Lecture 8: Forecasting (continued)

Bus 41910, Time Series Analysis, Mr. R. Tsay

In some applications, it is important to consider the model uncertainty in forecasting. We shall briefly mention two forecasting methods that can handle model uncertainty. The first method is the Bayesian forecasting through simulation. The second is an adaptive forecasting method.

A. Bayesian forecast: A good reference of this approach is Thompson and Miller (1986, JBES, pp. 427-436). For simplicity, this approach often focuses on AR models. Implementation of this approach to ARMA and MA models can be quite involved, because of the non-linear nature of the MA parameters. However, recent developments in Markov Chain Monte Carlo (MCMC) methods, e.g. the Gibbs sampler, have mitigated some of the difficulties in using this approach.
Consider the AR model

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t$$

Here the unknown parameters are $\mathbf{\Phi} = (\phi_0, \phi_1, \cdots, \phi_p)'$ and σ_a^2. Often it is more convenient to reparametrize σ_a^2 in term of the "precision" parameter τ such that $\tau = \sigma_a^{-2}$. A typical Bayesian analysis is then as follows:

- Specify a joint prior distribution of the parameters $\mathbf{\Phi}$ and τ.
- Assume the process Z_t is normally distributed.
- Obtain the posterior distribution of the parameters $\mathbf{\Phi}$ and τ.
- Consider the predictive distribution of the future observations on which the forecasts are based.

Let $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_n)'$ be the observed data. Under the diffuse prior

$$f(\mathbf{\Phi}, \tau) \approx \frac{1}{\tau}$$

Zellner (1971, p. 195) showed that the posterior distribution of τ is gamma and the conditional distribution of $\mathbf{\Phi}$ given $\boldsymbol{Z}$ and τ is normal. That is,

$$f(\tau|\boldsymbol{Z}) \sim \Gamma(v/2, 2/S(\hat{\mathbf{\Phi}}))$$

and

$$f(\mathbf{\Phi}|\boldsymbol{Z}, \tau) \sim N(\hat{\mathbf{\Phi}}, [\tau \boldsymbol{X}'\boldsymbol{X}]^{-1})$$

where $v = n - p - 1$ and $\boldsymbol{X}'\boldsymbol{X}$ is the design-matrix of the AR(p) model, $\hat{\mathbf{\Phi}}$ is the least squares estimate of $\mathbf{\Phi}$ and $S(\hat{\mathbf{\Phi}})$ denotes the residual sum of squares. [These results are

standard in Bayesian analysis. They are closely related to the case of *iid* normal with unknown mean and unknown variance and Jeffery's prior. See any textbook of Bayesian inference, e.g. DeGroot (1970) or Box and Tiao (1973).]
Turn to prediction. A joint ℓ-step ahead forecast requires the predictive distribution

$$\begin{aligned} f(Z_{n+1},\cdots,Z_{n+\ell}|\boldsymbol{Z}) &= \int f(Z_{n+1},\cdots,Z_{n+\ell},\tau,\boldsymbol{\Phi}|\boldsymbol{Z})d\boldsymbol{\Phi}d\tau \\ &= \int f(Z_{n+1},\cdots,Z_{n+\ell}|\boldsymbol{\Phi},\tau,\boldsymbol{Z})f(\boldsymbol{\Phi}|\tau,\boldsymbol{Z})f(\tau|\boldsymbol{Z})d\boldsymbol{\Phi}d\tau \end{aligned} \quad (1)$$

where the integration is over the region $-\infty < \phi_i < \infty$ and $\tau > 0$.
For an AR(p) model with $\ell = 1$, the corresponding predictive distribution is a Student t with $n-p-1$ degrees of freedom. For general ℓ, we have

$$f(Z_{n+1},\cdots,Z_{n+\ell}|\boldsymbol{Z}) = \prod_{i=1}^{\ell} f(Z_{n+i}|\boldsymbol{Z},Z_{n+1},\cdots,Z_{n+\ell-1}).$$

Unfortunately, products of t-distribution do not corresponding to a closed form distribution. Thus, to use the above predictive distribution, one often tries to match certain moments in order to produce point forecasts and forecast intervals.
On the other hand, consider the equation in (1). The last two term of the integrand are gamma and normal, respectively, whereas the first term of the integrand is the density of a sequence of ℓ random variables of an AR(p) model. It is, therefore, easy to "simulate" the predictive distribution.
A Simulation Procedure:

1. Choose a value of τ from the gamma distribution $\Gamma(v/2, 2/S(\hat{\boldsymbol{\Phi}}))$.
2. Choose a set of parameter $\boldsymbol{\Phi}$ from the conditional distribution $N(\hat{\boldsymbol{\Phi}}, [\tau\boldsymbol{X}'\boldsymbol{X}]^{-1})$.
3. Simulate a path of $Z_{n+1},\cdots,Z_{n+\ell}$ by using the AR(p) model, i.e., draw $a_{n+1},\cdots,a_{n+\ell}$ from $N(0,\tau^{-1})$ and use the parameters chosen in Steps 1 and 2.
4. Repeat Steps 1-3 for many times and use the collection of the paths to make forecasts and forecast intervals.

For illustration, see the analysis of U.S. quarterly unemployment rate in Thompson and Miller (1986).

B. Adaptive Forecasting: This approach is motivated by the idea that **ALL** statistical models are wrong. Thus, the principal of maximum likelihood is not applicable (strictly speaking). Thus, instead of selecting a most plausible model for a given data set, one simply entertains a "simple" model such as the exponential smoothing ARIMA(0,1,1) model or an ARMA(1,1) model for a non-seasonal time series. In practice, the selection of a model can be governed by theoretical as well as practical consideration.

Denote the unknown parameters of the entertained simple model by $\boldsymbol{\theta}$. Suppose that one is interested in the ℓ-step ahead forecasts. Then the parameter $\boldsymbol{\theta}$ is estimated by minimizing the sum of squares of the ℓ-step ahead forecast errors. Of course, for 1-step ahead forecast, this approach redices to the traditional least squares method. However, for multi-step ahead forecasts, it is different from the least squares or maximum likelihood method.
For discussion, suppose that the entertained model is ARMA(1,1) model

$$(1-\beta B)Z_t = (1-\eta B)b_t$$

where b_t might not be a white noise series, as we do not believe that the model is the "true" model. For this ARMA(1,1) model, the ψ-weights are $\psi_i = \beta^{i-1}(\beta-\eta)$. Therefore, the forecast errors are

$$e_t(\ell) = \begin{cases} b_{t+1} & \text{for } \ell = 1 \\ b_{t+2} + (\beta-\eta)b_{t+1} & \text{for } \ell = 2 \\ b_{t+3} + (\beta-\eta)b_{t+2} + \beta(\beta-\eta)b_{t+1} & \text{for } \ell = 3 \\ \vdots & \vdots \end{cases}$$

The parameters β and η are then estimated, for ℓ-step ahead forecasts, by minimizing

$$S(\ell,\beta,\eta) = \sum_{t=\ell+1}^{n-\ell} [e_t(\ell)]^2.$$

This is a non-linear optimization problem which can be solved by any package, e.g. IMSL or NAG subroutine.
For application of this adaptive approach in forecasting long-memory time series, see Tiao and Tsay (1994, JoF) and the references therein.

Finally, there are many other forecasting methods. For instance, Litterman (1980, 86, JBES) discussed Bayesian forecasting of vector AR models and Garcia-Ferrer, et al (1987, JBES) discussed pooled forecasts (or shrinkage estimates).