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 Nichèle (2007). We then discuss problems associated with maximum likelihood estimation of Allais and Nichèle (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
1. Logarithmic demand function

2. Linear expenditure system (Stone 1954)

3. Rotterdam model (Theil 1965)

   Second-order

   Taylor expansion

4. Translog model

   (Chistensen et al. 1975)

5. Almost ideal demand system

   (Deaton and Muellbauer 1980)


Figure 1: Demand System Model Genealogy
shows how the compensated or Hicksian demand function changes when price of \( i \)-th product changes. This change 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 Nichèle (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

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

where \( P_t \) is a price index defined by

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

and \( \alpha_{0,s_t}, \alpha_{i,s_t}, \gamma_{ij,s_t} \) and \( \beta_{i,s_t} \) \((i, j = 1, 2, \ldots, N)\) are regime-dependent.

The \( \beta_{i,s_t} \) parameters will be negative for necessities and positive for luxury goods.
The $\gamma_{kj,s_t}$ 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.

\[
\text{[Adding up]} \quad \sum_{i=1}^{N} \alpha_{i,s_t} = 1, \quad \sum_{i=1}^{N} \gamma_{ij,s_t} = 0, \quad \sum_{i=1}^{N} \beta_{i,s_t} = 0, \quad (3a) \\
\text{[Homogeneity]} \quad \sum_{j=1}^{N} \gamma_{ij,s_t} = 0, \quad (3b) \\
\text{[Symmetry]} \quad \gamma_{ij,s_t} = \gamma_{ji,s_t}. \quad (3c)
\]

"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 $\tau > 0$, a representative consumer has to increase his expenditure from $m_{0t}$ to $\tau m_{0t}$ to keep his utility level. "Symmetry" guarantees that the substitution effect in the Slutsky equation is symmetric, i.e., expenditure function $C(p, u_{0}^{*})$ satisfies $\frac{\partial^{2}C(p,u_{0}^{*})}{\partial p_{i}\partial p_{j}} = \frac{\partial^{2}C(p,u_{0}^{*})}{\partial p_{j}\partial p_{i}}$.

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 $\alpha_{i,s_t}$:

\[
\alpha_{i,s_t} = \tilde{\alpha}_{i,s_t} + \nu_{i,s_t} + \delta_{1,i} d_{1,t} + \delta_{2,i} d_{2,t} + \sum_{j=1}^{N} \phi_{ij} \bar{w}_{j,t-1} \quad (4)
\]

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

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

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

To satisfy "adding up" restriction on $\alpha_{i,s_t}$, 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 $\sum_{i=1}^{N} \tilde{\alpha}_{i,s_t} = 1$, $\sum_{i=1}^{N} \nu_{i,s_t} = \sum_{i=1}^{N} \delta_{1,i} = \sum_{i=1}^{N} \delta_{2,i} = 0$ 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:

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

and the transition matrix is defined as

\[
\Pi = \begin{bmatrix}
\pi_{11} & \pi_{21} & \ldots & \pi_{K1} \\
\pi_{12} & \pi_{22} & \ldots & \pi_{K2} \\
\vdots & \vdots & \ddots & \vdots \\
\pi_{1K} & \pi_{2K} & \ldots & \pi_{KK}
\end{bmatrix} \quad (6)
\]
where $\pi_{11} + \pi_{12} + \cdots + \pi_{iK} = 1, i = 1, 2, \ldots, K$. See Figure 2 for the schematic diagram of transition when $K = 3$.

**Likelihood function of MS-AIDS Model**

Let us denote $w_t$ as $(N-1) \times 1$ vector of budget shares at time $t$, $\bar{w}_{it}, i = 1, 2, \ldots, N-1$, $x_t$ as vector of explanatory variables at time $t$, $\theta_0$ as vector of regime-independent parameters, $\theta_s$ as vector of regime-dependent parameters, $\Omega_t$ as information set containing all observations obtained through time $t$, $\Omega_t \equiv \{w_t, w_{t-1}, \ldots, w_1, x_t, x_{t-1}, \ldots, x_1\}$, $\mathcal{Z}_t$ as vector of lags of $w_t$ and observable explanatory variables obtained through time $t$, $\mathcal{Z}_t \equiv \{w_{t-1}, w_{t-2}, \ldots, w_1, x_t, x_{t-1}, \ldots, x_1\}$, $\pi$ as vector of transition probabilities $\pi_{ij}$, $i = 1, 2, \ldots, K, j = 1, 2, \ldots, K-1$, $\Theta$ as a set of parameters such as $\Theta \equiv \{\theta, \pi\}$ where $\theta \equiv \{\theta_k\}_{k=0}^{K}, \{\Sigma_k\}_{k=1}^{K}$.

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

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

(7)
where

\[ p(w_t | \mathcal{Z}_t; \Theta) = \sum_{s_t=1}^{K} p(w_t | x_t, s_t; \theta) \Pr(s_t | \Omega_t; \Theta, \pi). \] (8)

Allais and Nichèle (2007) differentiate the above with respect to \( \theta \) and transition probabilities \( \pi_{ij} \) to obtain the score function.

According to Allais and Nichèle (2007), the score function with respect to \( \theta \) is

\[
\frac{\partial \log p(w_t | \mathcal{Z}_t; \Theta)}{\partial \theta} = \sum_{s_t=1}^{K} \frac{\partial \log p(w_t | x_t, s_t; \theta)}{\partial \theta} \Pr(s_t | \Omega_t; \Theta) \\
+ \sum_{\tau=1}^{t-1} \sum_{s_{\tau}=1}^{K} \frac{\partial \log p(w_\tau | x_\tau, s_\tau; \theta)}{\partial \theta} \{ \Pr(s_\tau | \Omega_t; \Theta) - \Pr(s_\tau | \Omega_{t-1}; \Theta) \} 
\]

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

\[
\frac{\partial \log p(w_1 | \mathcal{Z}_1; \Theta)}{\partial \theta} = \sum_{s_1=1}^{K} \frac{\partial \log p(w_1 | x_1, s_1; \theta)}{\partial \theta} \Pr(s_1 | \Omega_1; \Theta). 
\]

