Signal Extraction and Knowledge Discovery Based on Statistical Modeling

Genshiro Kitagawa

The Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku Tokyo 106-8569 Japan
kitagawa@ism.ac.jp
http://www.ism.ac.jp/~kitagawa

Abstract. In the coming post IT era, the problems of signal extraction and knowledge discovery from huge data sets will become very important. For this problem, the use of good model is crucial and thus the statistical modeling will play an important role. In this paper, we show two basic tools for statistical modeling, namely the information criteria for the evaluation of the statistical models and generic state space model which provides us with a very flexible tool for modeling complex and time-varying systems. As examples of these methods we shall show some applications in seismology and macro economics.

1 Importance of Statistical Modeling in Post IT Era

Once the model is specified, various types of inferences and prediction can be deduced from the model. Therefore, the model plays a curial role in scientific inference or signal extraction and knowledge discovery from data. In scientific research, it is frequently assumed that there exists a known or unknown "true" model. In statistical community as well, from the age of Fisher, the statistical theories are developed under the situation that we estimate the true model with small number of parameters based on limited number of data. However, in recent years, the models are rather considered as tools for extracting useful information from data. This is motivated by the information criterion AIC that revealed that in the estimation of model for prediction, we may obtain a good model by selecting a simple model even though it may have some bias.

On the other hand, if the model is considered as just a tool for signal extraction, the model cannot be uniquely determined and there exist many possible models depending on the viewpoints of the analysts. This means that the results of the inference and the decision depend on the used model. It is obvious that a good model yields a good result and a poor model yields a poor result. Therefore, in statistical modeling, the objective of the modeling is not to find out the unique "true" model, but to obtain a "good" model based on the characteristics of the object and the objective of the analysis.

To obtain a good model, we need a criterion to evaluate the goodness of the model. Akaike (1973) proposed to evaluate the model by the goodness of its

R. Gavaldà et al. (Eds.): ALT 2003, LNAI 2842, pp. 3–14, 2003.

predictive ability and to evaluate it by the Kullback-Leibler information. As is well-known, the minimization of the Kullback-Leibler information is equivalent to maximizing the expected log-likelihood of the model. Further, as a natural estimate of the expected log-likelihood, we can define the log-likelihood and thus can be lead to the maximum likelihood estimators. However, in comparing the models with parameters estimated by the method of maximum likelihood, there arises a bias in the log-likelihood as an estimate of the expected log-likelihood. By correcting for this bias, we obtain Akaike information criterion AIC. After the derivation of the AIC, various modifications or extensions of the AIC such as TIC, GIC and EIC are proposed.

The information criterion suggests various things that should be taken into account in modeling. Firstly, since the data is finite, the models with too large number of free parameters may have less ability for prediction. There are two alternatives to mitigate this difficulty. One way is to restrict the number of free parameters which is realized by minimizing the AIC criterion. The other way is to obtain a good model with huge number of parameters by imposing a restriction on the parameters. For this purpose, we need to combine the information not only from the data but also the one from the knowledge on the object and the objective of the analysis. Therefore, the Bayes models play important role, since the integration of information can be realized by the Bayes model with properly defined prior information and the data.

By the progress of the information technology, the information infrastructure in research area and society is being fully equipped, and the environment of the data has been changed very rapidly. For example, it becomes possible to obtain huge amount of data from moment to moment in various fields of scientific research and technology, for example the CCD image of the night sky, POS data in marketing, high frequency data in finance and the huge observations obtained in environmental measurement or in the study for disaster prevention. In contrast with the conventional well designed statistical data, the special feature of these data sets is that they can be obtained comprehensively. Therefore, it is one of the most important problem in post IT era to extract useful information or discover knowledge from not-so-well designed massive data.

For the analysis of such huge amount of data, an automatics treatment of the data is inevitable and a new facet of difficulty in modeling arises. Namely, in classical framework of modeling, the precision of the model increases as the increase of the data. However, in actuality, the model changes with time due to the change of the stricture of the object. Further, as the information criteria suggest, the complexity of the model increases as the increase of the data. Therefore, for the analysis of huge data set, it is necessary to develop a flexible model that can handle various types of nonstationarity, nonlinearity and non-Gaussianity. It is also important to remember that the information criteria are relative criteria. This means that the selection by any information criterion is nothing but the one within the pre-assigned model class. This suggests that the process of modeling is an everlasting improvement of the model based on the

increase of the data and knowledge on the object. Therefore, it is very important to prepare a flexible models that can fully utilize the knowledge of the analyst.

