CIRJE-F-764

Panel Data Analysis of Japanese Residential Water Demand Using a Discrete/Continuous Choice Approach

Koji Miyawaki
National Institute for Environmental Studies

Yasuhiro Omori
University of Tokyo

Akira Hibiki
National Institute for Environmental Studies

October 2010

Panel Data Analysis of Japanese Residential Water Demand Using a Discrete/Continuous Choice Approach

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

September 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).

Phone:+81-3-5841-5516. Fax:+81-3-5841-5521.

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

Abstract

Block rate pricing is often applied to income taxation, telecommunication services, and brand marketing in addition to its best-known application in public utility services. Under block rate pricing, consumers face piecewise-linear budget constraints. A discrete/continuous choice approach is usually used to account for piecewise-linear budget constraints for demand and price endogeneity. A recent study proposed a methodology to incorporate a separability condition that previous studies ignore, by implementing a Markov chain Monte Carlo simulation based on a hierarchical Bayesian approach. To extend this approach to panel data, our study proposes a Bayesian hierarchical model incorporating the random and fixed individual effects. In both models, the price and income elasticities are estimated to be negative and positive, respectively. Further, the

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

number of members and the number of rooms per household have positive relationship to the residential water demand when we apply the model with random individual effects, while they do not in the model with fixed individual effects.

Key words: Block rate pricing, Bayesian analysis, Panel data, Residential water demand.
JEL classification: C11, C23, C24, Q25.

1 Introduction

Block rate pricing usually has been applied to services in public utility sectors such as water, gas, and electricity.[1] However, block rate pricing is becoming common in areas such as local and wireless telephone services and brand marketing. Under block rate pricing, unit price changes with quantity consumed. When unit price increases with quantity consumed (Figure 9), such a price schedule is called the increasing block rate pricing. When unit price decreases with quantity consumed, it is called the decreasing block rate pricing. Then, under block pricing consumers maximize utility by selecting the unit price and the consumption amount. This circumstance leads to a utility-maximization problem under a piecewise-linear budget constraint.

As surveyed by Olmstead (2009), there are two types of estimation approaches that deal with this problem: reduced-form approaches, such as instrumental variables, and structural approaches. The structural approach solves a consumer's utility maximization problem in two steps. A consumer first decides appropriate consumption given each block's price, and then selects the block that maximizes consumer utility. This is also called a discrete/continuous choice approach because the block selection is discrete while the amount consumed is continuous. Its important feature is that the derived model explicitly addresses the relationship between the block choice and the amount consumed under block rate pricing.

As discussed in Olmstead (2009), the reduced-form approaches can incorporate only

[1]The other example where the same rate structure is applied is a progressive tax rate in income tax systems.

limited aspects of the piecewise-linear budget constraint, while the structural approaches account just for the particular implications of piecewise-linear budget constraints for demand as well as price endogeneity. Thus the latter approaches have two main advantages over the former ones: (1) the structural approaches can produce unbiased and consistent estimates of parameters of the price and the income[2]and (2) they are consistent with utility theory. Despite these advantages, most previous studies employ reduced-form approaches. Structural approaches are rare in demand analysis.[3] This is because the discrete/continuous choice approach had been applied only to the simplified block rate price structure—for example, the number of blocks is fixed at two.

While Pint (1999); Rietveld et al. (2000); Olmstead, Hanemann, and Stavins (2007); Olmstead (2009); Szabó (2009) considered multiple-block pricing through the classical approach (the maximum likelihood or the moment-based approach), Miyawaki, Omori, and Hibiki (2010) focused on the increasing block rate pricing and proposed a Bayesian approach. The latter has two following advantages over the former.

First, the Bayesian approach is a flexible estimation method to impose parameter constraints. As revealed by Miyawaki et al. (2010), the statistical model includes the so-called separability condition that tightly restricts elasticity parameters. From microeconomic theory point of view, the separability condition is a condition that guarantees the single-valued demand function and is one of sufficient conditions for the underlying preference relation to be strictly convex (see Hurwicz and Uzawa (1971) for the sufficient conditions). Despite its importance, previous literature generally ignores the condition because the parameter region becomes tightly restrained, making numerical maximization of the likelihood function difficult.

[2]Previous studies suggested that water demand is price inelastic. However, as is suggested in the meta-analysis (Dalhuisen, Florax, de Groot, and Nijkamp, 2003), the choice of the approach may affect the estimates, since the water demand is price inelastic in previous studies employing the reduced form approach, however, are price elastic in the discrete/continuous choice approach.

[3]Olmstead (2009) reported that, between 1963 and 2004, there were only three studies on water demand (Hewitt and Hanemann, 1995; Pint, 1999; Rietveld, Rouwendal, and Zwart, 2000) that adopted the discrete/continuous choice approach.

Second, when we analyze the panel data, the model often includes the individual effect to control the effect that is time-invariant. Its typical parameterizations are the random and the fixed individual effects (see, e.g., Chamberlain (1984)). Because our statistical model includes corner solutions as a part of the demand function, the latter specification is difficult to estimate its model parameters by simply differencing or demeaning. In this case, the Bayesian approach provides a straightforward estimation method, which is another advantage.

Thus, to extend Miyawaki et al. (2010) for the panel data analysis, this study proposes a Bayesian hierarchical model. It incorporates the random and fixed individual effects to estimate the residential water demand function under the separability condition using panel data of Japanese households. This is the first study that incorporates the individual effect in the discrete/continuous choice approach.

We organize this article as follows. Section 2 describes the proposed statistical model with the individual effect, which is an extension of Miyawaki et al. (2010) for panel data analysis. Then, 3 explains the Bayesian approach and discusses two specifications of the individual effect. Section 4 expresses the empirical data set, conducts the model comparison, and analyzes the Japanese residential water demand. Section 5 concludes.

2 Multinomial type V Tobit model

Let subscript i and t denote the observation and the time, respectively. We observe the residential water demand y_{it} under K_{it}-block increasing block rate pricing for the i-th observation at time t. By using the unit prices $\{P_{it,k}\}_{k=1}^{K_{it}}$, the upper quantity values $\{\bar{Y}_{it,k}\}_{k=1}^{K_{it}}$, and the fixed cost FC_{it}, the virtual incomes $\{Q_{it,k}\}_{k=1}^{K_{it}}$ are constructed. Then, by introducing the unobserved variables (w_{it}^*, s_{it}^*), which will be explained later, the statistical model is given as follows.[4]

$$y_{it} = y_{it}^* + u_{it}, \quad u_{it} \sim \text{i.i.d. } N(0,\sigma_u^2), \tag{1}$$

where

$$y_{it}^* = \begin{cases} y_{it,k} + w_{it}^* = \boldsymbol{x}'_{it,k}\boldsymbol{\beta} + w_{it}^*, & \text{if } s_{it}^* = 2k-1 \text{ and } k = 1,\ldots,K_{it}, \\ \bar{y}_{it,k}, & \text{if } s_{it}^* = 2k \text{ and } k = 1,\ldots,K_{it}-1, \end{cases} \tag{2}$$

$$s_{it}^* = \begin{cases} 2k-1, & \text{if } w_{it}^* \in R_{it,2k-1} \text{ and } k = 1,\ldots,K_{it}, \\ 2k, & \text{if } w_{it}^* \in R_{it,2k} \text{ and } k = 1,\ldots,K_{it}-1, \end{cases} \tag{3}$$

$$w_{it}^* = z'_{it}\boldsymbol{\delta}_i + v_{it}, \quad v_{it} \sim \text{i.i.d. } N\left(0,\sigma_v^2\right), \tag{4}$$

$$y_{it,k} = \boldsymbol{x}'_{it,k}\boldsymbol{\beta}, \quad \boldsymbol{x}_{it,k} = \left(p_{it,k}, q_{it,k}\right)', \quad \boldsymbol{\beta} = (\beta_1,\beta_2)', \quad k = 1,\ldots,K_{it}, \tag{5}$$

where K_{it} is the number of blocks, $(y_{it}, y_{it,k}, \bar{y}_{it,k}, p_{it,k}, q_{it,k})$ are the logarithm of demand, the logarithm of demand conditional on the k-th block, the logarithm of the upper quantity of the k-th block, the logarithm of the unit price for the k-th block, the logarithm of the virtual income for the k-th block, for observation i at time t. The conditional demand $y_{it,k}$ is specified below and the virtual income is defined in Appendix A.1 (see equation (A.26)).

This statistical model is based on the utility maximization behavior of a consumer. The problem and its solution (the demand function) is briefly described in Appendix A.1. From a statistical point of view, above statistical model is classified into a multinomial type V Tobit model (see Chapter 10 of Amemiya (1985)).

There are three main features in this statistical model. First, we assume the functional

[4]The statistical model above includes two popular models in the panel data analysis. When $\boldsymbol{z}_{it}$ includes $y_{i,t-1}$ as an explanatory variable, the model becomes the dynamic panel data model. On the other hand, when w_{it}^* has the AR(1) serial correlation, this model is interpreted as the AR(1) error component model (see also footnote 10 for the brief empirical results on this model). The AR(1) process specification can be further extended with a heteroskedastic variance structure, $\boldsymbol{w}_i^* \sim N(\boldsymbol{Z}_i\boldsymbol{\delta}_i, \Sigma)$, where $\boldsymbol{w}_{it}^* = (w_{i1}^*, \ldots, w_{iT}^*)'$ and $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iT})'$.

form of the conditional demand function to be log-linear, that is,

$$\log Y_{it,k} = \beta_1 \log P_{it,k} + \beta_2 \log Q_{it,k} \Longleftrightarrow y_{it,k} = \beta_1 p_{it,k} + \beta_2 q_{it,k}. \tag{6}$$

The log-linear demand function is one of the most popular functional forms in the water demand analysis (see, e.g., Hewitt and Hanemann (1995); Olmstead et al. (2007); Miyawaki et al. (2010)).

