Times Series and Forecasting (IX)

Chapter 9. Forecasting

Jianhua Hu

hu.jianhua@mail.shufe.edu.cn

Spring 2013, SHUFE

9.1. Introduction

- In this course, we have discussed two general types of statistical models for time series data, namely,
  - deterministic trend models (Chapter 3) of the form

    $$Y_t = \mu_t + X_t,$$

    where $\{X_t\}$ is a zero mean stochastic process, and
  - ARIMA(p, d, q) models of the form

    $$\phi(B)(1 - B)^d Y_t = \theta(B)e_t.$$

- For both types of models, we have studied model specification, model fitting, and diagnostic procedures to assess model fit.

Forecasting problem

- We now switch our attention to forecasting.
- Suppose that we have a sample of process values up until time t, say, $Y_1, Y_2, \cdots, Y_t$.
- Forecasting refers to the technique of predicting the future values

$$Y_{t+1}, Y_{t+2}, \cdots.$$

Notation for forecasting

- Let Y_{t+l} be the value of the process at time $t + l$, where $l \geq 1$.
- In this context, we refer to
  - t as the forecast origin,
  - l as the lead time and
  - the value Y_{t+l} as "l steps ahead" of the most recently observed value Y_t.

Forecasting method

- We need a formal strategy to determine model forecasts.
- Given a sample $Y_1, \cdots, Y_t$, we would like to use a statistic $h(Y_1, \cdots, Y_t)$, written as $Y_t(l)$, to predict Y_{t+l}.
- The mean squared error of prediction is defined by

$$\mathrm{MSEP}(Y_t(l)) = \mathrm{E}\left\{(Y_{t+l} - Y_t(l))^2\right\}.$$

- Write

$$\hat{Y}_t(l) = \mathrm{argmin}_{Y_t(l) \in \mathbb{R}} \mathrm{E}\left\{(Y_{t+l} - Y_t(l))^2\right\}.$$

- The $\widehat{Y}_t(l)$ is said to be the minimum mean squared error (MMSE) forecast.

Forecasting method

- So, our prediction problem is to choose a $\hat{Y}_t(l)$ such that MSEP attains the minimum at $Y_t(l) = \hat{Y}_t(l)$.

- The optimal solution is

$$\hat{Y}_t(l) = \mathsf{E}\left(Y_{t+l} | Y_1, \cdots, Y_t\right),$$

  the conditional expectation of Y_{t+l}, given the observed data $Y_1, \cdots, Y_t$.

9.2. Deterministic trend forecasting

- Consider the model

$$Y_t = \mu_t + X_t,$$

where

  - μ_t is a deterministic (non-random) trend function
  - $\{X_t\}$ is a white noise process with $\mathsf{E}(X_t) = 0$ and $\mathsf{var}(X_t) = \gamma_0$ (constant), for all t.

9.2. Deterministic trend forecasting

- The MMSE forecast is given by direct calculation,

$$\begin{aligned}\widehat{Y}_t(l) &= \mathsf{E}\left(Y_{t+l}|Y_1, Y_2, \cdots, Y_t\right)\\ &= \mathsf{E}\left(\mu_{t+l} + X_{t+l}|Y_1, Y_2, \cdots, Y_t\right)\\ &= \mathsf{E}\left(\mu_{t+l}|Y_1, Y_2, \cdots, Y_t\right) + \mathsf{E}\left(X_{t+l}|Y_1, Y_2, \cdots, Y_t\right)\\ &= \mu_{t+l} + \mathsf{E}\left(X_{t+l}\right)\\ &= \mu_{t+l}.\end{aligned}$$

- Namely, μ_{t+l} is the MMSE forecast.

