

Department of Social Systems and Management

Discussion Paper Series

No.1289

**Estimating the Markov-Switching
Almost Ideal Demand Systems:
a Bayesian Approach**

by

Satoshi KABE and Yuichiro KANAZAWA

March 2012

UNIVERSITY OF TSUKUBA

Tsukuba, Ibaraki 305-8573

JAPAN

Estimating the Markov-Switching Almost Ideal Demand Systems: A Bayesian approach

Satoshi KABE * Yuichiro KANAZAWA †

Abstract

Allais and Nichèle (2007) proposed a Markov-switching almost ideal demand system (MS-AIDS) model by extending the idea of Hamilton (1989). This model enables us to determine when the regime shifts occurred and to estimate parameters characterized across the different regimes. Moreover, degree of belongingness to each of the regimes and transitions between regimes are quantified by the probabilities. In this paper, we propose a Bayesian estimation for MS-AIDS model and illustrate applicability of our proposed method. The Bayesian estimation has some important advantages. First, it enables us to avoid the singularity problem suggested by Hamilton (1990, 1991). Second, our proposed Bayesian estimation ensures that transition probabilities lie between zero and one. Third, Bayesian estimation is able to avoid the messy calculations entailed in the score functions of log-likelihood. We then run a simulation study to confirm the validity of the proposed method. In the empirical study on the Japanese meat market, we found that our Bayesian estimation improves the mean squared errors for all meat products over the maximum likelihood estimation, while successfully capturing the regime shifts of meat demand coinciding with the timing of Bovine Spongiform Encephalopathy (BSE) cases in Japan and U.S.

Keywords: Bayesian estimation, Gibbs sampler, Markov-switching model,
Almost ideal demand system, Japanese meat market

JEL codes: C11, C51, D12, Q11

*Doctoral Student, Graduate School of Systems and Information, University of Tsukuba. E-mail: k0420214@sk.tsukuba.ac.jp.

†Professor of Statistics, Department of Social Systems and Management, University of Tsukuba. Email: kanazawa@sk.tsukuba.ac.jp.

1 Introduction

In empirical demand analysis, the almost ideal demand system (AIDS) model proposed by Deaton and Muellbauer (1980) has been widely used. This model assumes that market demand can be thought of as if it were the outcome of decisions by a rational representative consumer. The AIDS model is able to test the homogeneity and symmetry restrictions of demand theory as the Rotterdam model (Theil, 1965) is able to and applies the second-order approximation to any arbitrary functions in the PIGLOG expenditure function such as the Translog model (Christensen et al., 1975). Several researchers made modifications to the original AIDS model. For instance, Cooper and McLaren (1992) modified the model so that the range restrictions placed on the budget share can be reflected on the model. Banks et al. (1997) also made modification to the linear relationship between the expenditure and the demand. These models have also been employed extensively.

In the previous studies on the structural shifts in demand, Moosa and Baxter (2002) developed the time-varying coefficient AIDS model. They introduced the stochastic trend and seasonality terms into the linear approximated AIDS (LA-AIDS) model so that it can be applied to unstable demand structure in the alcoholic beverage market in the U.K. Ishida et al. (2006, 2010) employed the gradual switching AIDS model. They set the transition functions into the AIDS model to capture the gradual shifts following Bovine Spongiform Encephalopathy (henceforth BSE) and bird flu outbreaks in the Japanese meat market. The latter model assumes that researchers know the structural change points in demand in advance.

Obviously modeling abrupt changes in demand caused by the unique exogenous events and estimating these change points are the next step. Allais and Nichèle (2007) seem to be the first to propose a Markov-switching almost ideal demand system (MS-AIDS) model extending the idea of Hamilton (1989)¹. This model enables us to determine when the regime shifts occurred and to estimate parameters characterized across the different regimes. Moreover, degree of belongingness to each of the regimes and transitions between regimes are quantified by the probabilities. They analyzed the French meat and fish demands over the period 1991 - 2002 and detected the abrupt changes due to the two BSE crises in France. Kabe and Kanazawa (2012) also assessed the structural change points in the Japanese meat market during 1998 - 2006 via MS-AIDS model. They found the structural change point coincid-

¹Hamilton (1989) proposed the Markov-switching model to date the timing of recessions and booms with real gross national product (GNP) data in U.S. He found that the regime shift from positive to negative growth rate has a recurrent feature of the U.S. business cycle.

ing with the timing of first reported case of BSE, but not of bird flu. In both of these instances, MS-AIDS model is found to be quite effective in detecting abrupt changes in demand using monthly aggregate data.

In Allais and Nichèle (2007), they estimate the parameters including transition probabilities via maximum likelihood (ML) estimation. However, when the variance-covariance matrices differ between regimes, a singularity problem arises when the determinant of variance-covariance matrix is close to zero, sending log-likelihood of MS-AIDS model to infinity, and making numerical optimization methods (e.g., Newton-Raphson method) break down. This problem is well-known in the literature on estimation of mixture of normal distributions. They also estimate the transition probabilities via ML estimation without any constraints, although the transition probabilities have to lie between zero and one inclusive.

To avoid the singularity problem, subjective judgment is required in deciding what constitutes a suitable region for plausible value of the variance-covariance matrices, so that Hamilton (1990, 1991) suggests the Bayesian estimation as a simple solution of the singularity problem. Hamilton (1991, p.37) stated that “the [Bayesian] approach is intuitively appealing and trivial to implement. Monte Carlo analysis suggests that this approach can consistently improve the MSE’s for a wide variety of underlying models.”

Bayesian estimation enables us to incorporate the prior information on the variance-covariance matrices to the conjugate prior distributions. Moreover, Bayesian estimation can provide us with the posterior distributions of transition probabilities in the unit intervals, avoiding problems associated with unconstrained ML estimation. Finally, Bayesian estimation is able to replace the messy calculations entailed in the score functions of log-likelihood for MS-AIDS model with computationally simple Gibbs sampler. To the best of our knowledge, no Bayesian estimation method is proposed to solve these problems associated with the MS-AIDS model.

In this paper, we propose a Bayesian method to estimate parameters in MS-AIDS model along with the transition probabilities. We then run a simulation study to confirm the validity of the proposed Bayesian method. To illustrate its applicability, we take the proposed method to the Japanese meat market data and examine the regime shifts caused by the food safety concerns such as BSE and bird flu.

The rest of this paper is organized as follows. Section 2 briefly describes the Markov-switching AIDS model and introduces the necessary notations. Then section 3 proposes the Bayesian estimation, and we illustrate our proposed method by simulation in section 4. Section 5 presents the empirical study on the Japanese meat market via the proposed Bayesian estimation method. Finally, section 6 discusses the merits of the proposed Bayesian

method relative to the ML estimation we employed in Kabe and Kanazawa (2012), and then we point to future directions of the research.

2 Markov-Switching AIDS model

Suppose that s_t is an unobserved random variable that takes an integer value in $1, 2, \dots, K$ to express “regime” or “state” at time t , then budget share of i -th product at time t , \bar{w}_{it} which is defined as $p_{it}q_{it}/m_{0t}$ with price p_{it} , quantity q_{it} and expenditure (or budget) m_{0t} ($= \sum_i p_{it}q_{it}$) takes the following form:

$$\bar{w}_{it} = \alpha_{i,s_t} + \sum_{j=1}^N \gamma_{ij,s_t} \log p_{jt} + \beta_{i,s_t} \log \left(\frac{m_{0t}}{P_t} \right) \quad (2.1)$$

where P_t is a price index which is defined by

$$\log P_t = \alpha_{0,s_t} + \sum_{k=1}^N \alpha_{k,s_t} \log p_{kt} + \frac{1}{2} \sum_{k=1}^N \sum_{j=1}^N \gamma_{kj,s_t} \log p_{kt} \log p_{jt} \quad (2.2)$$

and α_{0,s_t} , α_{i,s_t} , γ_{ij,s_t} and β_{i,s_t} ($i, j = 1, 2, \dots, N$) are regime-dependent parameters.

The parameters in (2.1) and (2.2) have the theoretical constraints² as follows

$$[\text{Adding up}] \quad \sum_{i=1}^N \alpha_{i,s_t} = 1, \quad \sum_{i=1}^N \gamma_{ij,s_t} = 0, \quad \sum_{i=1}^N \beta_{i,s_t} = 0, \quad (2.3a)$$

$$[\text{Homogeneity}] \quad \sum_{j=1}^N \gamma_{ij,s_t} = 0, \quad (2.3b)$$

$$[\text{Symmetry}] \quad \gamma_{ij,s_t} = \gamma_{ji,s_t}. \quad (2.3c)$$

Following the previous studies (Rickertsen, 1996; Allais and Nichèle, 2007; Ishida et al., 2010), we include a trend effect, seasonal effect and habit effect into the intercept term α_{i,s_t} as

$$\alpha_{i,s_t} = \bar{\alpha}_{i,s_t} + \nu_{i,s_t} t + \delta_{1,i} d_{1,t} + \delta_{2,i} d_{2,t} + \sum_{j=1}^N \phi_{ij} \bar{w}_{j,t-1} \quad (2.4)$$

²“Adding up” guarantees that the total expenditure is equal to the sum of expenditures on the category of products under consideration. “Homogeneity” guarantees that if prices of products increase to $\tau p_{1t}, \dots, \tau p_{Nt}$ for a scalar $\tau > 0$, representative consumer has to increase his expenditure from m_{0t} to τm_{0t} to keep his utility level. “Symmetry” guarantees that the substitution effect in the Slutsky equation is symmetric.

where $d_{1,t}$ and $d_{2,t}$ are dummy variables

$$d_{1,t} = \begin{cases} 1 & \text{if } t \text{ is August} \\ 0 & \text{otherwise} \end{cases} \quad d_{2,t} = \begin{cases} 1 & \text{if } t \text{ is December} \\ 0 & \text{otherwise.} \end{cases}$$

As for seasonal effect, we set the dummy variables to adjust the seasonality in budget shares. The budget shares for meat and fish are considered to shift due to the seasonal habits (e.g., summer camp, gift-giving tradition, year-end party and so forth) in August and December. Furthermore, we include a habit effect which is defined as a linear function of one-lagged budget shares (Rickertsen, 1996; Allais and Nichèle, 2007). In order to satisfy the adding up condition, we impose the restriction $\sum_{i=1}^N \bar{\alpha}_{i,st} = 1$, $\sum_{i=1}^N \nu_{i,st} = 0$, $\sum_{i=1}^N \delta_{1,i} = \sum_{i=1}^N \delta_{2,i} = 0$ and $\sum_{i=1}^N \phi_{ij} = 0$. We also impose the restriction $\sum_{j=1}^N \phi_{ij} = 0$ to avoid the identification problem.

Using the theoretical constraints in (2.3a), (2.3b) and (2.3c), the MS-AIDS model (2.1) can be rewritten as

$$\bar{w}_{it} = \alpha_{i,st} + \sum_{j=1}^{N-1} \gamma_{ij,st} \log \left(\frac{p_{jt}}{p_{Nt}} \right) + \beta_{i,st} \log \left(\frac{m_{0t}}{P_t} \right) \quad (2.5)$$

where $i = 1, 2, \dots, N-1$. Imposing the restriction $\sum_{j=1}^N \phi_{ij} = 0$, intercept term $\alpha_{i,st}$ in (2.5) is expressed as

$$\alpha_{i,st} = \bar{\alpha}_{i,st} + \nu_{i,st}t + \delta_{1,i}d_{1,t} + \delta_{2,i}d_{2,t} + \sum_{j=1}^{N-1} \phi_{ij}(\bar{w}_{j,t-1} - \bar{w}_{N,t-1}).$$

The MS-AIDS model employs the Markov switching mechanism which is developed by Hamilton (1989). The Markov switching mechanism can express switching of regimes by using the unobserved random variables that follow the Markov process. To apply the Markov switching mechanism, we assume that transitions between regimes are governed by a K -state Markov chain with transition probabilities:

$$\Pr(s_t = j | s_{t-1} = i) = \pi_{ij}, \quad i, j = 1, 2, \dots, K \quad (2.6)$$

and the transition matrix is defined as

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{21} & \dots & \pi_{K1} \\ \pi_{12} & \pi_{22} & \dots & \pi_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{1K} & \pi_{2K} & \dots & \pi_{KK} \end{bmatrix} \quad (2.7)$$

where $\pi_{i1} + \pi_{i2} + \dots + \pi_{iK} = 1$, $i = 1, 2, \dots, K$.

