

Forecasting Dynamic Market Share Relationships *

Nobuhiko Terui[†]

Faculty of Economics, Tohoku University, Sendai 980-77, Japan

E-mail:terui@econ.tohoku.ac.jp, Voice:+81-22-217-6311, Fax:+81-22-217-6321

March, 1997(First vesion: February, 1997)

Abstract

In market share analysis, it is well recognized that we have often inadmissible predicted market share, which means that some of predictors take the values outside the range $[0, 1]$ and the total sum of predicted shares is not always one, so called "logical inconsistency". In this article, based on Bayesian VAR model, I propose a dynamic market share model with logical consistency. The proposed method makes it possible to forecast not only the values of market share by themselves, but also various dynamic market share relations across different brands or companies. The daily scanner data are analyzed by the proposed method.

*This researchw as done while the author was visiting the Tinbergen Institute, Rotterdam, Erasmus University. FORTRAN code and data set used in this article is available up on request via E-mail.

[†]The author acknowledges the financial supports from NWO(Netherlands Organization for Scientific Research) and JSPS(Japan Society for Promotion of Science).

1 Introduction

Many kinds of market share models have been proposed mainly for measuring the effect of marketing mix variables. The following conditions, known as *the logical-consistency requirement*, are essential for market share models;

- Each estimated market shares is non-negative and belongs to the range between zero and one(or 100%) (We call it *the range condition*).
- The sum of estimated market shares must be equal to one(or 100%) (We call it *the summing up condition*).

Some of market share models do not always satisfy the above conditions. For example, following Naert and Weverbergh (1981) and Cooper and Nakanishi(1988), market share models proposed so far can be classified into five types; linear models, multiplicative models, exponential models, MCI(multiplicative competitiveness interaction) models and MNL(multinomial logit) models. The first three models take the same form after transformation of variables and these models do not satisfy the logical-consistency requirement. On the other hand, MCI and MNL models, which can be interpreted as the normalized models of multiplicative and exponential models respectively, are defined so that the requirement is satisfied by construction. Their modeling and statistical properties are discussed in Nakanishi and Cooper(1974) and Cooper and Nakanishi(1988).

Apart from these normalized models, by using general regression based model, the statistical conditions for the requirements have been investigated by Naert and Bultez(1973), Beckwith(1973) and Bultes and Naert(1975). However, as a practical implementation, it is often the case to employ an adaptive method, where $k - 1$ (k is the number of brands or companies in a market) shares are estimated in the first by the same dimensional equation system and the remaining share is determined through the summing up condition. The justification of this method is given by Barten(1969), which showed that the maximum likelihood estimation provides the invariant parameter estimates with respect to the excluded equation. Ghosh, Neslin and Shoemaker (1984), Leeflang and Rueyl (1984) and Brodie and de Kluyer (1984) conducted the extensive empirical studies and compared the predictive performances of these models.

Market share analysis so far has been essentially based on regression models, where market shares are explained by marketing mix variables and some environmental variables(causal data), and the predictions by these models are conducted by assuming the future values of these explanatory variables. The specification of future values for marketing variables of competing brands(or companies) can lead to distinctive prediction errors. The predicted share by these models is not automatically generated by dynamical mechanism of the model, and in this sense, these regression based models are the *static* models. Although regression based approach has an important meaning of measuring the effects of marketing mix variables on market shares, a dynamic statistical model as the automatic forecasting system can be useful and desirable for real decision making situation.

On the other hand, time series analysis defines the *dynamic* system for data generating process and its usefulness has been recognized in many disciplines especially for short-term predictions. Multivariate(or vector valued) time series models employed here define the current market share not only by the past values of its own, but also by the past values of other competing brands(or companies).

Another background to motivate the prediction system for market share analysis is the fact that recent drastic development of computer network technology and its diffusion make it easy for marketing research to get precise instantaneous information about market share, sales or some marketing mix in the form of scanner data at the daily base. For example, "InfoScan/PromotionScan" by Information Resources, Inc.(IRI), and A.C.Nielsen provide a service of supplying national tracking data recorded in the major retail scanner data base in the U.S. and Japanese Nikkei "NEEDS-POS" information system provides the customers with the same kinds of daily information. For the company, the most important advantage of using this scanner data is its promptitude and make a short-term prediction of market share or sales by using their accumulated knowledge(data) instantaneously without risky specification of marketing strategy of competing brands(or companies).

In this article, as a dynamic model, I employ VAR(vector autoregressive) model to deal with market share time series and propose an automatic short-term prediction system, where the logical-consistency problem is solved by the use of Bayesian approach. The requirement brings non-standard problem to the classical statistics because the prediction space is restricted in the form of inequality, which is very hard for classical statistics.

On the other hand, Bayesian approach provides a powerful tool for solving this problem. The proposed method provides a logically consistent predictors of market share and also accommodate the prediction of complicated dynamic relations across market shares which can be useful for marketing strategy. Bayesian time series analysis has been well developed with the progress of efficient numerical integration techniques such as Markov Chain Monte Carlo, Gibbs sampling and Metropolis-Hastings algorithm. Several problems which are hard for classical approach are successfully solved by Bayesian approach, for example, in Geweke(1989a,b), Geweke and Terui(1991, 1993), Terui(1992) and so on. The proposed method in this article can be extended by including marketing mix variables to VARX(vector autoregressive with exogenous variables). However, this extension suffers from the same problem of specification of future values for explanatory variables as the regression based market share models. In this article, I aim to propose an automatic prediction system and leave this extended model for future research.

In the next section, as an important motivation of our analysis, I first mention some examples of interesting dynamic relations across market shares, which can be predicted by the proposed method. In section 3, I describe Bayesian VAR model, the concept of predictive density and the efficient algorithm for the evaluation of predictive density. Then I propose a dynamic market share model with logical consistency and discuss its implementation method. In section 4, I apply the proposed method to scanner data extracted from Nikkei "NEEDS-POS information" system and examine its performances. It is shown that the proposed method works well even when the sample size is not so large. Finally I briefly conclude in section 5.

