{\delta},0}\right), & \sigma_v^2 &\sim IG\left(\frac{n_{v,0}}{2}, \frac{S_{v,0}}{2}\right). \end{aligned} \tag{17}$$

Conditional on σ_u^2, β_j follows the truncated normal distribution with mean $\mu_{\beta_j,0}$, variance $\sigma_u^2\sigma_{\beta_j,0}^2$, and support $B_j = [l_j, m_j]$ $(j = 1,2)$. Conditional on σ_v^2, $\boldsymbol{\delta}$ follows the d-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}_{\boldsymbol{\delta},0}$ and covariance matrix $\sigma_v^2\boldsymbol{\Sigma}_{\boldsymbol{\delta},0}$. The parameter σ_j follows the inverse gamma distribution with parameters $n_{j,0}/2$ and $S_{j,0}/2$ $(j = u,v)$. Its mean and variance are $S_{j,0}/(n_{j,0}-2)$ for $n_{j,0} > 2$ and $2S_{j,0}^2/\{(n_{j,0}-2)^2(n_{j,0}-4)\}$ for $n_{j,0} > 4$, respectively. The support of β_j $(j = 1,2)$ reflects our prior knowledge. To elicit the prior distribution, one can make use of knowledge based on demand theory or utilize the estimates obtained from a similar population (see Subsection 6.2).

[7] By using this data generating process, it is straightforward to conduct the simulation study. Given true parameter values, we generate the gas demand. Then, given the generated gas demand as the dataset, the efficient MCMC simulator that will be described in Section 4 is applied to draw samples from the posterior distribution. We conducted the simulation study and found that true parameter values were recovered. However, due to the page limitation, we omit the details.

Let $\pi(\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2)$ be the prior density function of $(\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2)$. Then, it is straightforward to derive the posterior density function, which is given by

$$\begin{aligned}\pi\left(\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2,\boldsymbol{s}^*,\boldsymbol{w}^* \mid \boldsymbol{y}\right) &\propto \pi\left(\boldsymbol{\beta},\boldsymbol{\delta},\sigma_u^2,\sigma_v^2\right) \\ &\times (\sigma_u\sigma_v)^{-n}\exp\left[-\frac{1}{2}\left\{\sigma_u^{-2}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{w}^*)'(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{w}^*)+\sigma_v^{-2}(\boldsymbol{w}^*-\boldsymbol{Z\delta})'(\boldsymbol{w}^*-\boldsymbol{Z\delta})\right\}\right] \\ &\times \prod_{i=1}^{n}\left\{I\left(w_i^*\in R_{is_i^*}\right)\prod_{k=2}^{K_i-1} I(RL_{ik}\le RU_{ik})\right\}, \quad (18)\end{aligned}$$

where $\boldsymbol{y}=(y_1,y_2,\ldots,y_n)'$, $\boldsymbol{X}=(\boldsymbol{x}_{1s_1^*},\ldots,\boldsymbol{x}_{ns_n^*})'$, $\boldsymbol{s}^*=(s_1^*,s_2^*,\ldots,s_n^*)'$, $\boldsymbol{w}^*=(w_1^*,w_2^*,\ldots,w_n^*)'$, and $\boldsymbol{Z}=(\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_n)'$.

To draw samples of model parameters from this posterior density function, we use an efficient Gibbs sampler, the details of which are given in the next subsection and Appendix A.2.

4.2 Sampling β_1 with an efficient blanket

The full conditional distribution of β_1 is the truncated normal distribution, $TN_{C_1}(\mu_{\beta_1,1},\sigma_u^2\sigma_{\beta_1,1}^2)$, where

$$\sigma_{\beta_1,1}^{-2}=\sigma_{\beta_1,0}^{-2}+\sum_{i=1}^{n}\left(p_{is_i^*}\right)^2, \quad (19)$$

$$\mu_{\beta_1,1}=\sigma_{\beta_1,1}^2\left[\sigma_{\beta_1,0}^{-2}\mu_{\beta_1,0}+\sum_{i=1}^{n}p_{is_i^*}\left(y_i-\beta_2 q_{is_i^*}-w_i^*\right)\right], \quad (20)$$

$$C_1=\left[\bigcap_{i=1}^{n}\bigcap_{j=1,j\neq s_i^*}^{K_i}\left\{\beta_1 \mid V_{i,s_i^*}>V_{ij}\right\}\right]\cap\left[\bigcap_{i=1}^{n}\bigcap_{k=2}^{K_i-1}\{\beta_1 \mid RL_{ik}\le RU_{ik}\}\right]\cap[l_1,m_1]. \quad (21)$$

Because C_1 is difficult to calculate, we use rejection sampling. However, as revealed in the next subsection, a simple blanket, the envelope function in rejection sampling, is not efficient in the sense that the acceptance rate of the proposed candidate is extremely low. Therefore, we closely approximate C_1 by $\tilde{C}_1$, which is derived by using two properties of convex functions (the Hermite-Hadamard integral inequality and the power-mean inequality), thus improving our sampling efficiency.

First, without loss of generality, we assume that the support of the prior for β_1 is $B_1 = [l_1, 0]$. Then, we decompose C_1 into a set of larger sets and approximate them to obtain $\tilde{C}_1$. More precisely,

$$C_1 \subset \bigcap_{i=1}^{n} \bigcap_{j=1, j\neq s_i^*}^{K_i} C^{1i}_{s_i^* j} \subset \bigcap_{i=1}^{n} \bigcap_{j=1, j\neq s_i^*}^{K_i} \tilde{C}^{1i}_{s_i^* j} \equiv \tilde{C}_1, \tag{22}$$

where $C^{1i}_{kj} = \{\beta_1 \mid V_{ik} > V_{ij}\} \cap [l_1, 0]$. Third, we construct the interval $\tilde{C}^{1i}_{kj}$ ($\supset C^{1i}_{kj}$) using the following three steps.

Step 1. Apply the Hermite-Hadamard integral inequality. The Hermite-Hadamard integral inequality[8]and $\beta_1 \in [l_1, 0]$ imply

$$\int_{P_{ij}}^{P_{ik}} x^{\beta_1} dx \geq \begin{cases} \left(P_{ik} - P_{ij}\right)\left(\frac{P_{ik}+P_{ij}}{2}\right)^{\beta_1}, & \text{if } k < j, \\ \left(P_{ik} - P_{ij}\right)\frac{P_{ik}^{\beta_1}+P_{ij}^{\beta_1}}{2}, & \text{if } k > j. \end{cases} \tag{23}$$

Using this inequality, we have

$$V_{ik} > V_{ij} \Longleftrightarrow a_1 > \int_{P_{ij}}^{P_{ik}} x^{\beta_1} dx \Longrightarrow a_1 > \text{(the right hand side of equation (23))}, \tag{24}$$

where $a_1 = \exp(-w_i^*)(1-\beta_2)^{-1}(Q_{ik}^{1-\beta_2} - Q_{ij}^{1-\beta_2})$.

Step 2. Apply the power-mean inequality. The power-mean inequality and $\beta_1 \in [l_1, 0]$ imply

$$\left(\frac{P_{ik}^{l_1} + P_{ij}^{l_1}}{2}\right)^{1/l_1} < \left(\frac{P_{ik}^{\beta_1} + P_{ij}^{\beta_1}}{2}\right)^{1/\beta_1} \Longleftrightarrow \frac{P_{ik}^{\beta_1} + P_{ij}^{\beta_1}}{2} < \left(\frac{P_{ik}^{l_1} + P_{ij}^{l_1}}{2}\right)^{\beta_1/l_1}.^{9} \tag{26}$$

[8]Let $f : [a,b] \to \mathbb{R}$ be a convex function. Then,

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a}\int_a^b f(x)\,dx \leq \frac{f(a)+f(b)}{2}. \tag{25}$$

See, for example, Niculescu and Persson (2003) for a proof. Niculescu and Persson (2003) also noted that the first (or last) inequality can define the convex function itself.