Second, the heterogeneity of preferences w_{it}^* is introduced. This is a unobserved stochastic term that models consumers' characteristics and is assumed to be the sum of the linear combination of d-dimensional vector $\boldsymbol{z}_{it}$ and the error term v_{it} that is independently and identically distributed with the normal distribution of mean 0 and variance σ_v^2. To capture the individual effect, we allow the coefficient of $\boldsymbol{z}_{it}$ varying across observations. The $\boldsymbol{\delta}_i$ can be "fixed" or "random" depending on the choice of the prior distribution (see the next section).

Based on this w_{it}^*, the consumer's optimal block/kink s_{it}^* is determined. The s_{it}^* is a unobserved discrete random variable that indicates which block or kink is potentially optimal for the consumer based on the heterogeneity and augments the model parameter space so that we exploit the data augmentation method to estimate parameters (see, e.g., Tanner and Wong (1987) for the description of the data augmentation). More precisely, if w_{it}^* is included in the heterogeneity interval $R_{it,2k-1}$, the k-th conditional demand with the heterogeneity is optimal, while it is in $R_{it,2k}$, the upper limit of the k-th block is optimal, where

$$R_{it,2k-1} = \left(\bar{y}_{it,k-1} - \boldsymbol{x}_{it,k}'\boldsymbol{\beta},\ \bar{y}_{it,k} - \boldsymbol{x}_{it,k}'\boldsymbol{\beta}\right), \quad R_{it,2k} = \left(\bar{y}_{it,k} - \boldsymbol{x}_{it,k}'\boldsymbol{\beta},\ \bar{y}_{it,k} - \boldsymbol{x}_{it,k+1}'\boldsymbol{\beta}\right). \tag{7}$$

It is possible to allow the correlation between w_{it}^* and $\{p_{it,k}, q_{it,k}\}_{k=1,\ldots,K_{it}}$ (see e.g., Mundlak (1978); Chamberlain (1980, 1984)). However, because we only have two time points $(t = 1, 2)$ and the variability of unit prices and virtual incomes are small in our empirical data set, we do not pursue such a specification in this article.

Finally, the actual demand is observed with measurement error u_{it} that is independently and identically distributed with the normal distribution of mean 0 and σ_u^2. It represents the measurement error as well as the optimization error and the model misspecification error (see Hausman (1985)).

3 Bayesian analysis

There are two prior specifications: one that assumes $\boldsymbol{\delta}_i$ to be "random" and the other that assumes it to be "fixed" (see, e.g., Lindley and Smith (1972); Smith (1973)). Such a distinction (random or fixed) comes from the non-Bayesian approach and all model parameters are random in the Bayesian context. However, for convenience, we use the term "random" and "fixed" for $\boldsymbol{\delta}_i$ to distinct its prior specifications.

First, we explain the prior specification for the model with the random individual effects. The model parameters are $(\boldsymbol{\beta}, \{\boldsymbol{\delta}_i\}_{i=1}^n, \sigma_u^2, \sigma_v^2)$ and we assume the following proper hierarchical prior distributions on them.

$$\boldsymbol{\beta}|\sigma_u^2 \sim N_2\left(\boldsymbol{\mu}_{\boldsymbol{\beta},0}, \sigma_u^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta},0}\right), \quad \sigma_u^2 \sim IG\left(\frac{n_{u,0}}{2}, \frac{S_{u,0}}{2}\right), \quad \sigma_v^2 \sim IG\left(\frac{n_{v,0}}{2}, \frac{S_{v,0}}{2}\right), \tag{8}$$

$$\boldsymbol{\delta}_i|\sigma_v^2, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}} \sim \text{i.i.d. } N_d\left(\boldsymbol{\mu}_{\boldsymbol{\delta}}, \sigma_v^2 \boldsymbol{\Sigma}_{\boldsymbol{\delta}}\right), \quad \text{for } i = 1, \ldots, n, \tag{9}$$

$$\boldsymbol{\mu}_{\boldsymbol{\delta}} \sim N_d\left(\boldsymbol{\mu}_{\bar{\boldsymbol{\delta}},0}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},0}\right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\delta}} \sim IW_d\left(n_{\bar{\boldsymbol{\delta}},0}, \boldsymbol{S}_{\bar{\boldsymbol{\delta}},0}\right), \tag{10}$$

where $\boldsymbol{\mu}_{\boldsymbol{\beta},0} = (\mu_{\beta_1,0}, \mu_{\beta_2,0})'$ is a 2×1 known vector, $\boldsymbol{\Sigma}_{\boldsymbol{\beta},0} = \text{diag}(\sigma^2_{\beta_1,0}, \sigma^2_{\beta_2,0})$ is a 2×2 known diagonal matrix with positive diagonal elements $(\sigma^2_{\beta_1,0}, \sigma^2_{\beta_2,0})$, $(n_{u,0}, S_{u,0}, n_{v,0}, S_{v,0})$ are known positive constants, $\boldsymbol{\mu}_{\bar{\boldsymbol{\delta}},0}$ is a $d\times 1$ known vector, $\boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},0}$ and $\boldsymbol{S}_{\bar{\boldsymbol{\delta}},0}$ are known $d\times d$ positive definite matrices, and $n_{\bar{\boldsymbol{\delta}},0} > d-1$ is a known constant. The $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $IG(a,b)$, and $IW_p(m, \boldsymbol{\Psi})$ represent the k-dimensional multivariate normal distribution of mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, the inverse gamma distribution with shape parameter a and scale parameter b,[5] and the p-dimensional

[5]The mean and variance of $IG(a,b)$ (a and b positive constants) are $b/(a-1)$ for $a > 1$ and $b^2/\{(a-1)^2(a-2)\}$ for $a > 2$, respectively.

inverse Wishart distribution with m degrees of freedom and parameter matrix Ψ,[6] respectively.

In this prior setting, we assume that $\{\boldsymbol{\delta}_i\}_{i=1}^n$ are independently and identically drawn from the normal distributions (9), of which mean and variance are common across observations and are distributed by the hyperprior distributions (10). The iid assumption implies the exchangeability among $\{\boldsymbol{\delta}_i\}_{i=1}^n$, which is acceptable because we have no prior information available to distinguish them.

Next, the prior specification for the model with the fixed individual effects is described. Among the model parameters, we assume the same prior distributions on $(\boldsymbol{\beta}, \sigma_u^2, \sigma_v^2)$ (see equations (8)) and assume the following proper prior distributions on $\{\boldsymbol{\delta}_i\}_{i=1}^n$.

$$\boldsymbol{\delta}_i | \sigma_v^2 \sim N_d\left(\boldsymbol{\mu}_{\boldsymbol{\delta}_i,0}, \sigma_v^2 \boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,0}\right), \quad \text{for } i = 1, \ldots, n, \tag{15}$$

where $\boldsymbol{\mu}_{\boldsymbol{\delta}_i,0}$ is a $d \times 1$ known vector and $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,0}$ is a $d \times d$ know positive definite matrix. No hyperprior distributions are assumed. In this specification, $\{\boldsymbol{\delta}_i\}_{i=1}^n$ are treated in the same way with other parameters, and their prior means and variances are allowed to differ among observations.

However, we will assume the same prior mean and variance for all $\boldsymbol{\delta}_i$ in the empirical analysis because we have no prior information available to distinguish them (Subsection 4.3). Then, such a prior specification can be viewed as a special case of the one for the

---

[6]When $\boldsymbol{V}$ follows $IW_p(m, \boldsymbol{\Psi})$, its mean and variance are given by

$$E(\boldsymbol{V}) = \frac{\boldsymbol{\Psi}}{m-2p-2}, \quad (m-2p-2>0), \tag{11}$$

$$Var(v_{ii}) = \frac{2\psi_{ii}^2}{(m-2p-2)^2(m-2p-4)}, \quad (m-2p-4>0), \tag{12}$$

$$Var\left(v_{ij}\right) = \frac{\psi_{ii}\psi_{jj} + \frac{m-2p}{m-2p-2}\psi_{ij}^2}{(m-2p-1)(m-2p-2)(m-2p-4)}, \quad (m-2p-4>0), \tag{13}$$

$$Cov\left(v_{ij}, v_{kl}\right) = \frac{\frac{2}{m-2p-2}\psi_{ij}\psi_{kl} + \psi_{ik}\psi_{jl} + \psi_{il}\psi_{kj}}{(m-2p-1)(m-2p-2)(m-2p-4)}, \quad (m-2p-4>0), \tag{14}$$

where v_{ij} and ψ_{ij} are the (i,j) element of $\boldsymbol{V}$ and $\boldsymbol{\Psi}$, respectively (see Gupta and Nagar (2000)).

random individual effects. That is, the former specification is the same with the latter when we exclude the hyperprior distributions (10). Section 4.2 will conduct the model comparison by using the deviance information criteria (DIC) and determine which prior assumption is favored in terms of the empirical data.