2 Dynamic Relations of Future Market Share

Now consider the case where there are three brands in a market and denote the vector of market shares at the time t by $\mathbf{z}_t = (z_1(t), z_2(t), z_3(t))'$, where we assume that the first 2 brands (z_1, z_2) belong to one company and the third z_3 is the brand of competing company. Then, in addition to logically consistent forecasting of \mathbf{z}_t , it is very often the case that forecasting various dynamical relations across market shares is useful and desirable. Some of these relations are;

- $\Pr\{z_1(t+h) + z_2(t+h) > r \mid \mathcal{Y}\}$
 \Rightarrow The sum of share of one company's brands(the brands 1 and 2) at h step ahead exceeds $r \times 100\%$.
- $\Pr\{\frac{\sum_{j=1}^h z_1(t+j)}{h} > r \mid \mathcal{Y}\}$
 \Rightarrow The average share of brand 1 for next h periods exceeds $r \times 100\%$.
- $\Pr\{z_1(t+h) > z_3(t+h) \mid \mathcal{Y}\}$
 \Rightarrow At h step ahead, the share of the brand 1 exceeds that of the competing band 3.
- $\Pr\{\frac{\sum_{j=1}^h z_1(t+j)}{h} > \frac{\sum_{j=1}^h z_3(t+j)}{h} \mid \mathcal{Y}\}$
 \Rightarrow Over the next h periods, the mean of the brand 1's shares exceeds that of the competing brand 3.
- $\Pr\{\cap_{j=1}^h (z_1(t+j) > z_3(t+j)) \mid \mathcal{Y}\}$
 \Rightarrow The share of the brand 1 keeps greater than that of the brand 3 during h consecutive periods.
- $\Pr\{z_1(t+h) + z_2(t+h) > z_3(t+h) \mid \mathcal{Y}\}$
 \Rightarrow At h step ahead, the total share of one company exceeds that of the competing company.
- $\Pr\{\frac{\sum_{j=1}^h (z_1(t+j) + z_2(t+j))}{h} > \frac{\sum_{j=1}^h z_3(t+j)}{h} \mid \mathcal{Y}\}$
 \Rightarrow The total share of one company keeps greater in the average than the competing company's share during h succeeding periods.

where $\Pr\{\cdot \mid \mathcal{Y}\}$ means the posterior probability given the data \mathcal{Y} available up to t . In the following, I propose a logically consistent forecasting method to evaluate these kinds of posterior probabilities as well as the prediction of market shares z_t by themselves. Forecasting other interesting events can be done in just the same way by the proposed method.

3 Logically Consistent Dynamic Market Share Model

In this section, I first describe the framework of Bayesian VAR model and discuss the construction of predictive density. Then I propose a dynamic market share model with logical consistency and provide the algorithms for generating logically consistent forecasts as well as for evaluating dynamic market share relationships across the brands or companies.

3.1 Bayesian VAR Model and Joint Predictive Distribution

For k dimensional stationary time series $\mathbf{y}_t = (y_1(t), y_2(t), \dots, y_k(t))'$, VAR model with the order p is defined as

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}_0 + \mathbf{B}_1 \mathbf{y}_{t-1} + \dots + \mathbf{B}_p \mathbf{y}_{t-p} + \mathbf{e}_t \\ &= \mathbf{B}_0 + \sum_{j=1}^p \mathbf{B}_j \mathbf{y}_{t-j} + \mathbf{e}_t, \quad t = 1, \dots, T. \end{aligned} \quad (1)$$

where \mathbf{B}_0 is k dimensional vector of intercept, $\mathbf{B}_i, i = 1, \dots, p$ are $k \times k$ coefficients matrices, and \mathbf{e}_t is k dimensional innovation vector assumed to follow serially independent multivariate normal distribution with mean \mathbf{O} and covariance matrix $\mathbf{\Sigma}$, that is, $\mathbf{e}_t \sim N_k(\mathbf{O}, \mathbf{\Sigma})$.

Set $\mathbf{\Gamma} = [\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_p]'$ and define the data matrix by

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_T \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & \mathbf{y}'_0 & \mathbf{y}'_{-1} & \dots & \mathbf{y}'_{1-p} \\ 1 & \mathbf{y}'_1 & \mathbf{y}'_0 & \dots & \mathbf{y}'_{2-p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{y}'_{T-1} & \mathbf{y}'_{T-2} & \dots & \mathbf{y}'_{T-p} \end{pmatrix} \quad (2)$$

given the vector of initial values $\{\mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{1-p}\}$, and set non-informative standard Jeffrey's prior on the parameters

$$\pi(\mathbf{\Gamma}, \mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-(p+1)/2} \quad (3)$$

then the posterior distribution of $(\mathbf{\Gamma}, \mathbf{\Sigma})$ can be derived as (see, for example, Zellner(1971), Ch.8)

$$p(\mathbf{\Gamma}, \mathbf{\Sigma} | \mathcal{Y}) \propto |\mathbf{\Sigma}|^{-(T+k+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{S} + (\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}) \mathbf{X}' \mathbf{X} (\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}) \right] \mathbf{\Sigma}^{-1} \right\} \quad (4)$$

where $\mathcal{Y} = \{\mathbf{y}_{1-p}, \dots, \mathbf{y}_{-1}, \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_T\}$ means the total information of data, and $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Sigma}}$ mean the least squares estimates of coefficient matrix and covariance matrix of

innovation, respectively defined as

$$\widehat{\boldsymbol{\Gamma}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad \widehat{\boldsymbol{\Sigma}} = \mathbf{S}/(T - \nu) = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\Gamma}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\Gamma}})/(T - \nu).$$

Now standing on the point t , when we forecast h step ahead value of \mathbf{y} , the optimal predictor $\widehat{\mathbf{y}}_{t+h}$ in the sense of quadratic loss function can be given as the conditional expectation $E(\mathbf{y}_{t+h} | \mathcal{Y})$, which, in turn, given some initial information $\{\mathbf{y}_t, \mathbf{y}_{t-1}, \dots\}$ and $(\boldsymbol{\Gamma}, \boldsymbol{\Sigma})$, is defined recursively as

$$\widehat{\mathbf{y}}_{t+h} = \begin{cases} \mathbf{y}_{t+h} & \text{if } h \leq 0 \\ \mathbf{B}_0 + \sum_{j=1}^p \mathbf{B}_j \widehat{\mathbf{y}}_{t+h-j} & \text{if } h = 1, 2, \dots \end{cases} \quad (5)$$

Then in order to discuss the statistical properties of $\widehat{\mathbf{y}}_{t+h}$, the following posterior density function of $\widehat{\mathbf{y}}_{t+h}$ has to be evaluated.

$$p(\widehat{\mathbf{y}}_{t+h} | \mathcal{Y}) = \int \cdots \int p(\widehat{\mathbf{y}}_{t+h}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma} | \mathcal{Y}) d\boldsymbol{\Gamma} d\boldsymbol{\Sigma}. \quad (6)$$

We call (6) *the predictive density* and denoting the successive sequence of predictor vector by $\widehat{\mathbf{Y}}_{t+h} = \{\widehat{\mathbf{y}}_{t+1}, \widehat{\mathbf{y}}_{t+2}, \dots, \widehat{\mathbf{y}}_{t+h}\}$, we call their joint distribution

$$p(\widehat{\mathbf{Y}}_{t+h} | \mathcal{Y}) = \int \cdots \int p(\widehat{\mathbf{y}}_{t+1}, \widehat{\mathbf{y}}_{t+2}, \dots, \widehat{\mathbf{y}}_{t+h}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma} | \mathcal{Y}) d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \quad (7)$$

the joint predictive density.