3 Bayesian Estimation

Let \mathbf{w}_t be a $(N-1) \times 1$ vector of budget shares at time t , \bar{w}_{it} ($i = 1, 2, \dots, N-1$) and we define the matrix of explanatory variables for regime-dependent parameters $\bar{\alpha}_{i,s_t}, \gamma_{i1,s_t}, \gamma_{i2,s_t}, \dots, \gamma_{iN-1,s_t}, \beta_{i,s_t}, \nu_{i,s_t}$ as $\mathbf{X}_t^{(1)}$ and for regime-independent parameters $\delta_{1,i}, \delta_{2,i}, \phi_{i1}, \phi_{i2}, \dots, \phi_{i,N-1}$ as $\mathbf{X}_t^{(0)}$.

Given the value of price index (2.2), the MS-AIDS model (2.5) can be first rewritten by separating the parts that depend on regimes and by including the error term ε_{it} as

$$\begin{aligned} \bar{w}_{it} = & \bar{\alpha}_{i,s_t} + \sum_{j=1}^{N-1} \gamma_{ij,s_t} \log\left(\frac{p_{jt}}{p_{Nt}}\right) + \beta_{i,s_t} \log\left(\frac{m_{0t}}{P_t}\right) + \nu_{i,s_t} t \\ & + \delta_{1,i} d_{1,t} + \delta_{2,i} d_{2,t} + \sum_{j=1}^{N-1} \phi_{ij} (\bar{w}_{j,t-1} - \bar{w}_{N,t-1}) + \varepsilon_{it} \end{aligned} \quad (3.1)$$

and thus can further be rewritten as the matrix form:

$$\mathbf{w}_t = \mathbf{X}_t^{(1)} \boldsymbol{\theta}_{s_t} + \mathbf{X}_t^{(0)} \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t \quad (3.2)$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{s_t})$ and $\boldsymbol{\Sigma}_{s_t}$ is also regime-dependent parameter such that $\boldsymbol{\Sigma}_{s_t} = \boldsymbol{\Sigma}_j$ if time t belongs to regime j . The size of the matrices $\mathbf{X}_t^{(1)}$ and $\mathbf{X}_t^{(0)}$ are $(N-1) \times [3(N-1) + N(N-1)/2]$ and $(N-1) \times (N-1)(N+1)$.

Example

Let us consider the case that the number of products N is four. Then 3×15 matrix $\mathbf{X}_t^{(1)}$ is defined as

$$\mathbf{X}_t^{(1)} \equiv [\mathbf{I}_3 \quad \mathbf{P}_t \quad \mathbf{M}_t \quad \mathbf{T}_t]$$

where \mathbf{I}_3 is a 3×3 identity matrix,

$$\mathbf{M}_t \equiv \begin{bmatrix} \log\left(\frac{m_{0t}}{P_t}\right) & & 0 \\ & \log\left(\frac{m_{0t}}{P_t}\right) & \\ 0 & & \log\left(\frac{m_{0t}}{P_t}\right) \end{bmatrix}, \quad \mathbf{T}_t \equiv \begin{bmatrix} t & 0 \\ & t \\ 0 & & t \end{bmatrix},$$

and

$$\mathbf{P}_t \equiv \begin{bmatrix} \log\left(\frac{p_{1t}}{p_{4t}}\right) & \log\left(\frac{p_{2t}}{p_{4t}}\right) & \log\left(\frac{p_{3t}}{p_{4t}}\right) & 0 & 0 & 0 \\ 0 & \log\left(\frac{p_{1t}}{p_{4t}}\right) & 0 & \log\left(\frac{p_{2t}}{p_{4t}}\right) & \log\left(\frac{p_{3t}}{p_{4t}}\right) & 0 \\ 0 & 0 & \log\left(\frac{p_{1t}}{p_{4t}}\right) & 0 & \log\left(\frac{p_{2t}}{p_{4t}}\right) & \log\left(\frac{p_{3t}}{p_{4t}}\right) \end{bmatrix}.$$

The 15 element parameter vector $\boldsymbol{\theta}_{s_t}$ is defined as

$$\boldsymbol{\theta}_{s_t} \equiv \begin{bmatrix} \bar{\boldsymbol{\alpha}}_{s_t} \\ \boldsymbol{\gamma}_{s_t} \\ \boldsymbol{\beta}_{s_t} \\ \boldsymbol{\nu}_{s_t} \end{bmatrix}$$

where $\bar{\boldsymbol{\alpha}}_{s_t} \equiv [\bar{\alpha}_{1,s_t} \bar{\alpha}_{2,s_t} \bar{\alpha}_{3,s_t}]'$, $\boldsymbol{\gamma}_{s_t} \equiv [\gamma_{11,s_t} \gamma_{12,s_t} \gamma_{13,s_t} \gamma_{22,s_t} \gamma_{23,s_t} \gamma_{33,s_t}]'$, $\boldsymbol{\beta}_{s_t} \equiv [\beta_{1,s_t} \beta_{2,s_t} \beta_{3,s_t}]'$ and $\boldsymbol{\nu}_{s_t} \equiv [\nu_{1,s_t} \nu_{2,s_t} \nu_{3,s_t}]'$.

The 3×15 matrix $\mathbf{X}_t^{(0)}$ is defined as

$$\mathbf{X}_t^{(0)} \equiv [\mathbf{D}_{1t} \quad \mathbf{D}_{2t} \quad \mathbf{W}_{1t} \quad \mathbf{W}_{2t} \quad \mathbf{W}_{3t}]$$

where

$$\mathbf{D}_{1t} \equiv \begin{bmatrix} d_{1,t} & 0 \\ & d_{1,t} \\ 0 & d_{1,t} \end{bmatrix}, \quad \mathbf{D}_{2t} \equiv \begin{bmatrix} d_{2,t} & 0 \\ & d_{2,t} \\ 0 & d_{2,t} \end{bmatrix},$$

and

$$\mathbf{W}_{jt} \equiv \begin{bmatrix} \bar{w}_{j,t-1} - \bar{w}_{4,t-1} & & 0 \\ & \bar{w}_{j,t-1} - \bar{w}_{4,t-1} & \\ 0 & & \bar{w}_{j,t-1} - \bar{w}_{4,t-1} \end{bmatrix}.$$

The 15 element parameter vector $\boldsymbol{\theta}_0$ is defined as

$$\boldsymbol{\theta}_0 \equiv \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \\ \boldsymbol{\phi}_3 \end{bmatrix}$$

where $\boldsymbol{\delta}_1 \equiv [\delta_{11} \delta_{12} \delta_{13}]'$, $\boldsymbol{\delta}_2 \equiv [\delta_{21} \delta_{22} \delta_{23}]'$ and $\boldsymbol{\phi}_j \equiv [\phi_{1j} \phi_{2j} \phi_{3j}]'$.

Since the MS-AIDS model is parameterized nonlinear due to the price index (2.2), the estimate cannot be written as closed form. Instead, we assume that value of price index is already known and draw samples of parameters via Gibbs sampler. Afterward, these samples are used to update the value of price index and then samples of parameters are generated with the new price index. We repeat this process until convergence.

To obtain the likelihood function of MS-AIDS model, we denote the set of variables obtained from $t = 1$ through time t as

$$\mathcal{Y}_t \equiv \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t\}, \quad \mathcal{S}_t \equiv \{s_1, s_2, \dots, s_t\}, \\ \mathcal{X}_t \equiv \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$$

where \mathbf{x}_t is a $1 \times (2N + 3)$ vector of explanatory variables at time t in (3.1): $\mathbf{x}_t \equiv [1 \ \log(p_{1t}/p_{Nt}) \ \cdots \ \log(p_{N-1t}/p_{Nt}) \ \log(m_{0t}/P_t) \ t \ d_{1,t} \ d_{2,t} \ \bar{w}_{1,t-1} - \bar{w}_{N,t-1} \ \cdots \ \bar{w}_{N-1,t-1} - \bar{w}_{N,t-1}]$.

Then likelihood function $\mathcal{L}(\cdot|\cdot)$ is defined as

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\pi}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) = \mathcal{L}(\boldsymbol{\pi}|\mathcal{S}_T)\mathcal{L}(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) \quad (3.3)$$

where $\boldsymbol{\theta} \equiv \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_K, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_K\}$ and $\boldsymbol{\pi} \equiv \{\pi_{ij} : i, j = 1, 2, \dots, K\}$. Given a prior distribution $p(\boldsymbol{\theta}, \boldsymbol{\pi}) = p(\boldsymbol{\theta})p(\boldsymbol{\pi})$ ³, we obtain the posterior distributions with respect to $\boldsymbol{\theta}$ and to $\boldsymbol{\pi}$ as

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\pi}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) &\propto \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\pi}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)p(\boldsymbol{\theta}, \boldsymbol{\pi}) \\ &= \mathcal{L}(\boldsymbol{\pi}|\mathcal{S}_T)p(\boldsymbol{\pi}) \times \mathcal{L}(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)p(\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\pi}|\mathcal{S}_T) \times p(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T). \end{aligned} \quad (3.4)$$

Now we compute each of the terms on the right hand side of (3.4). To be able to do this, we first need to generate discrete latent variables s_1, s_2, \dots, s_T to represent regimes.

3.1 Sampling of latent variables s_1, s_2, \dots, s_T

Since $\mathcal{S}_T \equiv \{s_1, s_2, \dots, s_T\}$ is a sequence of unobservable finite discrete random variables, we need to generate samples s_1, s_2, \dots, s_T to compute the posterior distributions in (3.4). To generate samples of latent variables s_1, s_2, \dots, s_T , we apply the multi-move sampler (e.g., Carter and Kohn, 1994; Chib, 1996): Given the data obtained through time t , $\boldsymbol{\Omega}_t \equiv \{\mathcal{Y}_t, \mathcal{X}_t\}$ and set of parameters $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}, \boldsymbol{\pi}\}$, we consider a joint distribution $f(\mathcal{S}_T|\boldsymbol{\Omega}_T, \boldsymbol{\Theta})$ as

$$\begin{aligned} f(\mathcal{S}_T|\boldsymbol{\Omega}_T, \boldsymbol{\Theta}) &= f(s_1, s_2, \dots, s_T|\boldsymbol{\Omega}_T, \boldsymbol{\Theta}) \\ &= \Pr(s_T|\boldsymbol{\Omega}_T, \boldsymbol{\Theta}) \Pr(s_{T-1}|s_T, \boldsymbol{\Omega}_{T-1}, \boldsymbol{\Theta}) \cdots \Pr(s_1|s_2, \boldsymbol{\Omega}_1, \boldsymbol{\Theta}) \\ &= \Pr(s_T|\boldsymbol{\Omega}_T, \boldsymbol{\Theta}) \prod_{t=1}^{T-1} \Pr(s_t|s_{t+1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}) \end{aligned} \quad (3.5)$$

if $\mathcal{S}_T \equiv \{s_1, s_2, \dots, s_T\}$ is assumed to follow Markov process. This usage of Markovian in reverse order is justified by virtue of Bayes theorem as we see below

$$\Pr(s_t|s_{t+1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}) = \frac{\Pr(s_{t+1}|s_t) \Pr(s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})}{\Pr(s_{t+1}|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})}$$

³That is, the prior of $\boldsymbol{\theta}$ and the prior of $\boldsymbol{\pi}$ are independent.

$$= \frac{\Pr(s_{t+1}|s_t) \Pr(s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})}{\sum_{s_t=1}^K \Pr(s_{t+1}|s_t) \Pr(s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})} \quad (3.6)$$

where $\Pr(s_{t+1}|s_t)$ is a transition probability. Notice that $\Pr(s_t|s_{t+1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta})$ in (3.5) can be computed from (3.6). The quantity $\Pr(s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})$ can be derived by using the Hamilton filter (Hamilton, 1989).

Example

Suppose that the number of regimes K is two, we generate the samples of s_1, s_2, \dots, s_T as follows: Drawing a random number from the uniform distribution between 0 and 1. If the generated number is less than or equal to the $\Pr(s_{T-1} = 1|s_T, \boldsymbol{\Omega}_{T-1}, \boldsymbol{\Theta})$ in (3.6) for $t = T - 1$, we set $s_{T-1} = 1$, otherwise, set $s_{T-1} = 2$. We can generate the sample of s_t for $t = T - 1, T - 2, \dots, 1$ backwards this way.

In order to calculate $\Pr(s_{T-1} = 1|s_T, \boldsymbol{\Omega}_{T-1}, \boldsymbol{\Theta})$, we use $s_T \sim \Pr(s_T|\boldsymbol{\Omega}_T, \boldsymbol{\Theta})$ from the Hamilton filter as stated before in both the numerator and the denominator of (3.6). We also need $\pi_{11} = \Pr(s_T = 1|s_{T-1} = 1)$ and $\pi_{22} = \Pr(s_T = 2|s_{T-1} = 2)$ for $s_t = 1$ and 2. Generate π_{11} and π_{22} from (3.10) and (3.11) respectively and calculate $\pi_{12} = 1 - \pi_{11}$, $\pi_{21} = 1 - \pi_{22}$ from (3.8) and (3.9) respectively.