Let $\theta_i = (\boldsymbol{\beta}, \boldsymbol{\delta}_i, \sigma_u^2, \sigma_v^2)$ and $\theta = (\boldsymbol{\beta}, \{\boldsymbol{\delta}_i\}_{i=1}^n, \sigma_u^2, \sigma_v^2)$. Let $\pi_{RE}(\theta, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}})$ and $\pi_{FE}(\theta)$ denote the respective joint prior probability density functions of the model with random and fixed individual effects. Then, the joint posterior probability density functions for these two prior specifications are given by

$$\pi_{RE}\left(\theta, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}, \{\boldsymbol{w}_i^*, \boldsymbol{s}_i^*\}_{i=1}^n \mid \{\boldsymbol{y}_i\}_{i=1}^n\right) \propto \pi_{RE}\left(\theta, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}\right) \prod_{i=1}^n \prod_{t=1}^T f\left(y_{it}, s_{it}^*, w_{it}^* \mid \theta_i\right), \tag{16}$$

$$\pi_{FE}\left(\theta, \{\boldsymbol{w}_i^*, \boldsymbol{s}_i^*\}_{i=1}^n \mid \{\boldsymbol{y}_i\}_{i=1}^n\right) \propto \pi_{FE}\left(\theta\right) \prod_{i=1}^n \prod_{t=1}^T f\left(y_{it}, s_{it}^*, w_{it}^* \mid \theta_i\right), \tag{17}$$

where

$$f\left(y_{it}, s_{it}^*, w_{it}^* \mid \theta_i\right) \propto \sigma_u^{-1} \sigma_v^{-1} \exp\left[-\frac{1}{2}\left\{\sigma_u^{-2}\left(y_{it} - y_{it}^*\right)^2 + \sigma_v^{-2}\left(w_{it}^* - \boldsymbol{z}_{it}'\boldsymbol{\delta}_i\right)^2\right\}\right]$$
$$\times I\left(w_{it}^* \in R_{it, s_{it}^*}\right) \prod_{k=1}^{K^{it}-1} I\left(\boldsymbol{x}_{it,k+1}'\boldsymbol{\beta} \le \boldsymbol{x}_{it,k}'\boldsymbol{\beta}\right), \tag{18}$$

$\pi_{RE}\left(\theta, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}, \{\boldsymbol{w}_i^*, \boldsymbol{s}_i^*\}_{i=1}^n \mid \{\boldsymbol{y}_i\}_{i=1}^n\right)$ and $\pi_{FE}\left(\theta, \{\boldsymbol{w}_i^*, \boldsymbol{s}_i^*\}_{i=1}^n \mid \{\boldsymbol{y}_i\}_{i=1}^n\right)$ are the respective joint posterior density functions of the model with random and fixed individual effects, $I(A)$ is the indicator function; $I(A) = 1$ if A is true and $I(A) = 0$ otherwise, $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\boldsymbol{w}_i^* = (w_{i1}^*, w_{i2}^*, \ldots, w_{iT}^*)'$, and $\boldsymbol{s}_i^* = (s_{i1}^*, s_{i2}^*, \ldots, s_{iT}^*)'$. To obtain posterior samples from these posterior density functions, we apply the simple Gibbs sampler. The algorithms are given in Appendix A.2 and A.3.

The product of $I(\boldsymbol{x}_{it,k+1}'\boldsymbol{\beta} \le \boldsymbol{x}_{it,k}'\boldsymbol{\beta})$ over $i = 1, \ldots, n$, $t = 1, \ldots, T$, and $k = 1, \ldots, K_{it} - 1$ is the separability condition that guarantees disjoint heterogeneity intervals (see $R_{it,2k}$ in equations (7)). Because $\boldsymbol{\beta}$ is a two-dimensional vector in our statistical modeling, this condition reduces to two inequality constraints:

$$\beta_2 \le \bar{r}\beta_1 \text{ and } \beta_2 \le \underline{r}\beta_1, \tag{19}$$

where $\bar{r} = \max_{i,t,k} -(p_{it,k+1} - p_{it,k})/(q_{it,k+1} - q_{it,k})$ and $\underline{r} = \min_{i,t,k} -(p_{it,k+1} - p_{it,k})/(q_{it,k+1} - q_{it,k})$. Further discussions can be found in Miyawaki et al. (2010).

4 Empirical analysis

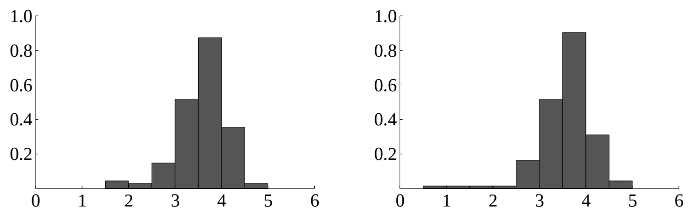
4.1 Data

We use the household-level dataset collected by internet surveys concerning household water and energy consumption and garbage emissions, which we conducted twice (in June 2006 and June 2007) for individuals in the Tokyo and Chiba prefectures in collaboration with INTAGE, Inc., a marketing research company (www.intage.co.jp/english), which has more than 1.3 million monitors all over Japan. As respondents, 1,687 monitors were randomly selected from all INTAGE monitors, 47,239, in this area who are between age 20 and 79. The numbers of respondents in June 2006 and June 2007 were 1,276 and 760, respectively. The number of respondents in both June 2006 and 2007 was 515. The individuals' answers concerned attributes of the household to which they belong, including the number of household members, household annual income, number of rooms and floor space of their house or apartment, and the household's monthly water and sewerage bills. Because water and sewerage are billed every second month in Japan, reported usage is considered to be a two-month usage. In the survey, these attributes are collected only once yearly, and we used respondents collected in June 2006 and April 2007. Since sewerage and water bills are also calculated based on water consumption, the amount of water consumption was calculated from the water and sewerage bills using the corresponding information on water charge schedules and sewerage charge schedules in each city. Every household faces increasing block rate pricing; the number of blocks varies from two to eleven, depending on cities where respondents live.

The number of observations used for the empirical analysis in the next subsection was reduced to 135 because of respondents' missing or inappropriate answers or for technical reasons as follows:

1. Consumption within the zero unit price block is observed.
2. Living in cities that have discontinuous parts in their price system.
3. Living in cities that changed rate tables in June 2006.[7]
4. Using a well for water use because of its special charge system.

The histograms of the amount of water consumption, the dependent variable for the empirical analysis in the next subsection, are shown in Figure 1. Other variables used as explanatory variables for the empirical analysis are listed in Table 1. In Figure 2, we summarize the block rate price structure. Each column of Figure 2 shows the relative frequencies of the number



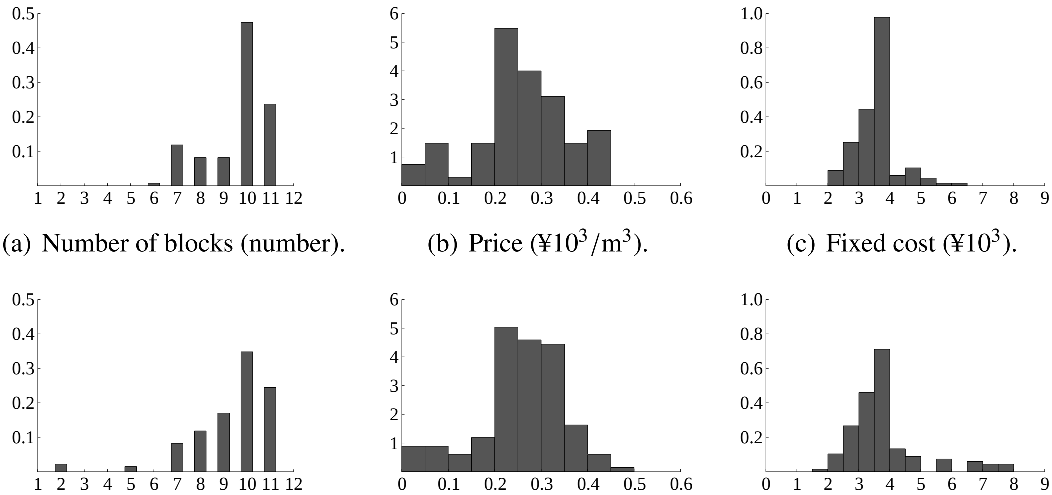
(a) June 2006. (b) June 2007.

Figure 1: Histograms of the amount of water consumption ($\log \mathrm{m}^3$).

Table 1: Explanatory variables used in the water demand function

Variable	*Coefficient*	*Description*
price	β_1	water+sewer (log ¥$10^3/\mathrm{m}^3$)
virtual income	β_2	income augmented by price (log ¥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)

[7] In June 2007, no cities changed the rate tables.



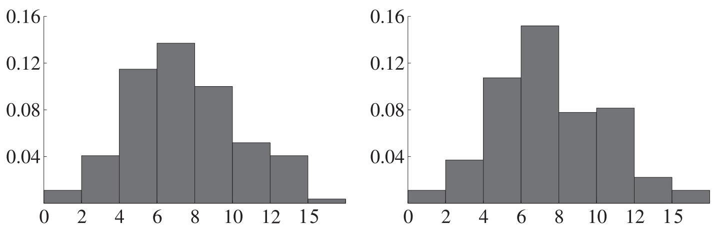
(a) Number of blocks (number). (b) Price (¥10^3/m^3). (c) Fixed cost (¥10^3).

(d) Number of blocks (number). (e) Price (¥10^3/m^3). (f) Fixed cost (¥10^3).