Step 3. Combine the above two-step results. By combining equations (24) and (26), and by rearranging these inequalities for β_1, we derive the closely approximated interval $\tilde{C}_{kj}^{1i} = \tilde{C}_{kj}^{\star 1i} \cap [l_1, 0]$, where

$$\tilde{C}_{kj}^{\star 1i} = \begin{cases} (-\infty, b_1/\bar{p}(1)), & \text{if } k < j \text{ and } \bar{p}(1) > 0, \\ (-\infty, \infty), & \text{if } k < j \text{ and } \bar{p}(1) = 0, \\ (b_1/\bar{p}(1), \infty), & \text{if } k < j \text{ and } \bar{p}(1) < 0, \\ (b_1/\bar{p}(l_1), \infty), & \text{if } k > j \text{ and } \bar{p}(l_1) > 0, \\ (-\infty, \infty), & \text{if } k > j \text{ and } \bar{p}(l_1) = 0, \\ (-\infty, b_1/\bar{p}(l_1),), & \text{if } k > j \text{ and } \bar{p}(l_1) < 0, \end{cases} \tag{27}$$

$b_1 = \log(a_1/(P_{ik} - P_{ij}))$[10], and $\bar{p}(x) = x^{-1} \log\{(P_{ik}^x + P_{ij}^x)/2\}$ $(x = 1, l_1)$. By construction, $C_{kj}^{1i} \subset \tilde{C}_{kj}^{1i}$. If $P_{iK_i} > 1$ is assumed, we have $\bar{p}(1) > \bar{p}(l_1) > 0$, which simplifies the above expression.

Finally, by using this interval $\tilde{C}_{kj}^{1i}$, we approximate C_1 by $\tilde{C}_1 = \cap_{i=1}^{n} \cap_{j=1, j \neq s_i^*}^{K_i} \tilde{C}_{s_i^* j}^{1i}$ as mentioned above. Figure 5 illustrates the relationships among C_1, $\tilde{C}_1$, and B_1.

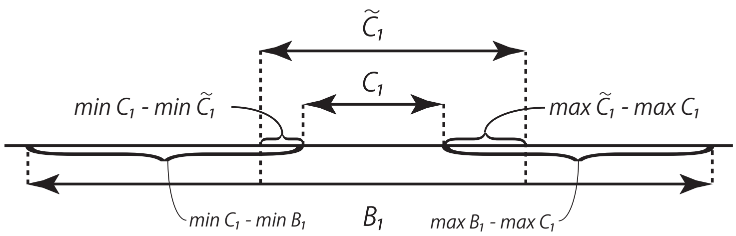


Figure 5: Relationships among C_1, $\tilde{C}_1$, and B_1.

With $\tilde{C}_1$, the sampling procedure for β_1 is implemented using the following two steps.

Step a. Generate β_1' from the uniform distribution on $\tilde{C}_1$ until it is in C_1.

[9]See, for example, Chapter 2 of Hardy, Littlewood, and Pólya (1952) for a proof of the power-mean inequality. This equivalence also uses the fact that $f(x) = x^{\beta_1}$ ($\beta_1 \in [l_1, 0]$) is decreasing as $x (> 0)$ increases.

[10]Because $a_1 \gtrless 0$ for all $k \lessgtr j$, $a_1/(P_{ik} - P_{ij}) > 0$ for any k and j ($k \neq j$).

Step b. Accept β_1' with the acceptance probability $\alpha(\beta_1,\beta_1')$; otherwise, retain β_1, where

$$\alpha\left(\beta_1,\beta_1'\right) = \min\left[1, \frac{\phi\left\{\left(\beta_1' - \mu_{\beta_1,1}\right)\sigma_u^{-1}\sigma_{\beta_1,1}^{-1}\right\}}{\phi\left\{\left(\beta_1 - \mu_{\beta_1,1}\right)\sigma_u^{-1}\sigma_{\beta_1,1}^{-1}\right\}}\right], \tag{28}$$

and $\phi(\cdot)$ is the probability density function of the standard normal distribution.

The sampling of β_2 is conducted in a similar manner. See Appendix A.2 for its full conditional distribution and Appendix A.3 for the derivation of its efficient blanket.

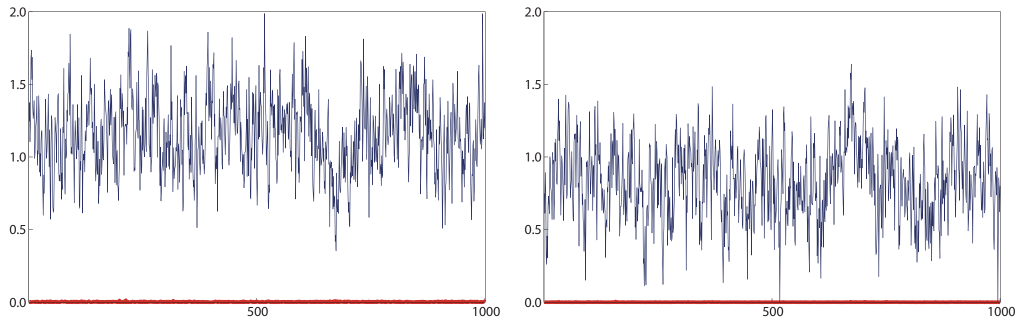
Joint sampling for (β_1,β_2) is an alternative sampling algorithm. The GHK simulator (proposed by Geweke (1991), Hajivassiliou and McFadden (1998), and Keane (1994)) is a method to draw samples from the truncated multivariate normal distribution. While using this simulator could improve the sampling efficiency, the GHK simulator has disadvantages. The support of the conditional posterior distribution for (β_1,β_2) is difficult to calculate because of the highly nonlinear indirect utility. Furthermore, its efficient two-dimensional blanket is also difficult to construct. One of the simplest blankets is $B_1 \times B_2$, which is the support of the joint prior distribution of (β_1,β_2). As we see in Figure 6 and Table 1 given in the next subsection, however, this blanket is extremely inefficient with respect to the empirical dataset.

4.3 Adequacy of the efficient blankets

In this subsection, we evaluate the adequacy of the efficient blanket in two respects by using the Japanese residential gas demand data. The first measure is the absolute differences, $\max \tilde{C}_j - \max C_j$ and $\min C_j - \min \tilde{C}_j$ $(j = 1,2)$, and the second measure is the adequacy ratio, $|C_j|/|\tilde{C}_j|$ $(j = 1,2)$, where $|A|$ is the area of the set A. Figure 5 is helpful in that it clarifies what these measures mean.

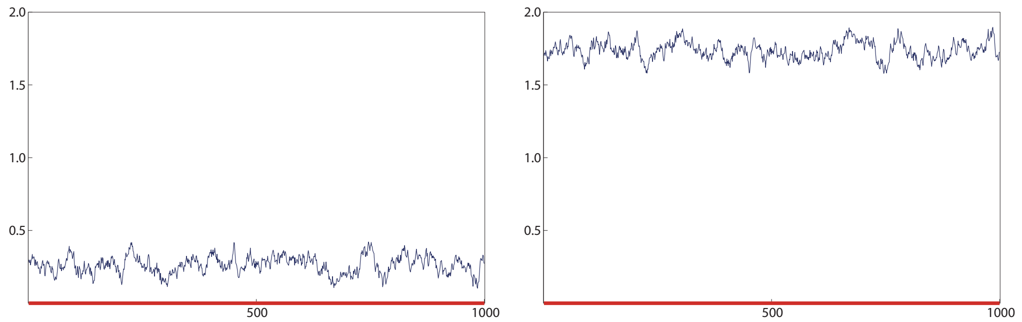
Because C_j is difficult to calculate, we obtain these measures via simulation. During each step in the MCMC iterations (see Appendix A.2), we obtain the approximated interval, $\tilde{C}_j$. Then, we compute 1,001 equispaced samples in this approximated interval and determine whether they belong to C_j. Among the samples that are in C_j, we obtain the maximum and the minimum to calculate the absolute differences. Furthermore, the ratio of the number of samples that belong to C_j to the number of those that do not is the adequacy ratio conditional on model parameters. These conditional adequacy ratios are averaged to calculate the adequacy ratio after the MCMC iterations are complete.

We calculate these two measures using the empirical dataset. The results are shown in Figure 6 and given in Table 1. Figure 6 presents time series plots of absolute differences. The



(a) $\min C_1 - \min \tilde{C}_1$ (red) and $\min C_1 - \min B_1$ (blue).

(b) $\max \tilde{C}_1 - \max C_1$ (red) and $\max B_1 - \max C_1$ (blue).