3.2 Sampling of transition probabilities π_{ij}

Given the samples of latent variables s_1, s_2, \dots, s_T , the likelihood function $\mathcal{L}(\boldsymbol{\pi}|\mathcal{S}_T)$ appears on the right hand side of (3.4) is defined as

$$\mathcal{L}(\boldsymbol{\pi}|\mathcal{S}_T) = \prod_{i=1}^K \prod_{j=1}^K \pi_{ij}^{n_{ij}}$$

where n_{ij} is the total number of transitions from i to j from $t = 1$ to $t = T$.

Suppose that the i -th column vector of transition matrix (2.7) is denoted by $\boldsymbol{\pi}_i = [\pi_{i1} \ \pi_{i2} \ \dots \ \pi_{iK}]'$ and let the prior distribution of $\boldsymbol{\pi}_i$, independently of $\boldsymbol{\pi}_j$ ($j \neq i$) be a K -dimensional Dirichlet distribution ⁴ :

$$\boldsymbol{\pi}_i \sim Dir(u_{i1}, u_{i2}, \dots, u_{iK}),$$

⁴The Dirichlet distribution for $\boldsymbol{\pi}_i$ is defined as

$$p(\boldsymbol{\pi}_i|u_{i1}, u_{i2}, \dots, u_{iK}) = \frac{\Gamma(u_{i0})}{\Gamma(u_{i1}) \cdots \Gamma(u_{iK})} \pi_{i1}^{u_{i1}-1} \cdots \pi_{iK}^{u_{iK}-1}$$

where $0 \leq \pi_{ij} \leq 1$, $\sum_{j=1}^K \pi_{ij} = 1$, $u_{ij} > 0$, and $u_{i0} = \sum_{j=1}^K u_{ij}$.

then posterior distribution of $\boldsymbol{\pi}_i$ is given as

$$\begin{aligned}
p(\boldsymbol{\pi}_i|\mathcal{S}_T) &\propto \mathcal{L}(\boldsymbol{\pi}_i|\mathcal{S}_T)p(\boldsymbol{\pi}_i) \\
&\propto \prod_{j=1}^K \pi_{ij}^{n_{ij}} \times (\pi_{i1}^{u_{i1}-1} \cdots \pi_{iK}^{u_{iK}-1}) \\
&= \pi_{i1}^{n_{i1}+u_{i1}-1} \pi_{i2}^{n_{i2}+u_{i2}-1} \cdots \pi_{iK}^{n_{iK}+u_{iK}-1}. \tag{3.7}
\end{aligned}$$

Therefore we have

$$\boldsymbol{\pi}_i|\mathcal{S}_T \sim Dir(n_{i1} + u_{i1}, n_{i2} + u_{i2}, \dots, n_{iK} + u_{iK}), \quad i = 1, 2, \dots, K.$$

This corresponds to the first term on the right hand side of (3.4).

Example

Suppose that the number of regimes K is two, likelihood function is defined as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\pi}|\mathcal{S}_T) &= \prod_{i=1}^2 \prod_{j=1}^2 \pi_{ij}^{n_{ij}} \\
&= \pi_{11}^{n_{11}} (1 - \pi_{11})^{n_{12}} \times \pi_{22}^{n_{22}} (1 - \pi_{22})^{n_{21}} \\
&= \mathcal{L}(\pi_{11}, \pi_{22}|\mathcal{S}_T)
\end{aligned}$$

because

$$\pi_{11} + \pi_{12} = 1, \tag{3.8}$$

and

$$\pi_{21} + \pi_{22} = 1. \tag{3.9}$$

Then we can use beta distribution as the conjugate prior distributions with respect to π_{11} and to π_{22} ⁵:

$$\pi_{11} \sim Beta(u_{11}, u_{12}), \quad \pi_{22} \sim Beta(u_{22}, u_{21}).$$

Since posterior distributions for π_{11} and π_{22} are given as

$$p(\pi_{11}, \pi_{22}|\mathcal{S}_T) \propto \mathcal{L}(\pi_{11}, \pi_{22}|\mathcal{S}_T)p(\pi_{11})p(\pi_{22}),$$

⁵The beta distribution for π_{11} is defined as

$$p(\pi_{11}|u_{11}, u_{12}) = \frac{\Gamma(u_{11} + u_{12})}{\Gamma(u_{11})\Gamma(u_{12})} \pi_{11}^{u_{11}-1} (1 - \pi_{11})^{u_{12}-1}$$

where $0 \leq \pi_{11} \leq 1$ and $u_{11}, u_{12} > 0$.

$$\propto \pi_{11}^{n_{11}+u_{11}-1}(1-\pi_{11})^{n_{12}+u_{12}-1} \times \pi_{22}^{n_{22}+u_{22}-1}(1-\pi_{22})^{n_{21}+u_{21}-1},$$

we have

$$\pi_{11}|\mathcal{S}_T \sim \text{Beta}(n_{11} + u_{11} - 1, n_{12} + u_{12} - 1), \quad (3.10)$$

and

$$\pi_{22}|\mathcal{S}_T \sim \text{Beta}(n_{22} + u_{22} - 1, n_{21} + u_{21} - 1). \quad (3.11)$$

Here n_{11} , n_{12} , n_{21} , and n_{22} are the numbers of transitions from regime 1 to 1, 1 to 2, 2 to 1, and 2 to 2 observed over the course of s_1, \dots, s_T , and these s_1, \dots, s_T are generated in Example in section 3.1.

3.3 Sampling of parameters $\boldsymbol{\theta}_0, \boldsymbol{\theta}_j, j = 1, 2, \dots, K$

To evaluate the posterior distribution $p(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)$ in (3.4) via Gibbs sampler, we need conditional distributions of $\{\boldsymbol{\theta}_j\}_{j=0}^K$ given $\{\boldsymbol{\Sigma}_j\}_{j=1}^K$ and of $\{\boldsymbol{\Sigma}_j\}_{j=1}^K$ given $\{\boldsymbol{\theta}_j\}_{j=0}^K$. To make the formulation clear, we assume that prior distribution $p(\boldsymbol{\theta})$ ⁶ can be written as

$$\begin{aligned} p(\boldsymbol{\theta}) &\equiv p(\{\boldsymbol{\theta}_j\}_{j=0}^K, \{\boldsymbol{\Sigma}_j\}_{j=1}^K) \\ &= p(\{\boldsymbol{\theta}_j\}_{j=0}^K)p(\{\boldsymbol{\Sigma}_j\}_{j=1}^K). \end{aligned}$$

Since posterior distribution $p(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)$ in (3.4) can be rewritten as

$$p(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) \equiv p(\{\boldsymbol{\theta}_j\}_{j=0}^K, \{\boldsymbol{\Sigma}_j\}_{j=1}^K|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T),$$

we have

$$\begin{aligned} &p(\{\boldsymbol{\theta}_j\}_{j=0}^K, \{\boldsymbol{\Sigma}_j\}_{j=1}^K|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) \\ &\propto \mathcal{L}(\{\boldsymbol{\theta}_j\}_{j=0}^K, \{\boldsymbol{\Sigma}_j\}_{j=1}^K|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)p(\{\boldsymbol{\theta}_j\}_{j=0}^K)p(\{\boldsymbol{\Sigma}_j\}_{j=1}^K). \end{aligned} \quad (3.12)$$

Given $\{\boldsymbol{\Sigma}_j\}_{j=1}^K, \mathcal{Y}_T, \mathcal{S}_T$, and \mathcal{X}_T , conditional posterior distributions for $\{\boldsymbol{\theta}_j\}_{j=0}^K$ is expressed by dividing both sides of (3.12) by $p(\{\boldsymbol{\Sigma}_j\}_{j=1}^K)$

$$\begin{aligned} &p(\{\boldsymbol{\theta}_j\}_{j=0}^K|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K) \\ &\propto \mathcal{L}(\{\boldsymbol{\theta}_j\}_{j=0}^K, \{\boldsymbol{\Sigma}_j\}_{j=1}^K|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)p(\{\boldsymbol{\theta}_j\}_{j=0}^K). \end{aligned} \quad (3.13)$$

Given the samples of latent variables s_1, s_2, \dots, s_T obtained in section 3.1 and variance-covariance matrices $\boldsymbol{\Sigma}_j$ ($j = 1, 2, \dots, K$) to be described in (3.17), we re-express the MS-AIDS model (3.2) as

$$\boldsymbol{w}_t = \boldsymbol{X}_t\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}_t \quad (3.14)$$

⁶In other words, we assume that the prior distribution's location parameters $\{\boldsymbol{\theta}_j\}_{j=0}^K$ and scale-like parameters $\{\boldsymbol{\Sigma}_j\}_{j=1}^K$ can be freely moved and form a K -dimensional rectangular parameter space.

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{s_t})$. The matrix \mathbf{X}_t in (3.14) is defined as

$$\mathbf{X}_t = \left[\mathbf{1}\{s_t = 1\}\mathbf{X}_t^{(1)} \quad \mathbf{1}\{s_t = 2\}\mathbf{X}_t^{(1)} \quad \cdots \quad \mathbf{1}\{s_t = K\}\mathbf{X}_t^{(1)} \quad \mathbf{X}_t^{(0)} \right]$$

with indicator function $\mathbf{1}\{s_t = j\}$ taking scalar value 1 if $s_t = j$, or 0 otherwise ⁷, and parameter vector $\boldsymbol{\theta}^*$ is defined as

$$\boldsymbol{\theta}^* \equiv \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_K \\ \boldsymbol{\theta}_0 \end{bmatrix}.$$

To generate samples of $\boldsymbol{\theta}^*$, we derive the posterior distribution of $\boldsymbol{\theta}^*$ conditional on $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_K$ from (3.13). Applying the multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ as a conjugate prior $p(\boldsymbol{\theta}^*)$, conditional posterior distribution of $\boldsymbol{\theta}^*$ in (3.13) is derived as

$$\begin{aligned} & p(\boldsymbol{\theta}^* | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K) \\ & \propto \prod_{t=1}^T \left[(2\pi)^{-\frac{N-1}{2}} |\boldsymbol{\Sigma}_{s_t}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{w}_t - \mathbf{X}_t \boldsymbol{\theta}^*)' \boldsymbol{\Sigma}_{s_t}^{-1} (\mathbf{w}_t - \mathbf{X}_t \boldsymbol{\theta}^*) \right\} \right] \\ & \quad \times |\mathbf{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) \right\} \\ & \propto \exp \{ (\boldsymbol{\theta}^* - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\theta}^* - \mathbf{b}) \} \end{aligned}$$

where

$$\mathbf{b} = \mathbf{B} \left(\sum_{t=1}^T \mathbf{X}_t' \boldsymbol{\Sigma}_{s_t}^{-1} \mathbf{w}_t + \mathbf{V}^{-1} \boldsymbol{\mu} \right), \quad \mathbf{B}^{-1} = \sum_{t=1}^T \mathbf{X}_t' \boldsymbol{\Sigma}_{s_t}^{-1} \mathbf{X}_t + \mathbf{V}^{-1}.$$

Then the conditional posterior distribution of $\boldsymbol{\theta}^*$ is

$$\boldsymbol{\theta}^* | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K \sim \mathcal{N}(\mathbf{b}, \mathbf{B}). \quad (3.15)$$

⁷That is, when $s_t = k$, the matrix \mathbf{X}_t consists of $k - 1$ of matrices of size $(N - 1) \times [3(N - 1) + N(N - 1)/2]$ whose elements are all zero, $\mathbf{X}_t^{(1)}$, and $K - k$ of matrices of size $(N - 1) \times [3(N - 1) + N(N - 1)/2]$ whose elements are all zero, and $\mathbf{X}_t^{(0)}$, all aligned from left to right.

