数理解析研究所講究録
第 1910 巻 2014 年 74-101

Estimating the Markov-switching almost ideal demand systems: maximum likelihood or a Bayesian estimation?

Satoshi KABE [1] and Yuichiro KANAZAWA [2]
Division of Policy and Planning Sciences,
Faculty of Engineering, Information and Systems,
University of Tsukuba

This talk is organized as follows: In Part I,we first examine structural change points due to the food safety concerns in the Japanese meat market via MS-AIDS model proposed in Allais and Nichéle (2007). We then discuss problems associated with maximum likelihood estimation of Allais and Nichéle (2007).

In Part II, we instead propose a Bayesian estimation of MS-AIDS model. We run two sets of simulations to confirm its validity. We take the proposed method to the same Japanese meat market data as above and examine the regime shifts caused by the food safety concerns. Finally we compare these two results and discuss their implications.

Part I
MS-AIDS model via maximum likelihood

Applied Demand Analysis

Demand analysis is an applied area in economics where there is a *very well developed consumer theory*, that implies several *parameter restrictions*. Various econometric models (**logarithmic demand** function, **linear expenditure system**, **Rotterdam model**, **translog model**, and **almost ideal demand system (AIDS)**) have been developed in which *it is possible to test at least some of these restrictions*. See Figure 1 for its genealogy.

From an econometric point of view, these models are interesting in that they involve *complete system estimation* methods, however some of them involve *nonlinearity*.

It is important to understand that even when we are only interested in the demand for a *single* good, there are still *two* goods involved: the good in which we are interested and "*all other goods.*"We generally model this by thinking of the choice problem as a choice between the good in question and money to be spend on all other goods. The change in the demand for that good caused by its price change, is the result of two effects: a *substitution effect*, the result of a change in the relative prices of that good and all other goods; and an *income effect*, the effect of a change in price resulting in a change in the consumer's purchasing power.

The *compensated* or *Hicksian* demand function tells us what consumption bundle achieves a target level of utility and minimum total expenditure. The *Slutsky equation*

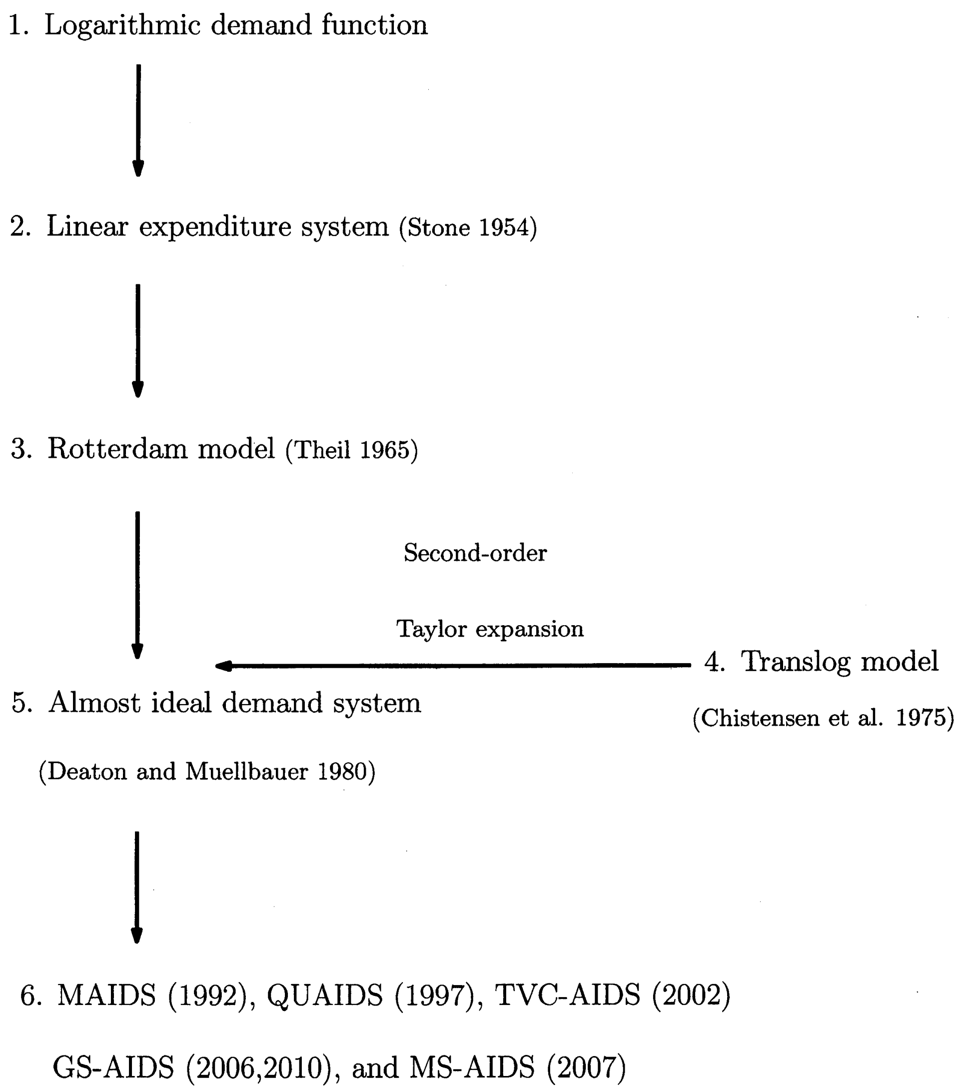

Figure 1: Demand System Model Genealogy

shows how the compensated or Hicksian demand function changes when price of i-th product changes. This changes is equal to the change in demand holding expenditure fixed *plus* the change in demand when income changes *times* how much income has to change to keep utility constant. In other words, the Slutsky equation decomposes the demand change induced by a price change into two separate effects: the *substitution effect* and the *income effect*.

Almost Ideal Demand System (AIDS) model

The Almost Ideal Demand System or AIDS was developed by Deaton and Muellbauer (1980). It has a *flexible functional form* that has the added advantage that it can be interpreted in terms of economic models of consumer behavior when estimated either with aggregated (macroeconomic) or household survey data. This model is generally considered to be the *best functional form* to estimate *static* systems of demand. *No fundamental advances have been made since 1980* on functional forms for demand systems, although some refinements have been made. One such refinement is made by Allais and Nichéle (2007).

Markov-Switching AIDS (MS-AIDS) model

Let us denote N be the number of products, q_i be a quantity of i-th product demanded by a representative consumer, p_i be a price of i-th product, m_0 be an expenditure (or budget) of a representative consumer defined as

$$\sum_{i=1}^{N} p_i q_i = m_0.$$

Also suppose $\bar{w}_i$ is a budget share of i-th product defined as

$$\bar{w}_i = \frac{p_i q_i}{m_0}.$$

Let s_t be an unobserved random variable that takes an integer value in $1, 2, \ldots, K$ to express "state" or "regime" at time t. The MS-AIDS model assumes that the budget share has the following form

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

where P_t is a price index defined by

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

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

The β_{i,s_t} parameters will be negative for *necessities* and positive for *luxury* goods.

The γ_{kj,s_t} parameters measure the change in i-th budget share following a proportional change to p_{jt} where real income as measured by m_{0t}/P_t is held constant.

The parameters in (1) and (2) will automatically satisfy "adding up" below, but both "homogeneity" and "symmetry" restrictions *can be tested.*

$$\textbf{[Adding up]} \qquad \sum_{i=1}^{N} \alpha_{i,s_t} = 1, \ \sum_{i=1}^{N} \gamma_{ij,s_t} = 0, \ \sum_{i=1}^{N} \beta_{i,s_t} = 0, \tag{3a}$$

$$\textbf{[Homogeneity]} \qquad \sum_{j=1}^{N} \gamma_{ij,s_t} = 0, \tag{3b}$$

$$\textbf{[Symmetry]} \qquad \gamma_{ij,s_t} = \gamma_{ji,s_t}. \tag{3c}$$

"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 $\tau p_{1t}, \ldots, \tau p_{Nt}$ 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, i.e., expenditure function $C(\boldsymbol{p}, u_0^*)$ satisfies $\frac{\partial^2 C(\mathbf{p}, u_0^*)}{\partial p_i \partial p_j} = \frac{\partial^2 C(\mathbf{p}, u_0^*)}{\partial p_j \partial p_i}$.

Following the previous studies (e.g., Ishida et al. 2010, Allais and Nichéle 2007), we include *trend* t, *seasonal effect* $d_{1,t}$ and $d_{2,t}$ and *habit effect* $\bar{w}_{j,t-1}$ into α_{i,s_t}:

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

where $d_{1,t}$ and $d_{2,t}$ are dummy variables, for instance, for meat market,

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

and habit effect is defined as a linear function of one-lagged budget shares.

To satisfy "adding up" restriction on α_{i,s_t}, we include the one-lagged budget shares of all products in the habit effect (Rickertsen 1996).

In order to satisfy the adding-up condition, we parameterize $\sum_{i=1}^{N} \bar{\alpha}_{i,s_t} = 1$, $\sum_{i=1}^{N} \nu_{i,s_t} = \sum_{i=1}^{N} \delta_{1,i} = \sum_{i=1}^{N} \delta_{2,i} = 0$ and $\sum_{i=1}^{N} \phi_{ij} = 0$. We also impose the restriction $\sum_{j=1}^{N} \phi_{ij} = 0$ to avoid identification problem.