The score function with respect to a transition probability \( \pi_{ij} \) is

\[
\frac{\partial \log p(w_t | \mathcal{Z}_t; \Theta)}{\partial \pi_{ij}} = \pi_{ij}^{-1} \Pr(s_t = j, s_{t-1} = i|\Omega_t; \Theta) - \pi_{iK}^{-1} \Pr(s_t = K, s_{t-1} = i|\Omega_t; \Theta) \\
+ \pi_{ij} \sum_{\tau=2}^{t-1} [\Pr(s_\tau = j, s_{\tau-1} = i|\Omega_t; \Theta) - \Pr(s_\tau = j, s_{\tau-1} = i|\Omega_{t-1}; \Theta)] \\
- \pi_{iK} \sum_{\tau=2}^{t-1} [\Pr(s_\tau = K, s_{\tau-1} = i|\Omega_t; \Theta) - \Pr(s_\tau = K, s_{\tau-1} = i|\Omega_{t-1}; \Theta)] \\
+ \sum_{s_1=1}^{K} \frac{\partial \log \Pr(s_1; \pi)}{\partial \pi_{ij}} [\Pr(s_1 | \Omega_t; \Theta) - \Pr(s_1 | \Omega_{t-1}; \Theta)] 
\]

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

\[
\frac{\partial \log p(w_1 | \mathcal{Z}_1; \Theta)}{\partial \pi_{ij}} = \sum_{s_1=1}^{K} \frac{\partial \log \Pr(s_1; \pi)}{\partial \pi_{ij}} \Pr(s_1 | \Omega_1; \Theta). 
\]

**Hamilton filter**

The Hamilton filter is one of the filtering algorithms to estimate the probability of discrete latent variables \( \Pr(s_t = j | \Omega_t; \Theta) \) based on the data obtained through time \( t, \Omega_t \) and set of parameters \( \Theta \). 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]

\[
\Pr(s_t = 1 | \Omega_t; \Theta) = \frac{\Pr(s_t = 1 | \Omega_{t-1}; \Theta)p(w_t | x_t, s_t = 1; \theta)}{\sum_{j} \Pr(s_t = j | \Omega_{t-1}; \Theta)p(w_t | x_t, s_t = j; \theta)}
\]

[Prediction Step]

\[
\Pr(s_{t+1} = 1 | \Omega_t; \Theta) = \Pr(s_t = 1 | \Omega_t; \Theta)\pi_{11} + \Pr(s_t = 2 | \Omega_t; \Theta)\pi_{21}.
\]

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

\[
g(w_t, s_t = j | \Omega_{t-1}, x_t; \Theta) = \Pr(s_t = j | \Omega_{t-1}; \Theta)p(w_t | x_t, s_t = j; \theta)
\]

for \( j = 1, 2, \ldots, K \). The density of \( w_t \) conditional on the past observed data set \( \Omega_{t-1} \) is

\[
f(w_t | x_t, \Omega_{t-1}; \Theta) = \sum_{j=1}^{K} \Pr(s_t = j | \Omega_{t-1}; \Theta)p(w_t | x_t, s_t = j; \theta).
\]

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

\[
\frac{g(w_t, s_t = j | \Omega_{t-1}, x_t; \Theta)}{f(w_t | x_t, \Omega_{t-1}; \Theta)} = \Pr(s_t = j | w_t, x_t, \Omega_{t-1}; \Theta)
\]

\[
= \Pr(s_t = j | \Omega_t; \Theta).
\]

Collecting the conditional probabilities \( \Pr(s_t = j | \Omega_{t-1}; \Theta) \), \( j = 1, 2, \ldots, K \), in a \( K \times 1 \) vector \( \hat{\xi}_{t|t-1} \) below and the conditional densities \( p(w_t | x_t, s_t = j; \theta) \), \( j = 1, 2, \ldots, K \), in a \( K \times 1 \) vector \( \eta_t \) below, (13) can be rewritten in a matrix form:

\[
\begin{bmatrix}
g(w_t, s_t = 1 | \Omega_{t-1}, x_t; \Theta) \\
g(w_t, s_t = 2 | \Omega_{t-1}, x_t; \Theta) \\
\vdots \\
g(w_t, s_t = K | \Omega_{t-1}, x_t; \Theta)
\end{bmatrix} =
\begin{bmatrix}
\Pr(s_t = 1 | \Omega_{t-1}; \Theta) \\
\Pr(s_t = 2 | \Omega_{t-1}; \Theta) \\
\vdots \\
\Pr(s_t = K | \Omega_{t-1}; \Theta)
\end{bmatrix}
\odot
\begin{bmatrix}
p(w_t | x_t, s_t = 1; \theta) \\
p(w_t | x_t, s_t = 2; \theta) \\
\vdots \\
p(w_t | x_t, s_t = K; \theta)
\end{bmatrix}
= \hat{\xi}_{t|t-1} \odot \eta_t
\]

and (14) can be rewritten as

\[
f(w_t | x_t, \Omega_{t-1}; \Theta) = \mathbf{1}_K' \cdot (\hat{\xi}_{t|t-1} \odot \eta_t)
\]

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

\[
\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \eta_t}{\mathbf{1}_K' \cdot (\hat{\xi}_{t|t-1} \odot \eta_t)}
\]
where

\[
\hat{\xi}_{t|t} = \begin{bmatrix}
\Pr(s_t = 1|\Omega_t; \Theta) \\
\Pr(s_t = 2|\Omega_t; \Theta) \\
\vdots \\
\Pr(s_t = K|\Omega_t; \Theta)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
g(w_t, s_t = 1|\Omega_{t-1}, x_t; \Theta) \\
g(w_t, s_t = 2|\Omega_{t-1}, x_t; \Theta) \\
\vdots \\
g(w_t, s_t = K|\Omega_{t-1}, x_t; \Theta)
\end{bmatrix}
\]

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

[Update Step ]

\[ \hat{\xi}_{t+1|t} = \Pi \cdot \hat{\xi}_{t|t} \]

[Prediction Step ]

\[ \hat{\xi}_{t+1|t} = \Pi \cdot \hat{\xi}_{t|t} \]

When 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 ]

\[ \Pr(s_t = 1|\Omega_t; \Theta) = \frac{\Pr(s_t = 1|\Omega_{t-1}; \Theta)p(w_t|x_t, s_t = 1; \theta)}{\sum_j \Pr(s_t = j|\Omega_{t-1}; \Theta)p(w_t|x_t, \mathcal{S}_t = j; \theta)} \]

[Prediction Step ]

\[ \Pr(s_{t+1} = 1|\Omega_t; \Theta) = \Pr(s_t = 1|\Omega_t; \Theta)\pi_{11} + \Pr(s_t = 2|\Omega_t; \Theta)\pi_{21}. \]

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

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

\[ \Downarrow \]

1') We reparameterize the transition probabilities \( \pi_{ij} \) as \( \lambda_{ij} = \log(\pi_{ij}/\pi_{iK}) \) \((-\infty < \lambda_{ij} < \infty)\), a **natural sufficient statistic**. Then we estimate the parameter \( \lambda_{ij} \) to calculate the transition probability \( \pi_{ij} \).
2') We calculate a maximum likelihood estimator $\hat{\Sigma}_{s_t}$ after estimating all parameters other than $\Sigma_{s_t}$.