Figure 2: Relative frequencies of the number of blocks, and the histograms of the price and the fixed cost. Top row is for June 2006 and bottom row is for June 2007.

of blocks, and the histograms of the unit price where the consumption is actually made and the minimum access charge for June 2006 and June 2007.

Regarding the income variable, it is a sensitive issue to ask households their exact annual income level. Therefore, in our survey, instead of the actual values, the household is asked to choose one of eight categories for the annual income in million yen; 0-2, 2-4, 4-6, 6-8, 8-10, 10-12, 12-15, over 15 million yen. The histograms for the income categories are shown in Figure 3. For the empirical analysis, we use the median of the interval of each categories



(a) June 2006. (b) April 2007.

Figure 3: Histograms for the income (¥10^6).

divided by six to estimate the two-month income for the households except of those who choose "over 15 million yen." Households whose annual incomes are over 15 million yen are asked to answer the value of their annual income.

Basic statistics for heterogeneity are given in Table 2. We calculate the correlation coefficients among explanatory variables for heterogeneity. All correlation coefficients are less than .6, except for the correlation between the number of rooms and total floor space, which is .68 in 2006 and .67 in 2007.

Table 2: Basic statistics of explanatory variables for heterogeneity

Variable	*Unit*	*Year*	*Mean*	*SD*	*Min.*	*Max.*
the number of members in a household (δ_1)	person	2006	3.18	1.20	1	7
		2007	3.21	1.23	1	8
the number of rooms in a house/apartment (δ_2)	room	2006	4.41	1.08	2	8
		2007	4.39	1.07	2	8
the total floor space of a house/apartment (δ_3)	50m^2	2006	1.68	0.72	0.24	4.60
		2007	1.68	0.72	0.24	4.60

4.2 Model comparison

The model, equations (5-1), is based on the discrete/continuous choice approach, more precisely, on the consumer's utility maximization problem. This section explains an alternative model that is not based on the consumer theory, but on the random choice.

$$y_{it,k} = \boldsymbol{x}'_{it,k}\boldsymbol{\beta}, \quad \boldsymbol{x}_{it,k} = (p_{it,k}, q_{it,k})', \quad \boldsymbol{\beta} = (\beta_1, \beta_2)', \quad k = 1, \dots, K_{it}, \tag{20}$$

$$w^*_{it} = \boldsymbol{z}'_{it}\boldsymbol{\delta}_i + v_{it}, \quad v_{it} \sim \text{i.i.d. } N\left(0, \sigma_v^2\right), \tag{21}$$

$$y^*_{it} = \begin{cases} y_{it,k} + w^*_{it} = \boldsymbol{x}'_{it,k}\boldsymbol{\beta} + w^*_{it}, & \text{with probability } \pi_{it,s} \text{ and } s = 1, 3, \dots, 2K_{it} - 1, \\ \bar{y}_{it,k}, & \text{with probability } \pi_{it,s} \text{ and } s = 2, 4, \dots, 2K_{it} - 2, \end{cases} \tag{22}$$

$$y_{it} = y^*_{it} + u_{it}, \quad u_{it} \sim \text{i.i.d. } N(0, \sigma_u^2), \tag{23}$$

where $\pi_{it,s}$ is a known constant such that $\sum_{s=1}^{2K_{it}-1} \pi_{it,s} = 1$. In this study, we assume $\pi_{it,s} = (2K_{it} - 1)^{-1}$ for all s.

The difference between the model (5-1) and above model is the block-choice rule. The former is based on the heterogeneity interval derived from the utility maximization problem, while the latter is based on the predetermined probability. When $\pi_{it,s} = 1$ for a specified s such that $\bar{y}_{it,(s-1)/2} \leq y_{it} < \bar{y}_{it,(s+1)/2}$, above model reduces to the linear model with two-error components.

There are two prior specifications for this model, similar to the model based on the discrete/continuous choice. More precisely, the random-individual-effects priors are distributions (8–10), and the fixed-individual-effects priors are distributions (8) with distributions (15). Table 3 summarizes two models and two prior specifications.

Table 3: Two different models and two different prior specifications

	Discrete/Continuous choice	*Random choice*
Random individual effects	$M1$	$M3$
Fixed individual effects	$M2$	$M4$

These four models are compared on the basis of DIC (see Spiegelhalter, Best, Carlin, and van der Linde (2002) for the discussion of DIC) by using the empirical data set. The results are given in Table 4. First, all standard errors are small enough to distinguish all four

Table 4: Model comparison

Model	*D*(θ)	p_D	*DIC* (*SE**)	*Rank*
$M1$	− 38.64	92.52	146.41 (2.76)	1
$M2$	111.70	94.17	300.03 (4.01)	2
$M3$	109.74	138.11	385.97 (1.32)	3
$M4$	281.08	104.29	489.66 (.65)	4

* The standard errors are the sample standard deviations of 20 DICs calculated from 20 independent replications.

models. Next, when we compare the model based on the discrete/continuous choice with the one based on the random choice, the former is better in terms of DIC under both prior specifications. Further, the random-effects model is superior to the fixed-effects model under both model specifications. Finally, the DIC result suggests that the random-individual-effects model based on the discrete/continuous choice approach ($M1$) is the most appropriate among these four models.

4.3 Estimation results of panel data models

This subsection first conducts the empirical analysis of Japanese residential water demand using $M1$, the random-effects model based on the discrete/continuous choice. It should be noted that use of two-period panel data conducted in June 2006 and June 2007 data is useful in removing the seasonality effect. The dependent variable is the amount of water consumption calculated from water and sewerage bills using the corresponding charge schedules. The explanatory variables are listed in Table 1. The separability condition on the parameter space of $\boldsymbol{\beta}$ implies

$$\beta_2 \leq -0.16\beta_1 \text{ and } \beta_2 \leq -3263.83\beta_1. \tag{24}$$

Prior distributions are parameterized by setting $\boldsymbol{\mu}_{\bar{\boldsymbol{\delta}},0} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},0} = 10\boldsymbol{I}_4$, $n_{\bar{\boldsymbol{\delta}},0} = 10$, $\boldsymbol{S}_{\bar{\boldsymbol{\delta}},0} = 10^{-1}\boldsymbol{I}_4$, $\boldsymbol{\mu}_{\boldsymbol{\beta},0} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta},0} = 10\boldsymbol{I}_2$, and $n_{u,0} = S_{u,0} = n_{v,0} = S_{v,0} = 0.1$. These priors are fairly flat to reflect that we do not have a sufficient prior information regarding parameters. We adopt the Gibbs sampler described in Appendix A.2. For Bayesian inferences, we generate 15 million samples after deleting the initial six million samples. The recorded values are reduced to $10,000$ samples by picking up every 1500-th value. Results are shown in Figure 4 and given in Table 5.

Each column of Table 5 represents the parameters, posterior means, posterior standard deviations, posterior 95% credible intervals, inefficiency factors, and p-value of convergence diagnostic statistics. The inefficiency factor is an indicator that measures the degree of autocorrelation of the Markov chain and is defined as $1 + 2\sum_{j=1}^{\infty}\rho(j)$, where $\rho(j)$ is the lag j sample autocorrelation. As pointed out in Chib (2001), this value is interpreted as the ratio of the variance of the sample mean obtained by the Markov chain to that of the sample mean

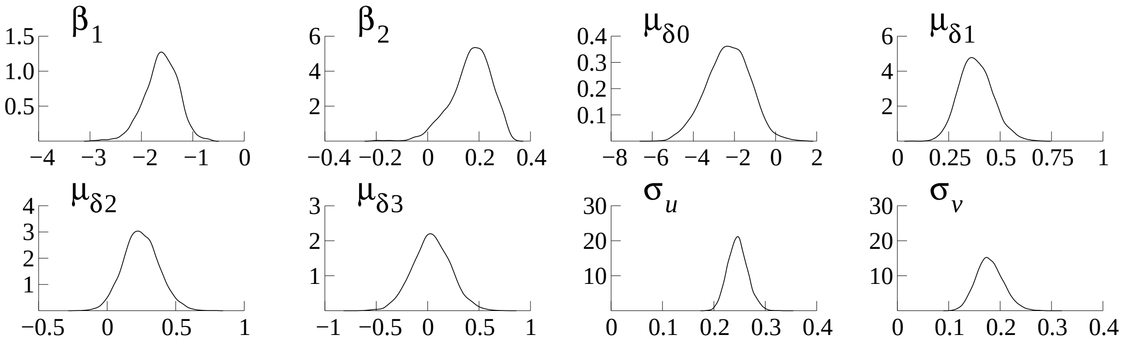


Figure 4: Estimated marginal posterior densities.

Table 5: Water demand function

Parameter	*Mean*	*SD*	*95% interval*	*INEF**	*CD**
β_1 (price)	−1.61	.33	[−2.30 −1.02]	125	.593
β_2 (income)	.17	.079	[− .00 .30]	157	.787
$\mu_{\delta 0}$ (constant)	−2.30	1.06	[−4.42 − .34]	134	.705
$\mu_{\delta 1}$ (num. of members)	.38	.082	[.23 .56]	20	.814
$\mu_{\delta 2}$ (num. of rooms)	.25	.13	[.00 .52]	6	.870
$\mu_{\delta 3}$ (floor space)	.039	.19	[− .33 .42]	2	.387
σ_u (measurement error)	.25	.019	[.21 .29]	2	.636
σ_v (heterogeneity)	.18	.027	[.13 .24]	9	.853

* "INEF" and "CD" denote the inefficiency factor and the p-value of convergence diagnostic statistic, respectively.

by uncorrelated draws. When it is close to one, the Markov chain would be as efficient as uncorrelated draws. When, on the other hand, it is much greater than one, say 10, we need to take a ten times longer Markov chain. The p-value is for the two-sided test of whether the convergence of the Markov chain is reached, proposed by Geweke (1992). The first 10% and last 50% MCMC samples are used to conduct this test as suggested by Geweke (1992).