The concept of the predictive distribution has been well recognized and investigated from non-Bayesian approach either. However, their results are not always general since the analysis is restricted to the case where the sufficient statistics are available(see for example, Hinkley (1979)). On the other hand, Bayesian approach constitutes the complete framework of predictive distribution. With respect to the evaluation of (7), the multiple integration does not have an analytical expression in a closed form and it is also hard to integrate analytically the joint predictive density over some region which defines the events mentioned in section 2. However, computationally efficient numerical integration techniques, for example, Markov Chain Monte Carlo(MCMC), Gibbs sampler, Metropolis-Hastings algorithm, have been well developed and provide numerical solutions for Bayesian analysis. In this article, I employ Monte Carlo integration via generating synthetic random numbers, which can be understood as one of MCMC. This method was investigated in

general way by Kloek and van Dijk(1978) and Hammersley and Handscomb(1979) and has been successfully applied in time series analysis, for examples, by Geweke(1989a,b), Geweke and Terui(1991, 1993) and Terui(1992).

Now using the decomposition of joint posterior density function of $(\mathbf{\Gamma}, \mathbf{\Sigma})$,

$$p(\mathbf{\Gamma}, \mathbf{\Sigma} \mid \mathcal{Y}) = p(\mathbf{\Gamma} \mid \mathbf{\Sigma}, \mathcal{Y})p(\mathbf{\Sigma} \mid \mathcal{Y}) \quad (8)$$

we can write (7) as

$$\begin{aligned} p(\widehat{\mathbf{Y}}_{t+h} \mid \mathcal{Y}) &= \int \cdots \int p(\widehat{\mathbf{Y}}_{t+h}, \mathbf{\Gamma}, \mathbf{\Sigma} \mid \mathcal{Y}) d\mathbf{\Gamma} d\mathbf{\Sigma} \\ &= \int \cdots \int p(\widehat{\mathbf{Y}}_{t+h} \mid \mathbf{\Gamma}, \mathbf{\Sigma}, \mathcal{Y}) p(\mathbf{\Gamma}, \mathbf{\Sigma} \mid \mathcal{Y}) d\mathbf{\Gamma} d\mathbf{\Sigma} \\ &= \int \cdots \int p(\widehat{\mathbf{Y}}_{t+h} \mid \mathbf{\Gamma}, \mathbf{\Sigma}, \mathcal{Y}) p(\mathbf{\Gamma} \mid \mathbf{\Sigma}, \mathcal{Y}) p(\mathbf{\Sigma} \mid \mathcal{Y}) d\mathbf{\Gamma} d\mathbf{\Sigma} \end{aligned} \quad (9)$$

where

$$p(\mathbf{\Sigma} \mid \mathcal{Y}) \propto |\mathbf{\Sigma}|^{-\nu/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{\Sigma}^{-1} \mathbf{S} \right] \right\} \quad (10)$$

and

$$p(\mathbf{\Gamma} \mid \mathbf{\Sigma}, \mathcal{Y}) \propto \exp \left\{ -\frac{1}{2} \text{tr} \left[(\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}}) \mathbf{X}' \mathbf{X} (\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}}) \right] \mathbf{\Sigma}^{-1} \right\}. \quad (11)$$

Next we vectorize the coefficient matrix and its estimate as $\boldsymbol{\beta} = \text{Vec}(\mathbf{\Gamma})$, $\widehat{\boldsymbol{\beta}} = \text{Vec}(\widehat{\mathbf{\Gamma}})$ and then we have

$$p(\boldsymbol{\beta} \mid \mathbf{\Sigma}, \mathcal{Y}) \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' (\mathbf{\Sigma}^{-1} \otimes \mathbf{X}' \mathbf{X}) (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \right\}. \quad (12)$$

Therefore we can see that $p(\mathbf{\Sigma} \mid \mathcal{Y})$ follows the inverted Wishart distribution $IW(\mathbf{S}, \nu)$ with the degrees of freedom ν and covariance matrix \mathbf{S} and $p(\boldsymbol{\beta} \mid \mathbf{\Sigma}, \mathcal{Y})$ has the kernel of m dimensional multivariate normal distribution $N_m(\widehat{\boldsymbol{\beta}}, (\mathbf{\Sigma}^{-1} \otimes \mathbf{X}' \mathbf{X})^{-1})$ with mean $\widehat{\boldsymbol{\beta}}$ and covariance matrix $(\mathbf{\Sigma}^{-1} \otimes \mathbf{X}' \mathbf{X})^{-1}$, where $m = pk + 1$, $\nu = T - m + k + 1$. The samplings from these distributions are easy to be drawn and Monte Carlo integrations via generating synthetic random numbers can be executed in the following algorithm;

[I] Generate the random number matrix $\mathbf{\Sigma}^{-1(i)}$ of $\mathbf{\Sigma}^{-1}$ from $IW(\mathbf{S}, \nu)$.

[II] Conditional on $\mathbf{\Sigma}^{-1(i)}$ in the step [I], generate the random number vector $\boldsymbol{\beta}^{(i)}$ of $\boldsymbol{\beta}$ from $N_m(\widehat{\boldsymbol{\beta}}, (\mathbf{\Sigma}^{-1(i)} \otimes \mathbf{X}' \mathbf{X})^{-1})$.

[III] By using $\beta^{(i)}$, generate $\widehat{\mathbf{Y}}_{t+h}^{(i)}$ by (5).

[IV] Iterate the steps [I], [II] and [III] n times ($i = 1, \dots, n$).

The empirical distribution of $\{\widehat{\mathbf{Y}}_{t+h}^{(1)}, \widehat{\mathbf{Y}}_{t+h}^{(2)}, \dots, \widehat{\mathbf{Y}}_{t+h}^{(n)}\}$ approximates the joint posterior distribution $p(\widehat{\mathbf{Y}}_{t+h} | \mathcal{Y})$. In particular, with respect to the numerical evaluation of the point forecast when we use the quadratic loss function, it is well known that

$$n \rightarrow \infty, \quad \frac{1}{n} \sum_{i=1}^n \widehat{\mathbf{Y}}_{t+h}^{(i)} \xrightarrow{a.s.} E(\widehat{\mathbf{Y}}_{t+h} | \mathcal{Y}). \quad (13)$$

The samplings from $p(\mathbf{\Gamma} | \mathbf{\Sigma}, \mathcal{Y})$ and $p(\mathbf{\Sigma} | \mathcal{Y})$ are discussed in Geweke(1988) and explained in this article as the appendix.

3.2 Dynamic Market Share Model with Logical Consistency

Let \mathbf{z}_t be the vector of market share of k brands in a market at the time t and then VAR model defines the dynamic market share model where the present share of each brand is determined not only by the past values of its own, but also by the past values of other brands. When we deal with market share time series, we must have the restrictions for logical consistency defined in section 1. The static market share models estimate the $k - 1$ equations by excluding the last equation which is redundant for most applications and determine the last by the summig up condition, $z_k(t) = 1 - \sum_{i=1}^{k-1} z_i(t)$. The justification of this method for maximal likelihood estimation is given by Barten(1969), which shows, in the regression models, that maximal likelihood estimates for $k - 1$ brands provide the invariant estimates with respect to the excluded variate. It is not difficult to show that the same invariance property holds for our model because the estimators by the proposed model has the same asymptotic property as maximum likelihood estimators. I adopt this approach for dynamic market share model.

