Constrained forecasting with time series models

Ilkka Karanta
VTT Information Technology, Information Systems
P.O. Box 1201 (street address: Tekniikantie 4B, Espoo)
FIN-02044 VTT, Finland
ilkka.karanta@vtt.fi

1. Introduction

In many practical forecasting tasks, different models are built for different time horizons. The forecasts produced on the basis of these models are used to help with different types of decisions. In many cases, it would be advantageous to merge the forecasts for different horizons: it is reasonable to expect significant accuracy improvement especially in the long term of the higher sample rate model.

This problem is handled in several methodologies such as combining forecasts (Bunn 1989), temporal disaggretagion (Guerrero 1990), and others.

This paper introduces a method of taking into account the aggregated/lower sample rate forecasts within the higher sample rate forecasts.

2. Some notation

Let z_t, $t = \ldots -1, 0, 1, 2, \ldots$ be an infinite realization of the random process of interest. The values of z_t are known from time instance 1 up to time instant T, and we want to know them for all time instants in the forecasting interval $\{T+1, \ldots, T+N\} \stackrel{\text{def}}{=} I$.

We assume that z_t can be modelled in linear regression form:

$$(1) \quad z_t = \sum_{k=1}^{m} b_t(k) x_t(k) + a_t,$$

where $b_t(k)$ are known coefficients. a_t is Gaussian noise, not necessarily white; it is assumed to have the covariance matrix $\mathbf{\Sigma}_a$. $x_t(k)$ are the explanatory variables.

These requirements are fulfilled for, e.g., regression models and ARMA models (where the regression form condition is fulfilled by converting the model to AR form), and transfer function models; they are not fulfilled by e.g. exponential smoothing models in general.

On the forecasting interval, we have observations Y_i (generated e.g. by making forecasts with other methods) of some linear combinations of z_t, $t \in I$:

$$(2) \quad Y_i = \sum_{j=1}^{N} c_{ij} z_{T+j} + e_i, \quad i = 1, \ldots, K.$$

e_i is Gaussian noise, and c_{ij} are some constants. The c_{ij} can be adapted to reflect different knowledge about the lower sample rate/aggregated values, such quarterly data in terms of monthly data, values at specific time instants (or even all time instants of interest) obtained by some other model than the one used, etc. In matrix form,

$$(3) \quad \mathbf{Y} = \mathbf{C}\mathbf{z} + \mathbf{e},$$

where the vector $\mathbf{Y} = (Y_1, \ldots, Y_K)'$, $\mathbf{C}$ is the matrix whose ith row is $(c_{i1}, \ldots, c_{iN})$, $\mathbf{z} = (z_{T+1}, \ldots, z_{T+N})'$, $\mathbf{e} = (e_1, e_2, \ldots e_K)'$. The covariance matrix of $\mathbf{e}$ is assumed to be $\mathbf{\Sigma}_e$.

3. The method

This method is based on the observation that both 1 and 2 define multivariate normal distributions with shared variables. This leads naturally to the thought that information in the two normal distributions be combined to form a third normal distribution which reflects the information from both of the original distributions.

Let us denote the forecasts obtained from 1 alone by $\hat{\mathbf{z}}$, and the more precise forecasts we are aiming at by $\breve{\mathbf{z}}$. Calculating $\breve{\mathbf{z}}$ as a result of conditioning the $\hat{\mathbf{z}}$ by Y (see Rao 1973 for the application of conditional probability to normally distributed variables), we get

$$(4) \quad \breve{\mathbf{z}} = \hat{\mathbf{z}} + \mathbf{\Sigma}_a \mathbf{C}^T (\mathbf{C}\mathbf{\Sigma}_a \mathbf{C}^T + \mathbf{\Sigma}_e)^{-1} (\mathbf{Y} - \mathbf{C}\hat{\mathbf{z}})$$

with error covariance matrix

$$(5) \quad \mathbf{P} = \mathbf{\Sigma}_a - \mathbf{\Sigma}_a \mathbf{C}^T (\mathbf{C}\mathbf{\Sigma}_a \mathbf{C}^T + \mathbf{\Sigma}_e)^{-1} \mathbf{C}\mathbf{\Sigma}_a.$$

This approach underlies the Kalman filter (see, e.g., Anderson and Moore 1979), and indeed it can be seen as a one-step Kalman filter. However, it has some advantages over the Kalman filter: no state space model needs to be formulated; and the method is applicable even when the matrix $\mathbf{C}$ is such that no meaningful observation matrix of the state space model can be formulated.

A downside of the method with respect to the Kalman filter is that it involves larger matrices. However, with current computing power this should pose no problem with reasonable forecasting spans.

REFERENCES

Anderson, B.D.O. and Moore, J.B. (1979). Optimal filtering. Prentice Hall, Inc., Englewood Cliffs, NJ.

Bunn, D.W. (1989). Forecasting with more than one model. Journal of Forecasting, 8(3):161–166.

Guerrero, V.M. (1990). Temporal disaggregation of time series: an arima-based approach. International Statistical Review, 58(1):29–46.

Rao, C.R. (1973). Linear statistical inference and its applications. John Wiley & Sons, New York–London–Sydney, Second edition.

FRENCH RÉSUMÉ

Notre probléme concerne la prise en compte des observations échantillonnées à faible fréquence (trimestriel) quaud on fait des prévisions des processes échantillonnés à une fréquence élevée (mensuel). Nous proposons une méthode baseé sur le conditionnement de la probabilité de distribution de l'échantillonnage à fréquence élevée avec la probabilité de distribution de l'échantillonnage à taux faible. La méthode est similaire au filtre de Kalman à un pas.