Obtained MCMC samples for all parameters can be considered to be those from the posterior distribution judging from the p-values of their convergence diagnostics. The inefficiency factors also suggest that we took a sufficiently long Markov chain to conduct inferences.

Table 5 shows several aspects of the Japanese residential water demand function. First, price and income elasticities are highly credible to be negative and positive, respectively, in terms of their 95% credible intervals.[8] These elasticities have theoretically correct signs. The absolute value of price elasticity is much larger than that of income elasticity. Thus, the water demand is less sensitive to the change in the individual income, while it is to the price change. Because the separability condition strongly restricts the parameter space, this result could be due to this condition (see also Miyawaki et al. (2010)).

Figure 5 shows the scatter plot of the posterior samples of elasticity parameters. The

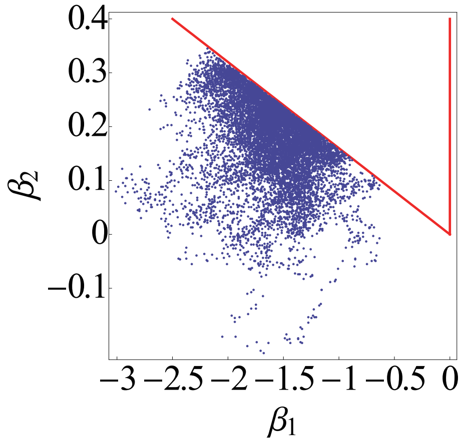


Figure 5: Scatter plot of the joint posterior density of (β_1, β_2).

diagonal and vertical lines represent $\beta_2 = -.16\beta_1$ and $\beta_2 = -3263.83\beta_1$, respectively, which are the boundaries of the separability condition (24). Due to this separability condition, the posterior samples are highly restricted to the north-east, which causes the slow convergence of the Markov chain to the posterior distribution.

Second, the coefficients of the heterogeneity variables are examined. Among the means of the heterogeneity coefficients, $\boldsymbol{\mu}_\delta$, $(\mu_{\delta_1}, \mu_{\delta_2})$ that correspond to the number of members in a household and the number of rooms in a household/apartment, respectively, have positive effects on water demand because their posterior probabilities $P(\mu_{\delta_j} > 0 \mid \text{Data}) > .95$ $(j = 1, 2)$.

[8]Precisely, the 95% credible interval for β_2 includes zero, which means that β_2 does not differ from zero in terms of the credible interval. However, the posterior probability $P(\beta_2 > 0 \mid \text{Data}) = .97$ implies that we have credible evidence for the positive income elasticity with more than 95% posterior probability.

In contrast, that for the total floor space in a household/apartment (μ_{δ_3}) has no effect on water demand in terms of its 95% credible interval. This result is partly influenced by the correlation between the number of rooms and total floor space, as noted at the end of the preceding subsection. Further, for most households, their posterior means of δ_{i1} and δ_{i2} are positive (see Figure 6).

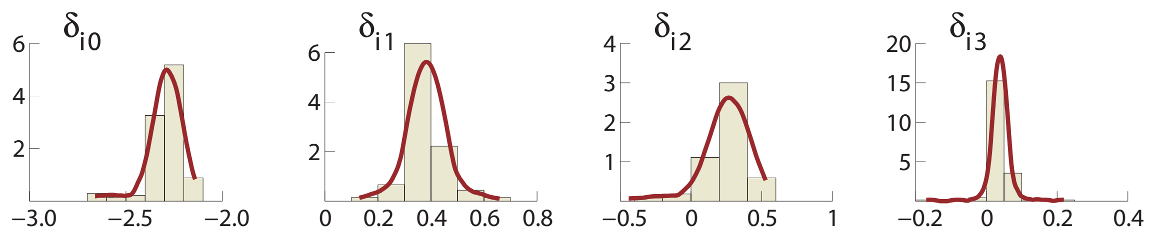


Figure 6: Histograms and kernel density estimates of the posterior means of $\boldsymbol{\delta}_i$.

Next, we analyze $M2$, the fixed-effects model based on the discrete/continuous choice. Prior distributions are parameterized by setting $\boldsymbol{\mu}_{\boldsymbol{\delta}_i,0} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,0} = 10\boldsymbol{I}_4$ for all i, $\boldsymbol{\mu}_{\boldsymbol{\beta},0} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta},0} = 10\boldsymbol{I}_2$, and $n_{u,0} = S_{u,0} = n_{v,0} = S_{v,0} = 0.1$. The Gibbs sampler described in Appendix A.3 are conducted and obtain 32×10^5 samples after generating 24×10^5 burn-in samples. Then, they are reduced to 10^4 samples by picking up every 320-th value to conduct Bayesian inferences. The results are given in Table 6. Compared with the results of $M1$, the posterior

Table 6: Water demand function ($M2$)

Parameter	*Mean*	*SD*	*95% interval*		*INEF**	*CD**
β_1 (price)	−1.42	.085	[−1.64	−1.31]	97	.058
β_2 (income)	.20	.015	[.16	.22]	93	.106
σ_u (measurement error)	.28	.034	[.22	.34]	10	.013
σ_v (heterogeneity)	.41	.041	[.33	.48]	5	.606

* "INEF" and "CD" denote the inefficiency factor and the p-value of convergence diagnostic statistic, respectively.

means are similar while the posterior standard deviations for variance parameters are larger.

The histograms of the posterior means of $\boldsymbol{\delta}_i$ are shown in Figure 7. In contrast to those of $M1$ (Figure 6), all coefficients are about zero in terms of their posterior means. There-

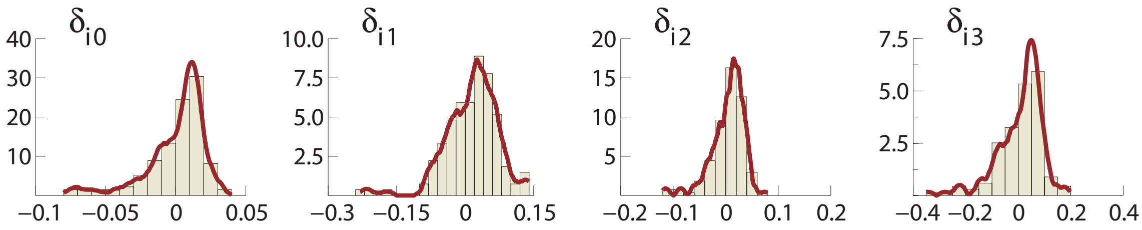


Figure 7: Histograms and kernel density estimates of the posterior means of $\boldsymbol{\delta}_i$ ($M2$).

fore, in $M2$, no heterogeneity variable would have positive or negative relations to water demand, which suggests that the fitting of $M2$ to the dataset is not appropriate in terms of the heterogeneity.

We compare our results with those obtained in previous studies,[9] all of which applied the maximum likelihood method to estimate the water demand function based on the discrete/continuous choice approach. We note that their statistical models to be estimated do not include the individual effect. Furthermore, the separability condition is also ignored in these studies.

Olmstead et al. (2007) used data from households in the United States and Canada. The household faces one of three kinds of price schedules: two-block increasing block rate pricing, four-block increasing block rate pricing, and uniform pricing. The estimated price and income elasticities (the coefficients of price and virtual income) are −.3407 and .1306, respectively, and their standard errors are .0298 and .0118, respectively. While their income elasticity is similar to ours, their price elasticity is smaller. They used 21 explanatory variables for heterogeneity, including number of residents per household, number of bathrooms, approximate are of the home, approximate area of its lot, and the approximate age of the home as household attributes. Coefficients of these variables are all significant at the 5% level. In particular, the coefficients of the number of residents per household and the approximate area of the home are .1960 and .1257, respectively.

[9]Pint (1999) estimated the water demand function during the California drought. Because Pint (1999) used the level of unit price as an explanatory variable for the conditional demand, its estimation result cannot be simply compared with ours.

Hewitt and Hanemann (1995) also estimated the residential water demand function under two-block increasing block rate pricing in Denton, Texas. The price and income elasticities are estimated to be -1.8989 and .1782, respectively, (their asymptotic t statistics are -6.421 and 1.864, respectively), which are similar to ours. Among variables for heterogeneity, they found that the number of bathrooms has a positive effect on water demand at the 5% significance level. They consider that the number of bathrooms would represent the number of members in a household, which would better explain the variation in residential water use.

Rietveld et al. (2000) analyzed the water demand function under four-block increasing block rate pricing in Indonesia. The price and income elasticities are estimated to be -1.280 and $.501 \times 10^{-6}$, respectively, with standard errors .235 and $.348 \times 10^{-6}$, respectively. The tendency for demand to be elastic with regard to price and inelastic with regard to income is coincident with the results of Hewitt and Hanemann (1995) and ours. Furthermore, the log of the number of members in a household has a positive effect on water demand at the 5% significance level.