The Markov switching mechanism switches regimes by the latent (unobserved) Markovian random variables. We assume that transitions between regimes are governed by a K-state Markov chain with transition probabilities:

$$\Pr(s_t = j | s_{t-1} = i) = \pi_{ij}, \quad i, j = 1, 2, \ldots, K, \tag{5}$$

and the transition matrix is defined as

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{21} & \cdots & \pi_{K1} \\ \pi_{12} & \pi_{22} & \cdots & \pi_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{1K} & \pi_{2K} & \cdots & \pi_{KK} \end{bmatrix} \tag{6}$$

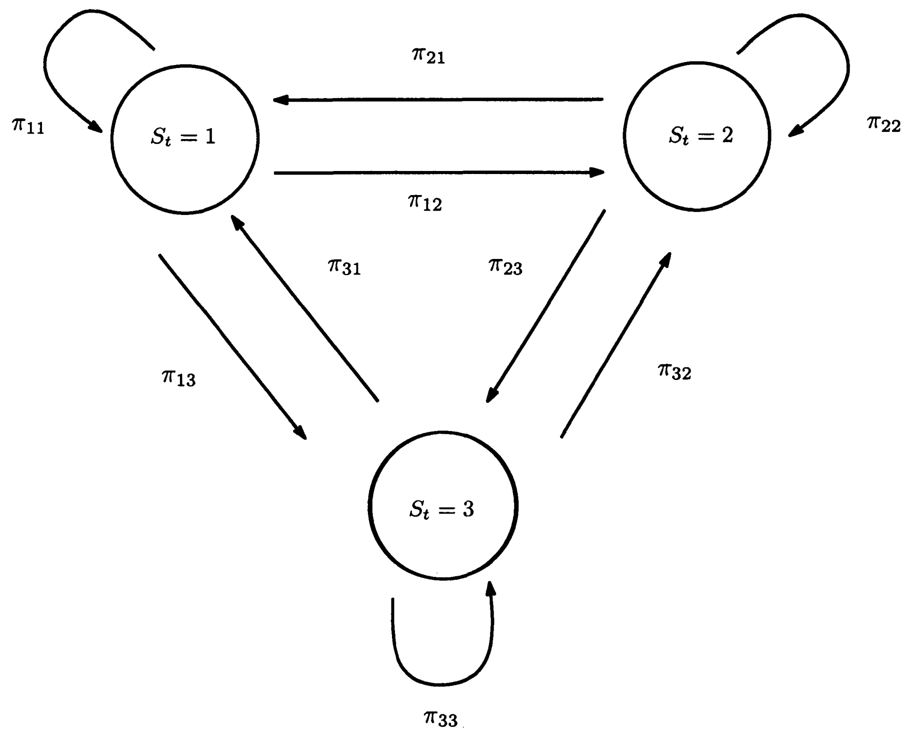

Figure 2: Schematic Diagram of Transition when $K = 3$

where $\pi_{i1} + \pi_{i2} + \cdots + \pi_{iK} = 1$, $i = 1, 2, \ldots, K$. See Figure 2 for the schematic diagram of transition when $K = 3$.

Likelihood function of MS-AIDS Model

Let us denote $\boldsymbol{w}_t$ as $(N-1) \times 1$ vector of budget shares at time t, $\bar{w}_{it}$, $i = 1, 2, \ldots, N-1$, $\boldsymbol{x}_t$ as vector of explanatory variables at time t, $\boldsymbol{\theta}_0$ as vector of regime-independent parameters, $\boldsymbol{\theta}_{s_t}$ as vector of regime-dependent parameters, $\boldsymbol{\Omega}_t$ as information set containing all observations obtained through time t, $\boldsymbol{\Omega}_t \equiv \{\boldsymbol{w}_t, \boldsymbol{w}_{t-1}, \ldots, \boldsymbol{w}_1, \boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_1\}$, $\boldsymbol{\mathcal{Z}}_t$ as vector of lags of $\boldsymbol{w}_t$ and observable explanatory variables obtained through time t, $\boldsymbol{\mathcal{Z}}_t \equiv \{\boldsymbol{w}_{t-1}, \boldsymbol{w}_{t-2}, \ldots, \boldsymbol{w}_1, \boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_1\}$, $\boldsymbol{\pi}$ as vector of transition probabilities π_{ij}, $i = 1, 2, \ldots, K, j = 1, 2, \ldots, K-1$, $\boldsymbol{\Theta}$ as a set of parameters such as $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}, \boldsymbol{\pi}\}$ where $\boldsymbol{\theta} \equiv \{\{\boldsymbol{\theta}_k\}_{k=0}^{K}, \{\Sigma_k\}_{k=1}^{K}\}$.

Suppose that distribution of $\boldsymbol{w}_t$ conditional on $\boldsymbol{x}_t$, s_t and $\boldsymbol{\theta}$ is defined as $p(\boldsymbol{w}_t | \boldsymbol{x}_t, s_t; \boldsymbol{\theta})$, the conditional log-likelihood function with respect to parameter set $\boldsymbol{\Theta}$ under all the observations $(\boldsymbol{w}_t, \boldsymbol{x}_t)$, $t = 1, 2, \ldots, T$ is

$$\ell(\boldsymbol{\Theta}) = \sum_{t=1}^{T} \log p(\boldsymbol{w}_t | \boldsymbol{\mathcal{Z}}_t; \boldsymbol{\Theta}) \tag{7}$$

where

$$p(\boldsymbol{w}_t|\boldsymbol{Z}_t;\boldsymbol{\Theta}) = \sum_{s_t=1}^{K} p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t;\boldsymbol{\theta}) \Pr(s_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi}). \tag{8}$$

Allais and Nichéle (2007) differentiate the above with respect to $\boldsymbol{\theta}$ and transition probabilities π_{ij} to obtain the score function.

According to Allais and Nichéle (2007), the score function with respect to $\boldsymbol{\theta}$ is

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{w}_t|\boldsymbol{Z}_t;\boldsymbol{\Theta})}{\partial \boldsymbol{\theta}} &= \sum_{s_t=1}^{K} \frac{\partial \log p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Pr(s_t|\boldsymbol{\Omega}_t;\boldsymbol{\Theta}) \\ &+ \sum_{\tau=1}^{t-1}\sum_{s_\tau=1}^{K} \frac{\partial \log p(\boldsymbol{w}_\tau|\boldsymbol{x}_\tau, s_\tau;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \{\Pr(s_\tau|\boldsymbol{\Omega}_t;\boldsymbol{\Theta}) - \Pr(s_\tau|\boldsymbol{\Omega}_{t-1};\boldsymbol{\Theta})\}\end{aligned} \tag{9}$$

for $t = 2, 3, \ldots, T$ and when $t = 1$,

$$\frac{\partial \log p(\boldsymbol{w}_1|\boldsymbol{Z}_1;\boldsymbol{\Theta})}{\partial \boldsymbol{\theta}} = \sum_{s_1=1}^{K} \frac{\partial \log p(\boldsymbol{w}_1|\boldsymbol{x}_1, s_1;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Pr(s_1|\boldsymbol{\Omega}_1;\boldsymbol{\Theta}). \tag{10}$$

The score function with respect to a transition probability π_{ij} is

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{w}_t|\boldsymbol{Z}_t;\boldsymbol{\Theta})}{\partial \pi_{ij}} &= \pi_{ij}^{-1} \Pr(s_t = j, s_{t-1} = i|\boldsymbol{\Omega}_t;\boldsymbol{\Theta}) - \pi_{iK}^{-1} \Pr(s_t = K, s_{t-1} = i|\boldsymbol{\Omega}_t;\boldsymbol{\Theta}) \\ &+ \pi_{ij}^{-1} \sum_{\tau=2}^{t-1} [\Pr(s_\tau = j, s_{\tau-1} = i|\boldsymbol{\Omega}_t;\boldsymbol{\Theta}) - \Pr(s_\tau = j, s_{\tau-1} = i|\boldsymbol{\Omega}_{t-1};\boldsymbol{\Theta})] \\ &- \pi_{iK}^{-1} \sum_{\tau=2}^{t-1} [\Pr(s_\tau = K, s_{\tau-1} = i|\boldsymbol{\Omega}_t;\boldsymbol{\Theta}) - \Pr(s_\tau = K, s_{\tau-1} = i|\boldsymbol{\Omega}_{t-1};\boldsymbol{\Theta})] \\ &+ \sum_{s_1=1}^{K} \frac{\partial \log \Pr(s_1;\boldsymbol{\pi})}{\partial \pi_{ij}} [\Pr(s_1|\boldsymbol{\Omega}_t;\boldsymbol{\Theta}) - \Pr(s_1|\boldsymbol{\Omega}_{t-1};\boldsymbol{\Theta})]\end{aligned} \tag{11}$$

for $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, K-1$ and $t = 2, 3, \ldots, T$, and when $t = 1$,

$$\frac{\partial \log p(\boldsymbol{w}_1|\boldsymbol{Z}_1;\boldsymbol{\Theta})}{\partial \pi_{ij}} = \sum_{s_1=1}^{K} \frac{\partial \log \Pr(s_1;\boldsymbol{\pi})}{\partial \pi_{ij}} \Pr(s_1|\boldsymbol{\Omega}_1;\boldsymbol{\Theta}). \tag{12}$$

Hamilton filter

The Hamilton filter is one of the filtering algorithms to estimate the probability of discrete latent variables $\Pr(s_t = j|\boldsymbol{\Omega}_t;\boldsymbol{\Theta})$ based on the data obtained through time t, $\boldsymbol{\Omega}_t$ and set of parameters $\boldsymbol{\Theta}$. The filtering algorithm obtains the estimates of unobserved variables from iterating the two steps ("update" and "prediction").