In this paper, we shall show two basic tools for statistical modeling. Namely, firstly we shall show various information criteria AIC, TIC, GIC and EIC. We shall then show a general state space model as a generic time series model for signal extraction. Finally, we shall show some applications in seismology and macro economics.

2 Information Criteria for Statistical Modeling

Assume that the observations are generated from an unknown "true" distribution function $G(x)$ and the model is characterized by a density function $f(x)$. In the derivation of AIC (Akaike (1973)), the expected log-likelihood $E_Y \log f(Y) = \int \log f(y) dG(y)$ is used as the basic measure to evaluate the similarity between two distributions, which is equivalent to the Kullback-Leibler information.

In actual situations, $G(x)$ is unknown and only a sample $X = \{X_1, \ldots, X_n\}$ is given. We then use the log-likelihood $\ell = n \int \log f(x) d\hat{G}_n(x) = \sum_{i=1}^{n} \log f(X_i)$ as a natural estimator of (n times of) the expected log-likelihood. Here $\hat{G}_n(x)$ is the empirical distribution function defined by the data.

When a model contains an unknown parameter θ and its density is of the form $f(x|\theta)$, it naturally leads to use the maximum likelihood estimator $\hat{\theta}$.

2.1 AIC and TIC

For a statistical model $f(x|\theta)$ fitted to the data, however, the log-likelihood $n^{-1}\ell(\theta) = n^{-1} \sum_{i=1}^{n} \log f(X_i|\theta) \equiv n^{-1} \log f(X|\theta)$ has a positive bias as an estimator of the expected log-likelihood, $E_G \log f(Y|\theta)$, and it cannot be directly used for model selection. By correcting the bias

$$b(G) = nE_X \left\{ \frac{1}{n} \log f(X|\theta(X)) - E_Y \log f(Y|\theta(X)) \right\}, \tag{1}$$

an unbiased estimator of the expected log-likelihood is given by

$$\mathrm{IC} = -2n \left\{ \frac{1}{n} \log f(X|\theta(X)) - \frac{1}{n} b(G) \right\} = -2 \log f(X|\theta(X)) + 2b(G). \tag{2}$$

Since it is very difficult to obtain the bias $b(G)$ in a closed form, it is usually approximated by an asymptotic bias. Akaike (1973) approximated $b(G)$ by the number of parameters, $b_{\mathrm{AIC}} = m$, and proposed the AIC criterion,

$$\mathrm{AIC} \ = \ -2 \log f(X|\hat{\theta}_{ML}) \ + \ 2m, \tag{3}$$

where $\hat{\theta}_{\mathrm{ML}}$ is the maximum likelihood estimate. On the other hand, Takeuchi (1976) showed that the asymptotic bias is given by $b_{\mathrm{TIC}} = \mathrm{tr}\{\hat{I}(G)\hat{J}(G)^{-1}\}$, where $\hat{I}(G)$ and $\hat{J}(G)$ are the estimates of the Fisher information and expected Hessian matrices, respectively .

2.2 General Information Criterion, GIC

The above method of bias correction for the log-likelihood can be extended to a general model constructed by using a statistical functional such as $\hat{\theta} = T(\hat{G}_n)$. For such a general statistical model, Konishi and Kitagawa (1996) derived the asymptotic bias

$$b_{\mathrm{GIC}}(G) = \mathrm{tr} E_Y \left\{ T^{(1)}(Y;G) \frac{\partial \log f(Y|T(G))}{\partial \theta'} \right\}, \tag{4}$$

and proposed GIC (Generalized Information Criterion). Here $T^{(1)}(Y;G)$ is the first derivative of the statistical functional $T(Y;G)$ which is usually called the influence function.