At the end of this section, we briefly discuss the model based on the random choice ($M3$) and show how the results are affected compared with the model based on the utility maximization ($M1$).[10] The prior distributions are parameterized in the same manner of $M1$. We generate 4×10^6 samples after deleting 16×10^5 samples and reduce them to 10^4 samples by picking up every 400-th value. Results are given in Table 7.

The price elasticity is much larger than that of $M1$ and its 95% credible interval includes zero. This is partly because $M3$ is free of the separability condition. As we see in Figure 5, the separability condition highly restricts elasticity parameters' space. In contrast to the model based on the discrete/continuous choice, $M3$ does not include heterogeneity intervals, and, hence, the separability condition. Figure 8 shows the contour plot of the joint posterior

[10] We further analyzed another panel data model, the AR(1) error component model. Because its results are found to be very similar to those obtained in Figure 4 and Table 5, their details are omitted. The parameter that represents the serial correlation is not credible to be positive or negative in the sense that its 95% credible interval includes zero. No serial correlation is also observed when we use the four-consecutive-months data—that is, the data from June 2006 to September 2006.

Table 7: Water demand function ($M3$)

Parameter	*Mean*	*SD*	*95% interval*		*INEF**	*CD**
β_1 (price)	−.022	.037	[−.096	.052]	1	.367
β_2 (income)	.24	.087	[.072	.42]	15	.275
$\mu_{\delta 0}$ (constant)	.88	.58	[−.31	2.01]	21	.343
$\mu_{\delta 1}$ (num. of members)	.17	.042	[.087	.25]	2	.426
$\mu_{\delta 2}$ (num. of rooms)	.079	.068	[−.054	.22]	3	.043
$\mu_{\delta 3}$ (floor space)	.027	.11	[−.18	.24]	1	.828
σ_u (measurement error)	.17	.020	[.13	.21]	2	.036
σ_v (heterogeneity)	.11	.012	[.088	.14]	1	.158

* "INEF" and "CD" denote the inefficiency factor and the p-value of convergence diagnostic statistic, respectively.

density of (β_1, β_2) for $M3$. The horizontal and vertical lines represent $\beta_2 = -.16\beta_1$ and

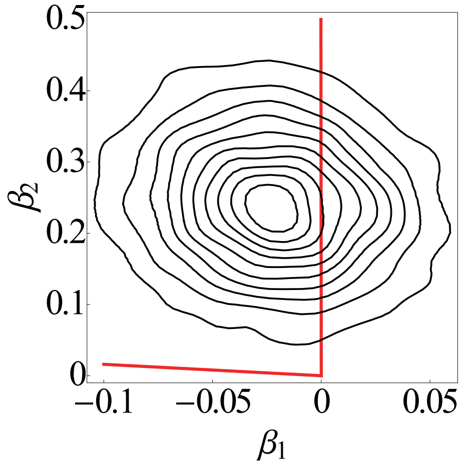


Figure 8: Contour plot of the joint posterior density of (β_1, β_2) ($M3$).

$\beta_2 = -3263.83\beta_1$, respectively, which are the boundaries of the separability condition. It is clear that the joint density is not constrained by this condition. Therefore, this estimate reveals how the separability condition affects parameter estimates.

The μ_{δ_i}s are estimated to be smaller except for μ_{δ_0} in terms of the posterior mean. In particular, μ_{δ_2} for the number of rooms in a household/apartment is smaller than that of $M1$ and has no positive relation on the residential water demand in terms of the 95% credible interval.

5 Conclusion

This paper proposed a Bayesian hierarchical model incorporating the random and fixed individual effects and conducted a structural analysis of the Japanese residential water demand using panel data. In both models, the price and income elasticities are estimated to be negative and positive, respectively. Further, the number of members and the number of rooms per household have positive relationship to the residential water demand when we apply the model with random individual effects, while they do not in the model with fixed individual effects.

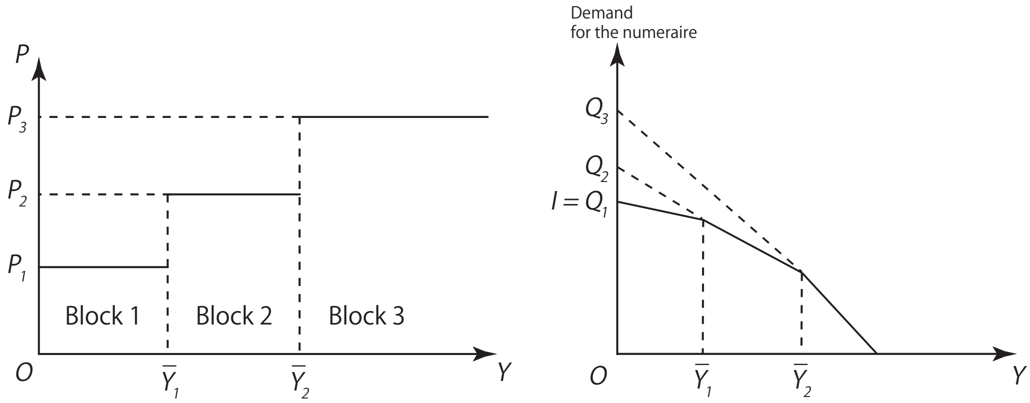
Finally, we note a possible application in the policy evaluation and a spatial extension for our model, which are left for our future work. First, the proposed model is useful for making policies that continue several periods. For example, the price and income elasticities play an important role when the policy makers make decisions on efficient use and allocation of water. This is especially important in developing countries and transition economies (see, e.g., da Motta, Huber, and Ruitenbeek (1998)). Furthermore, our model is beneficial to formulate the policy on population. The water and sewerage services are one of the factors that determine the population growth (see, e.g., Robinson (1997)).

Second, our model can incorporate a spatial dependency through the consumer heterogeneity. When we analyze the interregional residential water demand, it is important to control such a spatial dependency. The analysis of spatial dependency in the demand for public utilities would be a subject for future research.

Appendices

A.1 Increasing block rate pricing and its demand function

Figure 9(a) shows an example of a three-tier increasing block rate pricing where Y is the consumption of the good or service under increasing block rate pricing (such as water or



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

Figure 9: Three-Tier increasing block price structure.

electricity), P_k is the unit price of Y in block k ($k = 1, 2, 3$) and $\bar{Y}_k$ is the boundary quantity between block k and $k+1$, i.e., the upper limit of block k ($k = 1, 2$). Under this system, when consumption of Y exceeds $\bar{Y}_k$, the unit price jumps from P_k to P_{k+1}.

More generally, we consider the K-block increasing block rate pricing. Let P_k ($k = 1, \ldots, K$) and $\bar{Y}_k$ ($k = 0, \ldots, K$) be the unit price and the upper limit for the k-th block, noting that $P_k < P_{k+1}$ and $\bar{Y}_k < \bar{Y}_{k+1}$. We set $\bar{Y}_0 = 0$ and $\bar{Y}_K = \infty$ without loss of generality. In addition, the price schedule includes the fixed charge FC. A practical example of the fixed charge is a minimum access charge for water and electricity services.

When the consumption amount Y is within the k-th block ($\bar{Y}_{k-1} \leq Y < \bar{Y}_k$), the total payment for Y is given by

$$TC_k \equiv FC + \left\{ \sum_{j=1}^{k-1} P_j \left(\bar{Y}_j - \bar{Y}_{j-1} \right) \right\} + P_k \left(Y - \bar{Y}_k \right). \tag{A.25}$$

With the total payment, the budget constraint is given by

$$TC_k + Y_a \leq I \Longleftrightarrow P_k Y + Y_a \leq Q_k \equiv I - FC - \sum_{j=1}^{k-1} \left(P_j - P_{j+1} \right) \bar{Y}_j, \quad \text{if } \bar{Y}_{k-1} \leq Y < \bar{Y}_k, \tag{A.26}$$

for $k = 1, \ldots, K$, where I and Y_a are the total income and the demand for the numeraire, respectively. From this equation, it is clear that the budget line is piecewise-linear and $Q_k < Q_{k+1}$ (see also Figure 9(b)). The Q_k is called as the virtual income for the k-th block.

The demand function under increasing block rate pricing is derived by solving the utility maximization problem subject to the piecewise-linear budget constraint (A.26) through the discrete/continuous choice approach, and is given by

$$Y = \begin{cases} Y_k, & \text{if } \bar{Y}_{k-1} < Y_k < \bar{Y}_k \text{ and } k = 1, \ldots, K, \\ \bar{Y}_k, & \text{if } Y_{k+1} \le \bar{Y}_k \le Y_k \text{ and } k = 1, \ldots, K-1, \end{cases} \tag{A.27}$$

where Y_k is called as the conditional demand function for the k-th block and is the solution to the following utility maximization problem. Given P_k and Q_k,

$$\max_{Y_k, Y_a} U(Y_k, Y_a) \quad \text{s.t. } P_k Y_k + Y_a \le Q_k, \tag{A.28}$$

where $U(Y_k, Y_a)$ is the well-defined utility function. For the details of the derivation, see Moffitt (1986).

A.2 MCMC algorithm and its full conditional distributions for $M1$

The MCMC algorithm for the model with random individual effects is implemented in the following nine steps:

Step 1. Initialize $\boldsymbol{\beta}, \{\boldsymbol{\delta}_i, \boldsymbol{s}_i^*, \boldsymbol{w}_i^*\}_{i=1}^n, \sigma_u^2, \sigma_v^2, \boldsymbol{\mu}_{\boldsymbol{\delta}}$, *and* $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}$.