For instance, conditional probabilities of $s_t = 1$ $(t = 1, 2, \ldots, T)$ for two regimes system or $K = 2$ are iteratively estimated by the following two steps:

[Update Step]

$$\Pr(s_t = 1|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}) = \frac{\Pr(s_t = 1|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = 1; \boldsymbol{\theta})}{\sum_j \Pr(s_t = j|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = j; \boldsymbol{\theta})}$$

[Prediction Step]

$$\Pr(s_{t+1} = 1|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}) = \Pr(s_t = 1|\boldsymbol{\Omega}_t; \boldsymbol{\Theta})\pi_{11} + \Pr(s_t = 2|\boldsymbol{\Omega}_t; \boldsymbol{\Theta})\pi_{21}.$$

Let $\Pr(s_t = j|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta})$ be a conditional probability of being at regime j based on the data obtained through time $t-1$ given the parameter set $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}, \boldsymbol{\pi}\}$. The conditional joint density of $\boldsymbol{w}_t$ and s_t is given as

$$g(\boldsymbol{w}_t, s_t = j|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}) = \Pr(s_t = j|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = j; \boldsymbol{\theta}) \tag{13}$$

for $j = 1, 2, \ldots, K$. The density of $\boldsymbol{w}_t$ conditional on the past observed data set $\boldsymbol{\Omega}_{t-1}$ is

$$f(\boldsymbol{w}_t|\boldsymbol{x}_t, \boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) = \sum_{j=1}^{K} \Pr(s_t = j|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = j; \boldsymbol{\theta}). \tag{14}$$

From (13) and (14), we have by employing Bayes' theorem

$$\begin{aligned} \frac{g(\boldsymbol{w}_t, s_t = j|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta})}{f(\boldsymbol{w}_t|\boldsymbol{x}_t, \boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta})} &= \Pr(s_t = j|\boldsymbol{w}_t, \boldsymbol{x}_t, \boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) \\ &= \Pr(s_t = j|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}). \end{aligned} \tag{15}$$

Collecting the conditional probabilities $\Pr(s_t = j|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta})$, $j = 1, 2, \ldots, K$, in a $K \times 1$ vector $\hat{\boldsymbol{\xi}}_{t|t-1}$ below and the conditional densities $p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = j; \boldsymbol{\theta})$, $j = 1, 2, \ldots, K$, in a $K \times 1$ vector $\boldsymbol{\eta}_t$ below, (13) can be rewritten in a matrix form:

$$\begin{aligned} \begin{bmatrix} g(\boldsymbol{w}_t, s_t = 1|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}) \\ g(\boldsymbol{w}_t, s_t = 2|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}) \\ \vdots \\ g(\boldsymbol{w}_t, s_t = K|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}) \end{bmatrix} &= \begin{bmatrix} \Pr(s_t = 1|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) \\ \Pr(s_t = 2|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) \\ \vdots \\ \Pr(s_t = K|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) \end{bmatrix} \odot \begin{bmatrix} p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = 1; \boldsymbol{\theta}) \\ p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = 2; \boldsymbol{\theta}) \\ \vdots \\ p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = K; \boldsymbol{\theta}) \end{bmatrix} \\ &= \hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t \end{aligned} \tag{16}$$

and (14) can be rewritten as

$$f(\boldsymbol{w}_t|\boldsymbol{x}_t, \boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) = \mathbf{1}'_K \cdot \left(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t\right) \tag{17}$$

where $\mathbf{1}_K$ is a $K \times 1$ vector whose elements are 1 and the symbol $\odot$ represents element-by-element multiplication.

Using (16) and (17), (15) can be expressed in a matrix form as follows

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t}{\mathbf{1}'_K \cdot \left(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t\right)} \tag{18}$$

where

$$\hat{\boldsymbol{\xi}}_{t|t} = \begin{bmatrix} \Pr(s_t = 1|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}) \\ \Pr(s_t = 2|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}) \\ \vdots \\ \Pr(s_t = K|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}) \end{bmatrix} = \begin{bmatrix} \frac{g(\boldsymbol{w}_t, s_t=1|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta})}{f(\boldsymbol{w}_t|\boldsymbol{x}_t, \boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta})} \\ \frac{g(\boldsymbol{w}_t, s_t=2|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta})}{f(\boldsymbol{w}_t|\boldsymbol{x}_t, \boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta})} \\ \vdots \\ \frac{g(\boldsymbol{w}_t, s_t=K|\boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta})}{f(\boldsymbol{w}_t|\boldsymbol{x}_t, \boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta})} \end{bmatrix}.$$

Given an initial value $\hat{\boldsymbol{\xi}}_{1|0}$ and parameter set $\boldsymbol{\Theta}$, the optimal inference and forecasts of $\hat{\boldsymbol{\xi}}_{t|t}$ $(t = 1, 2, \ldots, T)$ can be found by iterating on the following pair of equations

[Update Step] $\quad \hat{\boldsymbol{\xi}}_{t|t} = \dfrac{\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t}{\mathbf{1}'_K \cdot \left(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}$

[Prediction Step] $\quad \hat{\boldsymbol{\xi}}_{t+1|t} = \boldsymbol{\Pi} \cdot \hat{\boldsymbol{\xi}}_{t|t}.$

Wen we consider the number of regimes $K = 2$, conditional probabilities of $s_t = 1$ $(t = 1, 2, \ldots, T)$ are estimated by iterating the following two steps:

[Update Step]

$$\Pr(s_t = 1|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}) = \frac{\Pr(s_t = 1|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = 1; \boldsymbol{\theta})}{\sum_j \Pr(s_t = j|\boldsymbol{\Omega}_{t-1}; \boldsymbol{\Theta}) p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t = j; \boldsymbol{\theta})}$$

[Prediction Step]

$$\Pr(s_{t+1} = 1|\boldsymbol{\Omega}_t; \boldsymbol{\Theta}) = \Pr(s_t = 1|\boldsymbol{\Omega}_t; \boldsymbol{\Theta})\pi_{11} + \Pr(s_t = 2|\boldsymbol{\Omega}_t; \boldsymbol{\Theta})\pi_{21}.$$

Minor improvement over Allais and Nichéle (2007)

We make the following minor improvements over Allais and Nichéle (2007) when we estimate MS-AIDS:

1) If transition probability π_{ij} $(0 \leq \pi_{ij} \leq 1, \sum_{j=1}^{K} \pi_{ij} = 1)$ has a boundary solution such as $\pi_{ij} = 0$ or 1, asymptotic normality of π_{ij} does not hold. However, Allais and Nichéle (2007) estimated π_{ij} **without any constraints**;

2) Allais and Nichéle (2007) estimated parameters σ^2_{ij,s_t} in variance-covariance matrix $\boldsymbol{\Sigma}_{s_t}$ with other parameters in MS-AIDS model. However, the maximum log-likelihood becomes infinite when $|\boldsymbol{\Sigma}_{s_t}|$ goes to zero (i.e., singularity problem) and a numerical maximization algorithm breaks down.

$$\Downarrow$$

1)′ We reparameterize the transition probabilities π_{ij} as $\lambda_{ij} = \log(\pi_{ij}/\pi_{iK})$ $(-\infty < \lambda_{ij} < \infty)$, a *natural sufficient statistic.* Then we estimate the parameter λ_{ij} to calculate the transition probability π_{ij}.

2)′ We calculate a maximum likelihood estimator $\widehat{\boldsymbol{\Sigma}}_{s_t}$ after estimating all parameters other than $\boldsymbol{\Sigma}_{s_t}$.

More concretely, as for **1)**′, we reparameterize the transition probability π_{ij} as follows

$$\lambda_{ij} = \log\left(\pi_{ij}/\pi_{iK}\right), \quad i = 1, 2, \ldots, K, \ j = 1, 2, \ldots, K-1$$

and estimate the parameter λ_{ij} instead of the transition probability π_{ij}. Since $\pi_{i1} + \pi_{i2} + \cdots + \pi_{iK} = 1$ and $\pi_{ij} = \pi_{iK}\exp(\lambda_{ij})$, we have

$$\pi_{iK}\exp(\lambda_{i1}) + \pi_{iK}\exp(\lambda_{i2}) + \cdots + \pi_{iK}\exp(\lambda_{iK-1}) + \pi_{iK} = 1.$$

The transition probabilities π_{iK} and π_{ij} are obtained as

$$\pi_{iK} = \frac{1}{1 + \exp(\lambda_{i1}) + \exp(\lambda_{i2}) + \cdots + \exp(\lambda_{iK-1})}$$

and

$$\begin{aligned}\pi_{ij} &= \pi_{iK}\exp(\lambda_{ij}) \\ &= \frac{\exp(\lambda_{ij})}{1 + \exp(\lambda_{i1}) + \exp(\lambda_{i2}) + \cdots + \exp(\lambda_{iK-1})}.\end{aligned}$$

To calculate the score function with respect to the parameter λ_{ij}, we apply the chain rule as follows

$$\frac{\partial \log p(\boldsymbol{w}_t|\boldsymbol{\mathcal{Z}}_t;\boldsymbol{\Theta})}{\partial \lambda_{ij}} = \frac{\partial \log p(\boldsymbol{w}_t|\boldsymbol{\mathcal{Z}}_t;\boldsymbol{\Theta})}{\partial \pi_{ij}} \times \frac{\partial \pi_{ij}}{\partial \lambda_{ij}}$$

and the partial derivative of π_{ij} with respect to λ_{ij} is obtained from

$$\begin{aligned}\frac{\partial \lambda_{ij}}{\partial \pi_{ij}} &= \frac{\partial}{\partial \pi_{ij}} \log\left(\pi_{ij}/\pi_{iK}\right) \\ &= \frac{\partial \log(\pi_{ij})}{\partial \pi_{ij}} - \frac{\partial \log(\pi_{iK})}{\partial \pi_{ij}} \\ &= \frac{\partial \log(\pi_{ij})}{\partial \pi_{ij}} - \frac{\partial \log(\pi_{iK})}{\partial \pi_{iK}} \times \frac{\partial \pi_{iK}}{\partial \pi_{ij}} \\ &= \frac{1}{\pi_{ij}} + \frac{1}{\pi_{iK}},\end{aligned}$$

where $i = 1, 2, \ldots, K, \ j = 1, 2, \ldots, K-1$.