(c) $\min C_2 - \min \tilde{C}_2$ (red) and $\min C_2 - \min B_2$ (blue).

(d) $\max \tilde{C}_2 - \max C_2$ (red) and $\max B_2 - \max C_2$ (blue).

Figure 6: Absolute differences.

Table 1: Adequacy ratios

Coefficient	$\lvert C_j\rvert/\lvert\tilde{C}_j\rvert = r_1$	$\lvert C_j\rvert/\lvert B_j\rvert = r_2$	*Efficiency ratio* (r_1/r_2)
β_1	.67 (.21)	.0037 (.0026)	181
β_2	1.00 (.00)	.0004 (.0003)	2,500

* Standard deviations in parentheses.

red lines represent time series plots of absolute differences that calculated from our efficient blankets, whereas the blue lines are those obtained using the simple method, where $\tilde{C}_j$ is replaced by B_j. The red lines are very close to the horizontal lines at zero, which implies that the proposed efficient blankets are sufficiently close to the true sets. Table 1 indicates the adequacy ratios in the first two columns and the efficiency ratio, the ratio of two adequacy ratios, in the third column. Although the adequacy ratios of the efficient blankets differ with respect to their parameters, they are much (about 200 to 2,500 times) higher than those of the simple blanket B_j. Therefore, based on the empirical dataset, our proposed method well approximates the true regions for both β_1 and β_2.

We also investigate how our method would be affected by the number of blocks. In the empirical dataset, there are 65 and 245 consumers under three-block and six-block decreasing block rate pricing, respectively (see also Panel 7(a) in Subsection 6.1). For these subsets of the empirical data, we calculate adequacy ratios $|C_j|/|\tilde{C}_j|$ for $j = 1,2$. The results are given in Table 2. The degrees of approximation decreases on average as the number of blocks

Table 2: Adequacy ratios as the number of blocks increases

Coefficient	*three blocks*	*six blocks*
β_1	.93 (.066)	.82 (.18)
β_2	1.00 (.0005)	1.00 (.0006)

* Standard deviations in parentheses.

increases in terms of the adequacy ratio but they are similar when we take into account their standard deviations.

5 Model without heterogeneity

As an alternative model, this section introduces the statistical model without heterogeneity, which is given by

$$y_i = \boldsymbol{x}'_{is_i^*}\boldsymbol{\beta} + u_i, \quad u_i \sim \text{i.i.d. } N\left(0, \sigma_u^2\right), \tag{29}$$

where $\boldsymbol{x}_{is_i^*} = (p_{is_i^*}, q_{is_i^*})'$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$,

$$s_i^* = k, \quad \text{if } V_{ik} \geq V_{ij} \text{ for } k \neq j \text{ and } k = 1, \ldots, K_i, \tag{30}$$

$$V_{ik} = -\frac{P_{ik}^{1+\beta_1}}{1+\beta_1} + \frac{Q_{ik}^{1-\beta_2}}{1-\beta_2}. \tag{31}$$

Then, the likelihood function for the i-th consumer is given by

$$f\left(y_i \mid \boldsymbol{\beta}, \sigma_u^2\right) = \frac{1}{\sqrt{2\pi}\sigma_u} \sum_{k=1}^{K_i} \exp\left\{-\frac{1}{2\sigma_u^2}\left(y_i - \boldsymbol{x}_{ik}'\boldsymbol{\beta}\right)^2\right\} \prod_{j \neq k} I\left(V_{ik} \geq V_{ij}\right). \tag{32}$$

With the same prior distributions as before (see equation (17)), the posterior distribution is derived as

$$\pi\left(\boldsymbol{\beta}, \sigma_u^2 \mid \boldsymbol{y}\right) \propto \pi\left(\boldsymbol{\beta}, \sigma_2^2\right) \sigma_u^{-n} \prod_{i=1}^{n} \sum_{k=1}^{K_i} \exp\left\{-\frac{1}{\sigma_u^2}\left(y_i - \boldsymbol{x}_{ik}'\boldsymbol{\beta}\right)^2\right\} \prod_{j \neq k} I\left(V_{ik} \geq V_{ij}\right), \tag{33}$$

where $\pi(\boldsymbol{\beta}, \sigma_2^2)$ is the prior probability density function associated with the prior distributions. We apply the Metropolis-Hastings within Gibbs algorithm to draw samples from the posterior distribution. See Appendix A.4 for its details.

6 Empirical analysis and policy evaluation of residential gas demand

6.1 Data description

This subsection describes the data to be used for the empirical study in the following two subsections. We conducted an online survey on the Internet from June 2006 to May 2008 that was designed to analyze the water and energy consumption and the garbage emission behavior of Japanese households. The population of this survey was comprised of the households living in the Tokyo and Chiba prefectures. There were about 8.4 million households as of January 2007. Among them, 47,239 individuals were registered to the survey company, INTAGE Inc. (http://www.intage.co.jp/english/). Out of 47,239 individuals, 1,687 individuals were randomly selected. Then, out of 1,687 individuals, 1,250 participated in our survey. They were asked for household attributes such as annual income, the number of members in the household, and so on in June 2006 and April 2007. They were also asked to record their water and energy consumptions and the garbage emission behavior every month.

For the empirical study, we used the attribute data in June 2006 and the gas consumption data in January 2007. The dependent variable is the amount of gas consumption ($\log \mathrm{m}^3$), which was calculated from the bill by using the corresponding gas price schedule that depends on the area in which the individuals were living. The list of independent variables and their corresponding coefficients is given in Table 3.

Table 3: Independent variables used in the gas demand function

Coefficient	*Variable*	*Attribute*
β_1	$(p_{i1},\ldots,p_{iK_i})$	log of monthly unit prices of gas (log ¥50 /m^3)
β_2	$(q_{i1},\ldots,q_{iK_i})$	log of monthly virtual incomes (log ¥50)
δ_1	z_{i1}	the constant
δ_2	z_{i2}	the number of members in a household (person)
δ_3	z_{i3}	the number of rooms in a home/apartment (room)
δ_4	z_{i4}	the total floor space of a home/apartment (50m^2)

The number of households is decreased from 1,250 to 473 for the reasons listed below.

- Dropped out of the survey before January 2007.
- Missing or incorrect data concerning household attributes or gas consumption.
- Use of liquefied petroleum gas. (Its price schedule is not publicly available.)

The sample selection problem will be examined at the end of this subsection.

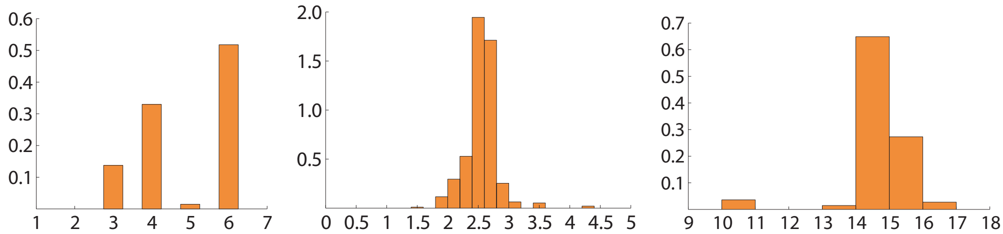
For these 473 households, we conducted an empirical study that is presented in the next subsection. The first row of Table 4 gives the summary statistics of the amount of gas consumption, which is the dependent variable. All these households faced decreasing block

Table 4: Summary statistics of the data used for the empirical study (the number of households is 473)

Variable	*Unit*	*Mean*	*SD*	*Min.*	*1st quartile*	*3rd quartile*	*Max.*
y_i	$\log \text{m}^3$	3.75	.78	.053	3.36	3.85	5.70
$\log I_i$	log ¥50	9.22	.56	7.42	9.03	9.61	10.82
z_{i2}	person	2.81	1.28	1	2	4	9
z_{i3}	room	4.09	1.10	1	4	5	8
z_{i4}	50m^2	1.54	.74	.20	1.10	1.80	8.00

* $Corr(z_{i1}, z_{i2}) = .49$, $Corr(z_{i1}, z_{i3}) = .38$, $Corr(z_{i2}, z_{i3}) = .71$.

rate pricing, and their price schedules differed depending on the cities in which they live. The price structures are shown in Figure 7, wherein the relative frequency of the number of blocks, the histogram of the unit price where the gas was actually consumed, and the histogram of the fixed gas service fee are illustrated.



