MATH3802
Chapter 5: Forecasting

Peter Thwaites

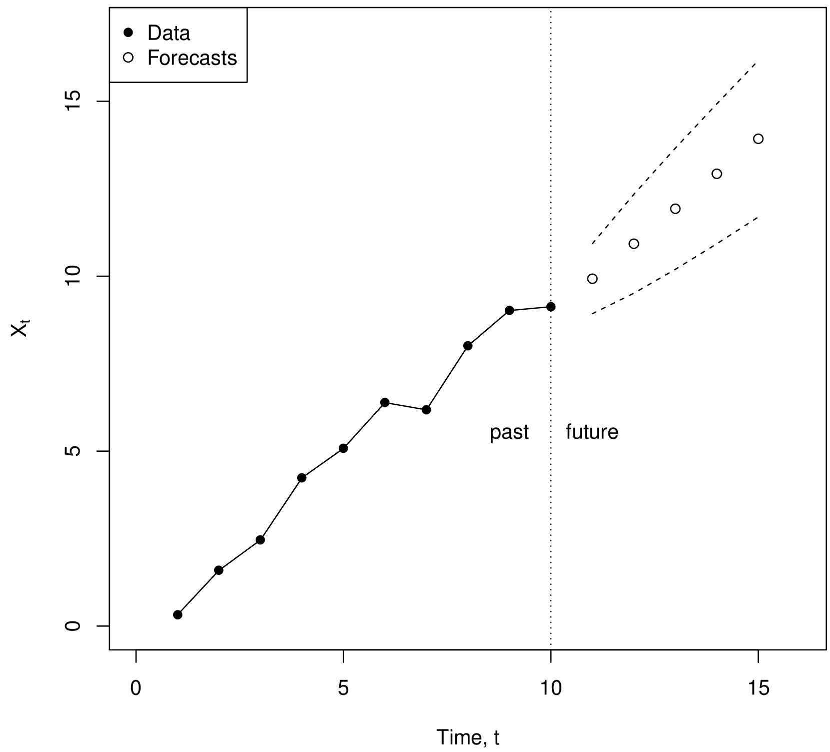
with thanks to Stuart Barber for previous years' slides and notes

March 2017

5.1 Introduction

Given data values $X_1, X_2, \ldots, X_n$, we want

- To predict a future value X_{n+l} which is l steps ahead, by a forecast $X_n(l)$ say, where l is the *lead time* of the forecast.
- To have confidence limits for $X_n(l)$.
- To ensure the *forecast error* $e_n(l) = X_{n+l} - X_n(l)$ is as small as possible.



5.2 Minimum mean square

Recall we want the *forecast error* $e_n(l) = X_{n+l} - X_n(l)$ to be small.

We might choose $X_n(l)$ to minimise

$$S = \mathbb{E}\left\{[X_{n+l} - X_n(l)]^2\right\}.$$

Suppose X_t can be written using a (possibly infinite) MA model

$$X_t = \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \beta_3\varepsilon_{t-3} + \cdots$$

Thus

$$\begin{aligned} X_{n+l} = \beta_0\varepsilon_{n+l} + \beta_1\varepsilon_{n+l-1} + \beta_2\varepsilon_{n+l-2} + \cdots + \beta_{l-1}\varepsilon_{n+1} \\ + \beta_l\varepsilon_n + \beta_{l+1}\varepsilon_{n-1} + \beta_{l+2}\varepsilon_{n-2} + \cdots . \end{aligned} \tag{1}$$

Let

$$X_n(l) = \beta'_0\varepsilon_n + \beta'_1\varepsilon_{n-1} + \beta'_2\varepsilon_{n-2} + \beta'_3\varepsilon_{n-3} + \cdots$$

The minimum mean square error forecast is thus

$$X_n(l) = \sum_{j=l}^{\infty} \beta_j\varepsilon_{n+l-j}. \tag{2}$$

Updating forecasts

In many situations an *updating forecast* is appropriate.