As for **2)**′, the maximum likelihood estimator of variance-covariance matrix $\boldsymbol{\Sigma}_{s_t}$ is derived as follows: From (7) and (8), a first derivative of log-likelihood function with respect to inverse variance-covariance matrix $\boldsymbol{\Sigma}_{s_t}^{-1}$ of regime $s_t = j$ is

$$
\begin{aligned}
&\frac{\partial \ell(\Theta)}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} \\
&= \sum_t \frac{1}{\sum_{s_t} p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t;\boldsymbol{\theta}) \Pr(s_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi})} \frac{\partial \sum_{s_t} p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t;\boldsymbol{\theta}) \Pr(s_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi})}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} \\
&= \sum_t \frac{1}{\sum_{s_t} p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t;\boldsymbol{\theta}) \Pr(s_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi})} \frac{\partial p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t=j;\boldsymbol{\theta}) \Pr(s_t=j|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi})}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} \\
&= \sum_t \frac{p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t=j;\boldsymbol{\theta}) \Pr(s_t=j|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi})}{\sum_{s_t} p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t;\boldsymbol{\theta}) \Pr(s_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi})} \frac{\partial \log\left[p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t=j;\boldsymbol{\theta}) \Pr(s_t=j|\boldsymbol{\Omega}_{t-1};\boldsymbol{\theta},\boldsymbol{\pi})\right]}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} \\
&= \sum_t \Pr(s_t=j|\boldsymbol{\Omega}_t;\Theta) \frac{\partial \log p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t=j;\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}}.
\end{aligned}
$$

Assuming that

$$p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t;\boldsymbol{\theta}) = (2\pi)^{-\frac{N-1}{2}} |\boldsymbol{\Sigma}_{s_t}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{\Sigma}_{s_t}^{-1}\boldsymbol{\varepsilon}_t\right),$$

we have

$$
\begin{aligned}
\frac{\partial \log p(\boldsymbol{w}_t|\boldsymbol{x}_t, s_t=j;\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} &= \frac{\partial}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} \left[-\frac{N-1}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Sigma}_{s_t}|^{-1} - \frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{\Sigma}_{s_t}^{-1}\boldsymbol{\varepsilon}_t\right] \\
&= \frac{1}{2}\frac{\partial \log|\boldsymbol{\Sigma}_{s_t=j}|^{-1}}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} - \frac{1}{2}\frac{\partial\left(\boldsymbol{\varepsilon}_t'\boldsymbol{\Sigma}_{s_t=j}^{-1}\boldsymbol{\varepsilon}_t\right)}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} \\
&= \frac{1}{2}\frac{\partial \log|\boldsymbol{\Sigma}_{s_t=j}^{-1}|}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} - \frac{1}{2}\frac{\partial \operatorname{tr}\left\{\boldsymbol{\Sigma}_{s_t=j}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right\}}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} \\
&= \frac{1}{2}\boldsymbol{\Sigma}_{s_t=j} - \frac{1}{2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'
\end{aligned}
$$

and

$$\frac{\partial \ell(\Theta)}{\partial \boldsymbol{\Sigma}_{s_t=j}^{-1}} = \sum_{t=1}^{T} \Pr(s_t=j|\boldsymbol{\Omega}_t;\Theta)\left[\frac{1}{2}\boldsymbol{\Sigma}_{s_t=j} - \frac{1}{2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right] = \mathbf{0}.$$

Therefore we have

$$\widehat{\boldsymbol{\Sigma}}_{s_t=j} = \frac{\sum_{t=1}^{T} \Pr(s_t=j|\boldsymbol{\Omega}_t;\Theta)\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'}{\sum_{t=1}^{T} \Pr(s_t=j|\boldsymbol{\Omega}_t;\Theta)}.$$

Estimation of parameters

The parameters of MS-AIDS model are estimated by iterating the following steps:

Step 0. Set the initial values of parameters and set $g = 0$.

Step 1. Given the parameters at g-th iteration, calculate the conditional probabilities $\Pr(s_t = j|\boldsymbol{\Omega}_t;\Theta)$ from the Hamilton Filter.

Step 2. Calculate the score functions with respect to parameters.

Step 2a. Choose the parameter set $\Theta^{(g+1)}$ such that $\ell(\Theta^{(g+1)}) \geq \ell(\Theta^{(g)})$ by using the BHHH algorithm (Berndt, Hall, Hall, and Hausman 1974).

Step 2b. Find the MLE $\hat{\Theta}$ $(= \Theta^{(g+1)})$ by using the random optimization method.

Step 3. Find the maximum likelihood estimates of parameters and set $g = g + 1$.

Step 4. Repeat Steps 1 - 3 until the log-likelihood does not change.

Food safety concerns in the Japanese meat market

In Japanese meat market, there have been serious food safety issues concerning the **Bovine Spongiform Encephalopathy (BSE)** and **Bird flu**. On September 2001, the Japanese government announced the first BSE case within the country. The consumption of beef infected by BSE is suspected to cause variant Creutzfeldt Jacob Disease (vCJD). On January 2004, the first infected case of bird with H5N1 virus in Japan was confirmed in Yamaguchi prefecture. People infected with the H5N1 virus have died in Southeast Asia (e.g., Vietnam, Indonesia, and Thailand etc.

Data

The Ministry of Internal Affairs and Communications in Japan provides us with the household expenditure survey data called **the Family Income and Expenditure Survey**. It includes the monthly time-series average expenditure and price of meat and fish products along with others. The number of products $N = 4$ (beef, pork, chicken, fish). The number of observations $T = 108$ months (Jan 1998 – Dec 2006).

Results

We estimate the following models and select a preferred model.

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

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

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

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

We calculate the average budget share of i-th product at each regime s_t as

$$\bar{w}_{i,s_t} = \frac{\sum_{t=1}^{T} \Pr(s_t | \boldsymbol{\Omega}_t; \widehat{\boldsymbol{\Theta}}) \bar{w}_{it}}{\sum_{t=1}^{T} \Pr(s_t | \boldsymbol{\Omega}_t; \widehat{\boldsymbol{\Theta}})}$$

and Table 2 shows the results of average budget shares.

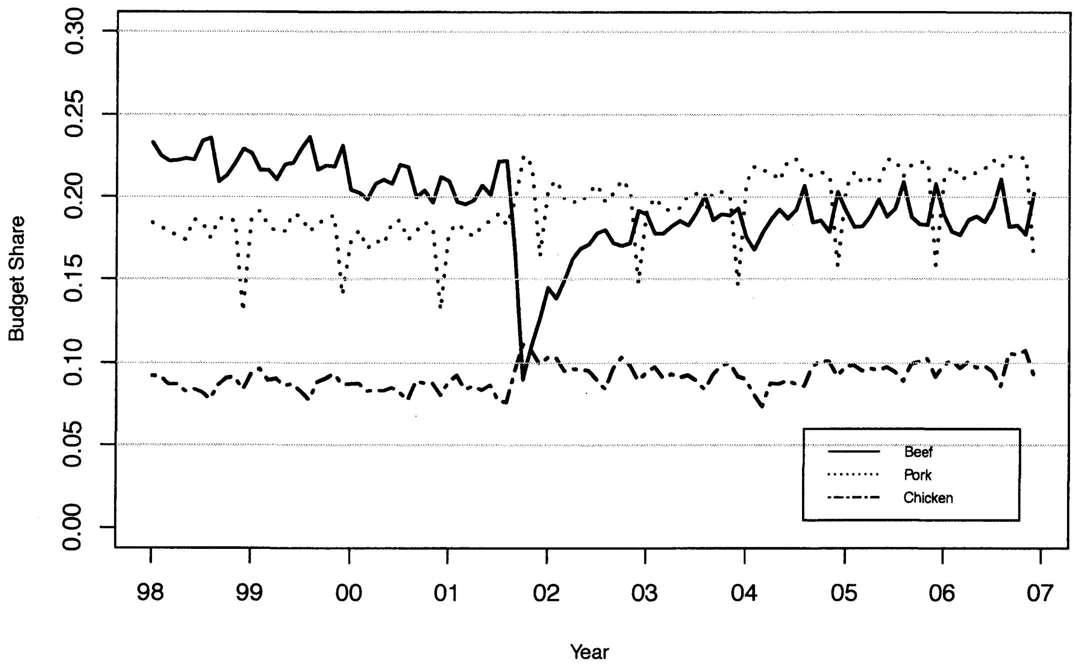

Figure 3: Plot of budget share data

Table 1: Model selection

Model	Log-likelihood	# of parameters	AIC	LR statics	d.f.	$\chi^2_{0.05}$(d.f.)
1	1160.12	38	-2244.24	211.95	21	32.67
2	1188.27	44	-2288.55	155.64	15	25.00
3	1243.39	53	-2380.77	45.41	13	12.59
4	1266.09	59	-2414.19	——	6	——

1) AIC : -2 log-likelihood + 2 (# of parameters)
2) LR statics : 2 (log-likelihood(null model) - log-likelihood(alternative model))
3) d.f. : degree of freedom
4) The number of regimes $K = 2$

Table 2: Estimated average budget share

	Regime 1	*Regime 2*
Beef	0.2155	0.1797
Pork	0.1781	0.2049
Chicken	0.0862	0.0955
Fish	0.5202	0.5199