(a) Number of blocks (number). (b) Unit price (¥50/m^3). (c) Fixed gas service fee (¥50).

Figure 7: Relative frequency of the number of blocks and histograms of the unit price and the fixed gas service fee in January 2007.

Because the exact annual income level is sensitive information to request, our survey divides annual income levels into eight categories: (in million yen) 0-2, 2-4, 4-6, 6-8, 8-10, 10-12, 12-15, and over 15. Then, we asked the household its income category. The monthly income variable to be used for the empirical study is estimated using the median of the recorded income category divided by 12. For the last category (over 15 million yen), the approximate annual income is also recorded, and we use this figure divided by 12 as the monthly income. The second row of in Table 4 summarizes the log of this income variable.

The summary statistics for the explanatory variables for heterogeneity are given in the third to the fifth rows of Table 4. At the bottom of this table, the correlation coefficients among the explanatory variables for heterogeneity are calculated. From these values, we can establish that there is a high positive correlation between the number of rooms and the total floor space, such that either of these variables could not explain the residential gas demand.

At the end of this subsection, we discuss the sample selection bias caused by our data reduction. Table 5 gives summary statistics based on the original data. In the third column, the numbers of households to calculate these statistics are also given. For example, there are 564 households who did not drop out our survey as of January 2007 and answered the questions about gas consumption properly. These statistics are mostly similar to those given

Table 5: Summary statistics of the original data

Variable	*Unit*	*the number of households*	*Mean*	*SD*	*1st quartile*	*3rd quartile*
y_i	$\log \mathrm{m}^3$	564	3.83	.93	3.40	4.30
$\log I_i$	log ¥50	1,103	9.17	.56	9.03	9.62
z_{i2}	person	1,230	2.87	1.36	2	4
z_{i3}	room	1,230	4.13	1.29	3	5
z_{i4}	$50\mathrm{m}^2$	1,230	1.56	.91	1.00	1.90

* $Corr(z_{i1}, z_{i2}) = .52$, $Corr(z_{i1}, z_{i3}) = .38$, $Corr(z_{i2}, z_{i3}) = .67$.

in Table 4. Therefore, it is natural to assume that the sample selection bias is small.

6.2 Residential gas demand function

The following two subsections are based on the model with heterogeneity. First, we assume the following prior distributions.

$$
\begin{aligned}
&\beta_1 \mid \sigma_u^2 \sim TN_{[-2,0]}\left(0, 100\sigma_u^2\right), \quad \sigma_u^2 \sim IG(0.01, 0.01), \\
&\beta_2 \mid \sigma_u^2 \sim TN_{[0,2]}\left(0, 100\sigma_u^2\right), \quad \sigma_v^2 \sim IG(0.01, 0.01), \\
&\boldsymbol{\delta} \mid \sigma_v^2 \sim N_4\left(\mathbf{0}, 100\sigma_v^2 \boldsymbol{I}\right),
\end{aligned} \tag{34}
$$

where $\boldsymbol{I}$ is the identity matrix. The truncation interval for β_j ($j = 1, 2$) is elicited as follows.[11] Because residential gas is one of the necessities for households, its demand is relatively inelastic with respect to price and income. Thus, we can expect the absolute values of β_1 and β_2 to be less than one. Furthermore, we assume negative price elasticity according to microeconomic demand theory (see, e.g., Mas-Colell et al. (1995)), and positive income elasticity according to the estimate taken from the Family Income and Expenditure Survey (FIES) conducted in 2008. The FIES survey is intended to analyze the Japanese households and estimated the expenditure elasticity for gas to be 0.29 (for households with more than two members) and away from zero at a 5% significant level. Thus, we assume the interval $[-2, 0]$ ($[0, 2]$) for β_1 (β_2), where -1 (1) is included to examine whether β_1 (β_2) is less than -1 (more than 1). Further analysis of our empirical dataset reveals that this prior truncation area for $\boldsymbol{\beta}$ is included in the area in which the separability condition is satisfied (see Figure 4).

Specified prior distributions are also evaluated by the method proposed by Chib and Ergashev (2009). For each random draw from the prior distributions, the optimal block s_i^* is computed. After 1000 draws, relative frequencies for each block are calculated, which are shown in Panel 8(a) for selected consumers. From this panel, specified prior distributions place relatively large weights on the first block.

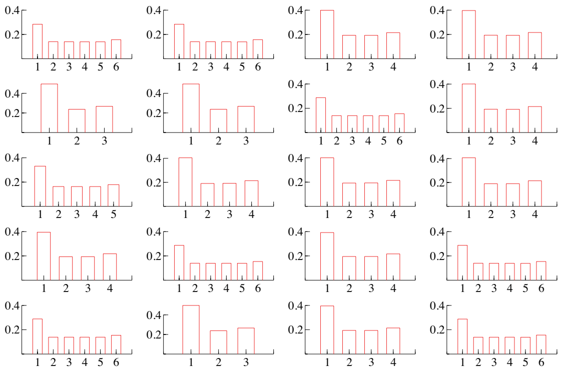
If we replace the conditional prior distribution for $\boldsymbol{\delta}$ by

$$TN_D\left(\boldsymbol{\mu}_{\boldsymbol{\delta},0}, \sigma_v^2 \boldsymbol{\Sigma}_{\boldsymbol{\delta},0}\right), \tag{35}$$

where $D = [0, \infty)^d$, the implied probability distributions become flatter than those with orig-

[11] We also conducted the analysis without the prior truncations for both β_j ($j = 1, 2$). Posterior means and standard deviations for elasticity parameters are -0.83 (0.28) for the price elasticity (β_1) and 0.27 (0.046) for the income elasticity (β_2). Their 95% credible intervals are $[-1.38, -0.28]$ for β_1 and $[0.17, 0.35]$ for β_2. Thus, price and income elasticities are highly credible to be negative and positive, respectively, in the sense that their 95% credible interval does not include zero. Furthermore, income elasticity is highly credible to be less than one. Other obtained results are very similar to those obtained with priors (34) specified above, which are omitted due to the page limitation. Thus, we conclude that the results given below are not sensitive to the prior truncation.

inal priors (see Panel 8(b)).


(a) Original

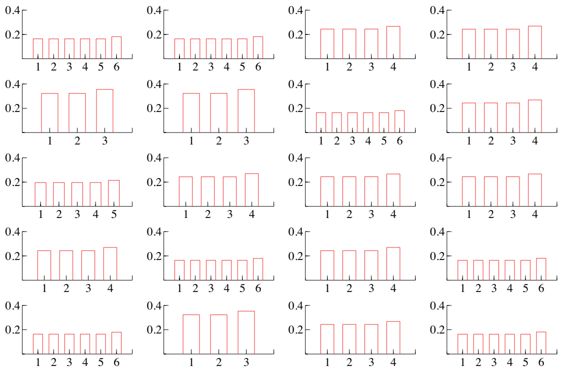

(b) Truncated

Figure 8: Implied probability distributions of the optimal block for selected consumers.

In terms of the implied probability distributions of the optimal block, it seems that the truncated prior for $\boldsymbol{\delta}$ would be plausible. However, such a prior is too restrictive because it excludes the possibility that some explanatory variables for heterogeneity have negative relation to the gas demand.[12] Thus, this article uses the prior distributions specified in equation (34).

With prior distributions specified in equation (34), the MCMC simulation was carried out to obtain 6×10^6 samples after deleting the first 6×10^5 samples. We reduced the obtained 6×10^6 samples to 2×10^4 samples by picking up every 300-th sample. The results are given in Table 6 and shown in Figure 9.[13]

Each column of the table represents the parameter names, the posterior means, the posterior standard deviations, the 95% credible intervals, and the estimated inefficiency factors. The inefficiency factor is defined as $1+2\sum_{j=1}^{\infty}\rho(j)$, where $\rho(j)$ is the sample autocorrelation

[12]We also conducted the Bayesian inference with the truncated prior for $\boldsymbol{\delta}$, and obtained results are mostly similar with those found in Table 6. However, the prior truncation affects the marginal posterior distribution for δ_4. The posterior mean for δ_4 is 0.086 with its 95% credible interval $[0.016, 0.17]$.