9.2. Deterministic trend forecasting

- For example, if $\mu_t = \beta_0 + \beta_1 t$, a linear trend model, then

$$\widehat{Y}_t(l) = \mu_{t+l} = \beta_0 + \beta_1(t+l).$$

- If $\mu_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$, a cosine trend model, then

$$\widehat{Y}_t(l) = \mu_{t+1} = \beta_0 + \beta_1 \cos[2\pi f(t+l)] + \beta_2 \sin[2\pi f(t+l)].$$

9.2. Deterministic trend forecasting

- Of course, MMSE forecasts must be estimated. For example, $\widehat{Y}_t(l) = \beta_0 + \beta_1(t+l)$ is estimated by

$$\widehat{\mu}_{t+l} = \widehat{\beta}_0 + \widehat{\beta}_1(t+l),$$

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the lse of β_0 and β_1, respectively.

- Also, $\widehat{Y}_t(l) = \beta_0 + \beta_1 \cos[2\pi f(t+l)] + \beta_2 \sin[2\pi f(t+l)]$ would be estimated by

$$\widehat{\mu}_{t+l} = \widehat{\beta}_0 + \widehat{\beta}_1 \cos[2\pi f(t+l)] + \widehat{\beta}_2 \sin[2\pi f(t+l)],$$

where $\widehat{\beta}_0$, $\widehat{\beta}_1$, and $\widehat{\beta}_2$ are the lse.

Figure 9.1. (a) the global temperature data and (b) beer sales data

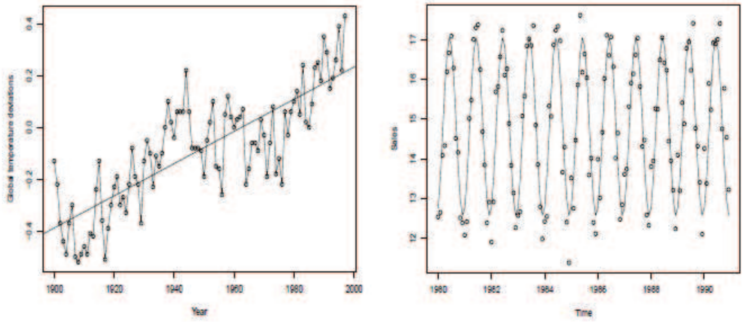


Figure 9.1: *Left: Global temperature data with straight line fit. Right: Beer sales data with cosine trend fit.*

Example 9.1(a). Forecasting for the global temperature data

Example

In Example 3.4, we fit a straight line trend model to the global temperature deviation data. The fitted model is

$$\widehat{Y}_t = -12.19 + 0.0062t,$$

where $t = 1900, \cdots, 1997$, depicted visually in Figure 9.1 (left). In 1997, we could have used the model to predict for 1998,

$$\widehat{\mu}_{1998} = \widehat{\mu}_{1997+1} = -12.19 + 0.0062(1997 + 1) \approx 0.198.$$

Example 9.1(a). Forecasting for the global data (con't)

- For 2005,

$$\widehat{\mu}_{2005} = \widehat{\mu}_{1997+8} = -12.19 + 0.0062(1997 + 8) \approx 0.241.$$

- For 2020,

$$\widehat{\mu}_{2020} = \widehat{\mu}_{1997+23} = -12.19 + 0.0062(1997 + 23) \approx 0.334.$$

- Predictions continue to increase linearly since $\widehat{\beta}_1 > 0$.

Example 9.1(b). Forecasting for the beer sales data

- In Example 3.6, we fit a cosine trend model to the monthly US beer sales data, which produced the fitted model

$$\widehat{Y}_t = 14.8 - 2.04\cos(2\pi t) + 0.93\sin(2\pi t),$$

where $t = 1980, 1980.083, 1980.166, \cdots, 1990.916$ (that is, the first value was in January, 1980 and the last value was in December, 1990).