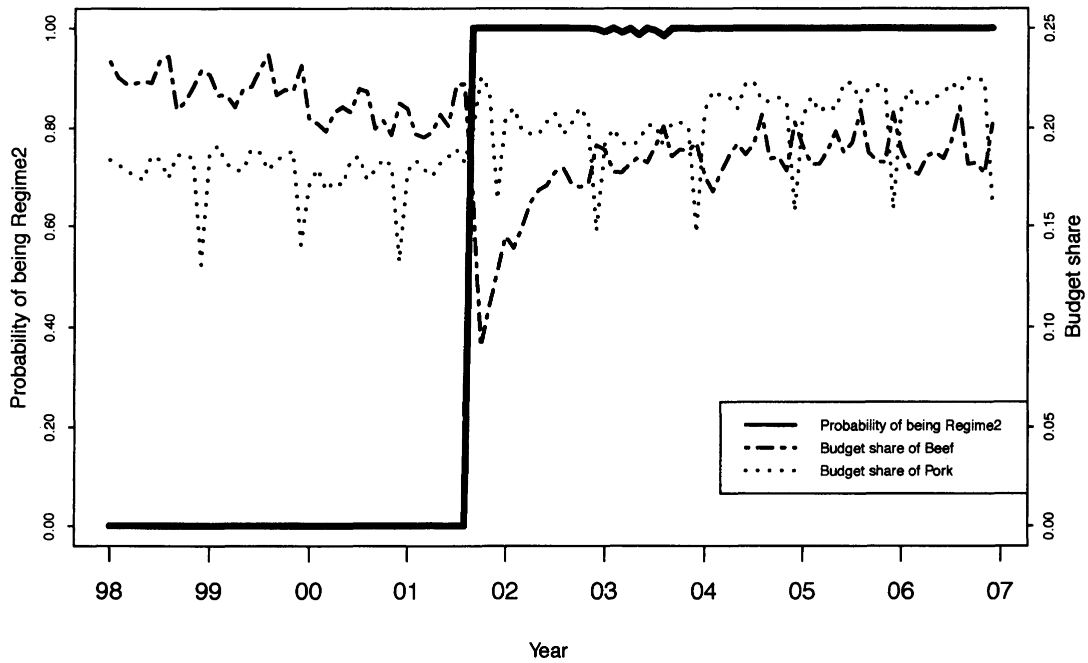

Figure 4: Probability of being regime 2, $\Pr(s_t = 2|\Omega_t, \widehat{\Theta})$ and budget share data of beef and pork under ML estimation.

Table 3: Estimated Parameters of MS-AIDS model

	Regime 1					*Regime 2*			
	Estimate	Sd. Error	t-value			Estimate	Sd. Error	t-value	
$\bar{\alpha}_1$	0.4359	0.1507	2.8920	***	$\bar{\alpha}_1$	0.2443	0.3114	0.7845	
$\bar{\alpha}_2$	0.4840	0.0754	6.4184	***	$\bar{\alpha}_2$	0.5429	0.1421	3.8199	***
$\bar{\alpha}_3$	0.2873	0.0944	3.0430	***	$\bar{\alpha}_3$	0.2299	0.1074	2.1408	**
γ_{11}	0.0068	0.0657	0.1035		γ_{11}	0.0552	0.0802	0.6879	
γ_{12}	-0.0174	0.0485	-0.3596		γ_{12}	-0.0197	0.0347	-0.5677	
γ_{13}	-0.0446	0.0384	-1.1592		γ_{13}	-0.0372	0.0277	-1.3400	
γ_{22}	0.0523	0.0364	1.4362		γ_{22}	0.0378	0.0301	1.2575	
γ_{23}	-0.0373	0.0348	-1.0715		γ_{23}	0.0065	0.0264	0.2467	
γ_{33}	0.1042	0.0404	2.5807	**	γ_{33}	0.0639	0.0297	2.1539	**
β_1	-0.0672	0.0325	-2.0659	**	β_1	-0.0404	0.0675	-0.5980	
β_2	-0.0687	0.0167	-4.1228	***	β_2	-0.0761	0.0303	-2.5097	**
β_3	-0.0126	0.0180	-0.6970		β_3	-0.0043	0.0231	-0.1875	

1) Significant level : ** 5%, *** 1%

Table 4: Estimated Parameters in Intercept term

Trend (Regime 1)

	Estimate	Sd. Error	t-value	
ν_1(beef)	-0.00039	0.00016	-2.4463	**
ν_2(pork)	-0.00001	0.00009	-0.1515	
ν_3(chicken)	-0.00006	0.00006	-1.1065	

Trend (Regime 2)

	Estimate	Sd. Error	t-value	
ν_1(beef)	0.00022	0.00032	0.6876	
ν_2(pork)	0.00017	0.00016	1.0501	
ν_3(chicken)	0.00015	0.00011	1.4211	

Seasonal (Aug)

	Estimate	Sd. Error	t-value	
δ_{11}(beef)	0.0115	0.0055	2.0894	**
δ_{12}(pork)	-0.0048	0.0028	-1.6942	
δ_{13}(chicken)	-0.0049	0.0029	-1.6673	

Seasonal (Dec)

	Estimate	Sd. Error	t-value	
δ_{21}(beef)	0.0186	0.0196	0.9480	
δ_{22}(pork)	-0.0254	0.0120	-2.1287	**
δ_{23}(chicken)	0.0079	0.0124	0.6329	

Habit effect

	Estimate	Sd. Error	t-value	
ϕ_{11}(beef)	0.1887	0.0709	2.6619	**
ϕ_{22}(pork)	0.0440	0.0831	0.5294	
ϕ_{33}(chicken)	0.3939	0.0747	5.2704	***

1) Significant level : ** 5%, *** 1%

Transition probability

	Estimate	Sd. Error	t-value	
λ_{11}	4.3297	0.1427	30.3483	***
λ_{21}	-5.1701	1.1361	-4.5509	***

Using the estimated parameters in Tables 3 and 4, we can calculate the Marshallian price elasticity η^P_{ij,s_t} and expenditure elasticity η^E_{i,s_t} at regime s_t as

$$\eta^P_{ij,s_t} = -\kappa_{ij} + \frac{\gamma_{ij,s_t}}{\bar{w}_{i,s_t}} - \frac{\beta_{i,s_t}}{\bar{w}_{i,s_t}}\left[\alpha_{j,s_t} + \sum_{k=1}^{N}\gamma_{kj,s_t}\log\bar{p}_{k,s_t}\right], \tag{19}$$

$$\eta^E_{i,s_t} = \frac{\beta_{i,s_t}}{\bar{w}_{i,s_t}} + 1 \tag{20}$$

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.

Conclusion of Part I

We capture *one abrupt* structural change point coinciding with timing of the first reported case of BSE in Japan, but not of bird flu (see Figure 4). We find that average budget share of beef declines after the first BSE case, while average budget share of pork increases significantly (see Table 2). We find that elasticities changed for beef and pork after the first BSE case: *Own-price elasticities of beef and pork* are *significant in pre-BSE period*, however own-price elasticity of beef in *post-BSE* period is *not* significant (see Table 5); This shift in regime 2 is likely to be the reflection of fact that *safety of beef* became very *important* for *some* Japanese consumers, but for *other* Japanese consumers who had *kept consuming beef*, price of beef *did not matter* very much.

Discussion of Part I

In the *Family Income and Expenditure Survey*, 8,076 two-or-more-person households and 745 one-person households from 168 cities, towns and villages are selected based on the three-stage—the municipality (i.e. city, town and village), the survey unit area and the household—stratified sampling method. The sampling units at three stages are: primarily

Table 5: Price elasticities and Expenditure elasticities

Regime 1	Price (η_{ij}^P)								Expenditure (η_i^E)	
	Beef		Pork		Chicken		Fish			
Beef	-0.815	**	0.062		-0.164		0.229		0.688	***
	(-2.311)		(0.227)		(-0.749)		(0.848)		(3.801)	
Pork	0.091		-0.529	***	-0.156		-0.020		0.615	***
	(0.384)		(-3.143)		(-0.981)		(-0.146)		(7.563)	
Chicken	-0.445		-0.366		0.228		-0.271		0.854	***
	(-1.132)		(-1.106)		(0.507)		(-1.539)		(4.529)	
Fish	-0.034		-0.126	*	-0.082	**	-1.043	***	1.285	***
	(-0.367)		(-1.947)		(-2.317)		(-11.017)		(18.381)	

Regime 2	Price (η_{ij}^P)								Expenditure (η_i^E)	
	Beef		Pork		Chicken		Fish			
Beef	-0.618		0.006		-0.181		0.016		0.776	**
	(-1.582)		(0.025)		(-1.078)		(0.059)		(2.064)	
Pork	0.032		-0.624	***	0.074		-0.110		0.629	***
	(0.181)		(-3.429)		(0.573)		(-0.893)		(4.241)	
Chicken	-0.373		0.091		-0.324		-0.347		0.953	***
	(-1.365)		(0.309)		(-1.012)		(-1.589)		(3.930)	
Fish	-0.076		-0.167	***	-0.090	**	-0.899	***	1.232	***
	(-1.185)		(-3.118)		(-2.350)		(-16.677)		(16.388)	

1) t-value in parentheses

2) Significant level : * 10%, ** 5%, *** 1%

the municipality (i.e. city, town and village), secondly the survey unit area and thirdly the household. It is a *good* but *expensive* survey available only monthly. Demand analysis of this sort *cannot* expect to be able to use *abundant* data. We believe meaningful analysis *should not* go too far back in time when the lifestyle was fundamentally different. We *hardly believe* the regime transition depicted in Figure 4: It was so *abrupt* with no tendency to revert to the original regime. Should we try something *different*?

Part II

Bayesian estimation of MS-AIDS model

Numerical optimization methods (e.g., Newton-Raphson method) have to depend on sensible selection of initial values of parameters to avoid singularity points on the parameter space and calculation of score functions is computationally intensive in maximum likelihood estimation of Allais and Nichéle (2007) and Kabe and Kanazawa (2012) as we saw in Part I.

In Bayesian estimation, we can incorporate prior information on variance-covariance matrices to avoid singularity problem (Hamilton 1991) and we can use of *conjugate* priors and *standard* algorithm such as Gibbs sampler to generate posterior (parameter) distributions in the *standard* formula (e.g., multivariate normal and inverse Wishart distributions).