Given data up to time n, we would predict X_{n+l} by $X_n(l)$. One time step later, we now know X_{n+1} and would predict X_{n+l} by $X_{n+1}(l-1)$.

Then it is easy to show, with $\beta_0 = 1$, that

$$X_{n+1}(l-1) = X_n(l) + \beta_{l-1}[X_{n+1} - X_n(1)].$$

Estimating residuals

To implement the MMSE forecast (2), we need to estimate the $\{\varepsilon_t\}$ by the residuals of our fitted models.

For an AR(p) process $X_t = \sum_{k=1}^{p} \alpha_k X_{t-k} + \varepsilon_t$. The residuals satisfy

$$e_t = X_t - \sum_{k=1}^{p} \widehat{\alpha}_k X_{t-k} \qquad t \geqslant p.$$

For an MA(q) process $X_t = \varepsilon_t + \sum_{k=1}^{q} \beta_k \varepsilon_{t-k}$, assume $\varepsilon_0 = \cdots = \varepsilon_{1-q} = 0$ and compute the e_t recursively.

Forecast error

The forecast error of (2) is

$$e_n(l) = \beta_0 \varepsilon_{n+l} + \beta_1 \varepsilon_{n+l-1} + \cdots + \beta_{l-1} \varepsilon_{n+1}.$$

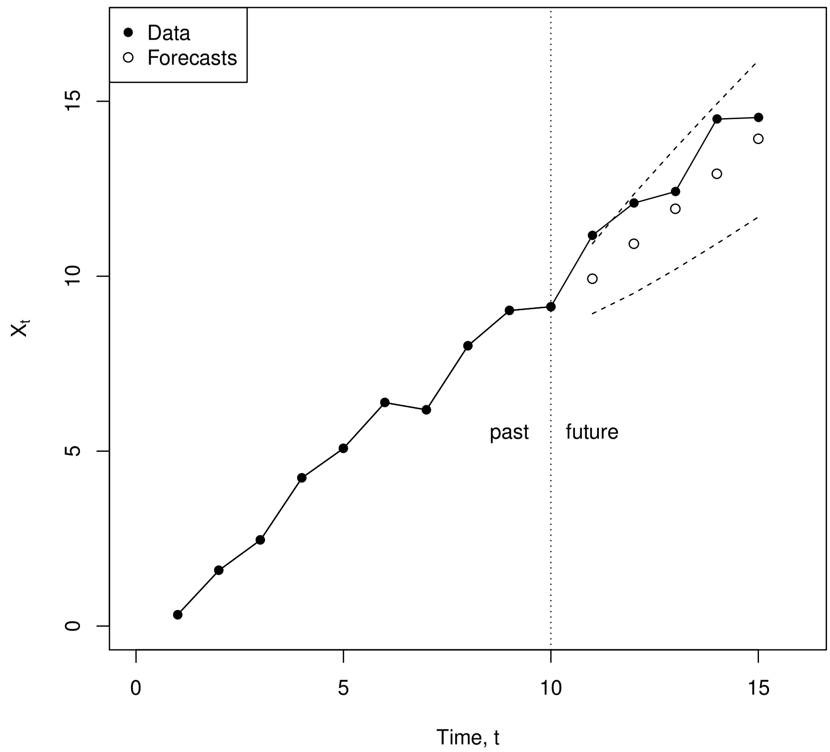
Note that

$$\mathbb{E}[e_n(l)] = 0 \qquad \text{and} \qquad \text{var}[e_n(l)] = \sigma_\varepsilon^2 \sum_{k=0}^{l-1} \beta_k^2.$$

Importantly, the forecast error is uncorrelated with the forecast itself and, for $m > 0$,

$$\text{cov}[e_n(l), e_n(l+m)] = \sigma_\varepsilon^2 \sum_{k=0}^{l-1} \beta_k \beta_{k+m}.$$

In practice, forecast errors tend to be on one side of actual future values.