Example 9.1(b). Forecasting for the beer sales data (con't)

- In December, 1990, we could have used the model to predict for January, 1991,

$$\widehat{Y}_{1991} = 14.8 - 2.04\mathsf{cos}[2\pi(1991)] + 0.93\mathsf{sin}[2\pi(1991)] \approx 12.76.$$

- For June, 1992,

$$\begin{aligned}\widehat{Y}_{1992.416} &= 14.8 - 2.04\mathsf{cos}[2\pi(1992.416)] + 0.93\mathsf{sin}[2\pi(1992.416)] \\ &\approx 17.03.\end{aligned}$$

Comments for the deterministic trend forecasting

- One of the drawbacks with this model is that predictions are based only on the deterministic trend fit.
- Therefore, the analyst is basically assuming that the fitted trend is applicable indefinitely into the future.
- That is, the forecast for Y_{t+l} ignores any correlation between Y_{t+l} and $Y_1, Y_2, \cdots, Y_t$.

Properties of deterministic trend forecasting

- For deterministic trend models of the form $Y_t = \mu_t + X_t$, where $\mathsf{E}(X_t) = 0$ and $\mathsf{var}(X_t) = \gamma_0$, the forecast error, with lead time l, is

$$e_t(l) = Y_{t+l} - \widehat{Y}_t(l) = \mu_{t+l} + X_{t+l} - \mu_{t+l} = X_{t+l}.$$

- For all $l \geq 1$,

$$\mathsf{E}[e_t(l)] = \mathsf{E}(X_{t+l}) = 0 \quad \mathsf{var}[e_t(l)] = \mathsf{var}(X_{t+l}) = \gamma_0.$$

- Forecasts are unbiased and the forecast error variance is constant for all lead times l.
- These facts will be useful when forming prediction intervals for future values.

9.3. ARIMA forecasting

- We now focus on forecasting methods for ARIMA models.
- Recall that an ARIMA(p, d, q) process can be written as

$$\phi(B)(1-B)^d Y_t = \theta_0 + \theta(B)e_t.$$

- We first focus on stationary ARMA(p, q) models; i.e., ARIMA(p, d, q) models with $d = 0$.

9.3.1. AR(1) forecasting

- Within the stationary ARMA(p, q) class, the most logical place to start is the AR(1) model, that is,

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t,$$

where $-1 < \phi < 1$ (stationarity condition), $\mu = \mathsf{E}(Y_t)$.

AR(1) forecasting

- The MMSE forecast of Y_{t+1} is

$$\begin{aligned}\widehat{Y}_t(1) &= \mathsf{E}\left(Y_{t+1}|Y_1, Y_2, \cdots, Y_t\right) \\ &= \mathsf{E}\left\{[\mu + \phi(Y_t - \mu) + e_{t+1}|Y_1, Y_2, \cdots, Y_t\right\} \\ &= \mu + \phi\mathsf{E}\left[(Y_t - \mu)|Y_1, \cdots, Y_t\right] + \mathsf{E}\left(e_{t+1}|Y_1, \cdots, Y_t\right) \\ &= \mu + \phi(Y_t - \mu).\end{aligned}$$

- The MMSE forecast of Y_{t+l}, for all lead times $l \geq 1$, is

$$\widehat{Y}_t(l) = \mu + \phi^l(Y_t - \mu).$$

- In practice, μ and ϕ have to be estimated.

AR(1) forecasting

- Forecasts will "converge" to the overall mean μ as the lead time l increases because $-1 < \phi < 1$.
- The one-step-ahead forecast error is

$$\begin{aligned} e_t(1) &= Y_{t+1} - \widehat{Y}_t(1) \\ &= \mu + \phi(Y_t - \mu) + e_{t+1} - [\mu + \phi(Y_t - \mu)] \\ &= e_{t+1}. \end{aligned}$$

- Forecasts are unbiased due to $\mathsf{E}[e_t(1)] = 0$, and

$$\mathsf{var}[e_t(1)] = \mathsf{var}(e_{t+1}) = \sigma_e^2.$$

AR(1) forecasting

- In general, the l-steps-ahead forecast error is defined to be

$$e_t(l) = Y_{t+l} - \widehat{Y}_t(l).$$

- Forecasts are unbiased, i.e., $\mathsf{E}[e_t(l)] = 0$, and

$$\mathsf{var}[e_t(l)] = \sigma_e^2 \left(\frac{1-\phi^{2l}}{1-\phi^2} \right),$$

  for any $l \geq 1$. See pp 196 (CC).