3.4 Sampling of parameters Σ_j , $j = 1, 2, \dots, K$

We assume that $\Sigma_1, \Sigma_2, \dots, \Sigma_K$ are independent, then conditional posterior distribution for $\{\Sigma_j\}_{j=1}^K$ is expressed as

$$\begin{aligned} p(\{\Sigma_j\}_{j=1}^K | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\theta_j\}_{j=0}^K) &= \prod_{j=1}^K p(\Sigma_j | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\theta_j\}_{j=0}^K) \\ &\propto \prod_{j=1}^K \mathcal{L}(\Sigma_j, \{\theta_j\}_{j=0}^K | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) p(\Sigma_j). \end{aligned} \quad (3.16)$$

To generate samples of Σ_j , $j = 1, 2, \dots, K$, we derive the conditional posterior distribution of Σ_j from (3.16). Applying the inverse Wishart distribution $\mathcal{IW}(\nu_j, \Lambda_j)$ as a conjugate prior $p(\Sigma_j)$, conditional posterior distribution of Σ_j is derived as

$$\begin{aligned} &p(\Sigma_j | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \theta^*) \\ &\propto \prod_{t \in \{t: s_t = j\}} \left[(2\pi)^{-\frac{N-1}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\varepsilon}_t' \Sigma_j^{-1} \boldsymbol{\varepsilon}_t\right) \right] \\ &\quad \times |\Sigma_j|^{-\frac{\nu_j + (N-1) + 1}{2}} \exp\left(-\frac{1}{2} \text{tr}\{\Sigma_j^{-1} \Lambda_j\}\right) \\ &\propto |\Sigma_j|^{-\frac{\nu_j + (N-1) + 1 + n_j}{2}} \exp\left(-\frac{1}{2} \text{tr}\left\{\Sigma_j^{-1} \left(\sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{1}\{s_t = j\} + \Lambda_j\right)\right\}\right) \end{aligned}$$

where n_j is the total number of time t belonging to regime j . Then conditional posterior distribution of Σ_j is

$$\Sigma_j | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \theta^* \sim \mathcal{IW}\left(\nu_j + n_j, \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{1}\{s_t = j\} + \Lambda_j\right) \quad (3.17)$$

where $j = 1, 2, \dots, K$.

From (3.15) and (3.17), we are able to construct Gibbs sampler algorithm by generating θ^* and substituting these into (3.17) and then generating Σ_j with the generated θ^* and substituting those back into (3.15).

4 Simulation study

In this section, we illustrate our proposed method by simulation. To generate simulation data, we consider the case that the number of products $N =$

4, the number of regimes $K = 2$, and the number of observations $T = 800$. The price data of four products p_{1t} , p_{2t} , p_{3t} , and p_{4t} generate from the uniform distributions such that $p_{1t} \sim \mathcal{U}(240, 380)$, $p_{2t} \sim \mathcal{U}(130, 150)$, $p_{3t} \sim \mathcal{U}(90, 100)$, and $p_{4t} \sim \mathcal{U}(120, 200)$. Total expenditure (or budget) on four products, m_{0t} , generates from $\mathcal{U}(7500, 16000)$. We specify unobserved latent variables $\mathcal{S}_T = \{s_1, s_2, \dots, s_T\}$ such that $s_t = 1$ if $1 \leq t \leq 400$, and $s_t = 2$ if $401 \leq t \leq 800$ ⁸.

To simplify the MS-AIDS model (2.1), parameter α_{i,s_t} in (2.4) is set to $\alpha_{i,s_t} = \bar{\alpha}_{i,s_t}$ and we estimate the parameters via the Gibbs sampling algorithm with (3.6), (3.10), (3.11), (3.15) and (3.17). The prior distributions are parameterized by setting $\boldsymbol{\mu} = \mathbf{0}$, $\mathbf{V} = 10^4 \mathbf{I}_{24}$, $\nu_j = 10$, $\boldsymbol{\Lambda}_j = 10^{-3} \mathbf{I}_3$ ($j = 1, 2$), $u_{11} = u_{22} = 5$, and $u_{12} = u_{21} = 2$. When we generate the samples of parameters via the Gibbs sampling algorithm, we restrict a priori that, in our simulated data, at least 40% of observations lie in each regime in order to avoid identification problem.

The Gibbs sampling algorithm is run so that the first 40,000 samples are discarded as burn-in and then the next 40,000 samples are recorded. To test the convergence of samples to the posterior distributions, we apply the Geweke (1992)'s convergence diagnostic. The first 10% and last 50% of the recorded simulated data are used to conduct this test, as suggested by Geweke (1992).

Tables 1 and 2 show the posterior means, posterior standard deviations (SD), 95% credible intervals (sometimes referred as 95% Bayesian confidence intervals as well), and Geweke's convergence diagnostic statistics (CD) for all parameters in regimes 1 and 2. We confirm that Geweke's convergence diagnostic statistics for all parameters are within ± 1.96 range. For α 's, γ 's, and β 's, their posterior distributions are all normal, so that the means and 50% percentiles in principle coincide for the same estimated, while this is not the case for σ 's because the posteriors for σ 's are inverse Wishart and they are not symmetric. The Geweke's convergence diagnostic statistics in Tables 1 and 2 show that we do not reject the null hypothesis for equality of the means of the first 10% and the last 50% of the recorded simulated data at the 5% significant level. Tables 1 and 2 also show that posterior means of all parameters are similar to those true values and 95% credible intervals include true values.

In Table 3, transition probabilities π_{11} and π_{22} pass the Geweke's convergence diagnostic and posterior means of transition probabilities indicate

⁸These number are set to mimic Japanese meat (beef, pork, chicken and fish) market from 1998 to 2006 to be analyzed in this paper, so that the prices and the expenditures are in Japanese yen. In the actual analysis, the number of observations are only available monthly with 108 observations.

Table 1: Estimated Parameters in Regime 1

	Mean	SD	2.5%	50%	97.5%	CD	True value
α_1	0.4101	0.0114	0.3879	0.4101	0.4322	1.3850	0.40
α_2	0.4955	0.0110	0.4740	0.4955	0.5170	-1.3797	0.50
α_3	0.2937	0.0109	0.2725	0.2937	0.3151	-0.2398	0.30
γ_{11}	0.0188	0.0035	0.0119	0.0188	0.0256	-1.2725	0.02
γ_{12}	-0.0177	0.0033	-0.0241	-0.0177	-0.0112	1.2060	-0.02
γ_{13}	-0.0437	0.0032	-0.0500	-0.0437	-0.0375	0.0054	-0.04
γ_{22}	0.0437	0.0071	0.0298	0.0437	0.0577	-0.2960	0.05
γ_{23}	-0.0290	0.0063	-0.0414	-0.0291	-0.0168	0.2734	-0.03
γ_{33}	0.0973	0.0071	0.0835	0.0973	0.1112	-0.7818	0.10
β_1	-0.0626	0.0025	-0.0674	-0.0626	-0.0577	-1.1847	-0.06
β_2	-0.0595	0.0024	-0.0642	-0.0595	-0.0549	1.2374	-0.06
β_3	-0.0182	0.0023	-0.0227	-0.0182	-0.0138	-0.0367	-0.02
σ_{11}^2	0.000108	0.000008	0.000094	0.000107	0.000124	1.8725	0.00010
σ_{12}	0.000007	0.000005	-0.000003	0.000007	0.000017	-1.0644	0.00000
σ_{13}	0.000006	0.000005	-0.000003	0.000006	0.000016	0.2284	0.00000
σ_{22}^2	0.000099	0.000007	0.000086	0.000099	0.000114	-1.2798	0.00010
σ_{23}	0.000001	0.000005	-0.000008	0.000001	0.000010	0.3664	0.00000
σ_{33}^2	0.000091	0.000006	0.000079	0.000090	0.000104	0.4843	0.00010

Table 2: Estimated Parameters in Regime 2

	Mean	SD	2.5%	50%	97.5%	CD	True value
α_1	0.2797	0.0130	0.2540	0.2797	0.3051	0.8683	0.30
α_2	0.4065	0.0140	0.3792	0.4064	0.4340	-1.7004	0.40
α_3	0.5098	0.0140	0.4821	0.5098	0.5370	-0.9716	0.50
γ_{11}	0.0217	0.0040	0.0140	0.0217	0.0296	-1.4201	0.02
γ_{12}	-0.0279	0.0039	-0.0356	-0.0279	-0.0202	0.4881	-0.03
γ_{13}	-0.0542	0.0039	-0.0619	-0.0542	-0.0467	0.2647	-0.05
γ_{22}	0.0531	0.0093	0.0348	0.0531	0.0715	0.6860	0.05
γ_{23}	-0.0223	0.0085	-0.0391	-0.0222	-0.0056	-0.9342	-0.02
γ_{33}	0.1092	0.0094	0.0907	0.1092	0.1279	0.7255	0.10
β_1	-0.0459	0.0027	-0.0511	-0.0459	-0.0406	-0.5835	-0.05
β_2	-0.0719	0.0029	-0.0777	-0.0719	-0.0663	1.5217	-0.07
β_3	-0.0305	0.0028	-0.0360	-0.0306	-0.0250	1.1880	-0.03
σ_{11}^2	0.000137	0.000010	0.000119	0.000136	0.000157	-1.5203	0.00015
σ_{12}	0.000003	0.000007	-0.000012	0.000003	0.000017	-1.5000	0.00000
σ_{13}	0.000000	0.000007	-0.000014	0.000000	0.000014	0.1945	0.00000
σ_{22}^2	0.000162	0.000011	0.000141	0.000161	0.000186	1.1823	0.00015
σ_{23}	-0.000002	0.000008	-0.000017	-0.000002	0.000013	-0.1355	0.00000
σ_{33}^2	0.000150	0.000011	0.000131	0.000150	0.000172	-0.7952	0.00015

that simulated budget share data tends to stay for a long time in the same regime. This result obviously reflects the specification of latent variables $\mathcal{S}_T = \{s_1, s_2, \dots, s_T\}$ such that $s_t = 1$ if $1 \leq t \leq 400$, and $s_t = 2$ if $401 \leq t \leq 800$. We also calculate the probability of being regime j , ($j = 1, 2$) at time t as

$$\Pr\{s_t = j\} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{s_t^{(i)} = j\} \quad (4.1)$$

where n is the number of recorded samples to summarize the posterior distributions, and $s_t^{(i)}$ is a i -th sample of latent variable s_t . Figure 1 shows that

Table 3: Estimated Transition Probability

	Mean	SD	2.5%	50%	97.5%	CD
π_{11}	0.9951	0.0035	0.9863	0.9958	0.9994	-0.8778
π_{22}	0.9975	0.0025	0.9908	0.9983	0.9999	-0.5904

regime shift from $s_t = 1$ to $s_t = 2$ is observed at time $t = 400$.

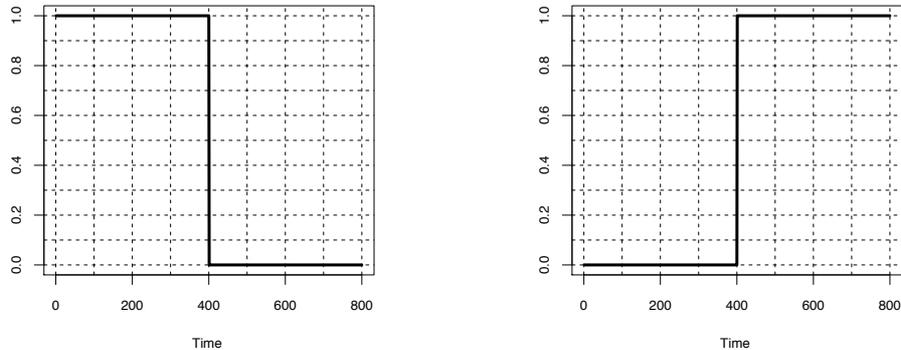


Figure 1: The left panel shows probability of $s_t = 1$, while the right panel shows the probability of $s_t = 2$.

5 Empirical study on Japanese meat market

In the previous studies on structural changes of demands in the Japanese meat market, for example, Jin and Koo (2003) identified a structural change point coinciding with the first BSE case in Japan by using the non-parametric tests. Peterson and Chen (2005) focused on the type and origin of beef products (i.e., wagyu⁹, dairy, U.S., and Australian beef) and showed the difference of impacts on the beef products due to the first BSE case in Japan. Ishida et al. (2006, 2010) examined not only the structural change due to the BSE but also the bird flu in Japan. Previous studies such as Peterson and Chen (2005) and Ishida et al. (2006, 2010) applied the gradual switching model proposed by Ohtani and Katayama (1986) to examine the gradual shifts in demands of meat products. This model utilizes a transition function to express a gradual shift pattern in demand under the assumption that starting-points of structural shift are already known. In this study, we examine the structural change points without any prior information about change points in the Japanese meat market via the proposed Bayesian MS-AIDS model. Also, we compute the elasticities of price and expenditure in each regime to examine the change of consumers' purchasing behavior pattern of meat products in Japan.

⁹Japanese native beef cattle

The Ministry of Internal Affairs and Communications in Japan provides us with the household expenditure survey data (i.e., Family Income and Expenditure Survey). The household expenditure survey data includes the monthly time series data about average expenditure and price of meat and fish products along with others. In this study, we used the average expenditure and price data of beef, pork, chicken and fish over January 1998 to December 2006 (108 months). Figure 2 plots the budget shares of meat products from January 1998 through December 2006.