Considering that it is sometimes appropriate to transform the variates by, for example, Box-Cox transformation, before accommodating market share time series in VAR model, I provide the model in the general way in the following. Now define $\mathbf{z}_t^{(k-1)}$ as the vector of the first $k - 1$ brand's market share at time t and set $\mathbf{z}_t = (\mathbf{z}_t^{(k-1)}, z_k(t))'$, then, for the transformed variates

$$\mathbf{y}_t = f(\mathbf{z}_t) = \left(f(\mathbf{z}_t^{(k-1)}), f(z_k(t)) \right)'$$

$$\equiv \left(\mathbf{y}_t^{(k-1)}, y_k(t) \right)', \quad (14)$$

the following system which includes $k - 1$ dimensional VAR model is proposed,

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_t^{(k-1)} \\ y_k(t) \end{pmatrix} = \begin{pmatrix} \mathbf{B}_0^{(k-1)} + \sum_{j=1}^p \mathbf{B}_j^{(k-1)} \mathbf{y}_{t-j}^{(k-1)} + \mathbf{D}^{(k-1)}(t) \\ f\left(1 - \sum_{j=1}^{k-1} f^{-1}(y_j(t))\right) \end{pmatrix} \quad (15)$$

where $\mathbf{B}_j, j = 1, 2, \dots, p$, are $(k - 1) \times (k - 1)$ coefficient matrices and $\mathbf{e}_t^{(k-1)}$ is $k - 1$ dimensional innovation vector assumed to be serially independent normal distribution ($\mathbf{e}_t^{(k-1)} \sim N_{k-1}(\mathbf{O}, \boldsymbol{\Sigma}^{(k-1)})$), and $\mathbf{D}^{(k-1)}(t)$ is the vector of deterministic trend function. The trend function is sometimes useful for evaluating dynamical relations across the market shares.

In order to evaluate the predictive density $p(\hat{\mathbf{z}}_{t+h} | \mathcal{Y}) = p(f^{-1}(\hat{\mathbf{y}}_{t+h}) | \mathcal{Y})$, we need to constitute the predictive distribution of transformed variate $\hat{\mathbf{y}}_{t+h}$ in the first. By using just the same logic mentioned in previous subsection, the h step ahead forecasts for transformed variates $\hat{\mathbf{y}}_{t+h}$ are generated by, for the first $k - 1$ brands,

$$\hat{\mathbf{y}}_{t+h}^{(k-1)} = \begin{cases} \mathbf{y}_{t+h}^{(k-1)} & \text{if } h \leq 0 \\ \mathbf{B}_0^{(k-1)} + \sum_{j=1}^p \mathbf{B}_j^{(k-1)} \hat{\mathbf{y}}_{t+h-j}^{(k-1)} + \mathbf{D}^{(k-1)}(t) & \text{if } h = 1, 2, \dots \end{cases} \quad (16)$$

and then the summing up condition and the inverse transformation lead to the predicted value of market share,

$$\hat{\mathbf{z}}_{t+h} = \begin{pmatrix} \hat{\mathbf{z}}_{t+h}^{(k-1)} \\ \hat{z}_k(t+h) \end{pmatrix} = \begin{pmatrix} f^{-1}(\hat{\mathbf{y}}_{t+h}^{(k-1)}) \\ f^{-1}\left(1 - \sum_{j=1}^{k-1} f^{-1}(\hat{y}_j(t+h))\right) \end{pmatrix}. \quad (17)$$

For the vector of h step ahead predictor $\hat{\mathbf{z}}_{t+h} = (\hat{\mathbf{z}}_{t+h}^{(k-1)}, \hat{z}_k(t+h))'$, the conditions for logical consistency restrict the prediction space for individual j ($0 \leq j \leq k$) and for all h (≥ 0) so that

$$\mathcal{R} = \{\hat{\mathbf{z}}_{t+h}; \text{ for } j(0 \leq j \leq k) \text{ and } h(\geq 0), \quad 0 \leq \hat{z}_j(t+h) \leq 1 \text{ and } \sum_{j=1}^k \hat{z}_j(t+h) = 1\}.$$

Because the summing up condition $\sum_{j=1}^k \hat{z}_j(t+h) = 1$ is automatically satisfied for our model by construction, the necessary restriction is

$$\mathcal{R}^* = \{\hat{\mathbf{z}}_{t+h}; \text{ for } j(0 \leq j \leq k) \text{ and } h(\geq 0), \quad 0 \leq \hat{z}_j(t+h) \leq 1\}. \quad (18)$$

Just now stacking the sequence of the vectors of unrestricted and logically consistent predictors from 1 through h , denoted by $\widehat{\mathbf{Z}}_{t+h} = \{\widehat{\mathbf{z}}_{t+1}, \widehat{\mathbf{z}}_{t+2}, \dots, \widehat{\mathbf{z}}_{t+h}\}$ and $\widehat{\mathbf{Z}}_{t+h}^* = \{\widehat{\mathbf{z}}_{t+1}^*, \widehat{\mathbf{z}}_{t+2}^*, \dots, \widehat{\mathbf{z}}_{t+h}^*\}$ respectively, the joint predictive density $p(\widehat{\mathbf{z}}_{t+h}^* | \mathcal{Y})$ with the logical consistency can be expressed in the form of multiple integral over restricted region \mathcal{R}^* ;

$$p(\widehat{\mathbf{Z}}_{t+h}^* | \mathcal{Y}) = \frac{p(\widehat{\mathbf{Z}}_{t+h} | \mathcal{Y}, \mathcal{R}^*)}{\int_{\mathcal{R}^*} \dots \int p(\widehat{\mathbf{Z}}_{t+h} | \mathcal{Y}) d\widehat{\mathbf{Z}}_{t+h}} \quad (19)$$

The analytical evaluation of this multiple integral is definitely hard and the numerical integration via efficient Monte Carlo integration provides a powerful tool for the evaluation. Noting that the denominator is only proportional constant and the restrictive predictive density $p(\widehat{\mathbf{Z}}_{t+h} | \mathcal{Y}, \mathcal{R}^*)$ can be expressed, in the same way as (9), by

$$\int \dots \int p(\widehat{\mathbf{Z}}_{t+h}, \mathbf{\Gamma}^{(k-1)}, \mathbf{\Sigma}^{(k-1)} | \mathcal{Y}, \mathcal{R}^*) p(\mathbf{\Gamma}^{(k-1)} | \mathbf{\Sigma}^{(k-1)}, \mathcal{Y}, \mathcal{R}^*) p(\mathbf{\Sigma}^{(k-1)} | \mathcal{Y}, \mathcal{R}^*) d\mathbf{\Gamma}^{(k-1)} d\mathbf{\Sigma}^{(k-1)}, \quad (20)$$

where $\mathbf{\Gamma}^{(k-1)} = [\mathbf{B}_0^{(k-1)}, \mathbf{B}_1^{(k-1)}, \dots, \mathbf{B}_p^{(k-1)}, \mathbf{D}^{(k-1)}]'$, we have the following steps for the construction of joint predictive distribution with the logical consistency;