More concretely, as for 1)', we reparameterize the transition probability $\pi_{ij}$ as follows

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

and estimate the parameter $\lambda_{ij}$ instead of the transition probability $\pi_{ij}$. 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 $\pi_{iK}$ and $\pi_{ij}$ are obtained as

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

and

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

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

$$\frac{\partial \log p(w_t|Z_t; \Theta)}{\partial \lambda_{ij}} = \frac{\partial \log p(w_t|Z_t; \Theta)}{\partial \pi_{ij}} \times \frac{\partial \pi_{ij}}{\partial \lambda_{ij}}$$

and the partial derivative of $\pi_{ij}$ with respect to $\lambda_{ij}$ is obtained from

$$\frac{\partial \lambda_{ij}}{\partial \pi_{ij}} = \frac{\partial}{\partial \pi_{ij}} \log \left( \frac{\pi_{ij}}{\pi_{iK}} \right) = \frac{\partial \log(\pi_{ij})}{\partial \pi_{ij}} - \frac{\partial \log(\pi_{iK})}{\partial \pi_{ij}}$$

$$= \frac{\partial \log(\pi_{ij})}{\partial \pi_{ij}} - \frac{\partial \log(\pi_{iK})}{\partial \pi_{iK}} \times \frac{\partial \pi_{iK}}{\partial \pi_{ij}}$$

$$= \frac{1}{\pi_{ij}} + \frac{1}{\pi_{iK}},$$

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

As for 2)', the maximum likelihood estimator of variance-covariance matrix $\Sigma_{s_t}$ is derived as follows: From (7) and (8), a first derivative of log-likelihood function with
respect to inverse variance-covariance matrix $\Sigma_{s_{t}}^{-1}$ of regime $s_{t} = j$ is

$$\frac{\partial \ell(\Theta)}{\partial \Sigma_{s_{t}=j}^{-1}} = \sum_{t} \frac{1}{\Sigma_{s_{t}} p(w_{t}|x_{t}, s_{t}; \theta) Pr(s_{t}|\Omega_{t-1}; \theta, \pi)} \frac{\partial \Sigma_{s_{t}} p(w_{t}|x_{t}, s_{t}; \theta) Pr(s_{t}|\Omega_{t-1}; \theta, \pi)}{\partial \Sigma_{s_{t}=j}^{-1}}$$

Assuming that

$$p(w_{t}|x_{t}, s_{t}; \theta) = (2\pi)^{-\frac{N-1}{2}} |\Sigma_{s_{t}}|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \epsilon_{t}'\Sigma_{s_{t}}^{-1}\epsilon_{t} \right),$$

we have

$$\frac{\partial \log p(w_{t}|x_{t}, s_{t}=j; \theta)}{\partial \Sigma_{s_{t}=j}^{-1}} = \frac{1}{2} \frac{\partial \log |\Sigma_{s_{t}=j}^{-1}|}{\partial \Sigma_{s_{t}=j}^{-1}} - \frac{1}{2} \frac{\partial (\epsilon_{t}'\Sigma_{s_{t}=j}^{-1}\epsilon_{t})}{\partial \Sigma_{s_{t}=j}^{-1}}$$

and

$$\frac{\partial \ell(\Theta)}{\partial \Sigma_{s_{t}=j}^{-1}} = \sum_{t=1}^{T} Pr(s_{t}=j|\Omega_{t}; \Theta) \left[ \frac{1}{2} \Sigma_{s_{t}=j}^{-1} - \frac{1}{2} \epsilon_{t}\epsilon_{t}' \right] = 0.$$

Therefore we have

$$\hat{\Sigma}_{s_{t}=j} = \frac{\sum_{t=1}^{T} Pr(s_{t}=j|\Omega_{t}; \Theta) \epsilon_{t}\epsilon_{t}'}{\sum_{t=1}^{T} Pr(s_{t}=j|\Omega_{t}; \Theta)}.$$

**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 $Pr(s_{t} = j|\Omega_{t}; \Theta)$ 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

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

and Table 2 shows the results of average budget shares.
Figure 3: Plot of budget share data

Table 1: Model selection

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

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

<table>
<thead>
<tr>
<th></th>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef</td>
<td>0.2155</td>
<td>0.1797</td>
</tr>
<tr>
<td>Pork</td>
<td>0.1781</td>
<td>0.2049</td>
</tr>
<tr>
<td>Chicken</td>
<td>0.0862</td>
<td>0.0955</td>
</tr>
<tr>
<td>Fish</td>
<td>0.5202</td>
<td>0.5199</td>
</tr>
</tbody>
</table>
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 3: Estimated Parameters of MS-AIDS model