The information criteria obtained so far can be generally expressed as $\log f(X|\hat{\theta}) - b_1(\hat{G}_n)$, where $b_1(\hat{G}_n)$ is the first order bias correction term such as (4). The second order bias-corrected information criterion can be defined by

$$\mathrm{GIC}_2 = -2 \log f(X|\hat{\theta}) + 2 \left\{ b_1(\hat{G}_n) + \frac{1}{n} \left(b_2(\hat{G}_n) - \Delta b_1(\hat{G}_n) \right) \right\}. \tag{5}$$

Here $b_2(G)$ is defined by the expansion

$$b(G) = E_X \left[\log f(X|\hat{\theta}) - n E_Y \log f(Y|\hat{\theta}) \right] = b_1(G) + \frac{1}{n} b_2(G) + O(n^{-2}), \tag{6}$$

and the bias of the first order bias correction term $\Delta b_1(G)$ is defined by

$$E_X \left[b_1(\hat{G}) \right] = b_1(G) + \frac{1}{n} \Delta b_1(G) + O(n^{-2}). \tag{7}$$

2.3 Bootstrap Information Criterion, EIC

The bootstrap method (Efron 1979) provides us with an alternative way of bias correction of the log-likelihood. In this method, the bias $b(G)$ in (1) is estimated by

$$b_B(\hat{G}_n) = E_{X^*} \left\{ \log f(X^*|\theta(X^*)) - \log f(X|T(X^*)) \right\}, \tag{8}$$

and the EIC (Extended Information Criterion) is defined by using this (Ishiguro et al. (1997)). In actual computation, the bootstrap bias correction term $b_B(\hat{G}_n)$ is approximated by bootstrap resampling.

The variance of the bootstrap estimate of the bias defined in (4) can be reduced automatically without any analytical arguments (Konishi and Kitagawa (1996), Ishiguro et al. (1997)). Let $D(X;G) = \log f(X|\hat{\theta}) - n E_Y[\log f(Y|\hat{\theta})]$. Then $D(X;G)$ can be decomposed into

$$D(X;G) = D_1(X;G) + D_2(X;G) + D_3(X;G) \tag{9}$$

where $D_1(X;G) = \log f(X|\hat{\theta}) - \log f(X|T(G))$, $D_2(X;G) = \log f(X|T(G)) - n E_Y[\log f(Y|T(G))]$ and $D_3(X;G) = n E_Y[\log f(Y|T(G))] - n E_Y[\log f(Y|\hat{\theta})]$.

For a general estimator defined by a statistical functional $\hat{\theta} = T(\hat{G}_n)$, it can be shown that the bootstrap estimate of $E_X[D_1 + D_3]$ is the same as that of $E_X[D]$, but $\mathrm{Var}\{D\} = O(n)$ and $\mathrm{Var}\{D_1 + D_3\} = O(1)$. Therefore by estimating the bias by

$$b_B^*(\hat{G}_n) = E_{X^*}[D_1 + D_3], \tag{10}$$

a significant reduction of the variance can be achieved for any estimators defined by statistical functional, especially for large n.

3 State Space Modeling

3.1 Smoothness Prior Modeling

A smoothing approach attributed to Whittaker [21], is as follows: Let

$$y_n = f_n + \varepsilon_n, \quad n = 1, ..., N \tag{11}$$

denote observations, where f_n is an unknown smooth function of n. ε_n is an i.i.d. normal random variable with zero mean and unknown variance σ^2. The problem is to estimate $f_n, n = 1, ..., N$ from the observations, $y_n, n = 1, ..., N$, in a statistically reasonable way. However, in this problem, the number of parameters to be estimated is equal to or even greater than the number of observations. Therefore, the ordinary least squares method or the maximum likelihood method yield meaningless results. Whittaker [21] suggested that the solution $f_n, n = 1, ..., N$ balances a tradeoff between infidelity to the data and infidelity to a smoothness constraint. Namely, for given tradeoff parameter λ^2 and the difference order k, the solution satisfies

$$\min_f \left[\sum_{n=1}^{N} (y_n - f_n)^2 + \lambda^2 \sum_{n=1}^{N} (\Delta^k f_n)^2 \right]. \tag{12}$$

Whittaker left the choice of λ^2 to the investigator.

3.2 State Space Modeling

It can be seen that the minimization of the criterion (12) is equivalent to assume the following linear-Gaussian model:

$$\begin{aligned} y_n &= f_n + w_n, \\ f_n &= c_1^k f_{n-1} + \cdots + c_k^k f_k + v_n, \end{aligned} \tag{13}$$

where $w_n \sim N(0, \sigma^2)$, $v_n \sim N(0, \tau^2)$, $\lambda^2 = \sigma^2/\tau^2$ and c_j^k is the j-th binomial coefficient.