Framework

Given the value of price index P_t in (2), the MS-AIDS model in (1) can be first rewritten by separating the parts that depend on regimes and by including the error term ε_{it} as

$$\begin{aligned}\bar{w}_{it} = \bar{\alpha}_{i,s_t} + \sum_{j=1}^{N-1} \gamma_{ij,s_t} \log\left(\frac{p_{jt}}{p_{Nt}}\right) + \beta_{i,s_t} \log\left(\frac{m_{0t}}{P_t}\right) + \nu_{i,s_t} t \\ + \delta_{1,i} d_{1,t} + \delta_{2,i} d_{2,t} + \sum_{j=1}^{N-1} \phi_{ij}(\bar{w}_{j,t-1} - \bar{w}_{N,t-1}) + \varepsilon_{it} \qquad (21)\end{aligned}$$

and thus can further be rewritten as the matrix form:

$$\boldsymbol{w}_t = \boldsymbol{X}_t^{(1)} \boldsymbol{\theta}_{s_t} + \boldsymbol{X}_t^{(0)} \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t \qquad (22)$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{s_t})$ 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 $\boldsymbol{X}_t^{(1)}$ and $\boldsymbol{X}_t^{(0)}$ are $(N-1) \times [\, 3(N-1) + N(N-1)/2 \,]$ and $(N-1) \times (N-1)(N+1)$.

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

$$\boldsymbol{X}_t^{(1)} \equiv \begin{bmatrix} \boldsymbol{I}_3 & \boldsymbol{P}_t & \boldsymbol{M}_t & \boldsymbol{T}_t \end{bmatrix}$$

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

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

and

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

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

$$\boldsymbol{X}_t^{(0)} \equiv \begin{bmatrix} \boldsymbol{D}_{1t} & \boldsymbol{D}_{2t} & \boldsymbol{W}_{1t} & \boldsymbol{W}_{2t} & \boldsymbol{W}_{3t} \end{bmatrix}$$

where

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

and

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

The 15 element parameter vector $\boldsymbol{\theta}_{s_t}$ 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 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 $\boldsymbol{\phi}_j \equiv [\ \phi_{1j}\ \phi_{2j}\ \phi_{3j}]'$.

Recall (22):

$$\boldsymbol{w}_t = \boldsymbol{X}_t^{(1)}\boldsymbol{\theta}_{s_t} + \boldsymbol{X}_t^{(0)}\boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{s_t})$$

and we can rewrite (22) as

$$\boldsymbol{w}_t = \boldsymbol{X}_t\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}_t. \tag{23}$$

The matrix $\boldsymbol{X}_t$ in (23) is defined as

$$\boldsymbol{X}_t = \begin{bmatrix} \mathbf{1}\{s_t = 1\}\boldsymbol{X}_t^{(1)} & \mathbf{1}\{s_t = 2\}\boldsymbol{X}_t^{(1)} & \cdots & \mathbf{1}\{s_t = K\}\boldsymbol{X}_t^{(1)} & \boldsymbol{X}_t^{(0)} \end{bmatrix}$$

and parameter vector $\boldsymbol{\theta}^*$ is defined as

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

We assume that likelihood function $\mathcal{L}(\cdot|\cdot)$ can be decomposed as

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\pi}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) = \mathcal{L}(\boldsymbol{\pi}|\mathcal{S}_T)\mathcal{L}(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) \tag{24}$$

where

$$\begin{aligned} \boldsymbol{\theta} &\equiv \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_K\}, \\ \boldsymbol{\pi} &\equiv \{\pi_{ij} : i, j = 1, 2, \ldots, K\}. \end{aligned}$$

Given a prior distribution of the form $p(\boldsymbol{\theta}, \boldsymbol{\pi}) = p(\boldsymbol{\theta})p(\boldsymbol{\pi})$, we obtain the posterior distributions with respect to $\boldsymbol{\theta}$ and to $\boldsymbol{\pi}$ as

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\pi}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T) &\propto \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\pi}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)p(\boldsymbol{\theta}, \boldsymbol{\pi}) \\ &= \mathcal{L}(\boldsymbol{\pi}|\mathcal{S}_T)p(\boldsymbol{\pi}) \times \mathcal{L}(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T)p(\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\pi}|\mathcal{S}_T) \times p(\boldsymbol{\theta}|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T). \end{aligned} \tag{25}$$

Gibbs sampler algorithm

Gibbs sampler with conjugate priors except for **Step 1** in each stage is set up as follows:

Step 0. Set the initial values of parameters and $g = 0$.

Step 1. Generate $\{s_t^{(g+1)}\}_{t=1}^T$ from $p(\{s_t\}_{t=1}^T \mid \{\boldsymbol{\theta}_j^{(g)}\}_{j=0}^K, \{\boldsymbol{\Sigma}_j^{(g)}\}_{j=1}^K, \{\pi_{ij}^{(g)}\}_{i,j=1}^K)$.

Step 2. Generate $\{\pi_{ij}^{(g+1)}\}_{i,j=1}^K$ from $p(\{\pi_{ij}\}_{i,j=1}^K \mid \{s_t^{(g+1)}\}_{t=1}^T)$.

Step 3. Generate $\{\boldsymbol{\theta}_j^{(g+1)}\}_{j=0}^K$ from $p(\{\boldsymbol{\theta}_j\}_{j=0}^K \mid \{s_t^{(g+1)}\}_{t=1}^T, \{\boldsymbol{\Sigma}_j^{(g)}\}_{j=1}^K)$.

Step 4. Generate $\{\boldsymbol{\Sigma}_j^{(g+1)}\}_{j=1}^K$ from $p(\{\boldsymbol{\Sigma}_j\}_{j=1}^K \mid \{s_t^{(g+1)}\}_{t=1}^T, \{\boldsymbol{\theta}_j^{(g+1)}\}_{j=0}^K)$.

Step 5. Repeat **Steps 1 - 4** with $g = g + 1$.

Now we are in a position to describe in more detail how we carry out each of these steps.

Step 1 : Sampling of latent variables $\{s_t\}_{t=1}^T$

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 carry out the following steps:

Step 1-0 : $s_T \sim \Pr(s_T|\boldsymbol{\Omega}_T, \boldsymbol{\Theta})$

Step 1-1 : For $t = T-1, T-2, \ldots, 1$,

$$s_t \sim \Pr(s_t|s_{t+1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}) = \frac{\Pr(s_{t+1}|s_t)\Pr(s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})}{\sum_{s_t=1}^K \Pr(s_{t+1}|s_t)\Pr(s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})}. \tag{26}$$

Note that a) $\Pr(s_{t+1}|s_t)$ is a transition probability in (5), b) $\Pr(s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta})$ can be derived from the **Hamilton filter** (Hamilton 1989), and c) this sampling algorithm is called **multi-move sampler** (Carter and Kohn 1994, Chib 1996).

Step 2 : Sampling of transition probabilities π_{ij}

Given the latent variables $s_1, s_2, \ldots, s_T$, posterior distribution of $\boldsymbol{\pi}_i = [\pi_{i1}\ \pi_{i2} \cdots \pi_{iK}]'$ is derived from

$$\begin{aligned} p(\boldsymbol{\pi}_i|\mathcal{S}_T) &\propto \prod_{j=1}^K \pi_{ij}^{n_{ij}} \times (\pi_{i1}^{u_{i1}-1} \cdots \pi_{iK}^{u_{iK}-1}) \\ &= \pi_{i1}^{n_{i1}+u_{i1}-1}\pi_{i2}^{n_{i2}+u_{i2}-1} \cdots \pi_{iK}^{n_{iK}+u_{iK}-1}. \end{aligned} \tag{27}$$

where n_{ij} is the total number of transitions from i to j, $\boldsymbol{\pi}_i = [\pi_{i1}\ \pi_{i2} \cdots \pi_{iK}]'$, n_{ij} is the total number of transitions from i to j, the prior is $\boldsymbol{\pi}_i \sim Dir(u_{i1}, u_{i2}, \ldots, u_{iK})$, and the posterior is $\boldsymbol{\pi}_i|\mathcal{S}_T \sim Dir(n_{i1}+u_{i1}, n_{i2}+u_{i2}, \ldots, n_{iK}+u_{iK})$.

Step 3 : Sampling of parameters $\boldsymbol{\theta}_0$**,** $\boldsymbol{\theta}_j$**,** $j = 1, 2, \ldots, K$

Posterior distribution of $\boldsymbol{\theta}^*$ conditional on $\{\boldsymbol{\Sigma}_j\}_{j=1}^K$ is derived from

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

where the prior is $\boldsymbol{\theta}^* \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{V})$, the posterior is $\boldsymbol{\theta}^*|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \{\boldsymbol{\Sigma}_j\}_{j=1}^K \sim \mathcal{N}(\boldsymbol{b}, \boldsymbol{B})$ with $\boldsymbol{b} = \boldsymbol{B}\left(\sum_{t=1}^T \boldsymbol{X}_t'\boldsymbol{\Sigma}_{s_t}^{-1}\boldsymbol{w}_t + \boldsymbol{V}^{-1}\boldsymbol{\mu}\right)$ and $\boldsymbol{B}^{-1} = \sum_{t=1}^T \boldsymbol{X}_t'\boldsymbol{\Sigma}_{s_t}^{-1}\boldsymbol{X}_t + \boldsymbol{V}^{-1}$.

Step 4 : Sampling of parameters $\boldsymbol{\Sigma}_j$**,** $j = 1, 2, \ldots, K$

Posterior distribution of $\boldsymbol{\Sigma}_j$ conditional on $\boldsymbol{\theta}^*$ is derived from

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

where n_j is the total number of time t belonging to regime j, the prior is $\boldsymbol{\Sigma}_j \sim \mathcal{IW}(\nu_j, \boldsymbol{\Lambda}_j)$, the posterior is $\boldsymbol{\Sigma}_j|\mathcal{Y}_T, \mathcal{S}_T, \mathcal{X}_T, \boldsymbol{\theta}^* \sim \mathcal{IW}\left(\nu_j + n_j, \sum_{t=1}^{T}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{1}\{s_t=j\}+\boldsymbol{\Lambda}_j\right)$.