- When l is large, since $\phi^{2l} \approx 0$,

$$\mathsf{var}[e_t(l)] \approx \frac{\sigma_e^2}{1-\phi^2} = \gamma_0 = \mathsf{var}(Y_t).$$

AR(2) forecasting

- The AR(2) model is,

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + e_t,$$

  where (ϕ_1, ϕ_2) belongs to the stationarity region in the ϕ_1, ϕ_2 plan, $\mu = \mathsf{E}(Y_t)$.

AR(2) forecasting

- The MMSE forecast of Y_{t+l}, for all lead times $l \geq 3$, is

$$\widehat{Y}_t(l) = \hat{\mu} + \hat{\phi}_1(\widehat{Y}_t(l-1) - \hat{\mu}) + \hat{\phi}_2(\widehat{Y}_t(l-2) - \hat{\mu}).$$

  with

  - $\widehat{Y}_t(1) = \hat{\mu} + \hat{\phi}_1(Y_t - \hat{\mu}) + \hat{\phi}_2(Y_{t-1} - \hat{\mu})$ and
  - $\widehat{Y}_t(2) = \hat{\mu} + (\hat{\phi}_1^2 + \hat{\phi}_2)(Y_t - \hat{\mu}) + \hat{\phi}_1\hat{\phi}_2(Y_{t-1} - \hat{\mu})$.

9.3.2. MA(1) forecasting

- Consider the invertible MA(1) process

$$Y_t = \mu + e_t - \theta e_{t-1},$$

where $-1 < \theta < 1$ (invertibility condition), $\mu = \mathsf{E}(Y_t)$.

- The MMSE prediction for Y_{t+1} is given by

$$\begin{aligned}
\widehat{Y}_t(1) &= \mathsf{E}\left(Y_{t+1}|Y_1, Y_2, \cdots, Y_t\right) \\
&= \mathsf{E}\left[(\mu + e_{t+1} - \theta e_t)|Y_1, Y_2, \cdots, Y_t\right] \\
&= \mu + \mathsf{E}\left(e_{t+1}|Y_1, Y_2, \cdots, Y_t\right) - \theta \mathsf{E}\left(e_t|Y_1, Y_2, \cdots, Y_t\right) \\
&= \mu - \theta e_t,
\end{aligned}$$

MA(1) forecasting

- The variance of the one-step-ahead prediction error $e_t(1)$ is given by

$$\begin{aligned}\text{var}[e_t(1)] &= \text{var}[Y_{t+1} - \widehat{Y}_t(1)] \\ &= \text{var}[(\mu + e_{t+1} - \theta e_t) - (\mu - \theta e_t)] \\ &= \text{var}(e_{t+1}) = \sigma_e^2,\end{aligned}$$

the white noise process variance.

MA(1) forecasting

- The MMSE prediction for Y_{t+l}, $l > 1$, is given by

$$\begin{aligned}
\widehat{Y}_t(l) &= \mathsf{E}\left(Y_{t+l}|Y_1, Y_2, \cdots, Y_t\right) \\
&= \mathsf{E}\left[(\mu + e_{t+l} - \theta e_{t+l-1})|Y_1, Y_2, \cdots, Y_t\right] \\
&= \mu + \mathsf{E}\left(e_{t+l}|Y_1, Y_2, \cdots, Y_t\right) - \theta \mathsf{E}\left(e_{t+l-1}|Y_1, Y_2, \cdots, Y_t\right) \\
&= \mu,
\end{aligned}$$

- Therefore, we have shown that, for the MA(1) model,

$$\widehat{Y}_t(l) = \begin{cases} \mu - \theta e_t, & l = 1 \\ \mu, & l > 1. \end{cases}$$

MA(1) forecasting

- The key feature of the MA(1) process is that observations one unit apart in time are correlated, whereas observations $l > 1$ units apart in time are not.
- For $l > 1$, there is no (auto)correlation to exploit in making a prediction; this is why a constant mean prediction is made.