Therefore, the models (13) can be expressed in a special form of the state space model

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \qquad \text{(system model)}, \\ y_n &= Hx_n + w_n \qquad \text{(observation model)}, \end{aligned} \tag{14}$$

where $x_n = (t_n, ..., t_{n-k+1})'$ is a k-dimensional state vector, F, G and H are $k \times k$, $k \times 1$ and $1 \times k$ matrices, respectively. For example, for $k = 2$, they are given by

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \end{bmatrix}, \quad F = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H = [1, 0]. \tag{15}$$

One of the merits of using this state space representation is that we can use computationally efficient Kalman filter for state estimation. Since the state vector contains unknown trend component, by estimating the state vector x_n, the trend is automatically estimated. Also unknown parameters of the model, such as the variances σ^2 and τ^2 can be estimated by the maximum likelihood method. In general, the likelihood of the time series model is given by

$$L(\theta) = p(y_1, \ldots, y_N|\theta) = \prod_{n=1}^{N} p(y_n|Y_{n-1}, \theta), \tag{16}$$

where $Y_{n-1} = \{y_1, \ldots, y_{n-1}\}$ and each component $p(y_n|Y_{n-1}, \theta)$ can be obtained as byproduct of the Kalman filter [6]. It is interesting to note that the tradeoff parameter λ^2 in the penalized least squares method (12) can be interpreted as the ratio of the system noise variance to the observation noise variance, or the signal-to-noise ratio.

The individual terms in (16) are given by, in general p-dimensional observation case,

$$p(y_n|Y_{n-1}, \theta) = \frac{1}{(\sqrt{2\pi})^p} \left|W_{n|n-1}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\varepsilon'_{n|n-1} W^{-1}_{n|n-1} \varepsilon_{n|n-1}\right\}, \tag{17}$$

where $\varepsilon_{n|n-1} = y_n - y_{n|n-1}$ is one-step-ahead prediction error of time series and $y_{n|n-1}$ and $V_{n|n-1}$ are the mean and the variance covariance matrix of the observation y_n, respectively, and are defined by

$$y_{n|n-1} = Hx_{n|n-1}, \quad W_{n|n-1} = HV_{n|n-1}H' + \sigma^2. \tag{18}$$

Here $x_{n|n-1}$ and $V_{n|n-1}$ are the mean and the variance covariance matrix of the state vector given the observations Y_{n-1} and can be obtained by the Kalman filter [6]. If there are several candidate models, the goodness of the fit of the models can be evaluated by the AIC criterion defined by

$$\text{AIC} = -2 \log L(\hat{\theta}) + 2(\text{number of parameters}). \tag{19}$$

3.3 General State Space Modeling

Consider a nonlinear non-Gaussian state space model for the time series y_n,

$$\begin{aligned} x_n &= F_n(x_{n-1}, v_n) \\ y_n &= H_n(x_n, w_n), \end{aligned} \tag{20}$$

where x_n is an unknown state vector, v_n and w_n are the system noise and the observation noise with densities $q_n(v)$ and $r_n(w)$, respectively. The first and the

second model in (20) are called the system model and the observation model, respectively. The initial state x_0 is assumed to be distributed according to the density $p_0(x)$. $F_n(x, v)$ and $H_n(x, w)$ are possibly nonlinear functions of the state and the noise inputs. This model is an extension of the ordinary linear Gaussian state space model (14).

The above nonlinear non-Gaussian state space model specifies the conditional density of the state given the previous state, $p(x_n|x_{n-1})$, and that of the observation given the state, $p(y_n|x_n)$. This is the essential features of the state space model, and it is sometimes convenient to express the model in this general form based on conditional distributions

$$\begin{aligned} x_n &\sim Q_n(\ \cdot\ |x_{n-1}) \\ y_n &\sim R_n(\ \cdot\ |x_n). \end{aligned} \tag{21}$$

With this model, it is possible to treat the discrete process such as the Poisson models.

3.4 Nonlinear Non-Gaussian Filtering

The most important problem in state space modeling is the estimation of the state vector x_n from the observations, $Y_t \equiv \{y_1, \ldots, y_t\}$, since many important problems in time series analysis can be solved by using the estimated state vector. The problem of state estimation can be formulated as the evaluation of the conditional density $p(x_n|Y_t)$. Corresponding to the three distinct cases, $n > t$, $n = t$ and $n < t$, the conditional distribution, $p(x_n|Y_t)$, is called the predictor, the filter and the smoother, respectively.

