STAT 804

Lecture 18 Notes

Forecast standard errors

You should remind yourself that the computations of conditional expectations we have made used the fact that the a's and b's are constants – the true parameter values. In fact we then replace the parameter values with estimates. The quality of our forecasts will be summarized by the forecast standard error:

$$\sqrt{\mathrm{E}[(X_t - \hat{X}_t)^2]} \,.$$

We will compute this ignoring the estimation of the parameters and then discuss how much that might have cost us.

If $\hat{X}_t = \mathrm{E}(X_t|X)$ then $\mathrm{E}(\hat{X}_t) + \mathrm{E}(X_t)$ so that our forecast standard error is just the variance of $X_t - \hat{X}_t$.

Consider first the case of an AR(1) and one step ahead forecasting:

$$X_T - \hat{X}_T = \epsilon_T \,.$$

The variance of this forecast is σ_ϵ^2 so that the forecast standard error is just σ_ϵ.

For forecasts further ahead in time we have

$$\hat{X}_{T+r} = a\hat{X}_{T+r-1}$$

and

$$X_{T+r} = aX_{T+r-1} + \epsilon_{T+r}$$

Subtracting we see that

$$\mathrm{Var}(X_{T+r} - \hat{X}_{T+r}) = \sigma_\epsilon^2 + \mathrm{Var}(X_{T+r-1} - \hat{X}_{T+r-1})$$

so that we may calculate forecast standard errors recursively. As $r \to \infty$ we can check that the forecast variance converges to

$$\sigma_\epsilon^2/(1-a^2)$$

which is simply the variance of individual Xs. When you forecast a *stationary* series far into the future the forecast error is just the standard deviation of the series.

Turn now to a general ARMA(p,q). Rewrite the process as the infinite order AR

$$X_t = \sum_{s>0} c_s X_{t-s} + \epsilon_t$$

to see that again, ignoring the truncation of the infinite sum in the forecast we have

$$X_T - \hat{X}_T = \epsilon_T$$

so that the one step ahead forecast standard error is again σ_ϵ.

Parallel to the AR(1) argument we see that

$$X_{T+r} - \hat{X}_{T+r} = \sum_{j=0}^{r-1} a_j (X_{T+j} - \hat{X}_{T+j}) + \epsilon_{T+r} \, .$$

The errors on the right hand side are not independent of one another so that computation of the variance requires either computation of the covariances or recognition of the fact that the right hand side is a linear combination of $\epsilon_T, \ldots, \epsilon_{T+r}$.

A simpler approach is to write the process as an infinite order MA:

$$X_t = \epsilon_t + \sum_{s>0} d_s \epsilon_{t-s}$$

for suitable coefficients d_s. Now if we treat conditioning on the data as being effectively equivalent to conditioning on all X_t for $t < T$ we are effectively conditioning on ϵ_t for all $t < T$. This means that

$$\begin{aligned} \mathrm{E}(X_{T+r} | X_{T-1}, X_{T-2}, \ldots) &= \mathrm{E}(X_{T+r} | \epsilon_{T-1}, \epsilon_{T-2}, \ldots) \\ &= \sum_{s>r} d_s \epsilon_{T+r-s} \end{aligned}$$

and the forecast error is just

$$X_{T+r} - \hat{X}_{T+r} = \epsilon_t + \sum_{s=1}^{r} d_s \epsilon_{T+r-s}$$

so that the forecast standard error is

$$\sigma_\epsilon \sqrt{1 + \sum_{s=1}^{r} d_s^2} \, .$$

Again as $r \to \infty$ this converges to σ_X.

Finally consider forecasting the ARIMA(p,d,q) process $(I - B)^d X = W$ where W is ARMA(p,q). The forecast errors in X can clearly be written as a linear combination of forecast errors for W permitting the forecast error in X to be written as a linear combination of the underlying errors ϵ_t. As an example consider first the ARIMA(0,1,0) process $X_t = \epsilon_t + X_{t-1}$. The forecast of ϵ_{T+r} is just 0 and so the forcast of X_{T+r} is just

$$\hat{X}_{T+r} = \hat{X}_{T+r-1} = \cdots = X_{T-1} \, .$$

The forecast error is

$$\epsilon_{T+r} + \cdots + \epsilon_T$$

whose standard deviation is $\sigma\sqrt{r+1}$. Notice that the forecast standard error grows to infinity as $r \to \infty$. For a general ARIMA$(p,1,q)$ we have

$$\hat{X}_{T+r} = \hat{X}_{T+r-1} + \hat{W}_{T+r}$$

and

$$X_{T+r} - \hat{X}_{T+r} = (W_{T+r} - \hat{W}_{T+r}) + \cdots + (W_T - \hat{W}_T)$$

which can be combined with the expression above for the forecast error for an ARMA(p, q) to compute standard errors.

Software

The S-Plus function *arima.forecast* can do the forecasting.

Comments

I have ignored the effects of parameter estimation throughout. In ordinary least squares when we predict the Y corresponding to a new x we get a forecast standard error of

$$\sqrt{Var(Y - x\hat{\beta})} = \sqrt{Var(\epsilon + x(\beta - \hat{\beta}))}$$

which is

$$\sigma\sqrt{1 + x(X^T X)^{-1} x^T}\,.$$

The procedure used here corresponds to ignoring the term $x(X^T X)^{-1} x^T$ which is the variance of the fitted value. Typically this value is rather smaller than the 1 to which it is added. In a 1 sample problem for instance it is simply $1/n$. Generally the major component of forecast error is the standard error of the noise and the effect of parameter estimation is unimportant.

In regression we sometimes compute perdiction intervals

$$\hat{Y} \pm c\hat{\sigma}_{\hat{Y}}$$

The multiplier c is adjusted to make the coverage probability $P(\frac{|Y - \hat{Y}|}{c} \le 1)$ close to a desired coverage probability such as 0.95. If the errors are normal then we can get c by taking $t_{0.025,n-p}s\sqrt{1 + x(X^T X)^{-1} x^T}$. When the errors are not normal, however, the error in $Y - \hat{Y}$ is dominated by ϵ which is not normal so that the coverage probability can be radically different from the nominal. Moreover, there is no particular theoretical justification for the use of t critical points. However, even for non-normal errors the prediction standard error is a useful summary of the accuracy of a prediction.