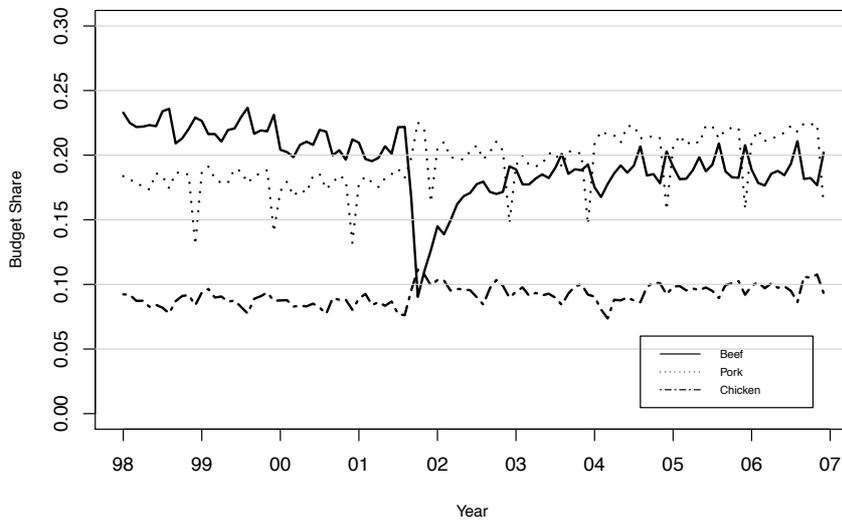


Figure 2: Plot of budget share data

In the Japanese meat market, there exists serious food safety concerns regarding to the BSE and bird flu during the period under this study. The BSE was initially recognized in Japan in September 2001. The budget share of beef dramatically declined after the first reported case of BSE in Japan, while those of pork and chicken increased (see Figure 2). After the first BSE case, budget share of beef gradually recovered, however it was still relatively small as compared with that of pork. During the study period, the first BSE case in U.S. was confirmed on December 2003. Although the U.S. was one of the largest exporters to Japan, Japanese government announced a ban on import of American beef immediately. On December 2005 the Japanese government resumed importing American beef under stricter conditions. Nevertheless banned specified-risk materials of beef products were found from the imported beef from the U.S. on January 2006. Afterward,

Japanese government imposed a ban on import of U.S. beef until July 2006.

The bird flu is an infectious disease caused by avian influenza viruses (e.g., H5N1 virus). The H5N1 virus is the most highly pathogenic strain. Many people infected with the H5N1 virus have died in Vietnam, Indonesia, Thailand and other Southeast Asian countries. On January 2004, the first H5N1-infected poultry were discovered in Japan. The budget share of chicken declined slightly and then it recovered within a short period of time.

5.1 Estimation Results

We estimate parameters of MS-AIDS model (2.1) with the intercept parameters in (2.4). The Gibbs sampling algorithm is run so that the first 5,000 samples are discarded as burn-in and then the next 25,000 samples are recorded. The prior distributions are parameterized by setting $\boldsymbol{\mu} = \mathbf{0}$, $\mathbf{V} = 10^4 \mathbf{I}_{45}$, $\nu_j = 10$, $\boldsymbol{\Lambda}_j = 10^{-3} \mathbf{I}_3$ ($j = 1, 2$), $u_{11} = u_{22} = 5$, and $u_{12} = u_{21} = 2$. We also restrict a priori that, in our observed data, at least 40% of observations lie in each regime in order to avoid identification problem within the Gibbs sampling algorithm.

In this study, we examine the following four models: model 1 only includes intercept parameter $\bar{\alpha}_{i,st}$. Model 2 includes seasonal effects on August and December into model 1, and model 3 adds a habit effect into model 2. Finally, model 4 further incorporates a trend effect into model 3.

Model 1 $\alpha_{i,st} = \bar{\alpha}_{i,st}$

Model 2 $\alpha_{i,st} = \bar{\alpha}_{i,st} + \delta_{1,i} d_{1,t} + \delta_{2,i} d_{2,t}$

Model 3 $\alpha_{i,st} = \bar{\alpha}_{i,st} + \delta_{1,i} d_{1,t} + \delta_{2,i} d_{2,t} + \sum_{j=1}^N \phi_{ij} \bar{w}_{j,t-1}$

Model 4 $\alpha_{i,st} = \bar{\alpha}_{i,st} + \nu_{i,st} t + \delta_{1,i} d_{1,t} + \delta_{2,i} d_{2,t} + \sum_{j=1}^N \phi_{ij} \bar{w}_{j,t-1}$

Table 4: Log-Marginal Likelihood and Log-Bayes Factor

	Model1	Model2	Model3	Model4
Model1	1098.416	—	—	—
Model2	18.927	1117.343	—	—
Model3	62.157	43.230	1160.573	—
Model4	63.452	44.525	1.295	1161.868

Table 4 shows the logarithmic marginal likelihood of model i , log-ML $_i$ ($i = 1, 2, 3, 4$) as diagonal elements and logarithmic Bayes factors (see Appendix

C), $\log\text{-BF}_{ij}$ for model i against model j as off-diagonal elements. To obtain the marginal likelihoods for candidate models, we use the method proposed by Newton and Raftery (1994). Although models 3 and 4 have large logarithmic marginal likelihoods relative to the other models, logarithmic Bayes factor for model 4 against model 3, $\log\text{-BF}_{43}$ ($= 1.295$) indicates “positive” (Kass and Raftery, 1995, p.777) evidence in favor of model 4. Therefore we conclude that model 4 fits the data best.

Table 5: Estimated Parameters of MS-AIDS model in Regime1

	Mean	SD	2.5%	50%	97.5%	CD
$\bar{\alpha}_1$	0.2866	0.1262	0.0416	0.2850	0.5373	0.6055
$\bar{\alpha}_2$	0.6046	0.0925	0.4118	0.6074	0.7780	-0.3348
$\bar{\alpha}_3$	0.3285	0.0794	0.1694	0.3290	0.4824	-0.7486
γ_{11}	0.0245	0.0613	-0.1131	0.0350	0.1229	-0.0015
γ_{12}	-0.0039	0.0437	-0.0701	-0.0129	0.0951	0.3383
γ_{13}	-0.0549	0.0335	-0.1315	-0.0534	0.0060	-0.4092
γ_{22}	0.0335	0.0473	-0.0693	0.0368	0.1174	-0.1208
γ_{23}	-0.0110	0.0403	-0.0947	-0.0096	0.0653	-0.7934
γ_{33}	0.0891	0.0480	-0.0020	0.0879	0.1892	0.7943
β_1	-0.0301	0.0265	-0.0817	-0.0301	0.0226	-0.6585
β_2	-0.0936	0.0193	-0.1320	-0.0936	-0.0553	-0.0383
β_3	-0.0195	0.0170	-0.0524	-0.0198	0.0147	0.7026
σ_{11}^2	0.000061	0.000013	0.000041	0.000060	0.000091	1.3920
σ_{12}	-0.000003	0.000008	-0.000018	-0.000003	0.000013	-0.5299
σ_{13}	-0.000008	0.000007	-0.000022	-0.000008	0.000006	-0.1079
σ_{22}^2	0.000035	0.000008	0.000024	0.000034	0.000053	0.4690
σ_{23}	0.000006	0.000005	-0.000004	0.000006	0.000016	0.2985
σ_{33}^2	0.000029	0.000006	0.000019	0.000028	0.000044	1.1468

Tables 5 and 6 show the results of parameters for beef, pork and chicken. The parameters for fish are estimated from the adding-up condition (2.3a). Tables 5 and 6 show the the posterior means, posterior standard deviations (SD), 95% credible intervals, and Geweke’s convergence diagnostic statistics (CD) for all parameters in MS-AIDS model (2.1) and variance-covariance matrices in regimes 1 and 2. To carry out the Geweke’s convergence diagnostic, we used the first 10% and last 50% of the recorded simulated data and Tables 5 and 6 show that all parameters pass the Geweke’s convergence diagnostic at 5% significant level. In the Bayesian framework, if a 95% credible interval does not include zero, estimated parameters are interpreted as

Table 6: Estimated Parameters of MS-AIDS model in Regime2

	Mean	SD	2.5%	50%	97.5%	CD
$\bar{\alpha}_1$	0.1926	0.1858	-0.1418	0.1817	0.5981	-0.4185
$\bar{\alpha}_2$	0.6542	0.1255	0.3993	0.6593	0.8891	-0.4349
$\bar{\alpha}_3$	0.2838	0.1017	0.0812	0.2843	0.4830	-0.6659
γ_{11}	0.1032	0.0682	-0.0267	0.1056	0.2270	-0.1384
γ_{12}	-0.0484	0.0355	-0.1164	-0.0489	0.0185	0.4018
γ_{13}	-0.0560	0.0299	-0.1154	-0.0555	0.0011	0.9316
γ_{22}	0.0432	0.0417	-0.0368	0.0425	0.1272	0.0669
γ_{23}	0.0236	0.0343	-0.0436	0.0234	0.0923	-0.3294
γ_{33}	0.0628	0.0399	-0.0146	0.0620	0.1429	-1.1898
β_1	-0.0329	0.0385	-0.1131	-0.0318	0.0392	0.8659
β_2	-0.0949	0.0246	-0.1421	-0.0954	-0.0454	0.5194
β_3	-0.0112	0.0210	-0.0529	-0.0111	0.0301	0.0316
σ_{11}^2	0.000161	0.000045	0.000096	0.000154	0.000267	-0.7579
σ_{12}	-0.000033	0.000018	-0.000074	-0.000031	-0.000004	0.2701
σ_{13}	-0.000029	0.000015	-0.000062	-0.000027	-0.000003	0.3091
σ_{22}^2	0.000045	0.000010	0.000030	0.000044	0.000069	-1.0053
σ_{23}	0.000004	0.000008	-0.000011	0.000004	0.000020	-0.0595
σ_{33}^2	0.000042	0.000010	0.000027	0.000041	0.000065	-1.3308

the significant parameters. Thus $\bar{\alpha}_1$, $\bar{\alpha}_2$, $\bar{\alpha}_3$, β_2 , σ_{11}^2 , σ_{22}^2 , and σ_{33}^2 in Table 5 and $\bar{\alpha}_2$, $\bar{\alpha}_3$, β_2 , σ_{11}^2 , σ_{12} , σ_{13} , σ_{22}^2 , and σ_{33}^2 in Table 6 are regarded significantly different from zero. These parameters in MS-AIDS model (2.1) are used to calculate the price and expenditure elasticities.

Next we show the results of estimated parameters in the intercept term (2.4). In Table 7, only trend effect of beef in regime 1 has a significant negative effect, while the other parameters include zero within the 95% credible intervals. As for seasonal effects in Table 8, beef has a significant positive effect, while chicken has a significant negative effect both in August. On the other hand, seasonal effects on December are not significant for all meat products. Finally, Table 9 shows that the budget shares of beef and pork for previous period have positive impacts on the current budget shares.

To compare our proposed Bayesian estimation with the ML estimation proposed in Allais and Nichèle (2007) and employed in Kabe and Kanazawa (2012), we calculate the mean squared errors (MSEs) for estimated budget shares. The MSEs of Bayesian estimation are evaluated by the posterior means of estimates of budget shares generated within the Gibbs sampler. The

Table 7: Trend Effect in MS-AIDS model

Trend (Regime 1)						
	Mean	SD	2.5%	50%	97.5%	CD
ν_{11} (beef)	-0.00038	0.00014	-0.00064	-0.00040	-0.00007	0.2165
ν_{21} (pork)	0.00003	0.00010	-0.00018	0.00004	0.00021	-0.9640
ν_{31} (chicken)	0.00002	0.00008	-0.00016	0.00002	0.00017	-0.6504
Trend (Regime 2)						
	Mean	SD	2.5%	50%	97.5%	CD
ν_{12} (beef)	0.00016	0.00022	-0.00029	0.00017	0.00058	0.9870
ν_{22} (pork)	0.00019	0.00012	-0.00005	0.00019	0.00044	-0.0507
ν_{32} (chicken)	0.00015	0.00012	-0.00009	0.00015	0.00038	-1.7246

Table 8: Seasonal Effect in MS-AIDS model

Seasonal (Aug)						
	Mean	SD	2.5%	50%	97.5%	CD
δ_{11} (beef)	0.0108	0.0040	0.0031	0.0108	0.0189	-0.1119
δ_{12} (pork)	-0.0040	0.0026	-0.0093	-0.0040	0.0010	0.8173
δ_{13} (chicken)	-0.0054	0.0024	-0.0101	-0.0054	-0.0006	0.7323
Seasonal (Dec)						
	Mean	SD	2.5%	50%	97.5%	CD
δ_{21} (beef)	0.0119	0.0124	-0.0111	0.0114	0.0375	0.0939
δ_{22} (pork)	-0.0142	0.0104	-0.0355	-0.0137	0.0045	-0.3565
δ_{23} (chicken)	0.0128	0.0079	-0.0028	0.0128	0.0281	-0.6845

Table 9: Habit Effect in MS-AIDS model

	Mean	SD	2.5%	50%	97.5%	CD
ϕ_{11} (beef)	0.3183	0.0772	0.1636	0.3197	0.4696	-0.0619
ϕ_{22} (pork)	0.0510	0.0776	-0.1010	0.0513	0.2030	0.0906
ϕ_{33} (chicken)	0.4454	0.1158	0.2136	0.4478	0.6683	-1.1871

results of MSEs with respect to our proposed Bayesian estimation (Bayes) and ML estimation (MLE) are given in Table 10. Our Bayesian estimation

improves the MSEs for all products over ML estimation. This result reflects the goodness of fit to the budget share data of our Bayesian estimation.