[13]Because our data are reduced from 1,250 to 473, such a large data reduction would influence the obtained results. To examine the effect from this reduction, we gathered households whose dependent variables are missing but whose explanatory variables are not. The number of households then became 759. Under the same MCMC setting, we estimated the residential gas demand function. Missing dependent variables were imputed within the MCMC simulation by using the data augmentation method. Obtained results are quite similar to those given in Table 6 and shown in Figure 9, and we omit the details. Thus, the data reduction in explanatory variables does not influence the estimates of model parameters.

Table 6: Gas demand function

Parameter	*Mean*	*SD*	*95% interval*		*INEF*
β_1 (price)	−.84	.26	[−1.35	−.32]	136
β_2 (income)	.26	.060	[.14	.38]	218
δ_1 (constant)	.84	.62	[− .32	2.06]	259
δ_2 (number of members)	.17	.026	[.12	.22]	11
δ_3 (number of rooms)	.18	.037	[.11	.25]	5
δ_4 (total floor space)	.038	.052	[− .067	.14]	6
σ_u (measurement error)	.55	.13	[.12	.65]	19
σ_v (heterogeneity error)	.17	.15	[.049	.58]	30

* "SD" and "INEF" denote the posterior standard deviation and the estimated inefficiency factor, respectively.

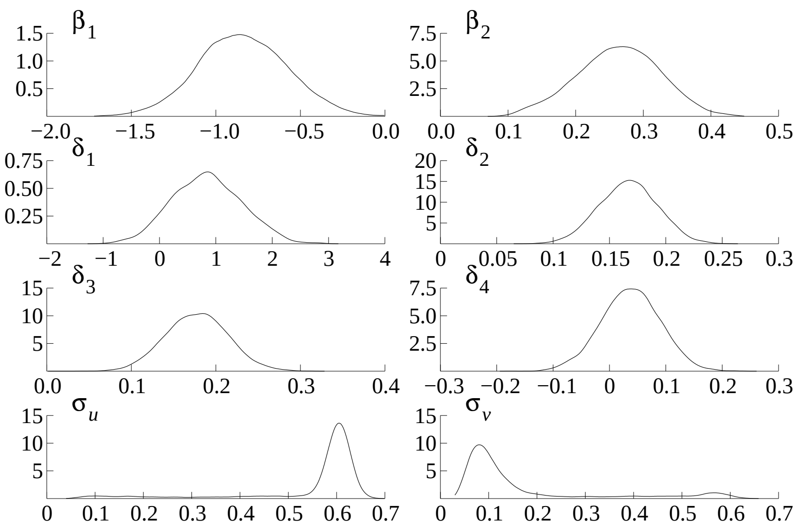


Figure 9: Marginal posterior densities.

at lag j, and is estimated using the spectral density. It can be interpreted as the ratio of the variance of the sample mean obtained by the MCMC draws to the variance of the sample mean by an uncorrelated Monte Carlo draw (see, e.g., Chib (2001)).

6.2.1 Estimates of price and income elasticities

From Table 6, we found that marginal posterior probabilities of price and income elasticities are $\Pr(\beta_1 < -0.32 \mid \boldsymbol{y}) > 0.975$ and $\Pr(\beta_2 > 0.14 \mid \boldsymbol{y}) > 0.975$. The estimated inefficiency factors of elasticity parameters (as well as that of δ_1) are much higher than other parameters. This is partly because of the tight restrictions on $\boldsymbol{\beta}$ and partly because of the high correlation between β_2 and δ_1 ($Corr(\beta_2, \delta_1) = -0.82$). The other correlation coefficients are less than 0.7 in their absolute values except for that between σ_u and σ_v ($Corr(\sigma_u, \sigma_v) = -0.93$). This high correlation between σ_u and σ_v is mainly because they are not fully identified.

We compared these estimates with those of previous studies. One of the classical studies of residential gas demand is the study by Balestra and Nerlove (1966). They analyzed the natural gas demand using a dynamic model with random effects. Their data are the state-level panel data for the United States during 1950 – 62. They estimated the (long-run) price and income elasticities to be −0.63 and 0.62, respectively, when the depreciation rate for gas appliances is unconstrained. While the estimated income elasticity calculated by these researchers using aggregate data is larger than ours, the estimated price elasticity is similar to ours.

Bloch (1980) also investigated residential gas demand by using the household-level data. This includes gas usage data for households living in Twin Rivers, New Jersey, during the winter months (November through April) from 1971 to 1976. The explanatory variables that Bloch (1980) used are the number of heating degree days, the price of natural gas, and the consumer price index. He found that the (long-run) price elasticity is estimated to be −0.596 or −0.224 depending on the functional form of the demand function. The former estimate is similar to our results.

6.2.2 Other parameters

Among the explanatory variables for heterogeneity, the number of members in a household and the number of rooms in a home are highly credible to be positive in terms of their 95% credible intervals. These factors should have a positive relationship with gas demand through water demand for the two following reasons: (1) these two variables are also credible to be positive in the Japanese residential water demand function (see Table 5 of Miyawaki et al.

(2013)); and (2) in Japan, residential gas is mainly used for boiling water.

6.3 Policy evaluation—the effect of price schedule changes

In this subsection, we conduct a welfare analysis and evaluate the effect of price schedule changes. As the suppositional price schedules, we use the following three uniform price systems, which differ in their unit price: (unit price, fixed service fee) = (¥50/m^3, ¥725), (¥120/m^3, ¥725), and (¥250/m^3, ¥725). These unit prices are less expensive, as high as, or more expensive than the unit price that most households are actually facing. The fixed service fee is set close to the actual fee for most households.

Figure 10 shows the effect of price changes on households in terms of compensating variation. Each boxplot is the predictive distribution of the compensating variation in one thousand yen for each household. The number of households is reduced to 90 by selecting every 5-th household. Boxplots are sorted in ascending order based on the number of members in a household.

These results are consistent with what we expect based on microeconomic theory. We observe the positive (negative) compensating variation when the unit price decreases (increases). That is, the unit price decrease (increase) implies welfare improvement (decline). However, uniform pricing itself does not seem to have a noticeable influence on compensating variation (see the panel of ¥120/m^3). Furthermore, the degree of improvement (decline) is affected by explanatory variables for heterogeneity. The above panels show that the more members there are in a household, the more the compensating variation is likely to change. Similar patterns are also found with other explanatory variables for heterogeneity.

6.4 Comparison with the model without heterogeneity

Table 7 reports estimation results of the model without heterogeneity. For these results, we specify prior distributions that are the same as ones for the model with heterogeneity. Then, 6×10^6MCMC samples are generated after discarding the first 6×10^5, and they are

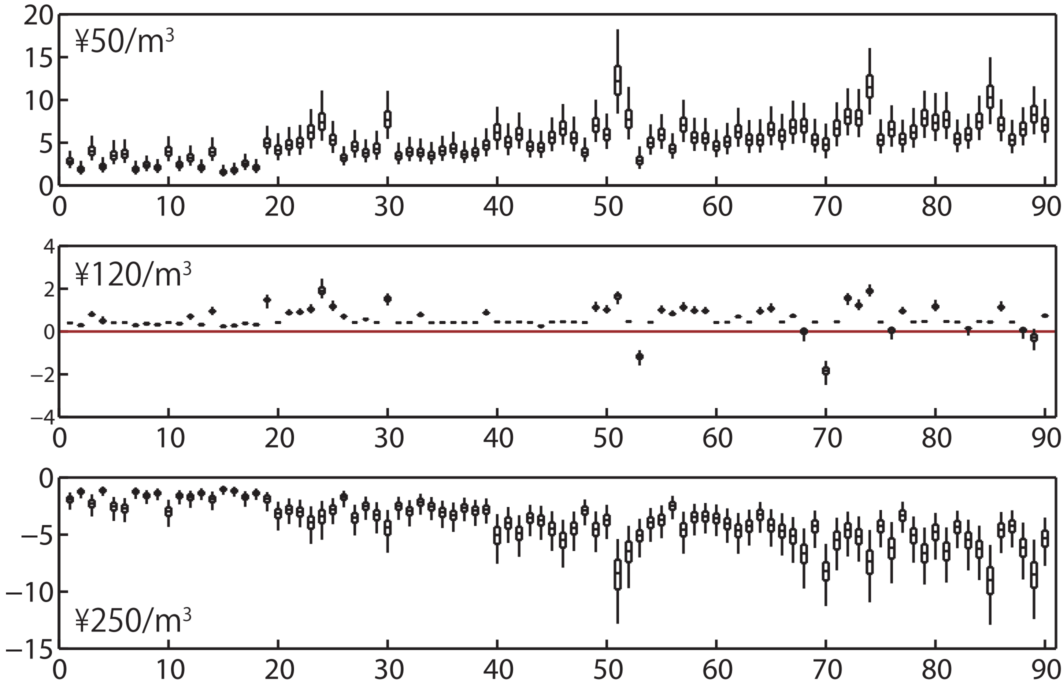


Figure 10: Boxplots of the predictive distribution of the compensating variation (¥10^3). Each box represents the range between the first and third quartiles. The upper and lower whiskers denote the 95-th and 5-th percentiles, respectively.