Step 2. Generate β_1 *given* $\beta_2, \left\{\boldsymbol{s}_i^*, \boldsymbol{w}_i^*\right\}_{i=1}^n, \sigma_u^2$.

Step 3. Generate β_2 *given* $\beta_1, \left\{\boldsymbol{s}_i^*, \boldsymbol{w}_i^*\right\}_{i=1}^n, \sigma_u^2$.

Step 4. Generate $\left(\sigma_v^2, \{\boldsymbol{\delta}_i\}_{i=1}^n\right)$ *given* $\left\{\boldsymbol{w}_i^*\right\}_{i=1}^n, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}$.

(a) Generate σ_v^2 *given* $\left\{\boldsymbol{w}_i^*\right\}_{i=1}^n, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}$.

(b) Generate $\boldsymbol{\delta}_i$ *given* $\left\{\boldsymbol{w}_i^*\right\}_{i=1}^n, \sigma_v^2, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}$ *for* $i = 1, \ldots, n$.

Step 5. Generate $\left(s_{it}^*, w_{it}^*\right)$ *given* $\boldsymbol{\beta}, \{\boldsymbol{\delta}_i\}_{i=1}^n, \sigma_u^2, \sigma_v^2$ *for* $i = 1, \ldots, n$ *and* $t = 1, \ldots, T$.

(a) Generate s_{it}^* given $\boldsymbol{\beta}, \boldsymbol{\delta}_i, \sigma_u^2, \sigma_v^2$.

(b) Generate w_{it}^* given $\boldsymbol{\beta}, \boldsymbol{\delta}_i, s_{it}^*, \sigma_u^2, \sigma_v^2$.

Step 6. Generate σ_u^2 given $\boldsymbol{\beta}, \left\{\boldsymbol{s}_i^*, \boldsymbol{w}_i^*\right\}_{i=1}^n$.

Step 7. Generate $\boldsymbol{\mu_\delta}$ given $\{\boldsymbol{\delta}_i\}_{i=1}^n, \boldsymbol{\Sigma_\delta}, \sigma_v^2$.

Step 8. Generate $\boldsymbol{\Sigma_\delta}$ given $\{\boldsymbol{\delta}_i\}_{i=1}^n, \boldsymbol{\mu_\delta}, \sigma_v^2$.

Step 9. Go to Step 2.

All full conditional distributions used in this algorithm are standard. Before describing full conditional distributions, we assume $p_{it,1} > 0$, $q_{it,1} > 0$, and $\bar{y}_{it,1} > 0$ to avoid tedious expressions depending on the sign of these variables without loss of generality. Let $k_{it} = \lceil s_{it}^*/2 \rceil$ and $\mathcal{A} = \{(i,t) \mid s_{it}^* \text{ is odd and equal to } 2k_{it} - 1 \text{ for } t = 1, \ldots, T\}$, where $\lceil x \rceil$ is the ceiling function returning the smallest integer that is larger than or equal to x. Then, the full conditional distributions are given by following each step of the algorithm.

Step 2. Generate β_1 given $\beta_2, \{\boldsymbol{s}_i^*, \boldsymbol{w}_i^*\}_{i=1}^n, \sigma_u^2$. The full conditional distribution for β_1 is the truncated normal distribution with mean μ_1, variance σ_1^2, and truncation interval R_1: $\beta_1 \sim TN_{R_1}(\mu_1, \sigma_1^2)$, where

$$\sigma_1^{-2} = \sigma_{\beta_1,0}^{-2} + \sum_{(i,t)\in\mathcal{A}} (p_{it,k_{it}})^2, \tag{A.29}$$

$$\mu_1 = \sigma_1^2 \left\{ \sigma_{\beta_1,0}^{-2} \mu_{\beta_1,0} + \sum_{(i,t)\in\mathcal{A}} p_{it,k_{it}} \left(y_{it} - \beta_2 q_{it,k_{it}} - w_{it}^* \right) \right\}, \tag{A.30}$$

$$R_1 = \left\{ \max_{i,t} \left(-\infty, BL_{it}^1 \right),\ \min_{i,t,k} \left(BU_{it}^1, -\beta_2 \frac{q_{it,k+1} - q_{it,k}}{p_{it,k+1} - p_{it,k}} \right) \right\}, \tag{A.31}$$

$$\left(BL_{it}^1, BU_{it}^1 \right) = \begin{cases} \left(\frac{\bar{y}_{it,k-1} - \beta_2 q_{it,k} - w_{it}^*}{p_{it,k}}, \frac{\bar{y}_{it,k} - \beta_2 q_{it,k} - w_{it}^*}{p_{it,k}} \right), & \text{if } (i,t) \in \mathcal{A}, \\ \left(\frac{\bar{y}_{it,k} - \beta_2 q_{it,k} - w_{it}^*}{p_{it,k}}, \frac{\bar{y}_{it,k} - \beta_2 q_{it,k+1} - w_{it}^*}{p_{it,k+1}} \right), & \text{otherwise.} \end{cases} \tag{A.32}$$

These (BL_{it}^1, BU_{it}^1) are constructed from the intervals R_{it,s_{it}^*} defined by equations (7) of Subsection 2.

Step 3. Generate β_2 given $\beta_1, \{\boldsymbol{s}_i^*, \boldsymbol{w}_i^*\}_{i=1}^n, \sigma_u^2$. The full conditional distribution for β_2 is the truncated normal distribution, $\beta_2 \sim TN_{R_2}(\mu_2, \sigma_2^2)$, where

$$\sigma_2^{-2} = \sigma_{\beta_2,0}^{-2} + \sum_{(i,t)\in\mathcal{A}} (q_{it,k_{it}})^2, \tag{A.33}$$

$$\mu_2 = \sigma_2^2 \left\{ \sigma_{\beta_2,0}^{-2}\mu_{\beta_2,0} + \sum_{(i,t)\in\mathcal{A}} q_{it,k_{it}} \left(y_{it} - \beta_1 p_{it,k_{it}} - w_{it}^* \right) \right\}, \tag{A.34}$$

$$R_2 = \left\{ \max_{i,t}\left(-\infty, BL_{it}^2\right),\ \min_{i,t,k}\left(BU_{it}^2, -\beta_1 \frac{p_{it,k+1} - p_{it,k}}{q_{it,k+1} - q_{it,k}} \right) \right\}, \tag{A.35}$$

$$\left(BL_{it}^2, BU_{it}^2\right) = \begin{cases} \left(\frac{\bar{y}_{it,k-1} - \beta_1 p_{it,k} - w_{it}^*}{q_{it,k}}, \frac{\bar{y}_{it,k} - \beta_1 p_{it,k} - w_{it}^*}{q_{it,k}} \right), & \text{if } (i,t) \in \mathcal{A}, \\ \left(\frac{\bar{y}_{it,k} - \beta_1 p_{it,k} - w_{it}^*}{q_{it,k}}, \frac{\bar{y}_{it,k} - \beta_1 p_{it,k+1} - w_{it}^*}{q_{it,k+1}} \right), & \text{otherwise.} \end{cases} \tag{A.36}$$

Step 4. Generate $(\sigma_v^2, \{\boldsymbol{\delta}_i\}_{i=1}^n)$ given $\{\boldsymbol{w}_i^*\}_{i=1}^n, \boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}$. Integrating the joint full conditional probability density of $(\sigma_v^2, \{\boldsymbol{\delta}_i\}_{i=1}^n)$ with respect to $\{\boldsymbol{\delta}_i\}_{i=1}^n$, we have the full conditional distribution of σ_v^2 as the inverse gamma distribution, $\sigma_v^2 \sim IG(n_{v,1}/2, S_{v,1}/2)$. Then, the full conditional distribution of $\boldsymbol{\delta}_i$ is the multivariate normal distribution, $\boldsymbol{\delta}_i | \sigma_v^2 \sim N_d(\boldsymbol{\mu}_{\boldsymbol{\delta}_i,1}, \sigma_v^2 \boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,1})$. Parameters of these full conditionals are $n_{v,1} = n_{v,0} + nT$,

$$S_{v,1} = S_{v,0} + n\boldsymbol{\mu}_{\boldsymbol{\delta}}' \boldsymbol{\Sigma}_{\boldsymbol{\delta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\delta}} + \sum_{i=1}^n \left(\boldsymbol{w}_i^{*\prime} \boldsymbol{w}_i^* - \boldsymbol{\mu}_{\boldsymbol{\delta}_i,1}' \boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,1}^{-1} \boldsymbol{\mu}_{\boldsymbol{\delta}_i,1} \right), \tag{A.37}$$

$$\boldsymbol{\mu}_{\boldsymbol{\delta}_i,1} = \boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,1} \left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\delta}} + \boldsymbol{Z}_i' \boldsymbol{w}_i^* \right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,1}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\delta}}^{-1} + \boldsymbol{Z}_i' \boldsymbol{Z}_i. \tag{A.38}$$