<table>
<thead>
<tr>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\sigma}_1)</td>
<td>0.4359</td>
</tr>
<tr>
<td>(\hat{\sigma}_2)</td>
<td>0.2873</td>
</tr>
<tr>
<td>(\hat{\gamma}_1)</td>
<td>0.0068</td>
</tr>
<tr>
<td>(\hat{\gamma}_2)</td>
<td>0.0348</td>
</tr>
<tr>
<td>(\hat{\beta}_1)</td>
<td>0.1042</td>
</tr>
<tr>
<td>(\hat{\beta}_2)</td>
<td>-0.0687</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Sd. Error</th>
<th>t-value</th>
<th>Estimate</th>
<th>Sd. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\sigma}_1)</td>
<td>0.1507</td>
<td>2.8920</td>
<td>***</td>
<td>(\hat{\sigma}_1)</td>
<td>0.3114</td>
</tr>
<tr>
<td>(\hat{\sigma}_2)</td>
<td>0.0754</td>
<td>6.4184</td>
<td>***</td>
<td>(\hat{\sigma}_2)</td>
<td>0.1421</td>
</tr>
<tr>
<td>(\hat{\sigma}_3)</td>
<td>0.0944</td>
<td>3.0430</td>
<td>***</td>
<td>(\hat{\sigma}_3)</td>
<td>0.1074</td>
</tr>
<tr>
<td>(\hat{\gamma}_1)</td>
<td>0.0657</td>
<td>0.1035</td>
<td></td>
<td>(\hat{\gamma}_1)</td>
<td>0.0552</td>
</tr>
<tr>
<td>(\hat{\gamma}_2)</td>
<td>0.0485</td>
<td>-0.3596</td>
<td></td>
<td>(\hat{\gamma}_2)</td>
<td>-0.0197</td>
</tr>
<tr>
<td>(\hat{\gamma}_3)</td>
<td>0.0384</td>
<td>-1.1592</td>
<td></td>
<td>(\hat{\gamma}_3)</td>
<td>-0.0372</td>
</tr>
<tr>
<td>(\hat{\gamma}_4)</td>
<td>0.0364</td>
<td>1.4362</td>
<td></td>
<td>(\hat{\gamma}_4)</td>
<td>0.0378</td>
</tr>
<tr>
<td>(\hat{\gamma}_5)</td>
<td>0.0348</td>
<td>-1.0715</td>
<td></td>
<td>(\hat{\gamma}_5)</td>
<td>0.0065</td>
</tr>
<tr>
<td>(\hat{\gamma}_6)</td>
<td>0.0404</td>
<td>2.5807</td>
<td>**</td>
<td>(\hat{\gamma}_6)</td>
<td>0.0639</td>
</tr>
<tr>
<td>(\hat{\beta}_1)</td>
<td>0.0325</td>
<td>-2.0659</td>
<td>**</td>
<td>(\hat{\beta}_1)</td>
<td>-0.0404</td>
</tr>
<tr>
<td>(\hat{\beta}_2)</td>
<td>0.0167</td>
<td>-4.1228</td>
<td>***</td>
<td>(\hat{\beta}_2)</td>
<td>-0.0761</td>
</tr>
<tr>
<td>(\hat{\beta}_3)</td>
<td>0.0180</td>
<td>-0.6970</td>
<td></td>
<td>(\hat{\beta}_3)</td>
<td>-0.0043</td>
</tr>
</tbody>
</table>

1) Significant level: ** 5%, *** 1%
Table 4: Estimated Parameters in Intercept term

<table>
<thead>
<tr>
<th>Trend (Regime 1)</th>
<th>Estimate</th>
<th>Sd. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu_1(\text{beef}) )</td>
<td>-0.00039</td>
<td>0.00016</td>
<td>-2.4463 **</td>
</tr>
<tr>
<td>( \nu_2(\text{pork}) )</td>
<td>-0.00001</td>
<td>0.00009</td>
<td>-0.1515</td>
</tr>
<tr>
<td>( \nu_3(\text{chicken}) )</td>
<td>-0.00006</td>
<td>0.00006</td>
<td>-1.1065</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Seasonal (Aug)</th>
<th>Estimate</th>
<th>Sd. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_{11}(\text{beef}) )</td>
<td>0.0115</td>
<td>0.0055</td>
<td>2.0984 **</td>
</tr>
<tr>
<td>( \delta_{12}(\text{pork}) )</td>
<td>-0.0048</td>
<td>0.0028</td>
<td>-1.6942</td>
</tr>
<tr>
<td>( \delta_{13}(\text{chicken}) )</td>
<td>-0.0049</td>
<td>0.0029</td>
<td>-1.6673</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Habit effect</th>
<th>Estimate</th>
<th>Sd. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_{11}(\text{beef}) )</td>
<td>0.1887</td>
<td>0.0709</td>
<td>2.6619 **</td>
</tr>
<tr>
<td>( \phi_{21}(\text{pork}) )</td>
<td>0.0440</td>
<td>0.0831</td>
<td>0.5294</td>
</tr>
<tr>
<td>( \phi_{21}(\text{chicken}) )</td>
<td>0.3939</td>
<td>0.0747</td>
<td>5.2704 **</td>
</tr>
</tbody>
</table>

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

Using the estimated parameters in Tables 3 and 4, we can calculate the Marshallian price elasticity \( \eta_{ij,s_t}^P \) and expenditure elasticity \( \eta_{i,s_t}^E \) at regime \( s_t \) as

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

(19)

\[
\eta_{i,s_t}^E = \frac{\beta_{i,s_t}}{\bar{w}_{i,s_t}} + 1
\]

(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

<table>
<thead>
<tr>
<th>Regime 1</th>
<th>Price ($\eta_{ij}^P$)</th>
<th>Expenditure ($\eta_{ij}^E$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beef</td>
<td>Pork</td>
</tr>
<tr>
<td>Beef</td>
<td>-0.815 **</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(-2.311)</td>
<td>(0.227)</td>
</tr>
<tr>
<td>Pork</td>
<td>0.091</td>
<td>-0.529 ***</td>
</tr>
<tr>
<td></td>
<td>(0.384)</td>
<td>(-3.143)</td>
</tr>
<tr>
<td>Chicken</td>
<td>-0.445</td>
<td>-0.366</td>
</tr>
<tr>
<td></td>
<td>(-1.132)</td>
<td>(-1.106)</td>
</tr>
<tr>
<td>Fish</td>
<td>-0.034</td>
<td>-0.126 *</td>
</tr>
<tr>
<td></td>
<td>(-0.367)</td>
<td>(-1.947)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regime 2</th>
<th>Price ($\eta_{ij}^P$)</th>
<th>Expenditure ($\eta_{ij}^E$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beef</td>
<td>Pork</td>
</tr>
<tr>
<td>Beef</td>
<td>-0.618</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(-1.582)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Pork</td>
<td>0.032</td>
<td>-0.624 ***</td>
</tr>
<tr>
<td></td>
<td>(0.181)</td>
<td>(-3.429)</td>
</tr>
<tr>
<td>Chicken</td>
<td>-0.373</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>(-1.365)</td>
<td>(0.309)</td>
</tr>
<tr>
<td>Fish</td>
<td>-0.076</td>
<td>-0.167 ***</td>
</tr>
<tr>
<td></td>
<td>(-1.185)</td>
<td>(-3.118)</td>
</tr>
</tbody>
</table>

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 Nichèle (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 \( \epsilon_{it} \) as

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

and thus can further be rewritten as the matrix form:

\[
w_t = X_t^{(1)} \theta_{s_t} + X_t^{(0)} \theta_0 + \epsilon_t
\]

where \( \epsilon_t \sim \mathcal{N}(0, \Sigma_{s_t}) \) and \( \Sigma_{s_t} \) is also regime-dependent parameter such that \( \Sigma_{s_t} = \Sigma_j \) if time \( t \) belongs to regime \( j \). The size of the matrices \( X_t^{(1)} \) and \( X_t^{(0)} \) are \((N - 1) \times \left[ 3(N - 1) + N(N - 1)/2 \right] \) and \((N - 1) \times (N - 1)(N + 1)\).