MA(1) forecasting

- Finally, the variance of the l-steps-ahead prediction error $e_t(l)$, for $l > 1$, is given by

$$\begin{aligned} \mathsf{var}[e_t(l)] &= \mathsf{var}[Y_{t+l} - \widehat{Y}_t(l)] = \mathsf{var}[(\mu + e_{t+l} - \theta e_{t+l-1}) - \mu] \\ &= \mathsf{var}(e_{t+l} - \theta e_{t+l-1}) = \sigma_e^2(1+\theta^2) = \gamma_0 = \mathsf{var}(Y_t). \end{aligned}$$

- Summarizing,

$$\mathsf{var}[e_t(l)] = \begin{cases} \sigma_e^2, & l = 1 \\ \sigma_e^2(1+\theta^2), & l > 1. \end{cases}$$

MA(q) forecasting

- Analogously to the MA(1) model, MMSE forecasts for the MA(q) process

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q},$$

are taken to be $\widehat{Y}_t(l) = \mu$, for all $l > q$.

9.3.3. ARMA(p, q)

- Consider the general ARMA(p, q) process

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q},$$

- MMSE forecasts are computed using the difference equation

$$\begin{aligned}\widehat{Y}_t(l) = \theta_0 &+ \phi_1 \widehat{Y}_t(l-1) + \phi_2 \widehat{Y}_t(l-2) + \cdots + \phi_p \widehat{Y}_t(l-p) \\ &- \theta_1 \mathsf{E}(e_{t+l-1} | Y_1, \cdots, Y_t) - \cdots - \theta_q \mathsf{E}(e_{t+l-q} | Y_1, \cdots, Y_t).\end{aligned}$$

ARMA(p, q)

- In this expression, for $j = 1, 2, \cdots, p$,

$$\widehat{Y}_t(l-j) = \mathsf{E}(Y_{t+l-j}|Y_1, Y_2, \cdots, Y_t),$$

- and for $k = 1, 2, \cdots, q$,

$$\mathsf{E}(e_{t+l-k}|Y_1, \cdots, Y_t) = \begin{cases} 0; & l-k > 0 \\ e_{t+l-k}, & l-k \leq 0 \end{cases}.$$

Remarks for ARMA(p, q) forecasting

- Computing $\widehat{Y}_t(l)$, in general, involves approximating conditional expectations $\mathsf{E}(Y_{t+l-k}|Y_1, Y_2, \cdots, Y_t)$, when $l - k \leq 0$, using infinite AR representations for invertible models (see pp 80, CC).
- This is only necessary for MMSE forecasts at early lags when q is larger than 1.
- Also, recursive formulas can be derived to compute $\mathsf{E}(Y_{t+l-j}|Y_1, Y_2, \cdots, Y_t)$, for $j = 1, 2, \cdots, p$.

ARMA(p, q)

- For example, consider the ARMA(1,1) process

$$Y_t = \theta_0 + \phi Y_{t-1} + e_t - \theta e_{t-1}.$$

- For $l = 1$, we have

$$\begin{aligned}\widehat{Y}_t(1) &= \mathsf{E}(Y_{t+1}|Y_1, \cdots, Y_t)\\ &= \mathsf{E}[(\theta_0 + \phi Y_t + e_{t+1} - \theta e_t)|Y_1, \cdots, Y_t]\\ &= \theta_0 + \phi Y_t - \theta e_t.\end{aligned}$$

- For $l = 2$, we have

$$\begin{aligned}\widehat{Y}_t(2) &= \mathsf{E}(Y_{t+2}|Y_1, \cdots, Y_t)\\ &= \mathsf{E}[(\theta_0 + \phi Y_{t+1} + e_{t+2} - \theta e_{t+1})|Y_1, \cdots, Y_t]\\ &= \theta_0 + \phi \widehat{Y}_t(1).\end{aligned}$$

Remarks for ARMA(p, q) forecasting

- Straightforward calculation shows that, for $l > 1$,

$$\widehat{Y}_t(l) = \theta_0 + \phi \widehat{Y}_t(l-1).$$

- Reparameterizing, the last expression can be written as, when $l > 1$,

$$\widehat{Y}_t(l) = \mu + \phi^l (Y_t - \mu) - \phi^{l-1} e_t.$$

- When l is large, $\phi^l (Y_t - \mu)$ and $\phi^{l-1} e_t$ are close to zero, and, hence, $\widehat{Y}_t(l) \approx \mu$ for large lead times l.