For the standard linear-Gaussian state space model, each density can be expressed by a Gaussian density and its mean vector and the variance-covariance matrix can be obtained by computationally efficient Kalman filter and smoothing algorithms [6].

For general state space models, however, the conditional distributions become non-Gaussian and their distributions cannot be completely specified by the mean vectors and the variance covariance matrices. Therefore, various types of approximations to the densities have been used to obtain recursive formulas for state estimation, e.g., the extended Kalman filter [6], the Gaussian-sum filter [5] and the dynamic generalized linear model [20].

However, the following non-Gaussian filter and smoother [11] can yield an arbitrarily precise posterior density.

[Non-Gaussian Filter]

$$\begin{aligned} p(x_n|Y_{n-1}) &= \int p(x_n|x_{n-1})p(x_{n-1}|Y_{n-1})dx_{n-1} \\ p(x_n|Y_n) &= \frac{p(y_n|x_n)p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})}, \end{aligned} \tag{22}$$

where $p(y_n|Y_{n-1})$ is defined by $\int p(y_n|x_n)p(x_n|Y_{n-1})dx_n$.

[Non-Gaussian Smoother]

$$p(x_n|Y_N) = p(x_n|Y_n)\int \frac{p(x_{n+1}|x_n)p(x_{n+1}|Y_N)}{p(x_{n+1}|Y_n)}dx_{n+1}. \tag{23}$$

However, the direct implementation of the formula requires computationally very costly numerical integration and can be applied only to lower dimensional state space models.

3.5 Sequential Monte Carlo Filtering

To mitigate the computational burden, numerical methods based on Monte Carlo approximation of the distribution have been proposed [9,12]. In the Monte Carlo filtering [12], we approximate each density function by many particles that can be considered as realizations from that distribution. Specifically, assume that each distribution is expressed by using m particles as follows: $\{p_n^{(1)}, \ldots, p_n^{(m)}\} \sim p(x_n|Y_{n-1})$ and $\{f_n^{(1)}, \ldots, f_n^{(m)}\} \sim p(x_n|Y_n)$. This is equivalent to approximate the distributions by the empirical distributions determined by m particles. Then it will be shown that a set of realizations expressing the one step ahead predictor $p(x_n|Y_{n-1})$ and the filter $p(x_n|Y_n)$ can be obtained recursively as follows.

[Monte Carlo Filter]

1. *Generate a random number* $f_0^{(j)} \sim p_0(x)$ *for* $j = 1, \ldots, m$.
2. *Repeat the following steps for* $n = 1, \ldots, N$.
   a) *Generate a random number* $v_n^{(j)} \sim q(v)$, *for* $j = 1, \ldots, m$.
   b) *Compute* $p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)})$, *for* $j = 1, \ldots, m$.
   c) *Compute* $\alpha_n^{(j)} = p(y_n|p_n^{(j)})$ *for* $j = 1, \ldots, m$.
   d) *Generate* $f_n^{(j)}$, $j = 1, \ldots, m$ *by the resampling of* $p_n^{(1)}, \ldots, p_n^{(m)}$. *with the weights proportional to* $\alpha_n^{(1)}, \ldots, \alpha_n^{(j)}$.

The above algorithm for Monte Carlo filtering can be extended to smoothing by a simple modification. The details of the derivation of the algorithm is shown in [12].

3.6 Self-Organizing State Space Model

If the non-Gaussian filter is implemented by the Monte Carlo filter, the sampling error sometimes renders the maximum likelihood method impractical. In this case, instead of estimating the parameter θ by the maximum likelihood method, we consider a Bayesian estimation by augmenting the state vector as $z_n = [x_n^T, \theta^T]^T$. The state space model for this augmented state vector z_n is given by

$$\begin{aligned} z_n &= F^*(z_{n-1}, v_n) \\ y_n &= H^*(z_n, w_n) \end{aligned} \tag{24}$$

where the nonlinear functions $F^*(z,v)$ and $H^*(z,w)$ are defined by $F^*(z,v) = [F(x,v),\theta]^T$, $H^*(z,w) = H(x,w)$.

Assume that we obtain the posterior distribution $p(z_n|Y_N)$ given the entire observations $Y_N = \{y_1, \cdots, y_N\}$. Since the original state vector x_n and the parameter vector θ are included in the augmented state vector z_n, it immediately yields the marginal posterior densities of the parameter and of the original state.