Table 10: Mean squared errors (MSEs)

	Beef	Pork	Chicken	Fish
Bayes	0.805×10^{-4}	0.172×10^{-4}	0.152×10^{-4}	0.603×10^{-4}
MLE	1.177×10^{-4}	0.264×10^{-4}	0.158×10^{-4}	0.657×10^{-4}

Figure 3 plots the probability of being regime 2 and budget share data of beef and pork from January 1998 through December 2006 under the proposed Bayesian method. We calculate the probability $\Pr\{s_t = 2\}$ from (4.1). In Figure 3, regime shift from $s_t = 1$ to $s_t = 2$ is observed at the timing of first BSE case in Japan in September 2001 and then the probability $\Pr\{s_t = 2\}$ gradually declines until the end of 2003 along with increase in budget share of beef. With the timing of first BSE case in U.S. in December 2003, we observe a high probability of being regime 2 once again. Since then, structure of budget share tends to stay in regime 2.

The first regime shift in Figure 3 reflects the switching of consumers' preference from beef to pork triggered by the BSE scare in Japan in September 2001. The second regime shift after the first BSE discovery in U.S. in December 2003 might have arisen due to the ban on import of American beef. Since the ban on importing U.S. beef led to the shortage of beef supply in the domestic meat market, consumers may have been forced to purchase more pork instead of beef. In addition, Table 11 shows that posterior means of transition probabilities π_{11} and π_{22} are relatively high. This implies that there is little chance for switching from regime 1 to regime 2 and from regime 2 to regime 1.

Figure 4 shows the results of probability of being regime 2, $\Pr(s_t = 2 | \Omega_t, \hat{\Theta})$ under the ML estimation. We estimated the probability from the Hamilton filter using data set obtained through time t , Ω_t , and ML estimates $\hat{\Theta}$ in model 4 (see Kabe and Kanazawa, 2012). The regime shift at the timing of first BSE case in Japan in September 2001 is observed in Figure 4. Nevertheless, when compared with the result of regime shift in Figure 3, the probability's gradual decline due to the recovery of beef budget share following the first BSE case in Japan observed in Figure 3 no longer can be observed in Figure 4. Unlike Figure 3, we cannot identify the regime shift at the timing of first BSE case in U.S. in December 2003 in Figure 4.

We calculate the average budget share of i -th product at regime $s_t = j$

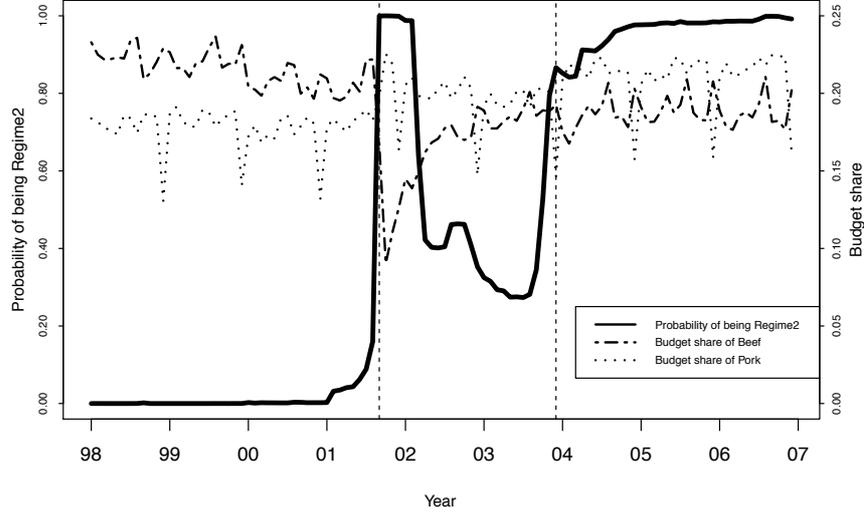


Figure 3: Probability of being regime 2, $\Pr\{s_t = 2\}$ and budget share data of beef and pork under the proposed Bayesian estimation. Two vertical dashed lines indicate the first BSE case in Japan on September 2001 and the first BSE case in U.S. on December 2003.

Table 11: Estimated Transition Probabilities

	Mean	SD	2.5%	50%	97.5%	CD
π_{11}	0.9520	0.0282	0.8830	0.9572	0.9911	-0.1400
π_{22}	0.9619	0.0300	0.8868	0.9689	0.9979	1.0742

as

$$\bar{w}_{i,s_t=j} = \frac{\sum_{t=1}^T \mathbf{1}\{s_t = j\} \bar{w}_{it}}{\sum_{t=1}^T \mathbf{1}\{s_t = j\}}.$$

Table 12 shows that regime 1 is characterized by a higher beef budget share relative to that of pork, while regime 2 is characterized by the reversal of these two budget shares.

Since substitution occurs mostly between beef and pork in regimes 1 and 2 (see Table 12), we focus on the price and expenditure elasticities for beef and pork. Using the estimated parameters in Tables 5, 6, 7, 8 and 9, we calculate the Marshallian price elasticity η_{ij,s_t}^P and expenditure elasticity η_{i,s_t}^E

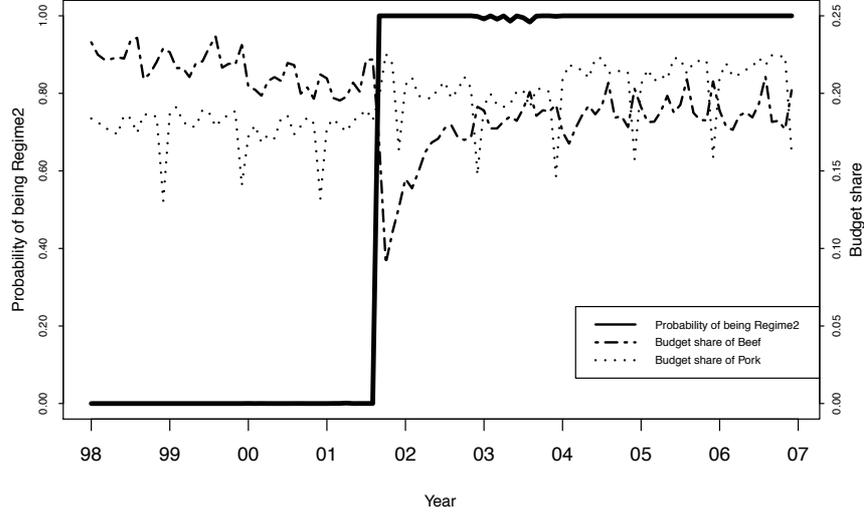


Figure 4: Probability of being regime 2, $\Pr(s_t = 2|\Omega_t, \hat{\Theta})$ and budget share data of beef and pork under ML estimation.

Table 12: Posterior mean of average budget share

	Regime 1	Regime 2
Beef	0.2075	0.1799
Pork	0.1825	0.2071
Chicken	0.0878	0.0961
Fish	0.5222	0.5170

at regime s_t for each 25,000 samples generated via Gibbs sampler as

$$\eta_{ij,s_t}^P = -\kappa_{ij} + \frac{\gamma_{ij,s_t}}{\bar{w}_{i,s_t}} - \frac{\beta_{i,s_t}}{\bar{w}_{i,s_t}} \left[\alpha_{j,s_t} + \sum_{k=1}^N \gamma_{kj,s_t} \log \bar{p}_{k,s_t} \right], \quad (5.1)$$

$$\eta_{i,s_t}^E = \frac{\beta_{i,s_t}}{\bar{w}_{i,s_t}} + 1, \quad (5.2)$$

where $\kappa_{ij} = 1$ for $i = j$ and $\kappa_{ij} = 0$ for $i \neq j$, and \bar{p}_{k,s_t} is an average price at regime s_t .

In Table 13, we show the posterior means and 95% credible intervals of price and expenditure elasticities for beef and pork. Although own-price elas-

ticities of pork have significant negative impacts in both regimes, own-price elasticity of beef in regime 2 includes zero within the 95% credible interval. Since American beef was banned and in short supply then, beef prices tended to increase in regime 2. Hence this price inelastic beef purchasing behavior in regime 2 in Table 13 implies that those who had kept purchasing beef in regime 2 did so regardless of its price.

Moreover substitution between beef and pork due to the change of their relative prices is rejected statistically because cross-price elasticities are not significant in both regimes. This consumer behavior represented in Table 13 shows that this insignificance stayed almost unchanged over the two regimes.

Finally, expenditure elasticities of beef and pork are as expected significantly positive in both regimes and they do not seem to have changed significantly between these regimes. In Japan, since beef unit price is almost as twice as high as pork unit price, it is not surprising that beef expenditure elasticity is much higher than pork expenditure elasticity.

Table 13: Price elasticities and Expenditure elasticities

Regime1	Price (η_{ij}^P)		Expenditure (η_{ij}^E)
	Beef	Pork	
Beef	-0.8171 (-1.4682, -0.3440)	0.0600 (-0.2839, 0.5459)	0.8552 (0.6088, 1.1092)
Pork	0.1483 (-0.2299, 0.7208)	-0.5142 (-1.0636, -0.0365)	0.4874 (0.2783, 0.6939)

Regime2	Price (η_{ij}^P)		Expenditure (η_{ij}^E)
	Beef	Pork	
Beef	-0.3357 (-1.0458, 0.3696)	-0.1679 (-0.5573, 0.2385)	0.8168 (0.3694, 1.2182)
Pork	-0.0961 (-0.4100, 0.2209)	-0.5050 (-0.9362, -0.0620)	0.5421 (0.3172, 0.7797)

1) 95% credible interval in parentheses

6 Conclusion

In this paper, we proposed the Bayesian estimation for MS-AIDS model proposed by Allais and Nichèle (2007) and illustrated the applicability of our

proposed method via simulated and real data. The proposed Bayesian estimation has some important advantages. First, it enables us to avoid the singularity problem suggested in Hamilton (1990, 1991). In the Bayesian framework, we can use conjugate prior distributions to incorporate the prior information about variance-covariance matrices in advance. On the other hand, ML estimation via numerical optimization methods (e.g., Newton-Raphson method) has to depend on sensible selection of initial values of parameters to avoid singularity points on the parameter space. Second, our proposed Bayesian estimation by design ensures that transition probabilities be located between zero and one by generating the samples from the beta distributions within the Gibbs sampler. Third, there is no need to calculate the score functions of log-likelihood, unlike ML estimation which employed in Allais and Nichèle (2007) and Kabe and Kanazawa (2012), is computationally very intensive. In our Bayesian estimation, posterior distributions of parameters are expressed as the standard formula (e.g., multivariate normal and inverse Wishart distributions). Thus each of parameters can be easily simulated via Gibbs sampler.

In the empirical study on the Japanese meat market, we found that our Bayesian estimation improves the mean squared errors for all meat products compared with the ML estimation. Moreover we found the regime shift in the budget shares of meat products in Figures 3 depicts much more sophisticated and realistic picture of regime transition than Figure 4. Specifically, in Figure 3, probability of being regime 2, $\Pr\{s_t = 2\}$, estimated via Bayesian estimation shows the regime shifts at the timing of first reported cases of BSE both in Japan in September 2001 and also in U.S. in December 2003. On the other hand, Figure 4 shows a single regime shift at the timing of first BSE case in Japan. Since ML parameter estimates are given as the point solutions, the probability of being regime 2, $\Pr(s_t = 2|\Omega_t, \hat{\Theta})$ in Figure 4 ignores the uncertainty about parameters and making the probability $\Pr(s_t = 2|\Omega_t, \hat{\Theta})$ in Figure 4 closer to zero or one. Perhaps as Scott (2002, p.345) observed, ignoring uncertainty about the parameter may have contributed to such result.