- [I]' Generate the random number matrix $\mathbf{\Sigma}^{-1(k-1)(i)}$ of $\mathbf{\Sigma}^{-1(k-1)}$ from the inverted Wishart distribution $IW(\mathbf{S}^{(k-1)}, \nu')$, where $\nu' = T - m' + k$, $m' = (k-1)p + 1$.
- [II]' Conditional on $\mathbf{\Sigma}^{-1(k-1)(i)}$, generate the random number vector $\boldsymbol{\beta}^{(k-1)(i)}$ of $\boldsymbol{\beta}^{(k-1)}$ from $m'-1$ dimensional normal distribution $N_{m'-1}(\widehat{\boldsymbol{\beta}}^{(k-1)}, (\mathbf{\Sigma}^{-1(k-1)} \otimes \mathbf{X}^{(k-1)'}\mathbf{X}^{(k-1)})^{-1})$.
- [III]' Using $\boldsymbol{\beta}^{(i)}$, generate $\widehat{\mathbf{Y}}_{t+h}^{(i)}$ by (16).
- [IV]' Transform $\widehat{\mathbf{Y}}_{t+h}^{(i)}$ into $\widehat{\mathbf{Z}}_{t+h}^{(i)}$ by (17).
- [V]' If $\{\widehat{\mathbf{Z}}_{t+h}^{(i)}\}$ satisfies the restriction (18), we take it as the logically consistent predictor into the joint predictive distribution and set $\widehat{\mathbf{Z}}_{t+h}^{*(i)} = \widehat{\mathbf{Z}}_{t+h}^{(i)}$. Otherwise, discard it and go back to [I]'
- [VI]' Iterate the steps [I]' through [V]' until we have the predetermined number n^* of iterations.

The empirical distribution constituted by the sequence of the predictor vector, $\{\widehat{\mathbf{Z}}_{t+h}^{*(1)}, \dots, \widehat{\mathbf{Z}}_{t+h}^{*(n^*)}\}$, approximates its posterior density function $p(\widehat{\mathbf{Z}}_{t+h}^* | \mathcal{Y})$ as the number of iteration goes to large.

With regard to the statistical inference on dynamic relation across market share time series, we can express the events mentioned in section 2 by using the smooth function of $\widehat{\mathbf{Z}}_{t+h}$, say $\phi(\widehat{\mathbf{Z}}_{t+h})$, in the way as $\{\phi(\widehat{\mathbf{Z}}_{t+h}) > r\}$. Then, the required posterior probability is written as

$$\Pr\{\phi(\widehat{\mathbf{Z}}_{t+h}^*) > r | \mathcal{Y}\} = \frac{\int_{\mathcal{R}^* \cap \mathcal{R}^\phi} \dots \int p(\widehat{\mathbf{Z}}_{t+h} | \mathcal{Y}) d\widehat{\mathbf{Z}}_{t+h}}{\int \dots \int p(\widehat{\mathbf{Z}}_{t+h} | \mathcal{Y}) d\widehat{\mathbf{Z}}_{t+h}} \quad (21)$$

where

$$\mathcal{R}^\phi = \{\widehat{\mathbf{Z}}_{t+h}; \text{ for } j(0 \leq j \leq k) \text{ and } h(\geq 0), \phi(\widehat{\mathbf{Z}}_{t+h}) > r\}. \quad (22)$$

This multiple integral also can not be expressed as a closed form in general. The algorithm for numerical evaluation of this integral is; in addition to $[I]' \sim [VI]'$,

$[VII]'$ for the sequence $\{\phi(\widehat{\mathbf{Z}}_{t+h}^{*(i)}), i = 1, \dots, n^*\}$, take the ratio between the number of times when the restriction $\phi(\widehat{\mathbf{Z}}_{t+h}^{*(i)}) > r$ is satisfied and n^* . The ratio approximates the required probability.

4 Empirical Application

In this section, we apply the proposed method to Japanese market share time series, which was extracted from Nikkei "NEEDS-POS" information system. This system supplies national tracking data scanned at 102 stores with the customers. Daily market share time series of the item "Butter" from July, 19 of 1996 to August, 27 of 1996 are analyzed. Market shares of 8 product makers are registered in scanner data during this period. However the top two makers' shares(z_1, z_2) occupy most part of market share(77.4 %). Thus the last 6 makers' shares are aggregated as the third series(z_3).

Figure 1 Time Series Plots and Estimates of Market Shares

I set up the dynamic market share model (15) for $\mathbf{z} = (z_1, z_2, z_3)'$. The first 35 samples were used for the estimation and the last 5 samples were used for the evaluation of forecasting. No transformation of variables was chosen for this data set. The maximum order

of VAR model was set as $p = 3$ and the minimum BIC(Bayesian Information Criteria) criteria selected $p = 1$. The estimated model is;

$$\left\{ \begin{array}{l} \mathbf{y}_t^{(2)} = \begin{pmatrix} 0.35464 \\ 0.25483 \end{pmatrix} + \begin{pmatrix} -0.00133 \\ -0.00054 \end{pmatrix} t + \begin{pmatrix} 0.42748 & 0.03614 \\ -0.18133 & 0.28447 \end{pmatrix} \mathbf{y}_{t-1}^{(2)} \\ y_3(t) = 1 - \sum_{j=1}^2 y_j(t), \quad \hat{\Sigma}^{(2)} = \begin{pmatrix} 0.00371 & -0.00050 \\ -0.00050 & 0.00246 \end{pmatrix}, \quad BIC = -10.802. \end{array} \right. \quad (23)$$

Figure 1 shows the time series plots of real observations of (z_1, z_2, z_3) used for estimation in the top panel and the estimates of each variable are drawn in the next 3 panels. We can observe good fittings for each market share except a few points. In particular, the 27th observation might be appropriate to be dealt with as outlier in our model.

Figure 2 Predictive Density for Market Share Time Series

Table 1 Summary Statistics for Predictive Density

Figure 2 shows the behaviors of the joint predictive densities for each market share variable and their 5%, 25 %, 50%, 75%, 95% points, posterior means, and real observations are drawn with the prediction steps. The number of iteration for Monte Carlo integration was fixed at $n^* = 2,000$ after confirming that the distributions do not change any more than 2,000 times. $\hat{z}_i(t+h)^{[q]}$ means the q th percentile of the empirical distribution of predictive density at the step h . Note that the posterior mean $E(\hat{z}_i(t+h) | \mathcal{Y})$ defines a Bayesian point forecast under the quadratic loss function. From the figures, it can be seen that the joint predictive densities for z_1 and z_2 are relatively symmetric, and the median and the posterior mean take the almost same values for every prediction step. On the other hand, the predictive density for z_3 is skewed downward. Some basic statistics for each density are tabulated in Table 1, where the number of irregularity is denoted in the last column. The irregularity here means the occurrence that the model produced inconsistent predictor which did not satisfy the condition (18) for logical consistency. Most of inconsistent predictors were generated in the last variable(z_3).