This method of Bayesian simultaneous estimation of the parameter and the state of the state space model can be easily extended to a time-varying parameter situation where the parameter $\theta = \theta_n$ evolves with time n. It should be noted that in this case we need a proper model for time evolution of the parameter.

4 Examples

4.1 Extraction of Seismic Waves

The earth's surface is under continuous disturbances due to a variety of natural forces and human induced sources. Therefore, if the amplitude of the earthquake signal is very small, it will be quite difficult to distinguish it from the background noise. In this section, we consider a method of extracting small seismic signals (P-wave and S-wave) from relatively large background noise [13], [17].

For the extraction of the small seismic signal from background noise, we consider the model

$$y_n = r_n + s_n + \varepsilon_n, \tag{25}$$

where r_n, s_n and ε_n denote the background noise, the signal and the observation noise, respectively. To separate these three components, it is assumed that the background noise r_n is expressed by the autoregressive model

$$r_n = \sum_{i=1}^{m} c_i r_{n-i} + u_n \tag{26}$$

where the AR order m and the AR coefficients c_i are unknown and u_n and ε_n are white noise sequences with $u_n \sim N(0, \tau_1^2)$ and $\varepsilon_n \sim N(0, \sigma^2)$.

The seismograms are actually records of seismic waves in 3-dimensional space and the seismic signal is composed of P-wave and S-wave. Hereafter East-West, North-South and Up-Down components are denoted as $y_n = [x_n, y_n, z_n]^T$. P-wave is a compression wave and it moves along the wave direction. Therefore it can be approximated by a one-dimensional model,

$$p_n = \sum_{j=1}^{m} a_j p_{n-j} + u_n. \tag{27}$$

On the other hand, S-wave moves on a plane perpendicular to the wave direction and thus can be expressed by 2-dimensional model,

$$\begin{bmatrix} q_n \\ r_n \end{bmatrix} = \sum_{j=1}^{\ell} \begin{bmatrix} b_{j11} & b_{j12} \\ b_{j21} & b_{j22} \end{bmatrix} \begin{bmatrix} q_{n-j} \\ r_{n-j} \end{bmatrix} + \begin{bmatrix} v_{n1} \\ v_{n2} \end{bmatrix}. \tag{28}$$

Therefore, the observed three-variate time series can be expressed as

$$\begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} = \begin{bmatrix} \alpha_{1n}\ \beta_{1n}\ \gamma_{1n} \\ \alpha_{2n}\ \beta_{2n}\ \gamma_{2n} \\ \alpha_{3n}\ \beta_{3n}\ \gamma_{3n} \end{bmatrix} \begin{bmatrix} p_n \\ q_n \\ r_n \end{bmatrix} + \begin{bmatrix} w_n^x \\ w_n^y \\ w_n^z \end{bmatrix}. \tag{29}$$

In this approach, the crucial problem is the estimation of time-varying wave direction, α_{jn}, β_{jn} and γ_{jn}. They can be estimated by the principle component analysis of the 3D data. These models can be combined in the state space model form.

Note that the variances of the component models corresponds to the amplitude of the seismic signals and are actually time varying. These variance parameters play the role of a signal to noise ratios, and the estimation of these parameters is the key problem for the extraction of the seismic signal. A self-organizing state space model can be applied to the estimation of the time-varying variances [13].

4.2 Seasonal Adjustment

The standard model for seasonal adjustment is given by

$$y_n = t_n + s_n + w_n, \tag{30}$$

where t_n, s_n and w_n are trend, seasonal and irregular components. A reasonable solution to this decomposition was given by the use of smoothness priors for both t_n and s_n [14]. The trend component t_n and the seasonal component s_n are assumed to follow

$$\begin{aligned} t_n &= 2t_{n-1} - t_{n-2} + v_n, \\ s_n &= -(s_{n-1} + \cdots + s_{n-11}) + u_n, \end{aligned} \tag{31}$$

where v_n, u_n and w_n are Gaussian white noise with $v_n \sim N(0, \tau_t^2)$, $u_n \sim N(0, \tau_s^2)$ and $w_n \sim N(0, \sigma^2)$.

However, by using a more sophisticated model, we can extract a more information from the data. For example, many of the economic time series related to sales or production are affected by the number of days of the week. Therefore, the sales of a department store will be strongly affected by the number of Sundays and Saturdays in each month. Such kind of effect is called the trading day effect.