Finally, we discuss the further extension of MS-AIDS model. Several studies have extended the Hamilton (1989)'s Markov-switching model. In particular, they focused on a useful modification of transition probabilities. For example, Diebold et al. (1994) introduced the time-varying transition probabilities into the Markov-switching model, and also they allowed the transition probabilities to evolve as logistic functions of economic variables. Alternatively, Markov-switching model assumes that latent variables controlling regime shifts are exogenous. Kim et al. (2008) relaxed this exogenous regime-switching assumption and proposed a Markov-switching regression

model with endogenous regime switching. The extensions to MS-AIDS model in these directions will be interesting for future research.

A Conditional posterior distributions of $\boldsymbol{\theta}^*$ and $\boldsymbol{\Sigma}_{s_t}$

To generate samples of $\boldsymbol{\theta}^*$, we derive the conditional posterior distributions of $\boldsymbol{\theta}^*$ from (3.13). Since $\boldsymbol{\Sigma}_{s_t}^{-1}$ is symmetric and positive definite, there exists a nonsingular matrix $\boldsymbol{\Sigma}_{s_t}^{-1/2}$ such that

$$\boldsymbol{\Sigma}_{s_t}^{-1} = \boldsymbol{\Sigma}_{s_t}^{-1/2'} \boldsymbol{\Sigma}_{s_t}^{-1/2}.$$

Multiplying $\boldsymbol{\Sigma}_{s_t}^{-1/2}$ on the both sides of (3.14), we have

$$\boldsymbol{\Sigma}_{s_t}^{-1/2} \mathbf{w}_t = \boldsymbol{\Sigma}_{s_t}^{-1/2} \mathbf{X}_t \boldsymbol{\theta}^* + \boldsymbol{\Sigma}_{s_t}^{-1/2} \boldsymbol{\varepsilon}_t. \quad (\text{A.1})$$

Let us denote $\tilde{\mathbf{w}}_t \equiv \boldsymbol{\Sigma}_{s_t}^{-1/2} \mathbf{w}_t$, $\tilde{\mathbf{X}}_t \equiv \boldsymbol{\Sigma}_{s_t}^{-1/2} \mathbf{X}_t$, and $\tilde{\boldsymbol{\varepsilon}}_t \equiv \boldsymbol{\Sigma}_{s_t}^{-1/2} \boldsymbol{\varepsilon}_t$. Applying the multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ as a conjugate prior $p(\boldsymbol{\theta}^*)$, conditional posterior distribution of $\boldsymbol{\theta}^*$ is derived as

$$\begin{aligned} p(\boldsymbol{\theta}^* | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K) &\propto \prod_{t=1}^T \left[\exp \left\{ -\frac{1}{2} (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*)' (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*) \right\} \right] \\ &\quad \times |\mathbf{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) \right\} \\ &= \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*)' (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*) \right\} \\ &\quad \times |\mathbf{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) \right\}. \end{aligned} \quad (\text{A.2})$$

Let us consider the OLS estimator $\hat{\boldsymbol{\theta}}^* = (\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t)^{-1} (\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{w}}_t)$, then (A.2) is written as

$$\begin{aligned} p(\boldsymbol{\theta}^* | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K) &\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^* + \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^* - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*)' (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^* + \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^* - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*) \right\} \\ &\quad \times |\mathbf{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) \right\}. \end{aligned} \quad (\text{A.3})$$

Since the orthogonality condition for OLS estimator $\hat{\boldsymbol{\theta}}^*$ implies that

$$\sum_{t=1}^T \tilde{\mathbf{X}}_t' (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^*) = \mathbf{0},$$

the first term in (A.3) is expressed as

$$\begin{aligned} & \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^*)' (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^*) - \frac{1}{2} \sum_{t=1}^T (\tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^* - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*)' (\tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^* - \tilde{\mathbf{X}}_t \boldsymbol{\theta}^*) \right\} \\ &= \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^*)' (\tilde{\mathbf{w}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\theta}}^*) - \frac{1}{2} \sum_{t=1}^T (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)' \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)' \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) \right\} \\ &= \exp \left\{ -\frac{1}{2} (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)' \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) \right\}. \end{aligned} \quad (\text{A.4})$$

Substituting (A.4) into (A.3), we have

$$\begin{aligned} & p(\boldsymbol{\theta}^* | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K) \\ &\propto \exp \left\{ -\frac{1}{2} (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)' \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) \right\} \\ &\quad \times |\mathbf{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)' \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) - \frac{1}{2} (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) \right\} \\ &= \exp \left\{ -\frac{1}{2} \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)' + \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \right] \right\} \end{aligned} \quad (\text{A.5})$$

where “tr” denotes the trace of matrix. We notice that

$$\begin{aligned} & \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)' + \mathbf{V}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\mu}) (\boldsymbol{\theta}^* - \boldsymbol{\mu})' \right] \\ &= \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) (\hat{\boldsymbol{\theta}}^* \hat{\boldsymbol{\theta}}^{*'} - \hat{\boldsymbol{\theta}}^* \boldsymbol{\theta}^{*'} - \boldsymbol{\theta}^* \hat{\boldsymbol{\theta}}^{*'} + \boldsymbol{\theta}^* \boldsymbol{\theta}^{*'}) \right] \end{aligned}$$

$$\begin{aligned}
& + \mathbf{V}^{-1} \left(\boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} - \boldsymbol{\theta}^* \boldsymbol{\mu}' - \boldsymbol{\mu} \boldsymbol{\theta}^{*\prime} + \boldsymbol{\mu} \boldsymbol{\mu}' \right) \Big] \\
= & \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) \left(\hat{\boldsymbol{\theta}}^* \hat{\boldsymbol{\theta}}^{*\prime} - 2 \hat{\boldsymbol{\theta}}^* \boldsymbol{\theta}^{*\prime} + \boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} \right) \right. \\
& \left. + \mathbf{V}^{-1} \left(\boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} - 2 \boldsymbol{\mu} \boldsymbol{\theta}^{*\prime} + \boldsymbol{\mu} \boldsymbol{\mu}' \right) \right] \\
= & \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \mathbf{V}^{-1} \right) \boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} - 2 \left\{ \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) \hat{\boldsymbol{\theta}}^* + \mathbf{V}^{-1} \boldsymbol{\mu} \right\} \boldsymbol{\theta}^{*\prime} \right. \\
& \left. + \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) \hat{\boldsymbol{\theta}}^* \hat{\boldsymbol{\theta}}^{*\prime} + \mathbf{V}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}' \right] \\
= & \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \mathbf{V}^{-1} \right) \boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} - 2 \left\{ \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) \hat{\boldsymbol{\theta}}^* + \mathbf{V}^{-1} \boldsymbol{\mu} \right\} \boldsymbol{\theta}^{*\prime} + \mathbf{C}_0 \right] \\
& \tag{A.6}
\end{aligned}$$

where

$$\mathbf{C}_0 = \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) \hat{\boldsymbol{\theta}}^* \hat{\boldsymbol{\theta}}^{*\prime} + \mathbf{V}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}'.$$

Substituting the OLS estimator $\hat{\boldsymbol{\theta}}^* = \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right)^{-1} \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\boldsymbol{w}}_t \right)$ into (A.6), we have

$$\begin{aligned}
& \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \mathbf{V}^{-1} \right) \boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} \right. \\
& \quad \left. - 2 \left\{ \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right) \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t \right)^{-1} \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\boldsymbol{w}}_t \right) + \mathbf{V}^{-1} \boldsymbol{\mu} \right\} \boldsymbol{\theta}^{*\prime} + \mathbf{C}_0 \right] \\
= & \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \mathbf{V}^{-1} \right) \boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} - 2 \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\boldsymbol{w}}_t + \mathbf{V}^{-1} \boldsymbol{\mu} \right) \boldsymbol{\theta}^{*\prime} + \mathbf{C}_0 \right] \\
= & \text{tr} \left[\left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \mathbf{V}^{-1} \right) \left\{ \boldsymbol{\theta}^* \boldsymbol{\theta}^{*\prime} - 2 \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \mathbf{V}^{-1} \right)^{-1} \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\boldsymbol{w}}_t + \mathbf{V}^{-1} \boldsymbol{\mu} \right) \boldsymbol{\theta}^{*\prime} \right\} \right. \\
& \left. + \mathbf{C}_0 \right]. \tag{A.7}
\end{aligned}$$

Let us denote

$$\mathbf{b} = \mathbf{B} \left(\sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{w}}_t + \mathbf{V}^{-1} \boldsymbol{\mu} \right), \quad (\text{A.8})$$

$$\mathbf{B}^{-1} = \sum_{t=1}^T \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \mathbf{V}^{-1}, \quad (\text{A.9})$$

then (A.7) can be rewritten as

$$\begin{aligned} & \text{tr} \left[\mathbf{B}^{-1} \left(\boldsymbol{\theta}^* \boldsymbol{\theta}^{*'} - 2\mathbf{b} \boldsymbol{\theta}^{*'} \right) + \mathbf{C}_0 \right] \\ &= \text{tr} \left[\mathbf{B}^{-1} (\boldsymbol{\theta}^* - \mathbf{b}) (\boldsymbol{\theta}^* - \mathbf{b})' - \mathbf{B}^{-1} \mathbf{b} \mathbf{b}' + \mathbf{C}_0 \right]. \end{aligned} \quad (\text{A.10})$$

Substituting (A.10) into (A.5), we have

$$\begin{aligned} & p(\boldsymbol{\theta}^* | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K) \\ & \propto \exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{B}^{-1} (\boldsymbol{\theta}^* - \mathbf{b}) (\boldsymbol{\theta}^* - \mathbf{b})' - \mathbf{B}^{-1} \mathbf{b} \mathbf{b}' + \mathbf{C}_0 \right] \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{B}^{-1} (\boldsymbol{\theta}^* - \mathbf{b}) (\boldsymbol{\theta}^* - \mathbf{b})' \right] \right\} \\ & = \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}^* - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\theta}^* - \mathbf{b}) \right\}. \end{aligned} \quad (\text{A.11})$$

Since $\tilde{\mathbf{w}}_t = \boldsymbol{\Sigma}_{s_t}^{-1/2} \mathbf{w}_t$ and $\tilde{\mathbf{X}}_t = \boldsymbol{\Sigma}_{s_t}^{-1/2} \mathbf{X}_t$, (A.8) and (A.9) can be rewritten as

$$\mathbf{b} = \mathbf{B} \left(\sum_{t=1}^T \mathbf{X}_t' \boldsymbol{\Sigma}_{s_t}^{-1} \mathbf{w}_t + \mathbf{V}^{-1} \boldsymbol{\mu} \right), \quad (\text{A.12})$$

$$\mathbf{B}^{-1} = \sum_{t=1}^T \mathbf{X}_t' \boldsymbol{\Sigma}_{s_t}^{-1} \mathbf{X}_t + \mathbf{V}^{-1}. \quad (\text{A.13})$$

From (A.11), (A.12) and (A.13), conditional posterior distribution of $\boldsymbol{\theta}^*$ is

$$\boldsymbol{\theta}^* | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K \sim \mathcal{N}(\mathbf{b}, \mathbf{B}). \quad (\text{A.14})$$

To generate samples of $\boldsymbol{\Sigma}_j$, $j = 1, 2, \dots, K$, we derive the conditional posterior distribution of $\boldsymbol{\Sigma}_j$ from (3.16). Applying the inverse Wishart distribution $\mathcal{IW}(\nu_j, \boldsymbol{\Lambda}_j)$ as a conjugate prior $p(\boldsymbol{\Sigma}_j)$, conditional posterior distribution of $\boldsymbol{\Sigma}_j$ is derived as

$$p(\boldsymbol{\Sigma}_j | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \boldsymbol{\theta}^*) \propto \prod_{t \in \{t: s_t = j\}} \left[(2\pi)^{-\frac{N-1}{2}} |\boldsymbol{\Sigma}_j|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \boldsymbol{\varepsilon}_t' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\varepsilon}_t \right) \right]$$

$$\begin{aligned}
& \times |\boldsymbol{\Sigma}_j|^{-\frac{\nu_j+(N-1)+1}{2}} \exp\left(-\frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Lambda}_j\}\right) \\
= & \prod_{t \in \{t: s_t=j\}} \left[(2\pi)^{-\frac{N-1}{2}} |\boldsymbol{\Sigma}_j|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t\}\right) \right] \\
& \times |\boldsymbol{\Sigma}_j|^{-\frac{\nu_j+(N-1)+1}{2}} \exp\left(-\frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Lambda}_j\}\right) \\
\propto & |\boldsymbol{\Sigma}_j|^{-\frac{n_j}{2}} \exp\left(-\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}_j^{-1}\sum_{t=1}^T \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t\mathbf{1}\{s_t=j\}\right\}\right) \\
& \times |\boldsymbol{\Sigma}_j|^{-\frac{\nu_j+(N-1)+1}{2}} \exp\left(-\frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Lambda}_j\}\right) \\
= & |\boldsymbol{\Sigma}_j|^{-\frac{\nu_j+(N-1)+1+n_j}{2}} \exp\left(-\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}_j^{-1}\left(\sum_{t=1}^T \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t\mathbf{1}\{s_t=j\} + \boldsymbol{\Lambda}_j\right)\right\}\right)
\end{aligned} \tag{A.15}$$

where n_j is the total number of time t belonging to regime j , and $\mathbf{1}\{s_t=j\}$ is indicator function which takes 1 if $s_t=j$, otherwise, takes 0. From (A.15), conditional posterior distribution of $\boldsymbol{\Sigma}_j$ is

$$\boldsymbol{\Sigma}_j | \mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \boldsymbol{\theta}^* \sim \mathcal{IW}\left(\nu_j + n_j, \sum_{t=1}^T \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t\mathbf{1}\{s_t=j\} + \boldsymbol{\Lambda}_j\right) \tag{A.16}$$

where $j = 1, 2, \dots, K$.