Next, in Table 2, the results of forecasting market share relations are reported. The top part of the table shows the probability of some market share relations for next 10 days

and we can expect with almost sure that z_1 will not be caught up by both z_2 and z_3 for next 10 days. And z_2 will take greater share than z_3 with probability 0.376 for next day, but this probability will be decreasing as the time goes on (the minimum is 0.1185 at the 6 days ahead).

Figure 3 Predictive Density for $z_2 - z_3$

Table 2 Probability for Dynamic Market Share Relations

The marginal predictive densities for $z_2 - z_3$ through the prediction steps $h = 1, 2, \dots, 5$ are drawn in Figure 3, where marginal densities are made from the histograms of the empirical distributions by connecting the ordinates at the mid points of each cell. The shaded area in each marginal density means the probability so that z_2 is greater than z_3 . Second part of the table shows the results of the same kind of relations for k consecutive days. The average properties are evaluated in the third part.

I confirmed that the different lags ($p = 2, 3$) for the model did not produce great difference on these results.

5 Concluding Remarks

In this article, I proposed an automatic forecasting system for market share time series, where the logical consistency requirement is solved by employing Bayesian approach. The proposed method provides multi-step forecasts, not only for the values of market share by themselves, but also for complicated dynamic market share relationships across different brands or companies. It was shown that the proposed model worked well even by using relatively small sample size used in the last section.

However it is possible to extend the proposed model in the following ways. In the first, the model can be extended to VARX (vector autoregressive with exogenous variable) models by including marketing mix variables. In that case, the future values of marketing mix must be specified in order to predict the market shares and dynamic relations between them. This means that the analysis would be the simulation studies or control theoretic approach rather than the automatic prediction. In the second, by considering that the outliers often exist in scanner data, we can extend the proposed model so as to accommodate the outliers. As a robust model for time series data, an autoregressive model with t

distributed innovations is proposed and Bayesian solution is discussed in Geweke(1993).
These extensions of the proposed model are left for future research.

Appendix:

Random number generation from $p(\boldsymbol{\Sigma} \mid \mathcal{Y})$ and $p(\boldsymbol{\Gamma} \mid \boldsymbol{\Sigma}, \mathcal{Y})$

With regard to the random numbers of $\boldsymbol{\Sigma}^{(i)}$ and $\boldsymbol{\beta}^{(i)}$, the following algorithms are available(See, Geweke(1988)).

Let the Cholesky decomposition of the inverse of \boldsymbol{S} defined in (4) be $\boldsymbol{S}^{-1} = \boldsymbol{L}\boldsymbol{L}'$, and take up the random matrix \boldsymbol{Q} from the Wishart distribution with orthogonal covariance matrix,

$$\boldsymbol{Q} \sim W(\boldsymbol{I}, T - m),$$

then we have

$$\boldsymbol{L}\boldsymbol{Q}\boldsymbol{L}' \sim W(\boldsymbol{L}\boldsymbol{L}', T - m) = W(\boldsymbol{S}^{-1}, T - m).$$

In the above, we can compose \boldsymbol{Q} from normal and chi square random numbers;

let \boldsymbol{U} be lower triangular matrix with their non-zero elements $u_{11}, u_{12}, \dots, u_{kk}$ follows independently

$$u_{ij} \sim N(0, 1) \quad \text{for } i > j, \quad u_{ii} \sim \chi^2(T - m - i + 1)$$

then we have

$$\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{Q} \sim W(\boldsymbol{I}, T - m)$$

Next denoting by \boldsymbol{R} the Cholesky decomposition of $\boldsymbol{\Sigma}$ and by the use of the relation,

$$\boldsymbol{\Sigma} = \boldsymbol{R}'\boldsymbol{R} \equiv (\boldsymbol{L}\boldsymbol{U})^{-1'}(\boldsymbol{L}\boldsymbol{U})^{-1}$$

we can generate the random number of $\boldsymbol{\Sigma}$ from \boldsymbol{U} and \boldsymbol{L} .

On the other hand, with respect to the generation of random number for the coefficient matrix $\boldsymbol{\Gamma}$ or equivalently $\boldsymbol{\beta}$, the Cholesky decomposition of

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \boldsymbol{A}\boldsymbol{A}'$$

leads to the expression of covariance matrix

$$(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})^{-1} = (\boldsymbol{R}' \otimes \boldsymbol{A})(\boldsymbol{R}' \otimes \boldsymbol{A})'.$$

Now let $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_k$ be independent random column vectors from $N_m(\boldsymbol{O}, \boldsymbol{I}_m)$ and set $\boldsymbol{\eta} = \text{Vec}(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_k)$ then we can generate the random vector of $\boldsymbol{\beta}$ by the following linear transformation,

$$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} + (\boldsymbol{R}' \otimes \boldsymbol{C}) \boldsymbol{\eta}.$$

References

- Beckwith, N.E.(1973), "Concerning the logical consistency of multivariate market share models", *Journal of Marketing Research*, **10**, 341-344.
- Brodie, R. and C.A.de Kluyer(1984), "Attraction versus linear and multiplicative market share models: an empirical evaluation", *Journal of Marketing Research*, **21**, 194-210.
- Bultes, A. and P.Naert(1975), "Consistent sum-constrained models", *Journal of American Statistical Association*, **70**, 529-535.
- Cooper, L.G. and M.Nakanishi(1988), *Market-Share Analysis*, Kluwer Academic Publishers, Boston.
- Geweke, J.(1989a) ,"Exact predictive densities for linear model with ARCH disturbances", *Journal of Econometrics*, **40**, 63-86.
- Geweke, J.(1989b), "Bayesian inference in Econometric models using Monte Carlo integrations", *Econometrica*, **57**, 1317-1339.
- Geweke, J.and N.Terui(1993) "Bayesian threshold autoregressive models for nonlinear time series", *Journal of Time Series Analysis*, **14**, 441-454.
- Geweke, J. and N.Terui(1991) "Threshold autoregressive models for macroeconomic time series: a Bayesian approach", *Proceedings of American Statistical Meeting(Business and Economics Section)*, 42-50.
- Hammersly,V. and D.C.Hanscomb(1979), *Monte Calro Methods*, Chapman and Hall, London.
- Ghosh, A., S. Neslin and R.Shoemaker(1984), "A comparison of market share models and estimation procedures", *Journal of Marketing Research*, **21**, 202-210.
- Hinkley, D.V.(1979) "Predictive likelihood" *Annals of Statistics*, **7**, 718-728.
- Kloek,T. and H.K.van Dijk(1978), "Bayesian estimate of equation system parameters: An application of integration by Monte Calro", *Econometrica*, **46**, 1-20.