Let us consider the case that the number of products \( N \) is four. Then \( 3 \times 15 \) matrix \( X_t^{(1)} \) is defined as

\[
X_t^{(1)} = \begin{bmatrix}
I_3 & P_t & M_t & T_t
\end{bmatrix}
\]
where $I_3$ is a $3 \times 3$ identity matrix,

$$M_t \equiv \begin{bmatrix}
\log\left(\frac{m_{P_t}}{P_t}\right) & 0 \\
0 & \log\left(\frac{m_{P_t}}{P_t}\right)
\end{bmatrix}, \quad T_t \equiv \begin{bmatrix}
t & 0 \\
0 & t
\end{bmatrix},$$

and

$$P_t \equiv \begin{bmatrix}
\log\left(\frac{P_{4t}}{P_t}\right) & \log\left(\frac{P_{4t}}{P_t}\right) & \log\left(\frac{P_{4t}}{P_t}\right) & 0 & 0 & 0 \\
0 & \log\left(\frac{P_{4t}}{P_t}\right) & 0 & \log\left(\frac{P_{4t}}{P_t}\right) & \log\left(\frac{P_{4t}}{P_t}\right) & 0 \\
0 & 0 & \log\left(\frac{P_{4t}}{P_t}\right) & 0 & \log\left(\frac{P_{4t}}{P_t}\right) & \log\left(\frac{P_{4t}}{P_t}\right)
\end{bmatrix}.$$

The $3 \times 15$ matrix $X_t^{(0)}$ is defined as

$$X_t^{(0)} \equiv \begin{bmatrix} D_{1t} & D_{2t} & W_t & W_{2t} & W_{3t} \end{bmatrix}$$

where

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

and

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

The 15 element parameter vector $\theta_{s_t}$ is defined as

$$\theta_{s_t} \equiv \begin{bmatrix}
\bar{\alpha}_{s_t} \\
\gamma_{s_t} \\
\beta_{s_t} \\
\nu_{s_t}
\end{bmatrix},$$

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

The 15 element parameter vector $\theta_0$ is defined as

$$\theta_0 \equiv \begin{bmatrix}
\delta_1 \\
\delta_2 \\
\phi_1 \\
\phi_2 \\
\phi_3
\end{bmatrix},$$

where $\delta_1 \equiv [\delta_{11}, \delta_{12}, \delta_{13}]'$, $\delta_2 \equiv [\delta_{21}, \delta_{22}, \delta_{23}]'$ and $\phi_j \equiv [\phi_{1j}, \phi_{2j}, \phi_{3j}]'$.

Recall (22):

$$w_t = X_t^{(1)} \theta_{s_t} + X_t^{(0)} \theta_0 + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_{s_t})$$

and we can rewrite (22) as

$$w_t = X_t \theta^* + \varepsilon_t.$$
The matrix $X_t$ in (23) is defined as

$$X_t = \begin{bmatrix} 1\{s_t = 1\}X_t^{(1)} & 1\{s_t = 2\}X_t^{(1)} & \cdots & 1\{s_t = K\}X_t^{(1)} & X_t^{(0)} \end{bmatrix}$$

and parameter vector $\theta^*$ is defined as

$$\theta^* = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \\ \theta_0 \end{bmatrix}.$$

We assume that likelihood function $L(\cdot)$ can be decomposed as

$$L(\theta, \pi|\mathcal{Y}_T, S_T, \mathcal{X}_T) = L(\pi|S_T)L(\theta|\mathcal{Y}_T, S_T, \mathcal{X}_T)$$

where

$$\theta \equiv \{\theta_0, \theta_1, \theta_2, \ldots, \theta_K, \Sigma_1, \Sigma_2, \ldots, \Sigma_K\},$$

$$\pi \equiv \{\pi_{ij}: i, j = 1, 2, \ldots, K\}.$$

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

$$p(\theta, \pi|\mathcal{Y}_T, S_T, \mathcal{X}_T) \propto L(\theta, \pi|\mathcal{Y}_T, S_T, \mathcal{X}_T)p(\theta, \pi)$$

$$= L(\pi|S_T)p(\pi) \times L(\theta|\mathcal{Y}_T, S_T, \mathcal{X}_T)p(\theta)$$

$$\propto p(\pi|S_T) \times p(\theta|\mathcal{Y}_T, S_T, \mathcal{X}_T).$$

**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 $p(\{s_t\}_{t=1}^T | \{\theta_j^{(g)}\}_{j=0}^K, \{\Sigma_j^{(g)}\}_{j=1}^K, \{\pi_{ij}^{(g)}\}_{i,j=1}^K)$.

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

**Step 3.** Generate $\{\theta_j^{(g+1)}\}_{j=0}^K$ from $p(\{\theta_j\}_{j=0}^K | \{s_t^{(g+1)}\}_{t=1}^T, \{\Sigma_j^{(g)}\}_{j=1}^K)$.

**Step 4.** Generate $\{\Sigma_j^{(g+1)}\}_{j=1}^K$ from $p(\{\Sigma_j\}_{j=1}^K | \{s_t^{(g+1)}\}_{t=1}^T, \{\theta_j^{(g+1)}\}_{j=0})$.

**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 \( t \), \( \Omega_t \equiv \{Y_t, X_t\} \) and set of parameters \( \Theta \equiv \{\theta, \pi\} \), we carry out the following steps:

**Step 1-0:** \( s_T \sim \Pr(s_T|\Omega_T, \Theta) \)

**Step 1-1:** For \( t = T-1, T-2 \), \( \ldots , 1 \),

\[
s_t \sim \Pr(s_t|s_{t+1}, \Omega_t, \Theta) = \frac{\Pr(s_{t+1}|s_t) \Pr(s_t|\Omega_t, \Theta)}{\sum_{s_t=1}^{K} \Pr(s_{t+1}|s_t) \Pr(s_t|\Omega_t, \Theta)}.
\] (26)

Note that a) \( \Pr(s_{t+1}|s_t) \) is a transition probability in (5), b) \( \Pr(s_t|\Omega_t, \Theta) \) 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 \( \pi_{ij} \)

Given the latent variables \( s_1, s_2, \ldots , s_T \), posterior distribution of \( \pi_i = [\pi_{i1}, \pi_{i2}, \ldots , \pi_{iK}]' \) is derived from

\[
p(\pi_i|S_T) \propto \prod_{j=1}^{K} \pi_{ij} \times (\pi_{i1}^{n_{i2}} \cdots \pi_{iK}^{n_{iK}-1})
= \prod_{j=1}^{K} \pi_{i1}^{n_{i1}+n_{i2}+\ldots+n_{iK}+u_{i1}-1} \cdots \pi_{iK}^{n_{iK}+u_{iK}-1}.
\] (27)

where \( n_{ij} \) is the total number of transitions from \( i \) to \( j \), \( \pi_i = [\pi_{i1}, \pi_{i2}, \ldots , \pi_{iK}]' \), \( n_{ij} \) is the total number of transitions from \( i \) to \( j \), the prior is \( \pi_i \sim \text{Dir}(u_{i1}, u_{i2}, \ldots , u_{iK}) \), and the posterior is \( \pi_i|S_T \sim \text{Dir}(n_{i1}+u_{i1}, n_{i2}+u_{i2}, \ldots , n_{iK}+u_{iK}) \).