reduced to 2×10^4 samples by picking up every 300-th sample. Compared with the model

Table 7: Gas demand function without heterogeneity

Parameter	*Mean*	*SD*	95% *interval*	*INEF*
β_1 (price)	−.96	.39	[−1.65 −.14]	1
β_2 (income)	.50	.040	[.42 .58]	1
σ_u (measurement error)	.71	.023	[.67 .76]	1

* "SD" and "INEF" denote the posterior standard deviation and the estimated inefficiency factor, respectively.

with heterogeneity, the price elasticity is similar, while the income elasticity and the standard deviation for the measurement error are not.

We calculate the log of marginal likelihoods for both models to compare their fits to the data. For this purpose, the method proposed by Chib and Jeliazkov (2005) is applied (see also Chib (1995) and Chib and Jeliazkov (2001) for the methods using the so-called basic marginal likelihood identity). The log of the marginal likelihood for the model with heterogeneity is approximately −9.48 with the numerical standard error 1.20, while −518.43 with 0.005 for the model without heterogeneity.

The difference mostly comes from the evaluation of the log of likelihood functions. Thus, explanatory variables included in the models cause the difference. The model without heterogeneity explains the log of gas demand by the log of prices and virtual incomes, while the model with heterogeneity uses variables for heterogeneity as well as the log of prices and virtual incomes. Therefore, the latter fits well to the data because it includes more relevant variables to explain the log of gas demand.

7 Concluding remarks

There are many previous studies that have used the discrete/continuous choice approach in the analysis of household behaviors under block rate pricing, transportation, housing, labor supply, etc. It should be noted that the indirect utility function becomes highly nonlinear, when the budget set is nonconvex, such as in the case of decreasing block rate pricing. However, previous studies (Burtless and Hausman (1978); Hausman (1980); Burtless and Moffitt (1985)) on decreasing block rate pricing do not address this problem. Blomquist and Newey (2002) proposed a nonparametric approach to address this problem, but their approach lacks the microeconomic theoretical background. This article proposes a new Bayesian estimation method for residential gas demand on the nonconvex budget set by extending the Bayesian approach taken by Miyawaki et al. (2013), which proposed a Bayesian estimation method to analyze consumer demand under increasing block rate pricing. The advantage of our method is not only that it addresses the nonlinearity problem associated with the nonconvex budget sets but also that it incorporates the (highly nonlinear) separability condition that is necessary for the demand model under multiple-block block decreasing block rate pricing and introduces normal disturbance into the multinomial choice model.

Finally, our method has the potential to estimate the multiple residential energy expenditure function. Previous studies have focused on the cross-elasticity of electricity and gas demand (see Beierlein, Dunn, and James C. McConnon (1981); Baker, Blundell, and Micklewright (1989); Lee and Singh (1994); Maddala, Trost, Li, and Joutz (1997); Vaage (2000); Mansur, Mendelsohn, and Morrison (2008)). However, they do not take into consideration the price structure of electricity and gas services. Japanese electricity services are provided under increasing block rate pricing where the unit price increases as the volume consumed increases. Thus, by combining the proposed method and the method of Miyawaki et al. (2013) to estimate the demand function under increasing block rate pricing, we could also construct a multivariate demand function under both increasing and decreasing block rate pricing in a natural manner and estimate the residential energy demand function using the Bayesian approach. We will leave this for our future work.

A Appendices

A.1 Heterogeneity interval

We derive the explicit bounds of the heterogeneity interval, which is given by

$$R_{ik} = \left\{w_i^* \mid V_{ik} > V_{ij} \text{ for } j \neq k\right\} = \bigcap_{j \neq k} \left\{w_i^* \mid V_{ik} > V_{ij}\right\}. \tag{36}$$

Let $D(x_1, x_0; \theta) = \theta^{-1}(x_1^\theta - x_0^\theta)$ $(x_0 > 0,\ x_1 > 0,\ \theta \neq 0)$. Then, $D(x_1, x_0; \theta) \gtreqless 0$ if $x_1 \gtreqless x_0$.[14]

With this function, we solve $V_{ik} > V_{ij}$ for w_i^*.

$$V_{ik} > V_{ij} \Longleftrightarrow -\exp\left(w_i^*\right) D\left(P_{ik}, P_{ij}; 1+\beta_1\right) > -D\left(Q_{ik}, Q_{ij}; 1-\beta_2\right) \tag{37}$$

[14] Suppose $x_1 > x_0 > 0$. Then, because x_l^θ $(l = 0, 1)$ is decreasing (increasing) with respect to x_l if $\theta < (>)0$, the numerator $x_1^\theta - x_0^\theta \lesseqgtr 0$ if $\theta \lesseqgtr 0$. Therefore, $D(x_1, x_0; \theta) > 0$ if $x_1 > x_0 > 0$. Similarly, $D(x_1, x_0; \theta) < 0$ if $x_0 > x_1 > 0$.

$$\Longleftrightarrow \begin{cases} w_i^* < \ln E_{kj}, & \text{if } k < j, \\ w_i^* > \ln E_{kj}, & \text{if } k > j, \end{cases} \tag{38}$$

where $E_{kj} = D(Q_{ik}, Q_{ij}; 1-\beta_2)/D(P_{ik}, P_{ij}; 1+\beta_1)$. The last equivalence makes use of the property of decreasing block rate pricing: $P_{ik} \gtrless P_{ij}$ and $Q_{ik} \gtrless Q_{ij}$ if $k \lessgtr j$. Both $P_{ik} > 0$ and $Q_{ik} > 0$ for all k because we assume the log-linear demand specification. Thus, $D(P_{ik}, P_{ij}; 1+\beta_1) \gtrless 0$ and $D(Q_{ik}, Q_{ij}; 1-\beta_2) \gtrless 0$ if $k \lessgtr j$.

Finally, we have

$$\begin{aligned} R_{i1} &= \left(-\infty, \min_{1<j} \ln E_{1j}\right), \\ R_{ik} &= \left(\max_{k>j} \ln E_{kj}, \min_{k<j} \ln E_{kj}\right), \quad k = 2, \ldots, K_i - 1, \\ R_{iK_i} &= \left(\max_{K_i>j} \ln E_{K_i j}, \infty\right). \end{aligned} \tag{39}$$

We note that $R_{ik} \cap R_{ij} = \emptyset$ $(k \neq j)$.

A.2 Gibbs sampler

The Gibbs sampler is implemented in seven steps.

Step 1. Set initial values to $(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{s}^*, \boldsymbol{w}^*, \sigma_u^2, \sigma_v^2)$.

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

See Subsection 4.2.