Remarks for ARMA(p, q) forecasting

- For $l \leq q$, MMSE forecasts will depend on both the AR and MA parts of the model.
- When $l > q$, the MA contributions vanish and forecasts will depend solely on the recursion identified in the AR part. That is, when $l > q$,

$$\widehat{Y}_t(l) = \theta_0 + \phi_1 \widehat{Y}_t(l-1) + \phi_2 \widehat{Y}_t(l-2) + \cdots + \phi_p \widehat{Y}_t(l-p).$$

- The last expression can be written equivalently, for $l > q$, as

$$\widehat{Y}_t(l) = \mu + \phi_1[\widehat{Y}_t(l-1) - \mu] + \cdots + \phi_p[\widehat{Y}_t(l-p) - \mu].$$

Remarks for ARMA(p, q) forecasting

- As a function of l, $\widehat{Y}_t(l) - \mu$ follows the same Yule-Walker recursion as the ACF ρ_k.
- The roots of $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p$ determine the behavior of $\widehat{Y}_t(l) - \mu$, when $l > q$.

Remarks for ARMA(p, q) forecasting

- For any stationary ARMA(p, q) process,

$$\lim_{l \to \infty} \widehat{Y}_t(l) = \mu,$$

  where $\mu = \mathsf{E}(Y_t)$.
- For large lead times l, MMSE predictions will be approximately equal to the process mean.
- This is an expected result, because "the dependence dies out as the time span between observations increases, and this dependence is the only reason we can improve on the naive forecast of using μ alone."

Remarks for ARMA(p, q) forecasting

- For any stationary ARMA(p, q) process, the l-steps-ahead forecast error is

$$e_t(l) = Y_{t+l} - \widehat{Y}_t(l).$$

- The variance of this forecast error, in general, can be obtained using the expressions on pp 201 (CC).
- Importantly, for any stationary ARMA(p, q) process,

$$\lim_{l \to \infty} \text{var}[e_t(l)] = \gamma_0,$$

  where $\gamma_0 = \text{var}(Y_t)$.
- For large lead times l, the variance of our prediction error will be close to the process variance.

9.3.4. Nonstationary models

- For invertible ARIMA(p, d, q) processes with $d \geq 1$, MMSE forecasts are computed using the same approach as before.
- Consider the ARIMA(p,1,q) model

$$\phi(B)(1 - B)Y_t = \theta(B)e_t,$$

where $(1 - B)Y_t = \nabla Y_t$ is the series of first differences.

Nonstationary models

- Note that

$$\begin{aligned}\phi(B)(1-B) &= \left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(1-B)\\ &= 1-(1+\phi_1)B-(\phi_2-\phi_1)B^2-\cdots-(\phi_p-\phi_{p-1})B^p+\phi_p B^{p+1}\\ &= \phi^*(B)\end{aligned}$$

- The ARIMA$(p,1,q)$ model

$$\phi(B)(1-B)Y_t=\theta(B)e_t$$

can be rewritten as

$$\phi^*(B)Y_t=\theta(B)e_t,$$

a nonstationary ARMA$(p+1,q)$ process.

Nonstationary models

- We determine MMSE forecasts the same way as in the stationary case.
- For example, we have a (nonstationary) ARIMA(1,1,1) process

$$(1-\phi B)(1-B)Y_t = (1-\theta B)e_t,$$

- Equivalently,

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + e_t - \theta e_{t-1},$$

is a nonstationary ARMA(2,1) process.