B Geweke (1992)'s convergence diagnostic

Geweke (1992) proposed a convergence diagnostic for Markov chains based on a test for equality of the means of the first and last part of a Markov chain after a burn-in period. Given the samples of parameter $\{\theta^{(j)}\}_{j=1}^n$, we consider the two sample averages $\bar{\theta}_A$ and $\bar{\theta}_B$ such that

$$\bar{\theta}_A = \frac{1}{n_A} \sum_{j=1}^{n_A} \theta^{(j)} \quad \text{and} \quad \bar{\theta}_B = \frac{1}{n_B} \sum_{j=n^*}^n \theta^{(j)} \tag{B.1}$$

where $n^* = n - n_B + 1$. From these values in (B.1), Geweke (1992) proposes the following statistic called convergence diagnostic (CD):

$$CD = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\mathbf{Var}(\bar{\theta}_A) + \mathbf{Var}(\bar{\theta}_B)}} \tag{B.2}$$

where $\mathbf{Var}(\bar{\theta}_A)$ and $\mathbf{Var}(\bar{\theta}_B)$ are variances of $\bar{\theta}_A$ and $\bar{\theta}_B$. If the sequence of $\theta^{(j)}$ is weak stationary, (B.2) converges in distribution to the standard normal from the central limit theorem (see Hamilton, 1994). Hence if absolute value of CD in (B.2) is greater than 1.96, null hypothesis $H_0 : \bar{\theta}_A = \bar{\theta}_B$ is rejected at 5% significant level and the convergence of Markov chain is judged not yet achieved.

To carry out the convergence diagnostic test proposed by Geweke (1992), we need to evaluate the variances $\mathbf{Var}(\bar{\theta}_A)$ and $\mathbf{Var}(\bar{\theta}_B)$ in (B.2). Let us define the autocovariance γ_{t-s} such that

$$\gamma_{t-s} = \mathbf{Cov}(\theta^{(s)}, \theta^{(t)}) \quad (\text{B.3})$$

then $\mathbf{Var}(\bar{\theta}_A)$ in (B.2) is

$$\begin{aligned} \mathbf{Var}(\bar{\theta}_A) &= \mathbf{Var}\left(\frac{1}{n_A} \sum_{j=1}^{n_A} \theta^{(j)}\right) \\ &= \frac{1}{n_A^2} \sum_{s=1}^{n_A} \sum_{t=1}^{n_A} \mathbf{Cov}(\theta^{(s)}, \theta^{(t)}) \\ &= \frac{1}{n_A^2} \sum_{s=1}^{n_A} \sum_{t=1}^{n_A} \gamma_{t-s}. \end{aligned} \quad (\text{B.4})$$

Here we can assume (B.3) because we assume the sequence of $\theta^{(j)}$ is weak stationary. Let $r = t - s$, ($1 - n_A \leq r \leq n_A - 1$), (B.4) can be rewritten as

$$\begin{aligned} \frac{1}{n_A^2} \sum_{s=1}^{n_A} \sum_{t=1}^{n_A} \gamma_{t-s} &= \frac{1}{n_A^2} \sum_{r=-(n_A-1)}^{n_A-1} (n_A - |r|) \gamma_r \\ &= \frac{1}{n_A} \sum_{r=-(n_A-1)}^{n_A-1} \left(1 - \frac{|r|}{n_A}\right) \gamma_r. \end{aligned} \quad (\text{B.5})$$

We define the population spectrum at frequency ω to be

$$f_A(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma_r \exp(-i^* \omega r) \quad (\text{B.6})$$

where $i^* = \sqrt{-1}$. To evaluate (B.5), we construct the sample analog of (B.6), which is known as the sample periodogram:

$$\hat{f}_A(\omega) = \frac{1}{2\pi} \sum_{r=-(n_A-1)}^{n_A-1} \hat{\gamma}_r \exp(-i^* \omega r). \quad (\text{B.7})$$

If the sample size n_A is large enough, then $\hat{\gamma}_r$ is close enough to γ_r and $\omega = 0$, then

$$\sum_{r=-(n_A-1)}^{n_A-1} \left(1 - \frac{|r|}{n_A}\right) \gamma_r \approx 2\pi \hat{f}_A(0). \quad (\text{B.8})$$

Therefore $\mathbf{Var}(\bar{\theta}_A)$ is estimated by $2\pi \hat{f}_A(0)/n_A$. In a similar fashion, $\mathbf{Var}(\bar{\theta}_B)$ is also estimated by $2\pi \hat{f}_B(0)/n_B$.

C Estimation of marginal likelihood

Given the parameter $\theta \in \Theta$, marginal likelihood for data $\mathbf{y} \equiv \{y_1, y_2, \dots, y_T\}$ conditional on model M is defined as

$$p(\mathbf{y}|M) = \int_{\Theta} \mathcal{L}(\theta|\mathbf{y}, M) p(\theta|M) d\theta \quad (\text{C.1})$$

where

$$\mathcal{L}(\theta|\mathbf{y}, M) = \prod_{t=1}^T p(y_t|\theta, M)$$

and $p(\theta|M)$ is a prior distribution for model M . To evaluate the marginal likelihood $p(\mathbf{y}|M)$ in (C.1), we suppose that $p(\theta|M)$ is a proper density. Then we notice that

$$\begin{aligned} 1 &= \int_{\Theta} p(\theta|M) d\theta \\ &= \int_{\Theta} \left[\frac{p(\mathbf{y}|M) p(\theta|\mathbf{y}, M)}{\mathcal{L}(\theta|\mathbf{y}, M) p(\theta|M)} \right] p(\theta|M) d\theta \\ &= p(\mathbf{y}|M) \int_{\Theta} \frac{1}{\mathcal{L}(\theta|\mathbf{y}, M)} p(\theta|\mathbf{y}, M) d\theta \end{aligned} \quad (\text{C.2})$$

where

$$\frac{p(\mathbf{y}|M) p(\theta|\mathbf{y}, M)}{\mathcal{L}(\theta|\mathbf{y}, M) p(\theta|M)} = \frac{p(\theta|\mathbf{y}, M)}{\frac{\mathcal{L}(\theta|\mathbf{y}, M) p(\theta|M)}{p(\mathbf{y}|M)}} = \frac{p(\theta|\mathbf{y}, M)}{p(\theta|\mathbf{y}, M)} = 1.$$

Given the samples $\{\theta^{(j)}\}_{j=1}^n$ from the posterior distribution $p(\theta|\mathbf{y}, M)$, Newton and Raftery (1994) estimates the marginal likelihood $p(\mathbf{y}|M)$ in (C.2) as

$$\hat{p}(\mathbf{y}|M) = \left[\frac{1}{n} \sum_{j=1}^n \frac{1}{\mathcal{L}(\theta^{(j)}|\mathbf{y}, M)} \right]^{-1} \quad (\text{C.3})$$

and the Bayes factor for model i against model j is obtained as

$$\text{BF}_{ij} = \frac{\hat{p}(\mathbf{y}|M_i)}{\hat{p}(\mathbf{y}|M_j)}. \quad (\text{C.4})$$

References

- Allais, O., and V. Nichèle (2007) ‘Capturing structural changes in french meat and fish demand over the period 1991–2002.’ *European review of agricultural economics* 34(4), 517–538
- Banks, J., R. Blundell, and A. Lewbel (1997) ‘Quadratic engel curves and consumer demand.’ *Review of Economics and Statistics* 79(4), 527–539
- Carter, C., and R. Kohn (1994) ‘On gibbs sampling for state space models.’ *Biometrika* 81(3), 541–553
- Chib, S. (1996) ‘Calculating posterior distributions and modal estimates in markov mixture models.’ *Journal of Econometrics* 75(1), 79–97
- Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1975) ‘Transcendental logarithmic utility functions.’ *American Economic Review* 65(3), 367–383
- Cooper, R. J., and K. R. McLaren (1992) ‘An empirically oriented demand system with improved regularity properties.’ *Canadian Journal of Economics* 25(3), 652–668
- Deaton, A., and J. Muellbauer (1980) ‘An almost ideal demand system.’ *The American Economic Review* 70(3), 312–326
- Diebold, F. X., J. H. Lee, and G. Weinbach (1994) ‘Regime switching with time-varying transition probabilities.’ in C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*. (Advanced Texts in Econometrics, CWJ Granger and G. Mizon, eds.), 283–302. Oxford: Oxford University Press.
- Geweke, J. (1992) ‘Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.’ In *Bayesian Statistics*, ed. A. P. Dawid J. M. Bernardo, J. O. Berger and A. F. M. Smith, 4 ed. (Oxford: Oxford University Press)

- Hamilton, J. D. (1989) ‘A new approach to the economic analysis of nonstationary time series and the business cycle.’ *Econometrica: Journal of the Econometric Society* 57(2), 357–384
- (1990) ‘Analysis of time series subject to changes in regime.’ *Journal of Econometrics* 45(1-2), 39–70
- (1991) ‘A quasi-bayesian approach to estimating parameters for mixtures of normal distributions.’ *Journal of Business & Economic Statistics* pp. 27–39
- (1994) *Time Series Analysis* (Princeton University Press)
- Ishida, T., N. Ishikawa, and M. Fukushige (2006) ‘Impact of bse and bird flu on consumers’ meat demand in japan.’ *DISCUSSION PAPER SERIES in Osaka university*
- Ishida, T., N. Ishikawa, and M. Fukushige (2010) ‘Impact of bse and bird flu on consumers’ meat demand in japan.’ *Applied Economics* 42(1), 49–56
- Jin, H. J., and W. W. Koo (2003) ‘The effect of the bse outbreak in japan on consumers’ preferences.’ *European Review of Agricultural Economics* 30(2), 173–192
- Kabe, S., and Y. Kanazawa (2012) ‘Another view of impact of bse crisis in japanese meat market through the almost ideal demand system model with markov switching.’ *Department of Social Systems and Management Discussion Paper Series No.1288 (University of Tsukuba), forthcoming in Applied Economics Letters (Accepted on December 2011)*. URL: <http://www.sk.tsukuba.ac.jp/SSM/libraries/list1276.php>
- Kass, R. E., and A. E. Raftery (1995) ‘Bayes factors.’ *The Journal of the American Statistical Association* 90(430), 773–795
- Kim, C. J., J. Piger, and R. Startz (2008) ‘Estimation of markov regime-switching regression models with endogenous switching.’ *Journal of Econometrics* 143(2), 263–273
- Moosa, I. A., and J. L. Baxter (2002) ‘Modeling the trend and seasonal within an aids model of demand for alcoholic beverage in the united kingdom.’ *Journal of Applied Econometrics* 17(2), 95–106
- Newton, M. A., and A. E. Raftery (1994) ‘Approximate bayesian inference with the weighted likelihood bootstrap.’ *Journal of the Royal Statistical Society, Series B* 56(1), 3–48

- Ohtani, K., and S. Katayama (1986) 'A gradual switching regression model with autocorrelated errors.' *Economics Letters* 21(2), 169–172
- Peterson, H. H., and Y. J. K. Chen (2005) 'The impact of bse on japanese retail meat demand.' *Agribusiness* 21(3), 313–327
- Rickertsen, K. (1996) 'Structural change and the demand for meat and fish in norway.' *European Review of Agricultural Economics* 23(3), 316–330
- Scott, S. L. (2002) 'Bayesian methods for hidden markov models.' *Journal of the American Statistical Association* 97(457), 337–351
- Theil, H. (1965) 'The information approach to demand analysis.' *Econometrica* 33(1), 67–87