Examples

Example 5.1 (One-step ahead forecasts) For example, consider the case $l = 1$. Assuming $\beta_0 = 1$,

$$X_n(1) = \beta_1\varepsilon_n + \beta_2\varepsilon_{n-1} + \beta_3\varepsilon_{n-2} + \cdots$$
$$e_n(1) = \varepsilon_{n+1}.$$

Example 5.2 (AR(1)) Let $X_t = \alpha X_{t-1} + \varepsilon_t$ with $|\alpha| < 1$ and $\{\varepsilon_t\}$ being a white noise process with variance σ_ε^2. Hence

$$X_n(l) = \alpha^l X_n,$$

and

$$\mathbb{E}(e_n(l)) = 0 \qquad \mathrm{var}(e_n(l)) \xrightarrow[l\to\infty]{} \gamma_0.$$

Conditional mean forecast

What is $\mathbb{E}[X_{n+l}|X_1,\ldots,X_n]$?

Given $\{\varepsilon_t : t \leqslant n\}$ we can in principle determine $\{X_t : t \leqslant n\}$ and *vice versa* so

$$\mathbb{E}[\varepsilon_t|X_1,\ldots,X_n] = \begin{cases} 0 & \text{if } t > n, \\ \varepsilon_t & \text{if } t \leqslant n. \end{cases}$$

From this, we get

$$\mathbb{E}[X_{n+l}|X_1,\ldots,X_n] = X_n(l)$$

and, assuming Gaussian white noise, the approximate 95% prediction interval limits

$$X_n(l) \pm 1.96\sqrt{(\beta_0^2 + \beta_1^2 + \cdots + \beta_{l-1}^2)\sigma_\varepsilon^2}.$$

5.3 Model based forecasting

An alternative approach: extend the model equation for time t after the last observation. Assume we have data X_t for times $t = 1,\ldots,n$.

Imagine we have fitted an AR(2) model assuming Gaussian white noise and obtained $\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\sigma}_\varepsilon^2$.

For time $n+1$, the model is

$$X_{n+1} = \alpha_1 X_n + \alpha_2 X_{n-1} + \varepsilon_{n+1},$$

So, an approximate 95% confidence interval for X_{n+1} is

$$\widehat{\alpha}_1 X_n + \widehat{\alpha}_2 X_{n-1} \pm 1.96\widehat{\sigma}_\varepsilon.$$

Iterating this procedure gives forecasts for longer times.

For example, consider the AR(1) process $X_t = \alpha X_{t-1} + \varepsilon_t$ with $\{\varepsilon_t\}$ being Gaussian white noise.

At time $n + l$,

$$\mathbb{E}(X_{n+l}|X_n) = \alpha^l X_n$$

and

$$\mathrm{var}(X_{n+l}|X_n) = (1 + \alpha^2 + \cdots + \alpha^{2l})\sigma_\varepsilon^2 \qquad .$$

For normally distributed errors, the model-based and minimum mean square error approaches will yield the same forecasts.

Iterating this procedure gives forecasts for longer times.

For example, consider the AR(1) process $X_t = \alpha X_{t-1} + \varepsilon_t$ with $\{\varepsilon_t\}$ being Gaussian white noise.

At time $n + l$,

$$\mathbb{E}(X_{n+l}|X_n) = \alpha^l X_n \xrightarrow[l \to \infty]{} 0$$

and

$$\mathrm{var}(X_{n+l}|X_n) = (1 + \alpha^2 + \cdots + \alpha^{2l})\sigma_\varepsilon^2 \quad \xrightarrow[l \to \infty]{} \quad \frac{1}{1-\alpha^2}\sigma_\varepsilon^2.$$

For long time forecasts, the mean and variance coincide with the values for the stationary process and the data do not provide any extra information.

For normally distributed errors, the model-based and minimum mean square error approaches will yield the same forecasts.