**Step 3:** Sampling of parameters \( \theta_0, \theta_j, j = 1, 2, \ldots , K \)

Posterior distribution of \( \theta^* \) conditional on \( \{\Sigma_j\}_{j=1}^{K} \) is derived from

\[
p(\theta^*|Y_T, S_T, X_T, \{\Sigma_j\}_{j=1}^{K}) \propto \prod_{t=1}^{T} [ (2\pi)^{-\frac{N-1}{2}} |\Sigma_{s_t}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (w_t - X_t\theta^*)' \Sigma_{s_t}^{-1} (w_t - X_t\theta^*) \right\} ]
\times |V|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\theta^* - \mu)'V^{-1}(\theta^* - \mu) \right\}
\propto \exp \left\{ (\theta^* - b)'B^{-1}(\theta^* - b) \right\}.
\]

where the prior is \( \theta^* \sim N(\mu, V) \), the posterior is \( \theta^*|Y_T, S_T, X_T, \{\Sigma_j\}_{j=1}^{K} \sim N(\mu, V) \) with \( b = B \left( \sum_{t=1}^{T} X_t' \Sigma_{s_t}^{-1} w_t + V^{-1} \mu \right) \) and \( B^{-1} = \sum_{t=1}^{T} X_t' \Sigma_{s_t}^{-1} X_t + V^{-1} \).

**Step 4:** Sampling of parameters \( \Sigma_j, j = 1, 2, \ldots , K \)
Posterior distribution of $\Sigma_j$ conditional on $\theta^*$ is derived from

$$p(\Sigma_j | Y_T, S_T, X_T, \theta^*) \propto \prod_{t \in \{t: s_t = j\}} \left[ (2\pi)^{-\frac{N-1}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp\left( -\frac{1}{2} \epsilon_t' \Sigma_j^{-1} \epsilon_t \right) \right]$$

$$\propto |\Sigma_j|^{-\frac{\nu_j+(N-1)+n_j}{2}} \exp\left( -\frac{1}{2} \text{tr} \left\{ \Sigma_j^{-1} \sum_{t=1}^{T} \epsilon_t \epsilon_t' 1\{s_t = j\} + \Lambda_j \right\} \right)$$

where $n_j$ is the total number of time $t$ belonging to regime $j$, the prior is $\Sigma_j \sim \mathcal{IW}(\nu_j, \Lambda_j)$, the posterior is $\Sigma_j | Y_T, S_T, X_T, \theta^* \sim \mathcal{IW}(\nu_j + n_j, \sum_{t=1}^{T} \epsilon_t \epsilon_t' 1\{s_t = j\} + \Lambda_j)$.

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 $p_{1t} \sim U(240, 380)$, $p_{2t} \sim U(130, 150)$, $p_{3t} \sim U(90, 100)$, and $p_{4t} \sim U(120, 200)$, the total expenditure (or budget) on four products, $m_{0t}$ generates from $U(7500, 16000)$, and the prior distributions are set up as

$$\theta^* \sim \mathcal{N}(0, 10^4 I_{24}), \quad \Sigma_j \sim \mathcal{IW}(10, 10^{-3} I_3), \quad \pi_{11} \sim \text{Beta}(5, 2), \quad \pi_{22} \sim \text{Beta}(5, 2).$$

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 $S_T = \{s_1, s_2, \ldots, s_T\}$ 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.
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 $S_T = \{s_1, s_2, \ldots, s_T\}$ 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

$$\theta^* \sim \mathcal{N}(0, 10^4 I_{45}), \quad \Sigma_j \sim \mathcal{IW}(10, 10^{-3} I_3), \quad \pi_{11} \sim \text{Beta}(5, 2), \quad \pi_{22} \sim \text{Beta}(5, 2).$$

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 5: The left panel shows probability of $s_t = 1$, while the right panel shows the probability of $s_t = 2$ (Case 1)
Empirical study on Japanese meat market: Results

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1098.416</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Model 2</td>
<td>18.927</td>
<td>1117.343</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Model 3</td>
<td>62.157</td>
<td>43.230</td>
<td>1160.573</td>
<td>—</td>
</tr>
<tr>
<td>Model 4</td>
<td>63.452</td>
<td>44.525</td>
<td>1.295</td>
<td>1161.868</td>
</tr>
</tbody>
</table>