- Leeflang, D.S.H. and J.C.Rueyl(1984), "On the predictive power of market share attraction models", *Journal of Marketing Research*, **21**, 211-215.
- Naert, P. and A.Bultes(1973), "Logically consistent market share model", *Journal of Marketing Research*, **10**, 334-340.
- Naert, P. and W.Waverbergh(1981), "On the prediction power of market share attraction models", *Journal of Marketing Research*, **18**, 146-153.
- Nakanishi, M. and L.G.Cooper(1974), "Parameter estimation for a multiplicative competitive interaction model - least squares approach", *Journal of Marketing Research*, **11**, 303-311.
- Press, W.H.,B.P.Flannery, S.A.Teukolsky and W.T.Vetterling(1989), *Numerical Recipes: The Art of Scientific Computing*, Cambridge Univ.Press, New York.
- Terui, N.(1992)., "Limit cycle and multistep prediction in an exponential AR model for nonlinear time series", working paper, Dept. of Statistics, Texas A&M University.
- Zellner, A.(1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Table 1: Summary Statistics of Predictive Density

h step ahead	$\hat{z}_1^{[5]}$	$E(\hat{z}_1 \mathcal{Y})$	$\hat{z}_1^{[95]}$	$S.D.(\hat{z}_1 \mathcal{Y})$	Irregularity
1	0.5124	0.5681	0.6225	0.0340	-
2	0.4964	0.5585	0.6239	0.0414	-
3	0.4836	0.5518	0.6269	0.0459	-
4	0.4735	0.5474	0.6268	0.0497	-
5	0.4659	0.5442	0.6256	0.0532	-
6	0.4585	0.5412	0.6266	0.0572	-
7	0.4523	0.5384	0.6277	0.0609	-
8	0.4449	0.5361	0.6288	0.0636	-
9	0.4372	0.5339	0.6313	0.0662	-
10	0.4316	0.5312	0.6327	0.0694	-
h step ahead	$\hat{z}_2^{[5]}$	$E(\hat{z}_2 \mathcal{Y})$	$\hat{z}_2^{[95]}$	$S.D.(\hat{z}_2 \mathcal{Y})$	Irregularity
1	0.1598	0.2058	0.2488	0.0272	-
2	0.1460	0.1938	0.2426	0.0300	-
3	0.1392	0.1904	0.2442	0.0325	-
4	0.1343	0.1896	0.2460	0.0347	-
5	0.1306	0.1893	0.2493	0.0368	1
6	0.1270	0.1890	0.2518	0.0391	1
7	0.1235	0.1887	0.2538	0.0413	-
8	0.1195	0.1885	0.2558	0.0432	-
9	0.1168	0.1882	0.2582	0.0451	1
10	0.1135	0.1879	0.2605	0.0471	1
h step ahead	$\hat{z}_3^{[5]}$	$E(\hat{z}_3 \mathcal{Y})$	$\hat{z}_3^{[95]}$	$S.D.(\hat{z}_3 \mathcal{Y})$	Irregularity
1	0.1607	0.2260	0.2925	0.0399	1
2	0.1740	0.2473	0.3092	0.0413	2
3	0.1803	0.2574	0.3244	0.0444	5
4	0.1820	0.2625	0.3345	0.0470	3
5	0.1834	0.2660	0.3412	0.0496	2
6	0.1829	0.2688	0.3464	0.0521	10
7	0.1815	0.2714	0.3516	0.0543	1
8	0.1803	0.2739	0.3568	0.0566	5
9	0.1791	0.2764	0.3622	0.0588	7
10	0.1783	0.2788	0.3685	0.0610	3

$\hat{z}_i^{[5]}$ and $\hat{z}_i^{[95]}$ indicate i th company's 5th and 95th order statistics respectively.

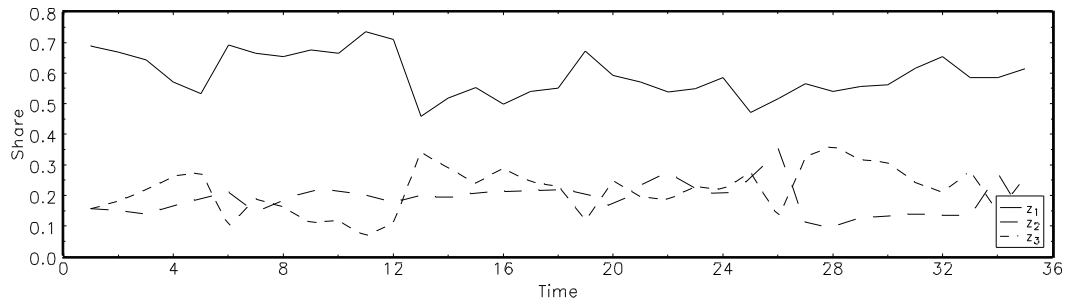
$E(\cdot | \mathcal{Y})$ means the posterior means.

$S.D.(\cdot | \mathcal{Y})$ means the posterior standard deviation.

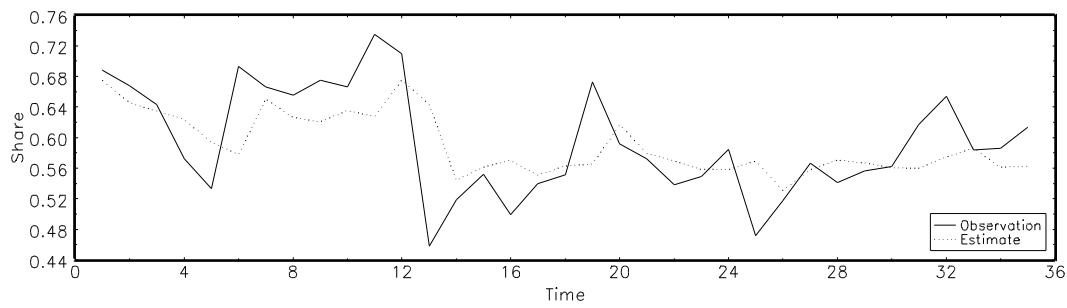
Table 2: Forecasting Market Share Relations

Prediction Step	$\widehat{z}_1(t+h) > \widehat{z}_2(t+h)$	$\widehat{z}_2(t+h) > \widehat{z}_3(t+h)$	$\widehat{z}_1(t+h) > \widehat{z}_3(t+h)$
1	1.0000	0.3760	1.0000
2	0.9995	0.1660	0.9995
3	0.9995	0.1305	0.9985
4	0.9995	0.1240	0.9980
5	0.9995	0.1195	0.9970
6	0.9990	0.1185	0.9940
7	0.9975	0.1220	0.9910
8	0.9970	0.1225	0.9890
9	0.9970	0.1225	0.9865
10	0.9965	0.1230	0.9830
Terms(k)	$\cap_{j=1}^k \{\widehat{z}_1(t+j) > \widehat{z}_2(t+j)\}$	$\cap_{j=1}^k \{\widehat{z}_2(t+j) > \widehat{z}_3(t+j)\}$	$\cap_{j=1}^k \{\widehat{z}_1(t+j) > \widehat{z}_3(t+j)\}$
2	0.9995	0.1645	0.9995
3	0.9995	0.1285	0.9985
4	0.9990	0.1190	0.9980
5	0.9990	0.1130	0.9970
Terms(k)	$(\sum_{j=1}^k \{\widehat{z}_1(t+j) > \widehat{z}_2(t+j)\})/k$	$(\sum_{j=1}^k \{\widehat{z}_2(t+j) > \widehat{z}_3(t+j)\})/k$	$(\sum_{j=1}^k \{\widehat{z}_1(t+j) > \widehat{z}_3(t+j)\})/k$
2	0.9995	0.1775	0.9995
3	0.9995	0.1387	0.9990
4	0.9995	0.1211	0.9985
5	0.9990	0.1185	0.9985