Simulation

We conduct two sets of simulations. They share the following common setting: the number of iterations : $80,000$ (Burn-in period : $40,000$); the number of products $N = 4$, the number of regimes $K = 2$; the price data of four products p_{1t}, p_{2t}, p_{3t}, and p_{4t} generate from the uniform $p_{1t} \sim \text{U}(240, 380)$, $p_{2t} \sim \text{U}(130, 150)$, $p_{3t} \sim \text{U}(90, 100)$, and $p_{4t} \sim \text{U}(120, 200)$, the total expenditure (or budget) on four products, m_{0t} generates from $\text{U}(7500, 16000)$, and the prior distributions are set up as

$$\boldsymbol{\theta}^* \sim \mathcal{N}(\mathbf{0}, 10^4\boldsymbol{I}_{24}), \quad \boldsymbol{\Sigma}_j \sim \mathcal{IW}(10, 10^{-3}\boldsymbol{I}_3),$$
$$\pi_{11} \sim Beta(5,2), \quad \pi_{22} \sim Beta(5,2).$$

To estimate the parameters for the different regimes, we find that we need a *sufficient* number of data in *each regime* to avoid identification problem within the Gibbs sampler. In the following two simulation studies, we find out that when the number of observation fell below 240 out of 800 and 270 out of 900 (30%) data points in each regime respectively, the generated samples from the Gibbs sampling algorithm did *not converge* in terms of the Geweke's convergence diagnostic at 5% significance level. On the other hand, when the number of observations exceeds 320 out of 800 and 360 out of 900 data points (40%) respectively, all parameters converged *successfully.* Hence 40% threshold restriction is imposed on the latent variables $s_t, t = 1, 2, ..., T$: If generated sample of s_t satisfied this restriction, we updated the latent variables within the Gibbs sampler; otherwise we did not update them.

Simulation Result: Case 1

Regime shift occurs once during the period from regime one to two

The number of observations is $T = 800$. We specify unobserved latent variables $\mathcal{S}_T = \{s_1, s_2, \ldots, s_T\}$ such that $s_t = 1$ if $1 \leq t \leq 400$, and $s_t = 2$ if $401 \leq t \leq 800$. See Figure 5 below.

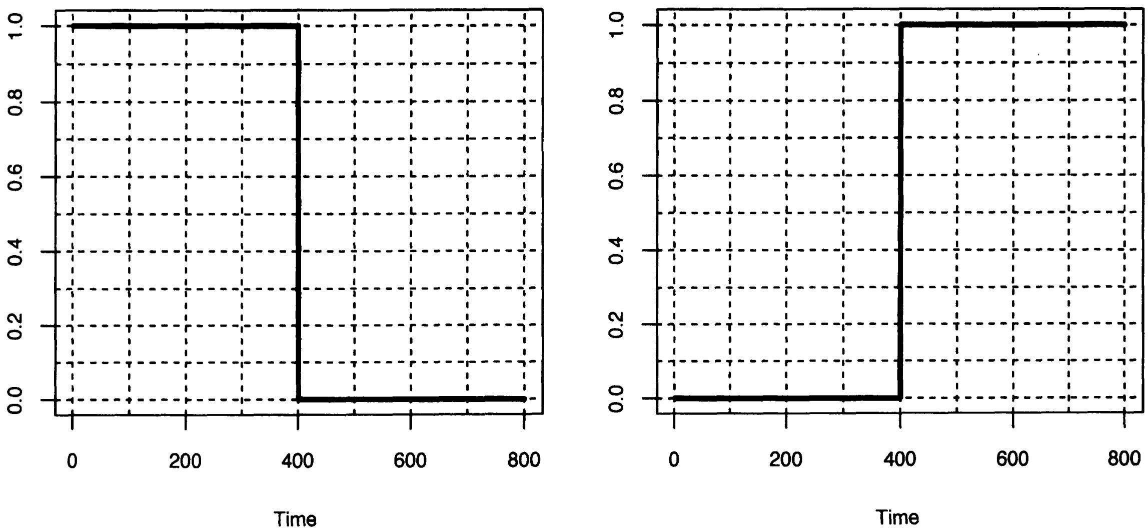

Figure 5: The left panel shows probability of $s_t = 1$, while the right panel shows the probability of $s_t = 2$ (Case 1)

Simulation Result: Case 2

Regime switches from one to two, and reverts to regime one again

The number of observations $T = 900$. We specify unobserved latent variables $\mathcal{S}_T = \{s_1, s_2, \ldots, s_T\}$ such that $s_t = 1$ if $1 \leq t \leq 200$, $s_t = 2$ if $201 \leq t \leq 600$, and $s_t = 1$ if $601 \leq t \leq 900$. See Figure 6.

Empirical study on Japanese meat market: Estimation Framework

We use the data in Part I with the proposed Bayesian estimation method. The number of iterations is 30,000 with Burn-in period of 5,000, the number of regimes is $K = 2$, and the following prior distributions

$$\begin{aligned} \boldsymbol{\theta}^* &\sim \mathcal{N}(\mathbf{0}, 10^4 \boldsymbol{I}_{45}), \quad \boldsymbol{\Sigma}_j \sim \mathcal{IW}(10, 10^{-3}\boldsymbol{I}_3), \\ \pi_{11} &\sim Beta(5,2), \qquad \pi_{22} \sim Beta(5,2). \end{aligned}$$

We set up four candidate models listed below:

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

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

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

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

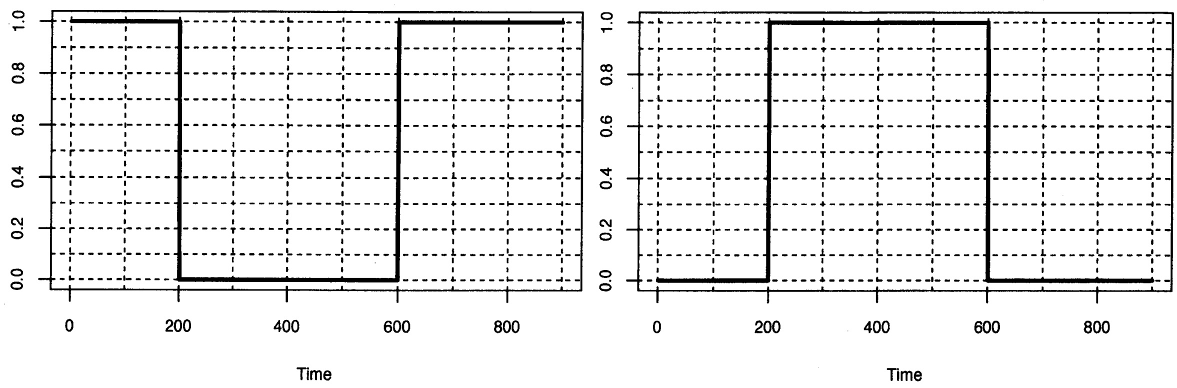

Figure 6: The left panel shows probability of $s_t = 1$, while the right panel shows the probability of $s_t = 2$ (Case 2)

Empirical study on Japanese meat market: Results

Table 6: Log-Marginal Likelihood and Log-Bayes Factor

	Model 1	*Model 2*	*Model 3*	*Model 4*
Model 1	1098.416	—	—	—
Model 2	18.927	1117.343	—	—
Model 3	62.157	43.230	1160.573	—
Model 4	63.452	44.525	1.295	**1161.868**

Table 6 shows the logarithmic marginal likelihood of model i, log-ML$_i$ ($i = 1, 2, 3, 4$) as diagonal elements and logarithmic Bayes factors, log-BF$_{ij}$ for model i against model j as off-diagonal elements. Logarithmic Bayes factor for model 4 against model 3, log-BF$_{43}$ (= 1.295) indicates "positive" (Kass and Raftery 1995) evidence in favor of model 4.

See Tables 7 and 8 for parameter estimates for regimes 1 and 2 respectively. We estimate the trend, seasonal, and habit effects to be as in Tables 9, 10, and 11 respectively as well.

To compare our proposed Bayesian estimation with the ML estimation employed in Part I, we calculate the mean squared errors (MSEs) for estimated budget shares in Table 12. Note that the MSEs of Bayesian estimation are evaluated by the posterior means of estimates of budget shares generated within the Gibbs sampler.

We calculate the average budget share of i-th product at regime $s_t = j$ 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 7: Estimated Parameters of MS-AIDS model in Regime 1

	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**

Table 8: Estimated Parameters of MS-AIDS model in Regime 2

	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**

Table 13 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.

Probability of being regime 2, $\Pr\{s_t = 2\}$ and budget share data of beef and pork under the proposed Bayesian framework is in Figure 7.