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

The full conditional distribution of β_2 is the truncated normal distribution, $TN_{C_2}(\mu_{\beta_2,2}, \sigma_u^2 \sigma_{\beta_2,2}^2)$, where

$$\sigma_{\beta_2,1}^{-2} = \sigma_{\beta_2,0}^{-2} + \sum_{i=1}^{n} \left(q_{ir_i^*}\right)^2, \tag{40}$$

$$\mu_{\beta_2,1} = \sigma_{\beta_2,1}^2 \left[\sigma_{\beta_2,0}^{-2} \mu_{\beta_2,0} + \sum_{i=1}^{n} q_{is_i^*} \left(y_i - \beta_1 p_{is_i^*} - w_i^*\right)\right], \tag{41}$$

$$C_2 = \left[\bigcap_{i=1}^{n} \bigcap_{j=1, j \neq s_i^*}^{K_i} \left\{ \beta_2 \mid V_{i,s_i^*} > V_{ij} \right\} \right] \cap \left[\bigcap_{i=1}^{n} \bigcap_{k=2}^{K_i - 1} \{ \beta_2 \mid RL_{ik} \leq RU_{ik} \} \right] \cap [l_2, m_2]. \tag{42}$$

The rejection sampling with an efficient blanket is applied to obtain samples of β_2. The efficient blanket $\tilde{C}_2$ will be derived in the next appendix. The acceptance probability is given by

$$\alpha\left(\beta_2, \beta_2'\right) = \min\left[1, \frac{\phi\left\{\left(\beta_2' - \mu_{\beta_2,1}\right)\sigma_u^{-1}\sigma_{\beta_2,1}^{-1}\right\}}{\phi\left\{\left(\beta_2 - \mu_{\beta_2,1}\right)\sigma_u^{-1}\sigma_{\beta_2,1}^{-1}\right\}}\right]. \tag{43}$$

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

By integrating the joint density function of $(\sigma_v^2, \boldsymbol{\delta})$ given $\boldsymbol{w}^*$ over $\boldsymbol{\delta}$, we have the full conditional distribution of σ_v^2 as the inverse gamma distribution, $IG(n_{v,1}/2, S_{v,1}/2)$, where $n_{v,1} = n_{v,0} + n$ and

$$S_{v,1} = S_{v,0} + \boldsymbol{\mu}_{\boldsymbol{\delta},0}' \boldsymbol{\Sigma}_{\boldsymbol{\delta},0}^{-1} \boldsymbol{\mu}_{\boldsymbol{\delta},0} + \boldsymbol{w}^{*\prime} \boldsymbol{w}^* - \boldsymbol{\mu}_{\boldsymbol{\delta},1}' \boldsymbol{\Sigma}_{\boldsymbol{\delta},1}^{-1} \boldsymbol{\mu}_{\boldsymbol{\delta},1}. \tag{44}$$

Then, given σ_v^2, the full conditional distribution of $\boldsymbol{\delta}$ is the multivariate normal distribution, $N_d(\boldsymbol{\mu}_{\boldsymbol{\delta},1}, \sigma_v^2 \boldsymbol{\Sigma}_{\boldsymbol{\delta},1})$, where

$$\boldsymbol{\mu}_{\boldsymbol{\delta},1} = \boldsymbol{\Sigma}_{\boldsymbol{\delta},1} \left(\boldsymbol{\Sigma}_{\boldsymbol{\delta},0}^{-1} \boldsymbol{\mu}_{\boldsymbol{\delta},0} + \boldsymbol{Z}' \boldsymbol{w}^* \right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\delta},1}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\delta},0}^{-1} + \boldsymbol{Z}' \boldsymbol{Z}. \tag{45}$$

Step 5. Generate $\{s_i^*, w_i^*\}_{i=1}^n$ given $\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2$.

The blocking technique is applied to draw samples of (s_i^*, w_i^*). The full conditional distribution of s_i^* is the multinomial distribution, the probability mass function of which is given by

$$\pi\left(s_i^* = s \mid \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2\right) \propto \left[\Phi\left\{\tau^{-1}\left(RU_{is} - \theta_{is}\right)\right\} - \Phi\left\{\tau^{-1}\left(RL_{is} - \theta_{is}\right)\right\}\right] \exp\left(-\frac{m_{is}}{2}\right), \tag{46}$$

for $s = 1, \ldots, K_i$, where $\tau^2 = (\sigma_u^{-2} + \sigma_v^{-2})^{-1}$ and

$$(m_{is}, \theta_{is}) = \left(\frac{(\sigma_u \sigma_v)^{-2} \left(y_i - \boldsymbol{x}_{is}' \boldsymbol{\beta} - \boldsymbol{z}_i' \boldsymbol{\delta} \right)^2}{\sigma_u^{-2} + \sigma_v^{-2}}, \ \frac{\sigma_u^{-2} \left(y_i - \boldsymbol{x}_{is}' \boldsymbol{\beta} \right) + \sigma_v^{-2} \boldsymbol{z}_i' \boldsymbol{\delta}}{\sigma_u^{-2} + \sigma_v^{-2}} \right). \tag{47}$$

Given $s_i^* = s$, the full conditional distribution of w_i^* is the truncated normal distribution, $TN_{R_{is}}(\theta_{is}, \tau^2)$.

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

The full conditional distribution of σ_u^2 is the inverse gamma distribution, $IG(n_{u,1}/2, S_{u,1}/2)$, where $n_{u,1} = n_{u,0} + n + 2$ and

$$S_{u,1} = S_{u,0} + \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta},0} \right)' \boldsymbol{\Sigma}_{\boldsymbol{\beta},0}^{-1} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta},0} \right) + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{w}^*)' (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{w}^*). \tag{48}$$

Step 7. Go to Step 2.

A.3 Efficient blanket for β_2

We assume that the support of the prior distribution for β_2 is $B_2 = [0, m_2]$ without loss of generality. Let

$$C_{kj}^{2i} = \left\{ \beta_2 \mid V_{ik} > V_{ij} \right\} \cap [0, m_2] \text{ and } a_2 = \exp(w_i^*)(1 + \beta_1)^{-1} \left(P_{ik}^{1+\beta_1} - P_{ij}^{1+\beta_1} \right). \tag{49}$$

Then, the Hermite-Hadamard integral inequality and $\beta_2 \in [0, m_2]$ derive

$$a_2 < \begin{cases} \left(Q_{ik} - Q_{ij} \right) \frac{Q_{ik}^{-\beta_2} + Q_{ij}^{-\beta_2}}{2}, & \text{if } k < j, \\ \left(Q_{ik} - Q_{ij} \right) \left(\frac{Q_{ik} + Q_{ij}}{2} \right)^{-\beta_2}, & \text{if } k > j. \end{cases} \tag{50}$$

By applying the power-mean inequality, we have $\tilde{C}^{2i}_{kj} = \tilde{C}^{\star 2i}_{kj} \cap [0, m_2]$ $(\supset C^{2i}_{kj})$, where

$$\tilde{C}^{\star 2i}_{kj} = \begin{cases} (-\infty, -b_2/\bar{q}(-m_2)), & \text{if } k < j \text{ and } \bar{q}(-m_2) > 0, \\ (-\infty, \infty), & \text{if } k < j \text{ and } \bar{q}(-m_2) = 0, \\ (-b_2/\bar{q}(-m_2), \infty), & \text{if } k < j \text{ and } \bar{q}(-m_2) < 0, \\ (-b_2/\bar{q}(1), \infty), & \text{if } k > j \text{ and } \bar{q}(1) > 0, \\ (-\infty, \infty), & \text{if } k > j \text{ and } \bar{q}(1) = 0, \\ (-\infty, -b_2/\bar{q}(1)), & \text{if } k > j \text{ and } \bar{q}(1) < 0, \end{cases} \tag{51}$$

$b_2 = \log(a_2/(Q_{ik} - Q_{ij}))$, and $\bar{q}(x) = x^{-1}\log\{(Q^x_{ik} + Q^x_{ij})/2\}$ $(x = 1, -m_2)$. If $Q_{iK_i} > 1$ is assumed, we have $\bar{q}(1) > \bar{q}(-m_2) > 0$, which simplifies the above expression. With this closely approximated interval $\tilde{C}^{2i}_{kj}$, we have $\tilde{C}_2 = \cap^n_{i=1} \cap^{K_i}_{j=1, j\neq s^*_i} \tilde{C}^{2i}_{s^*_i j}$, which includes C_2.

A.4 Metropolis-Hastings within Gibbs algorithm

To draw samples from the posterior distribution of the model without heterogeneity, we apply the Metropolis-Hastings within Gibbs algorithm, which is implemented in four steps.

Step 1. Set initial values to $(\boldsymbol{\beta}, \sigma^2_u)$.

Step 2. Generate $\boldsymbol{\beta}$ given σ^2_u.