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

$$\bar{w}_{i,s_t=j} = \frac{\sum_{t=1}^{T} 1\{s_t = j\} \bar{w}_t}{\sum_{t=1}^{T} 1\{s_t = j\}}.$$
Table 7: Estimated Parameters of MS-AIDS model in Regime 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>0.2866</td>
<td>0.1262</td>
<td>0.0416</td>
<td>0.2850</td>
<td>0.5373</td>
<td>0.6055</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.6046</td>
<td>0.0925</td>
<td>0.4118</td>
<td>0.6074</td>
<td>0.7780</td>
<td>-0.3348</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.3285</td>
<td>0.0794</td>
<td>0.1694</td>
<td>0.3290</td>
<td>0.4824</td>
<td>-0.7486</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.0245</td>
<td>0.0613</td>
<td>-0.1131</td>
<td>0.0350</td>
<td>0.1229</td>
<td>-0.0015</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>-0.0039</td>
<td>0.0437</td>
<td>-0.0701</td>
<td>-0.0129</td>
<td>0.0951</td>
<td>0.3383</td>
</tr>
<tr>
<td>$\gamma_{13}$</td>
<td>-0.0549</td>
<td>0.0335</td>
<td>-0.1315</td>
<td>-0.0534</td>
<td>0.0060</td>
<td>-0.4092</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.0335</td>
<td>0.0473</td>
<td>-0.0693</td>
<td>0.0368</td>
<td>0.1174</td>
<td>-0.1208</td>
</tr>
<tr>
<td>$\gamma_{23}$</td>
<td>-0.0110</td>
<td>0.0403</td>
<td>-0.0947</td>
<td>-0.0096</td>
<td>0.0653</td>
<td>-0.7934</td>
</tr>
<tr>
<td>$\gamma_{33}$</td>
<td>0.0891</td>
<td>0.0480</td>
<td>-0.0020</td>
<td>0.0879</td>
<td>0.1892</td>
<td>0.7943</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.0301</td>
<td>0.0265</td>
<td>-0.0817</td>
<td>-0.0301</td>
<td>0.0226</td>
<td>-0.6585</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.0936</td>
<td>0.0193</td>
<td>-0.1320</td>
<td>-0.0936</td>
<td>-0.0553</td>
<td>-0.0383</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.0195</td>
<td>0.0170</td>
<td>-0.0524</td>
<td>-0.0198</td>
<td>0.0147</td>
<td>0.7026</td>
</tr>
<tr>
<td>$\sigma_{11}^2$</td>
<td>0.000001</td>
<td>0.000013</td>
<td>0.000041</td>
<td>0.000060</td>
<td>0.000091</td>
<td>1.3920</td>
</tr>
<tr>
<td>$\sigma_{12}^2$</td>
<td>-0.000003</td>
<td>0.000008</td>
<td>-0.000018</td>
<td>-0.000003</td>
<td>0.000013</td>
<td>-0.5299</td>
</tr>
<tr>
<td>$\sigma_{13}^2$</td>
<td>-0.000008</td>
<td>0.000007</td>
<td>-0.000022</td>
<td>-0.000008</td>
<td>0.000006</td>
<td>-0.1079</td>
</tr>
<tr>
<td>$\sigma_{22}^2$</td>
<td>0.000035</td>
<td>0.000008</td>
<td>0.000024</td>
<td>0.000034</td>
<td>0.000053</td>
<td>0.4690</td>
</tr>
<tr>
<td>$\sigma_{23}^2$</td>
<td>0.000006</td>
<td>0.000005</td>
<td>-0.000004</td>
<td>0.000006</td>
<td>0.000016</td>
<td>0.2985</td>
</tr>
<tr>
<td>$\sigma_{33}^2$</td>
<td>0.000029</td>
<td>0.000006</td>
<td>0.000019</td>
<td>0.000028</td>
<td>0.000044</td>
<td>1.1468</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>0.1926</td>
<td>0.1858</td>
<td>-0.1418</td>
<td>0.1817</td>
<td>0.5981</td>
<td>-0.4185</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.6542</td>
<td>0.1255</td>
<td>0.3993</td>
<td>0.6593</td>
<td>0.8891</td>
<td>-0.4349</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.2838</td>
<td>0.1017</td>
<td>0.0812</td>
<td>0.2843</td>
<td>0.4830</td>
<td>-0.6659</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.1032</td>
<td>0.0682</td>
<td>-0.0267</td>
<td>0.1056</td>
<td>0.2270</td>
<td>-0.1384</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>-0.0484</td>
<td>0.0355</td>
<td>-0.1164</td>
<td>-0.0489</td>
<td>0.0185</td>
<td>0.4018</td>
</tr>
<tr>
<td>$\gamma_{13}$</td>
<td>-0.0560</td>
<td>0.0299</td>
<td>-0.1154</td>
<td>-0.0555</td>
<td>0.0011</td>
<td>0.9316</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.0432</td>
<td>0.0417</td>
<td>-0.0368</td>
<td>0.0425</td>
<td>0.1272</td>
<td>0.0669</td>
</tr>
<tr>
<td>$\gamma_{23}$</td>
<td>0.0236</td>
<td>0.0343</td>
<td>-0.0436</td>
<td>0.0234</td>
<td>0.0923</td>
<td>-0.3294</td>
</tr>
<tr>
<td>$\gamma_{33}$</td>
<td>0.0628</td>
<td>0.0399</td>
<td>-0.0146</td>
<td>0.0620</td>
<td>0.1429</td>
<td>-1.1898</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.0329</td>
<td>0.0385</td>
<td>-0.1131</td>
<td>-0.0318</td>
<td>0.0392</td>
<td>0.8659</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.0949</td>
<td>0.0246</td>
<td>-0.1421</td>
<td>-0.0954</td>
<td>-0.0454</td>
<td>0.5194</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.0112</td>
<td>0.0210</td>
<td>-0.0529</td>
<td>-0.0111</td>
<td>0.0301</td>
<td>0.0316</td>
</tr>
<tr>
<td>$\sigma_{11}^2$</td>
<td>0.000161</td>
<td>0.000045</td>
<td>0.000096</td>
<td>0.000154</td>
<td>0.000267</td>
<td>-0.7579</td>
</tr>
<tr>
<td>$\sigma_{12}^2$</td>
<td>-0.000033</td>
<td>0.000018</td>
<td>-0.000074</td>
<td>-0.000031</td>
<td>-0.00004</td>
<td>0.2701</td>
</tr>
<tr>
<td>$\sigma_{13}^2$</td>
<td>-0.000029</td>
<td>0.000015</td>
<td>-0.000062</td>
<td>-0.000027</td>
<td>-0.000003</td>
<td>0.3091</td>
</tr>
<tr>
<td>$\sigma_{22}^2$</td>
<td>0.000045</td>
<td>0.000010</td>
<td>0.000030</td>
<td>0.000044</td>
<td>0.000069</td>
<td>-1.0053</td>
</tr>
<tr>
<td>$\sigma_{23}^2$</td>
<td>0.000004</td>
<td>0.000008</td>
<td>-0.000011</td>
<td>0.000004</td>
<td>0.000020</td>
<td>-0.0595</td>
</tr>
<tr>
<td>$\sigma_{33}^2$</td>
<td>0.000042</td>
<td>0.000010</td>
<td>0.000027</td>
<td>0.000041</td>
<td>0.000065</td>
<td>-1.3308</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Regime1</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_{11}$ (beef)</td>
<td>-0.00038</td>
<td>0.00014</td>
<td>-0.00064</td>
<td>-0.00040</td>
<td>-0.00007</td>
<td>0.2165</td>
</tr>
<tr>
<td>$\nu_{21}$ (pork)</td>
<td>0.00003</td>
<td>0.00010</td>
<td>-0.00018</td>
<td>0.00004</td>
<td>0.00021</td>
<td>-0.9640</td>
</tr>
<tr>
<td>$\nu_{31}$ (chicken)</td>
<td>0.00002</td>
<td>0.00008</td>
<td>-0.00016</td>
<td>0.00002</td>
<td>0.00017</td>
<td>-0.6504</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regime2</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_{12}$ (beef)</td>
<td>0.00016</td>
<td>0.00022</td>
<td>-0.00029</td>
<td>0.00017</td>
<td>0.00058</td>
<td>0.9870</td>
</tr>
<tr>
<td>$\nu_{22}$ (pork)</td>
<td>0.00019</td>
<td>0.00012</td>
<td>-0.00005</td>
<td>0.00019</td>
<td>0.00044</td>
<td>-0.0507</td>
</tr>
<tr>
<td>$\nu_{32}$ (chicken)</td>
<td>0.00015</td>
<td>0.00012</td>
<td>-0.00009</td>
<td>0.00015</td>
<td>0.00038</td>
<td>-1.7246</td>
</tr>
</tbody>
</table>