Table 9: Trend Effect in MS-AIDS model

	Mean	*SD*	*2.5%*	*50%*	*97.5%*	*CD*
Regime1						
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
Regime2						
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 10: Seasonal Effect in MS-AIDS model

	Mean	*SD*	*2.5%*	*50%*	*97.5%*	*CD*
August						
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**
December						
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 11: 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**

Table 12: 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}

Table 14 shows that posterior means of transition probabilities π_{11} and π_{22} are relatively

Table 13: 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

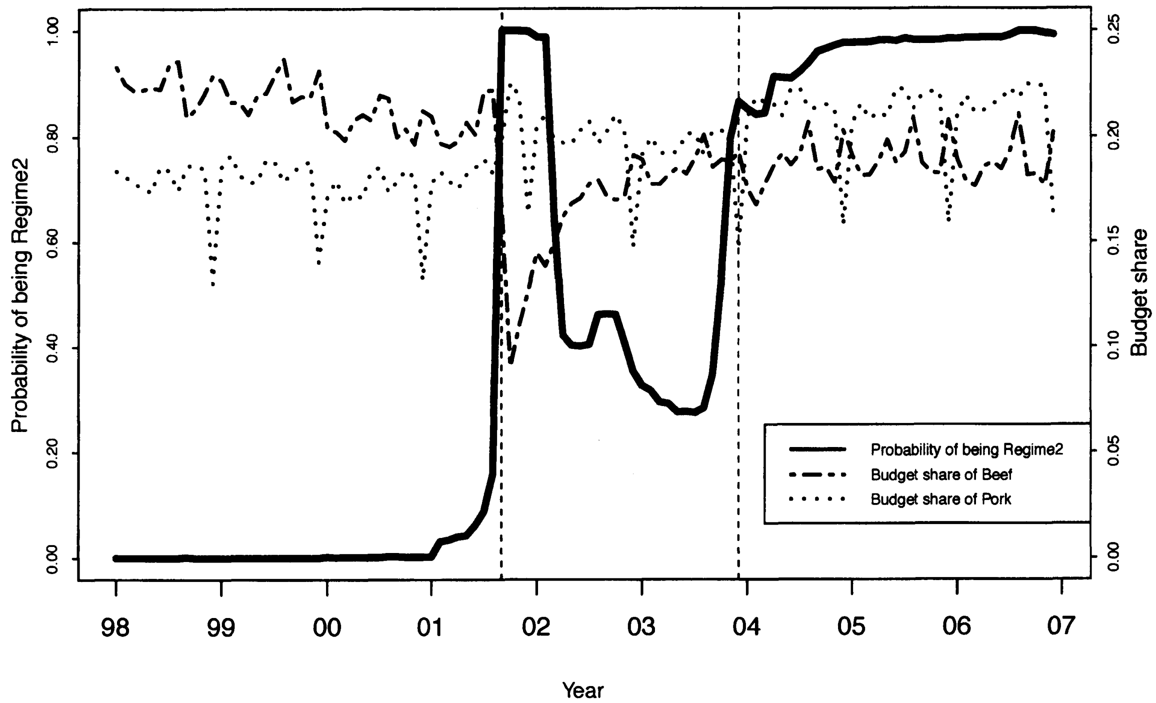

Figure 7: Probability of being regime 2, $\Pr\{s_t = 2\}$ and budget share data of beef and pork under the proposed Bayesian estimation. Two vertical dashed lines indicate the first BSE case in Japan on September 2001 and the first BSE case in U.S. on December 2003.

Table 14: 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**

high. This implies that there is little chance for switching from regime 1 to regime 2 and from regime 2 to regime 1. Also significant differences exist in terms of price elasticities between regimes 1 and 2 as seen in Table 15.

Table 15: Price elasticities and Expenditure elasticities

Regime 1	Price (η^P_{ij})				Expenditure (η^E_i)
	Beef	Pork	Chicken	Fish	
Beef	-0.817	0.060	-0.242	0.144	0.855
	(-1.468, -0.344)	(-0.284, 0.546)	(-0.593, 0.053)	(-0.175,0.511)	(0.609,1.109)
Pork	0.148	-0.514	0.027	-0.149	0.487
	(-0.230,0.721)	(-1.064,-0.036)	(-0.440,0.445)	(-0.449,0.175)	(0.278,0.694)
Chicken	-0.558	0.003	0.067	-0.291	0.779
	(-1.490,0.141)	(-0.976,0.878)	(-0.967,1.220)	(-0.803,0.325)	(0.405,1.169)
Fish	-0.029	-0.196	-0.093	-0.957	1.274
	(-0.168,0.126)	(-0.320,-0.070)	(-0.181,0.015)	(-1.176,-0.774)	(1.151,1.392)

Regime 2	Price (η^P_{ij})				Expenditure (η^E_i)
	Beef	Pork	Chicken	Fish	
Beef	-0.336	-0.168	-0.292	-0.021	0.817
	(-1.046,0.370)	(-0.557,0.238)	(-0.618,0.028)	(-0.452,0.431)	(0.369,1.218)
Pork	-0.096	-0.505	0.181	-0.122	0.542
	(-0.410,0.221)	(-0.936,-0.062)	(-0.146,0.517)	(-0.383,0.136)	(0.317,0.780)
Chicken	-0.559	0.319	-0.310	-0.333	0.884
	(-1.189,0.030)	(-0.400,1.056)	(-1.113,0.524)	(-0.778,0.113)	(0.451,1.314)
Fish	-0.089	-0.199	-0.099	-0.882	1.269
	(-0.262,0.069)	(-0.335,-0.081)	(-0.186,-0.015)	(-1.053,-0.700)	(1.128,1.412)

Notes: 95% credible interval in parentheses

Conclusion of Part II

We find via two simulations—the case where *regime shift occurs once* during the period from regime one to two and the case where *regime switches from one to two*, and *reverts to the original regime once again*—that the proposed Bayesian estimation method *works.* We find that the proposed *Bayesian estimation improves upon the ML estimation* in terms of the MSEs for all meat products (see Table 12).

Between the ML and Bayesian estimations, we find both contrasting and similar findings: *Contrasting* finding is that the probability of being regime 2 estimated via Bayesian estimation shows a **nuanced two step** regime shifts—the first and second waves arrive when the first BSE cases were reported in Japan in September 2001 and in U.S. in December 2003 respectively (see Figure 7), while a **one step** abrupt regime shift via ML estimation (see Figure 4); *Similar* findings include: posterior mean of average **budget share of beef declines after the first BSE** case, while **that of pork increases** significantly (see Table 13); The own-price elasticities of beef and pork are significant in regime 1, however **own-price elasticity of beef in regime 2 includes zero** within 95% credible interval (see Table 15).

Discussion of Part II

In Table 6, we employ logarithmic **Bayes factor** to justify Model 4. We can employ Bayes factor to find covariate selection for **linear regression** because *variable selection via Bayes factors employing Zellner's* g**-prior** *(1986) is consistent* from Fernández et al. (2001) and Liang et al. (2008). However, as noted in Liang et al. (2008), large spread of the prior induced by the **noninformative choice of** g **forces the Bayes factor to favor the smallest model**, regardless of the information in the data,"—"**Bartlett's paradox.**" **Bayes factors are** known to be **sensitive to the choice of the prior** on the parameters within each model. Even asymptotically, the influence of prior does not vanish (see Kass and Raftery, 1995; Fernández et al., 2001).

Future Direction

I came away from O-Bayes 2013 thinking that a *deep philosophical difference* exists between **prediction-based** or log-score model selection such as AIC and DIC—models are just a convenient tools to uncover quantifiable relationships—vis-a-vis **consistency based model selection** such as BIC and Bayes factor—a true model exists and a good model selection should be able to choose that model at least asymptotically. My *organizational behavior* and *criminology* research belongs to the former, while *demand analysis* and *discrete choice problems* belong to the latter. For this problem, I am leaning towards finding a model selection criterion based on the Generalized BIC by Konishi et. al. (2004) because with demand analysis, one always has to work with few data and thus small sample characteristics are very important. However, remember AIDS and its variants including MS-AIDS employed here are *close to being linear*, but *not quite* because of the price index terms.

References

[1] Allais, O. and Nichéle, V. (2007) "Capturing structural changes in French meat and fish demand over the period 1991 2002 ", *European review of agricultural economics.*

[2] Banks, J. and Blundell, R. and Lewbel, A. (1997) "Quadratic Engel curves and consumer demand", *Review of Economics and Statistics*, Vol.79, No.4, pp.527-539.

[3] Carter, C. and Kohn, R. (1994) "On gibbs sampling for state space models", *Biometrika*, 81(3), 541-553.

[4] Chib, S. (1996) "Calculating posterior distributions and model estimates in markov mixture models", *Journal of Econometrics*, 75(1), 79-97.

[5] Christensen, L.R. and Jorgenson, D.W. and Lau, L.J. (1975) "Transcendental logarithmic utility functions", *The American Economic Review*, pp.367-383.

[6] Cooper, R.J. and McLaren, K.R. (1992) "An empirically oriented demand system with improved regularity properties", *Canadian Journal of Economics*, Vol.25, No.3, pp.652-668.

[7] Deaton, A. and Muellbauer, J. (1980) "An almost ideal demand system", *The American Economic Review*, Vol.70, No.3, pp.312-326.

[8] 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*, Vol.57, No.2, pp.357-384.

[9] Diebold, F.X., Lee, J.H., and Weinbach, G. (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.

[10] Fernández, C., Ley, E., and Steel, M. F. (2001). "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, 100(2):381-427.

[11] 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. Bernado, J.O. Berger and A. F. M. Smith, 4 ed. (Oxford: Oxford University Press).

[12] 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.

[13] Hamilton, J.D. (1990) "Analysis of time series subject to changes in regime", *Journal of Econometrics*, 45(1-2), 39-70.

[14] Hamilton, J.D. (1991) "A quasi-Bayesian approach to estimating parameters for mixtures of normal distributions", *Journal of Business & Economic Statistics* , pp.27-39.

[15] Hamilton, J.D. (1996) "Specification testing in Markov-switching time-series models", *Journal of Econometrics*, Vol.70, No.1, pp.127-157.

[16] Ishida, T. and Ishikawa, N. and Fukushige, M. (2006) "Impact of BSE and Bird Flu on Consumers' Meat Demand in Japan", DISCUSSION PAPER SERIES, Osaka University.

Division of Policy and Planning Sciences
Faculty of Engineering, Information and Systems,
University of Tsukuba
Tsukuba 305-8573
JAPAN
E-mail address: `k0420214@sk.tsukuba.ac.jp`
E-mail address: `kanazawa@sk.tsukuba.ac.jp`

筑波大学・システム情報系・博士特別研究員　加部　哲史
筑波大学・システム情報系・社会工学域　金澤　雄一郎