Let s^*_i be the optimal block for the i-th consumer, i.e., $V_{is^*_i} \geq V_{ij}$ for $j \neq s^*_i$ and $j = 1, \ldots, K_i$, given the current value of $\boldsymbol{\beta}$. Generate a candidate $\boldsymbol{\beta}^\dagger$ from the following bivariate normal distribution, $N(\boldsymbol{\mu}_{\beta,1}, \sigma^2_u \boldsymbol{\Sigma}_{\beta,1})$, where

$$\boldsymbol{\Sigma}_{\beta,1} = \left(\boldsymbol{\Sigma}^{-1}_{\beta,0} + \sum_{i=1}^{n} \boldsymbol{x}_{i,s^*_i} \boldsymbol{x}'_{i,s^*_i} \right)^{-1}, \quad \boldsymbol{\mu}_{\beta,1} = \boldsymbol{\Sigma}_{\beta,1} \left(\boldsymbol{\Sigma}^{-1}_{\beta,0} \boldsymbol{\mu}_{\beta,0} + \sum_{i=1}^{n} \boldsymbol{x}_{i,s^*_i} y_i \right), \tag{52}$$

and calculate the optimal block $s^\dagger_i$ given $\boldsymbol{\beta}^\dagger$. Then, accept the candidate with the following probability,

$$\min\left[1, \prod_{i=1}^{n} \frac{\exp\left\{-\frac{1}{2\sigma_u^2}\left(y_i - \boldsymbol{x}'_{i,s_i^\dagger}\boldsymbol{\beta}^\dagger\right)^2\right\}}{\exp\left\{-\frac{1}{2\sigma_u^2}\left(y_i - \boldsymbol{x}'_{i,s_i^*}\boldsymbol{\beta}^\dagger\right)^2\right\}}\right]. \tag{53}$$

If the candidate is accepted, update $\boldsymbol{\beta}$ by $\boldsymbol{\beta}^\dagger$. Otherwise, retain the current value of $\boldsymbol{\beta}$.

Step 3. Generate σ_u^2 given $\boldsymbol{\beta}$.

The full conditional distribution for σ_u^2 is the inverse Gamma distribution, which is given by

$$IG\left\{\frac{n_{u,0}+2+n}{2}, \frac{S_{u,0} + \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\beta,0}\right)' \boldsymbol{\Sigma}_{\beta,0}^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\beta,0}\right) + \sum_{i=1}^{n}\left(y_i - \boldsymbol{x}'_{i,s_i^*}\boldsymbol{\beta}\right)^2}{2}\right\}, \tag{54}$$

where s_i^* is the i-th consumer's optimal block given $\boldsymbol{\beta}$.

Step 4. Go to Step 2.

Acknowledgement

This work is supported by the Grants-in-Aid for Scientific Research 18330039 and by the Grant-in-Aid for Young Scientists (Start-up) 21830115 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. The authors thank K. Fukumoto (Gakushuin University), M. Fukushige (Osaka University), K. Kakamu (Chiba University), K. Oya (Osaka University), A. C. Harvey (University of Cambridge), M. Kitahara (Tokyo Metropolitan University), N. Terui (Tohoku University), and H. K. van Dijk (Erasmus Universiteit Rotterdam) for their helpful comments. The computational results were calculated by using Ox (see Doornik (2002)).

References

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*(422), 669–679.

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

Baker, P., R. Blundell, and J. Micklewright (1989). Modelling household energy expenditures using micro data. *The Economic Journal 99*, 720–738.

Balestra, P. and M. Nerlove (1966). Cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica 34*(3), 585–612.

Beierlein, J. G., J. W. Dunn, and J. James C. McConnon (1981). The demand for electricity and natural gas in the northeastern united states. *Review of Economics and Statistics 63*(3), 403–408.

Bernard, J.-T., D. Bolduc, and D. Bélanger (1996). Quebec residential electricity demand: a microeconometric approach. *The Canadian Journal of Economics 29*(1), 92–113.

Bloch, F. E. (1980). Residential demand for natural gas. *Journal of Urban Economics 7*, 371–383.

Blomquist, S. and W. Newey (2002). Nonparametric estimation with nonlinear budget sets. *Econometrica 70*(6), 2455–2480.

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.

Burtless, G. and R. A. Moffitt (1985). The joint choice of retirement age and postretirement hours of work. *Journal of Labor Economics 3*(2), 207–236.

Carpio, C. E., M. K. Wohlgenant, and C. D. Safley (2008). A structural econometric model of joint consumption of goods and recreational time: An application to pick-your-own fruit. *American Journal of Agricultural Economics 90*(3), 644–657.

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

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association 90*(432), 1313–1321.

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 B. Ergashev (2009). Analysis of multifactor affine yield curve models. *Journal of the American Statistical Association 104*(488), 1324–1337.

Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the metropolishastings output. *Journal of the American Statistical Association 96*(453), 270–281.

Chib, S. and I. Jeliazkov (2005). Acceptreject metropolishastings sampling and marginal likelihood estimation. *Statistica Neerlandica 59*(1), 30–44.

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.

Dubin, J. A. and D. L. McFadden (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica 52*(2), 345–362.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In E. M.

Keramidas (Ed.), *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 571–578. Fairfax, VA: Interface Foundation of North America.

Goldberg, P. K. (1998). The effects of the corporate average fuel efficiency standards in the us. *The Journal of Industrial Economics 46*(1), 1–33.

Hajivassiliou, V. A. and D. L. McFadden (1998). The method of simulated scores for the estimation of ldv models. *Econometrica 66*(4), 863–896.

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

Hardy, G. H., J. E. Littlewood, and G. Pólya (1952). *Inequalities* (2nd ed.). Cambridge: Cambridge University Press.

Hausman, J. A. (1980). The effect of wages, taxes, and fixed costs on women's labor force participation. *Journal of Public Economics 14*(2), 161–194.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. *American Economic Review 71*(4), 662–676.

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

Hensher, D. A. and F. W. Milthorpe (1987). Selectivity correction in discrete-continuous choice analysis: With empirical evidence for vehicle choice and use. *Regional Science and Urban Economics 17*(1), 123–150.

Herriges, J. A. and K. K. King (1994). Residential demand for electricity under inverted block rates: Evidence from a controlled experiment. *Journal of Business & 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*(2), 173–192.

Hurwicz, L. and H. Uzawa (1971). On the integrability of demand functions. In J. S. Chipman, L. Hurwicz, M. K. Richter, and H. F. Sonnenschein (Eds.), *Preferences, Utility, and Demand*, The Harbrace series in business and economics, Chapter 6, pp. 114–148. New York: Harcourt, Brace, Jovanovich.

Keane, M. P. (1994). A computationally practical simulation estimator for panel data. *Econometrica 62*(1), 95–116.

King, M. A. (1980). An econometric model of tenure choice and demand for housing as a joint decision. *Journal of Public Economics 14*(2), 137–159.

Lee, L.-F. and R. P. Trost (1978). Estimation of some limited dependent variable models with application to housing demand. *Journal of Econometrics 8*(3), 357–382.

Lee, R.-S. and N. Singh (1994). Patterns in residential gas and electricity consumption: An econometric analysis. *Journal of Business & Economic Statistics 12*(2), 233–241.

Maddala, G. S., R. P. Trost, H. Li, and F. Joutz (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics 15*(1), 90–100.

Mannering, F. and C. Winston (1985). A dynamic empirical analysis of household vehicle ownership and utilization. *RAND Journal of Economics 16*(2), 215–236.

Mansur, E. T., R. Mendelsohn, and W. Morrison (2008). Climate change adaptation: A study of fuel choice and consumption in the US energy sector. *Journal of Environmental Economics and Management 55*(2), 175–193.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microecnomic Theory*. New York: Oxford University Press.

Miyawaki, K., Y. Omori, and A. Hibiki (2013). Exact estimation of demand functions under block-rate pricing. Forthcoming at *Econometric Reviews*.

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

Niculescu, C. P. and L.-E. Persson (2003). Old and new on the Hermite-Hadamard inequality. *Real Analysis Exchange 29*(2), 663–685.

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.

Szabó, A. (2009). The value of free water: Analyzing south africa's free basic water policy. mimeo.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*(398), 528–540.

Train, K. E. (1991). *Optimal Regulation: the Economic Theory of Natural Monopoly*. Cambridge, Mass.: MIT Press.

Vaage, K. (2000). Heating technology and energy use: a discrete/continuous choice approach to Norwegian household energy demand. *Energy Economics 22*(6), 649–666.

West, S. E. (2004). Distributional effects of alternative vehicle pollution control policies. *Journal of Public Economics 88*(3-4), 735–757.