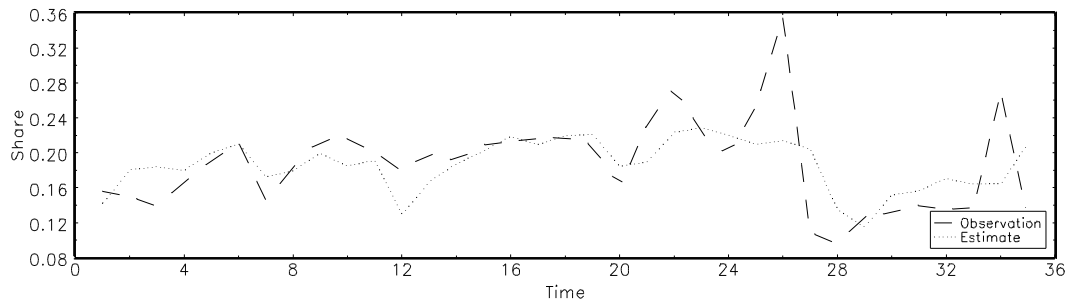
Figure 1: Market Share Time Series and Estimates
Market Share Time Series from Scanner Data



Estimates of z_1 : Inside



Estimates of z_2 : Inside



Estimates of z_3 : Inside

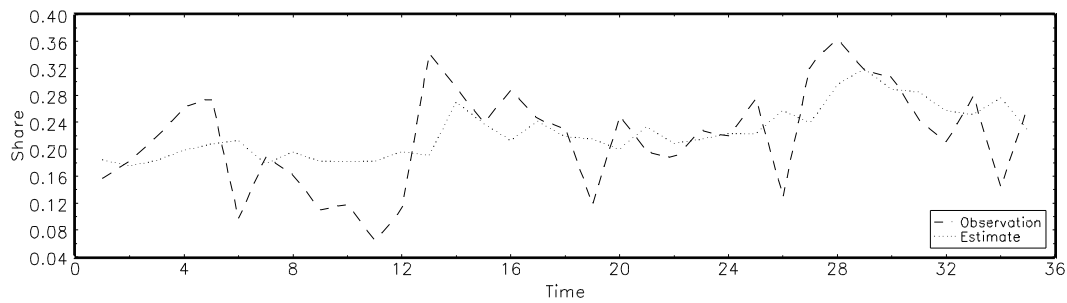
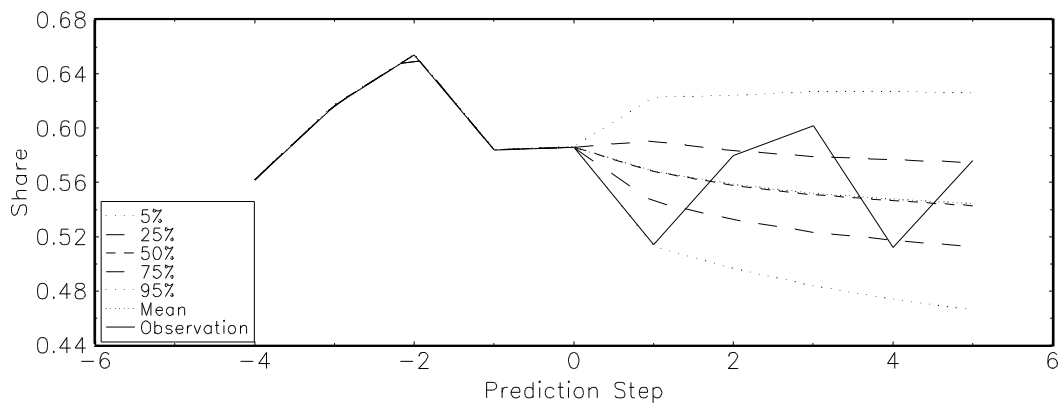
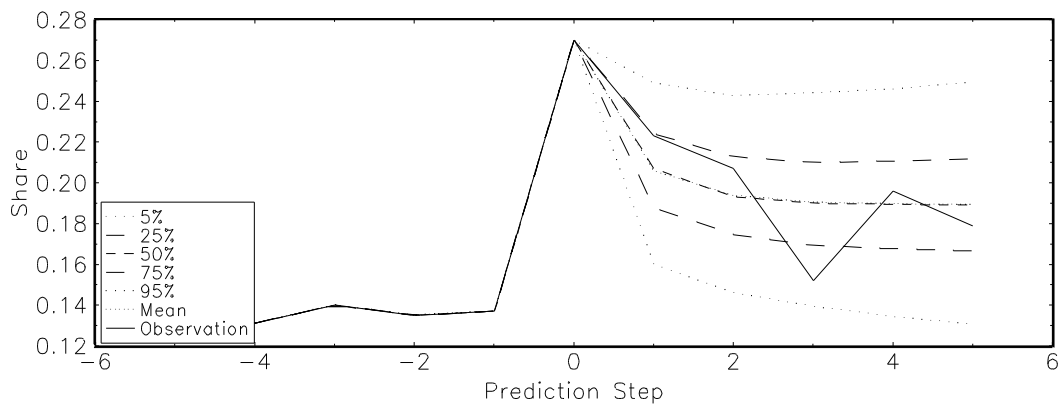


Figure 2: Predictive Density for Market Share Time Series

Predictive Density of z_1



Predictive Density of z_2



Predictive Density of z_3

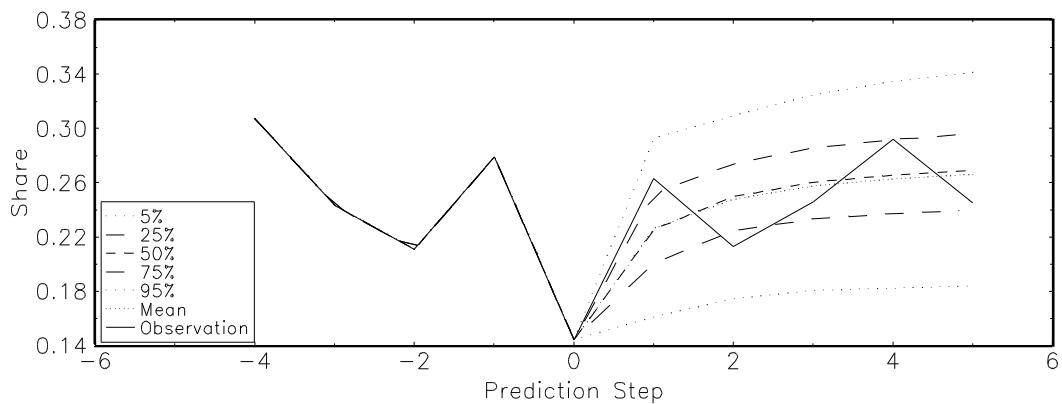


Figure 3: Predictive Density for " $z_2 - z_3$ "