Nonstationary models

- If $l = 1$, then

$$\begin{aligned}\widehat{Y}_t(1) &= \mathsf{E}\left(Y_{t+1}|Y_1, Y_2, \cdots, Y_t\right)\\ &= \mathsf{E}\left\{\left[(1+\phi)Y_t - \phi Y_{t-1} + e_{t+1} - \theta e_t\right]|Y_1, Y_2, \cdots, Y_t\right\}\\ &= (1+\phi)Y_t - \phi Y_{t-1} - \theta e_t.\end{aligned}$$

- If $l = 2$, then

$$\begin{aligned}\widehat{Y}_t(2) &= \mathsf{E}\left(Y_{t+2}|Y_1, Y_2, \cdots, Y_t\right)\\ &= \mathsf{E}\left\{\left[(1+\phi)Y_{t+1} - \phi Y_t + e_{t+2} - \theta e_{t+1}\right]|Y_1, \cdots, Y_t\right\}\\ &= (1+\phi)\widehat{Y}_t(1) - \phi Y_t.\end{aligned}$$

Nonstationary models

- For $l > 2$,

$$
\begin{aligned}
\widehat{Y}_t(l) &= \mathsf{E}\left(Y_{t+l}|Y_1, \cdots, Y_t\right) \\
&= \mathsf{E}\left\{\left[(1+\phi)Y_{t+l-1} - \phi Y_{t+l-2} + e_{t+l} - \theta e_{t+l-1}\right]|Y_1, \cdots, Y_t\right\} \\
&= (1+\phi)\widehat{Y}_t(l-1) - \phi\widehat{Y}_t(l-2).
\end{aligned}
$$

- Writing recursive expressions MMSE forecasts with any invertible ARIMA(p, d, q) model can be done similarly.

Nonstationary models

- The l-steps ahead forecast error $e_t(l)$ for any invertible ARIMA(p, d, q) model has the following characteristics:

$$\mathsf{E}[e_t(l)] = 0 \quad \mathsf{var}[e_t(l)] = \sigma_e^2 \sum_{j=0}^{l-1} \Psi_j^2,$$

  where the Ψ weights correspond to those in the truncated linear process representation on pp 200 (CC).

- MMSE ARIMA forecasts are unbiased.

Nonstationary models

- For nonstationary models, the Ψ weights do not "die out" like they do ("die out") with stationary models.
- For nonstationary models, the variance of the forecast error increases as l does. This would be expected because the distant future of a nonstationary process is hard to estimate precisely.

9.4. Prediction intervals

- We have already demonstrated that MMSE forecasts in ARIMA models are unbiased, and we have quantified the variation in the prediction error.
- The next step is to form prediction intervals for a future response Y_{t+l}.

9.4.1. Prediction intervals of deterministic trend models

- A $100(1-\alpha)$ percent prediction interval for Y_{t+l} is an interval $\left(\widehat{Y}_{t+l}^{(L)}, \widehat{Y}_{t+l}^{(U)}\right)$ which satisfies

$$Pr\left(\widehat{Y}_{t+l}^{(L)} < Y_{t+l} < \widehat{Y}_{t+l}^{(U)}\right) = 1-\alpha.$$

Prediction intervals of deterministic trend models

- Consider the model

$$Y_t = \mu_t + X_t,$$

where μ_t is a deterministic (non-random) trend function and where $X_t \sim N(0, \gamma_0)$.

- We have already shown the following:

$$\widehat{Y}_t(l) = \mu_{t+l}, \quad \mathsf{E}[e_t(l)] = 0 \quad \text{and} \quad \mathsf{var}[e_t(l)] = \gamma_0,$$

where $e_t(l) = Y_{t+l} - \widehat{Y}_t(l)$ is the l-steps ahead prediction error.