Table 10: Seasonal Effect in MS-AIDS model

<table>
<thead>
<tr>
<th>August</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_{11}$ (beef)</td>
<td>0.0108</td>
<td>0.0040</td>
<td>0.0031</td>
<td>0.0108</td>
<td>0.0189</td>
<td>-0.1119</td>
</tr>
<tr>
<td>$\delta_{12}$ (pork)</td>
<td>-0.0040</td>
<td>0.0026</td>
<td>-0.0093</td>
<td>-0.0040</td>
<td>0.0010</td>
<td>0.8173</td>
</tr>
<tr>
<td>$\delta_{13}$ (chicken)</td>
<td>-0.0054</td>
<td>0.0024</td>
<td>-0.0101</td>
<td>-0.0054</td>
<td>-0.0006</td>
<td>0.7323</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>December</th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_{21}$ (beef)</td>
<td>0.0119</td>
<td>0.0124</td>
<td>-0.0111</td>
<td>0.0114</td>
<td>0.0375</td>
<td>0.0939</td>
</tr>
<tr>
<td>$\delta_{22}$ (pork)</td>
<td>-0.0142</td>
<td>0.0104</td>
<td>-0.0355</td>
<td>-0.0137</td>
<td>0.0045</td>
<td>-0.3565</td>
</tr>
<tr>
<td>$\delta_{23}$ (chicken)</td>
<td>0.0128</td>
<td>0.0079</td>
<td>-0.0028</td>
<td>0.0128</td>
<td>0.0281</td>
<td>-0.6845</td>
</tr>
</tbody>
</table>

Table 11: Habit Effect in MS-AIDS model

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$ (beef)</td>
<td>0.3183</td>
<td>0.0772</td>
<td>0.1636</td>
<td>0.3197</td>
<td>0.4696</td>
</tr>
<tr>
<td>$\phi_{22}$ (pork)</td>
<td>0.0510</td>
<td>0.0776</td>
<td>-0.1010</td>
<td>0.0513</td>
<td>0.2030</td>
</tr>
<tr>
<td>$\phi_{33}$ (chicken)</td>
<td>0.4454</td>
<td>0.1158</td>
<td>0.2136</td>
<td>0.4478</td>
<td>0.6683</td>
</tr>
</tbody>
</table>

Table 12: Mean squared errors (MSEs)

<table>
<thead>
<tr>
<th>Beef</th>
<th>Pork</th>
<th>Chicken</th>
<th>Fish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes</td>
<td>0.805 x 10^{-4}</td>
<td>0.172 x 10^{-4}</td>
<td>0.152 x 10^{-4}</td>
</tr>
<tr>
<td>MLE</td>
<td>1.177 x 10^{-4}</td>
<td>0.264 x 10^{-4}</td>
<td>0.158 x 10^{-4}</td>
</tr>
</tbody>
</table>

Table 14 shows that posterior means of transition probabilities $\pi_{11}$ and $\pi_{22}$ are relatively...
Table 13: Posterior mean of average budget share

<table>
<thead>
<tr>
<th></th>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef</td>
<td>0.2075</td>
<td>0.1799</td>
</tr>
<tr>
<td>Pork</td>
<td>0.1825</td>
<td>0.2071</td>
</tr>
<tr>
<td>Chicken</td>
<td>0.0878</td>
<td>0.0961</td>
</tr>
<tr>
<td>Fish</td>
<td>0.5222</td>
<td>0.5170</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_{11} )</td>
<td>0.9520</td>
<td>0.0282</td>
<td>0.8830</td>
<td>0.9572</td>
<td>0.9911</td>
<td>-0.1400</td>
</tr>
<tr>
<td>( \pi_{22} )</td>
<td>0.9619</td>
<td>0.0300</td>
<td>0.8868</td>
<td>0.9689</td>
<td>0.9979</td>
<td>1.0742</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Regime 1</th>
<th>Price ($\eta_{i}^{P}$)</th>
<th>Expenditure ($\eta_{i}^{E}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beef</td>
<td>Pork</td>
</tr>
<tr>
<td>Beef</td>
<td>-0.817</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>(-1.468, -0.344)</td>
<td>(-0.284, 0.546)</td>
</tr>
<tr>
<td>Pork</td>
<td>0.148</td>
<td>-0.514</td>
</tr>
<tr>
<td></td>
<td>(-0.230,0.721)</td>
<td>(-1.064,-0.036)</td>
</tr>
<tr>
<td>Chicken</td>
<td>-0.558</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(-1.490,0.141)</td>
<td>(-0.976,0.878)</td>
</tr>
<tr>
<td>Fish</td>
<td>-0.029</td>
<td>-0.196</td>
</tr>
<tr>
<td></td>
<td>(-0.168,0.126)</td>
<td>(-0.320,-0.070)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regime 2</th>
<th>Price ($\eta_{i}^{P}$)</th>
<th>Expenditure ($\eta_{i}^{E}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beef</td>
<td>Pork</td>
</tr>
<tr>
<td>Beef</td>
<td>-0.336</td>
<td>-0.168</td>
</tr>
<tr>
<td></td>
<td>(-1.046,0.370)</td>
<td>(-0.557,0.238)</td>
</tr>
<tr>
<td>Pork</td>
<td>-0.096</td>
<td>-0.505</td>
</tr>
<tr>
<td></td>
<td>(-0.410,0.221)</td>
<td>(-0.936,-0.062)</td>
</tr>
<tr>
<td>Chicken</td>
<td>-0.559</td>
<td>0.319</td>
</tr>
<tr>
<td></td>
<td>(-1.180,0.030)</td>
<td>(-0.400,1.056)</td>
</tr>
<tr>
<td>Fish</td>
<td>-0.089</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>(-0.262,0.069)</td>
<td>(-0.335,-0.081)</td>
</tr>
</tbody>
</table>

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 must to 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


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