To extract the trading day effect, we consider the decomposition

$$y_n = t_n + s_n + td_n + w_n, \tag{32}$$

where t_n, s_n and w_n are as above and the trading day effect component, td_n, is assumed to be expressed as

$$td_n = \sum_{j=1}^{7} \beta_j d_{jn}, \tag{33}$$

where d_{jn} is the number of j-th day of the week (e.g., j=1 for Sunday and j=2 for Monday, etc.) and β_j is the unknown trading day effect coefficient. To assure the identifiability, it is necessary to put constraint that $\beta_1 + \cdots + \beta_7 = 0$.

Since the numbers of day of the week are completely determined by the calendar, if we obtain good estimates of the trading day effect coefficients, then it will greatly contribute to the increase of the precision of the prediction.

4.3 Analysis of Exchange Rate Data

We consider the multivariate time series of exchange rate between US dollars and other foreign currencies. By using proper smoothness prior models, we try to decompose the change of the exchange rate into two components, one expresses the effect of US economy and the other the effect of other country. By this decomposition, it is possible to determine, for example, whether the decrease of the Yen/USD exchange rate at a certain time is due to weak Yen or strong US dollar.

References

1. Akaike, H.: Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N., Csaki, F. (eds.): 2nd International Symposium in Information Theory. Akademiai Kiado, Budapest, (1973) 267–281.
2. Akaike, H.: A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**, 716–723 (1974)
3. Akaike, H.: Likelihood and the Bayes procedure (with discussion), In Bayesian Statistics, edited by J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith, University press, Valencia, Spain, 143–166 (1980)
4. Akaike, H., and Kitagawa, G. eds.: The Practice of Time Series Analysis, Springer-Verlag New York (1999)
5. Alspach, D.L., Sorenson, H.W.: Nonlinear Bayesian Estimation Using Gaussian Sum Approximations. IEEE Transactions on Automatic Control, **AC-17** (1972) 439–448.
6. Anderson, B.D.O., Moore, J.B.: Optimal Filtering, New Jersey, Prentice-Hall (1979).
7. Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer-Verlag, NewYork (2000).
8. Efron, B.: Bootstrap methods: Another look at the jackknife. Ann. Statist. **7**, (1979) 1–26.
9. Gordon, N.J., Salmond, D.J., Smith, A.F.M., Novel approach to nonlinear /non-Gaussian Bayesian state estimation, IEE Proceedings-F, **140**, (2) (1993) 107–113.
10. Ishiguro, M., Sakamoto, Y.,Kitagawa, G.: Bootstrapping log-likelihood and EIC, an extension of AIC. Annals of the Institute of Statistical Mathematics, **49** (3), (1997) 411–434.
11. Kitagawa, G.: Non-Gaussian state-space modeling of nonstationary time series. Journal of the American Statistical Association, **82** (1987) 1032–1063.
12. Kitagawa, G.: Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics, **5** (1996) 1–25.

13. Kitagawa, G.: Self-organizing State Space Model. Journal of the American Statistical Association. **93** (443) (1998) 1203–1215.
14. Kitagawa, G., Gersch, W.: A Smoothness Priors-State Space Approach to the Modeling of Time Series with Trend and Seasonality. Journal of the American Statistical Association, **79** (386) (1984) 378–389.
15. Kitagawa, G. and Gersch, W.: Smoothness Priors Analysis of Time Series, Lecture Notes in Statistics, No. 116, Springer-Verlag, New York (1996).
16. Kitagawa, G. and Higuchi, T.: Automatic transaction of signal via statistical modeling, The Proceedings of The First International Conference on Discovery Science, Springer-Verlag Lecture Notes in Artificial Intelligence Series, 375–386 (1998).
17. Kitagawa, G., Takanami, T., Matsumoto, N.: Signal Extraction Problems in Seismology, Intenational Statistical Review, **69** (1), (2001) 129–152.
18. Konishi, S., Kitagawa, G.: Generalised information criteria in model selection. Biometrika, **83**, (4), (1996) 875–890.
19. Sakamoto, Y., Ishiguro, M. and Kitagawa, G.: Akaike Information Criterion Statistics, D-Reidel, Dordlecht, (1986)
20. West, M., Harrison, P.J., Migon,H.S.: Dynamic generalized linear models and Bayesian forecasting (with discussion). Journal of the American Statistical Association. **80** (1985) 73–97.
21. Whittaker, E.T: On a new method of graduation, Proc. Edinborough Math. Assoc., **78**, (1923) 81–89.