Step 5. Generate (s_{it}^*, w_{it}^*) given $\boldsymbol{\beta}, \{\boldsymbol{\delta}_i\}_{i=1}^n, \sigma_u^2, \sigma_v^2$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The full conditional distribution of s_i^* is the multinomial distribution. Its probability mass function is given by

$$\pi\left(s_{it}^* = s \mid \boldsymbol{\beta}, \{\boldsymbol{\delta}\}_{i=1}^n, \sigma_u^2, \sigma_v^2\right) \propto \tau_s \left[\Phi\left\{ \tau_s^{-1} \left(RU_{it,s} - \theta_{it,s} \right) \right\} - \Phi\left\{ \tau_s^{-1} \left(RL_{it,s} - \theta_{it,s} \right) \right\} \right] \exp\left(-\frac{m_{it,s}}{2} \right), \tag{A.39}$$

for $s = 1,\ldots,2K^{it}-1$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, $RU_{it,s}$ and $RL_{it,s}$ denote the respective upper and lower limits of $R_{it,s}$ (see (7)), and

$$\left(m_{it,s},\theta_{it,s},\tau_s^2\right)=\begin{cases}\left(\dfrac{\sigma_u^{-2}\sigma_v^{-2}\left(y_{it}-\boldsymbol{x}'_{it,k}\boldsymbol{\beta}-\boldsymbol{z}'_{it}\boldsymbol{\delta}_i\right)^2}{\sigma_u^{-2}+\sigma_v^{-2}},\ \dfrac{\sigma_u^{-2}\left(y_{it}-\boldsymbol{x}'_{it,k}\boldsymbol{\beta}\right)+\sigma_v^{-2}\boldsymbol{z}'_{it}\boldsymbol{\delta}_i}{\sigma_u^{-2}+\sigma_v^{-2}},\ \left(\sigma_v^{-2}+\sigma_u^{-2}\right)^{-1}\right), & \\ \qquad\qquad \text{if } s=2k-1 \text{ and } k=1,\ldots,K^{it}, & \\ \left(\sigma_u^{-2}\left(y_{it}-\bar{y}_{it,k}\right)^2,\ \boldsymbol{z}'_{it}\boldsymbol{\delta}_i,\ \sigma_v^2\right),\quad \text{if } s=2k \text{ and } k=1,\ldots,K^{it}-1. &\end{cases} \tag{A.40}$$

Given $s^*_{it}=s$, we generate w^*_{it} from the truncated normal distribution, $w^*_{it}|s^*_{it}=s\sim TN_{R_{it,s}}(\theta_{it,s},\tau_s^2)$.

Step 6. Generate σ_u^2 given $\boldsymbol{\beta},\{\boldsymbol{s}^*_i,\boldsymbol{w}^*_i\}_{i=1}^n$. The full conditional distribution of σ_u^2 is the inverse gamma distribution, $\sigma_u^2\sim IG(n_{u,1}/2,S_{u,1}/2)$, where $n_{u,1}=n_{u,0}+2+nT$ and

$$S_{u,1}=S_{u,0}+\left(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta},0}\right)'\boldsymbol{\Sigma}_{\boldsymbol{\beta},0}^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta},0}\right)+\sum_{i=1}^n\left(\boldsymbol{y}_i-\boldsymbol{y}^*_i\right)'\left(\boldsymbol{y}_i-\boldsymbol{y}^*_i\right). \tag{A.41}$$

Step 7. Generate $\boldsymbol{\mu}_{\boldsymbol{\delta}}$ given $\{\boldsymbol{\delta}_i\}_{i=1}^n,\boldsymbol{\Sigma}_{\boldsymbol{\delta}},\sigma_v^2$. The full conditional distribution of $\boldsymbol{\mu}_{\boldsymbol{\delta}}$ is the multivariate normal distribution, $\boldsymbol{\mu}_{\boldsymbol{\delta}}\sim N_d(\boldsymbol{\mu}_{\bar{\boldsymbol{\delta}},1},\boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},1})$, where

$$\boldsymbol{\mu}_{\bar{\boldsymbol{\delta}},1}=\boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},1}\left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},0}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{\delta}},0}+\sigma_v^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\delta}}^{-1}\sum_{i=1}^n\boldsymbol{\delta}_i\right),\quad \boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},1}^{-1}=\boldsymbol{\Sigma}_{\bar{\boldsymbol{\delta}},0}^{-1}+n\sigma_v^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\delta}}^{-1}. \tag{A.42}$$

Step 8. Generate $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}$ given $\{\boldsymbol{\delta}_i\}_{i=1}^n,\boldsymbol{\mu}_{\boldsymbol{\delta}},\sigma_v^2$. The full conditional distribution of $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}$ is the inverse Wishart distribution, $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}\sim IW_d(n_{\bar{\boldsymbol{\delta}},1},\boldsymbol{S}_{\bar{\boldsymbol{\delta}},1})$, where $n_{\bar{\boldsymbol{\delta}},1}=n_{\bar{\boldsymbol{\delta}},0}+n$ and

$$\boldsymbol{S}_{\bar{\boldsymbol{\delta}},1}^{-1}=\boldsymbol{S}_{\bar{\boldsymbol{\delta}},0}^{-1}+\sigma_v^{-2}\sum_{i=1}^n(\boldsymbol{\delta}_i-\boldsymbol{\mu}_{\boldsymbol{\delta}})(\boldsymbol{\delta}_i-\boldsymbol{\mu}_{\boldsymbol{\delta}})'. \tag{A.43}$$

A.3 MCMC algorithm and its full conditional distributions for $M2$

The MCMC algorithm for the model with fixed individual effects is almost similar with the one with random individual effects given in the previous subsection. Step 1 and 4 are modified and Step 7 and 8 are removed according to this prior specification.

Step 1. Initialize $\boldsymbol{\beta}, \left\{\boldsymbol{\delta}_i, \boldsymbol{s}_i^*, \boldsymbol{w}_i^*\right\}_{i=1}^n, \sigma_u^2$, *and* σ_v^2.

Step 4. Generate $\left(\sigma_v^2, \{\boldsymbol{\delta}_i\}_{i=1}^n\right)$ *given* $\left\{\boldsymbol{w}_i^*\right\}_{i=1}^n$.

(a) Generate σ_v^2 *given* $\left\{\boldsymbol{w}_i^*\right\}_{i=1}^n$.

(b) Generate $\boldsymbol{\delta}_i$ *given* $\left\{\boldsymbol{w}_i^*\right\}_{i=1}^n, \sigma_v^2$ *for* $i = 1, \ldots, n$.

The full conditional distributions for Step 4 are derived in a similar manner. The the full conditional distributions of σ_v^2 and $\boldsymbol{\delta}_i$ are the inverse gamma distribution, $\sigma_v^2 \sim IG(n_{v,1}/2, S_{v,1}/2)$, and the multivariate normal distribution, $\boldsymbol{\delta}_i|\sigma_v^2 \sim N_d(\boldsymbol{\mu}_{\boldsymbol{\delta}_i,1}, \sigma_v^2\boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,1})$, where $n_{v,1} = n_{v,0} + nT$,

$$S_{v,1} = S_{v,0} + n\boldsymbol{\mu}_{\boldsymbol{\delta}_i,0}'\boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,0}^{-1}\boldsymbol{\mu}_{\boldsymbol{\delta}_i,0} + \sum_{i=1}^n \left(\boldsymbol{w}_i^{*\prime}\boldsymbol{w}_i^* - \boldsymbol{\mu}_{\boldsymbol{\delta}_i,1}'\boldsymbol{\Sigma}_{\boldsymbol{\delta}_i,1}^{-1}\boldsymbol{\mu}_{\boldsymbol{\delta}_i,1}\right), \tag{A.44}$$

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

Acknowledgment

We highly appreciate beneficial comments and suggestions from two anonymous referees. This work is supported by the Grants-in-Aid for Scientific Research 21243018 from the Japanese Ministry of Education, Science, Sports, Culture and Technology and by the Academic Frontier Project of the Nihon University Population Research Institute. The authors thank H. Kozumi (Kobe University) and K. Kakamu (Chiba University) for their valuable comments. All computational results were obtained using Ox for Linux (see Doornik, 2002).

References

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

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies 47*, 225–238.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, Chapter 22, pp. 1247–1318. Amsterdam: North-Holland.

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.

da Motta, R. S., R. M. Huber, and H. J. Ruitenbeek (1998). Market based instruments for environmental policymaking in latin america and the caribbean: lessons from eleven countries. *Environment and Development Economics 4*, 177–201.

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.

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

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.

Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distributions*, Volume 104 of *Monographs and Surveys in Pure and Applied Mathematics*. Boca Raton: Chapman and Hall.

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

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.

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.

Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological) 34*(1), 1–41.

Miyawaki, K., Y. Omori, and A. Hibiki (2010). Bayesian estimation of demand functions under block rate pricing. University of Tokyo CIRJE Discussion Paper Series CIRJE-F-712, Revised version of CIRJE-F-424, CIRJE-F-568, and CIRJE-F-631.

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

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Olmstead, S. M. (2009). Reduced-form vs. structural models of water demand under nonlinear prices. *Journal of Business and Economic Statistics 87*(1), 84–94.

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.

Pint, E. M. (1999). Household responses to increased water rates during the California drought. *Land Economics 75*, 246–266.

Rietveld, P., J. Rouwendal, and B. Zwart (2000). Block rate pricing of water in Indonesia: An analysis of welfare effects. *Bulletin of Indonesian Economic Studies 36*(3), 73–92.

Robinson, J. A. (1997). Long-term consequences of population growth: Technological change, natural resources, and the environment. In M. R. Rosenzweig and O. Stark (Eds.), *Handbook of Population and Family Economics*, Volume 1B, Chapter 21, pp. 1175–1298. Amsterdam: North-Holland.

Smith, A. F. M. (1973). A general bayesian linear model. *Journal of the Royal Statistical Society. Series B (Methodological) 35*(1), 67–75.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measure of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 64*(4), 583–639.

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 American Statistical Association 82*(398), 528–540.