Prediction intervals of deterministic trend models

- Under normality, the random variable

$$Z = \frac{Y_{t+l} - \widehat{Y}_t(l)}{\sqrt{\mathsf{var}[e_t(l)]}} = \frac{Y_{t+l} - \widehat{Y}_t(l)}{\mathsf{se}[e_t(l)]} = \frac{X_{t+l}}{\sqrt{\gamma_0}} \sim N(0,1).$$

- Therefore, Z is a pivotal quantity and

$$Pr\left(-z_{\alpha/2} < \frac{Y_{t+l} - \widehat{Y}_t(l)}{\mathsf{se}[e_t(l)]} < z_{\alpha/2}\right) = 1 - \alpha.$$

Prediction intervals of deterministic trend models

- Rearranging the event inside the probability symbol, we have

$$Pr\left(\widehat{Y}_t(l) - z_{\alpha/2}\mathsf{se}[e_t(l)] < Y_{t+l} < \widehat{Y}_t(l) + z_{\alpha/2}\mathsf{se}[e_t(l)]\right) = 1-\alpha,$$

  which shows that

$$\left(\widehat{Y}_t(l) - z_{\alpha/2}\mathsf{se}[e_t(l)], \widehat{Y}_t(l) + z_{\alpha/2}\mathsf{se}[e_t(l)]\right)$$

  is a $100(1-\alpha)$ percent prediction interval for Y_{t+l}.

- Of course, $\widehat{Y}_t(l)$ and se$[e_t(l)]$ must be estimated.

Remarks for the prediction intervals

- Importantly, assume that X_t is normally distributed with constant variance in obtaining prediction intervals from deterministic trend models.
- This may or may not be true in practice, but the validity of our predictive inference requires it to be true.
- Since

$$z_{\alpha/2}\text{se}[e_t(l)] = z_{\alpha/2}\sqrt{\gamma_0}$$

  is free of the lead time l, prediction intervals will have the same width indefinitely into the future.
- If a different zero mean model is assumed for $\{X_t\}$, such as an AR(1), then this will no longer be true, especially at early lags.

9.4.2. Prediction intervals of ARIMA models

- For any model in the ARIMA(p, d, q) family, assume that the white noise terms $\{e_t\}$ are normally distributed, then

$$Z = \frac{Y_{t+l} - \widehat{Y}_t(l)}{\text{var}[e_t(l)]} = \frac{Y_{t+l} - \widehat{Y}_t(l)}{\text{se}[e_t(l)]} \sim N(0, 1),$$

which implies that

$$\left(\widehat{Y}_t(l) - z_{\alpha/2}\text{se}[e_t(l)], \widehat{Y}_t(l) + z_{\alpha/2}\text{se}[e_t(l)]\right)$$

is a $100(1 - \alpha)$ percent prediction interval for Y_{t+l}.

Remarks for the prediction intervals

- In practice, SAS gives the (estimated) MMSE forecasts and standard errors, that is, estimates of $\widehat{Y}_t(l)$ and se$[e_t(l)]$, so computing prediction intervals with ARIMA(p, d, q) models is almost automatic.

- It is again important to emphasize that normality is assumed.

9.5. Forecasting transformed series: nonstationary model predictions

- Forecasting future values from nonstationary ARIMA processes (i.e., with $d \geq 1$) poses no additional methodological challenges when compared to stationary ARMA processes.

- It is easy to take this fact for granted, because, as we have already seen, SAS automates the entire forecasting process for stationary and nonstationary models.

9.5.1. Nonstationary model predictions

- Consider forecasting observations from an IMA(1,1) model

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1}.$$

- The one-step ahead MMSE forecast is

$$\begin{aligned}\widehat{Y}_t(1) &= \mathsf{E}(Y_{t+1}|Y_1, Y_2, \cdots, Y_t) \\ &= \mathsf{E}[(Y_t + e_{t+1} - \theta e_t)|Y_1, Y_2, \cdots, Y_t] = Y_t - \theta e_t.\end{aligned}$$

- For $l > 1$, the MMSE forecast is

$$\begin{aligned}\widehat{Y}_t(l) &= \mathsf{E}(Y_{t+l}|Y_1, Y_2, \cdots, Y_t) \\ &= \mathsf{E}[(Y_{t+l-1} + e_{t+l} - \theta e_{t+l-1})|Y_1, Y_2, \cdots, Y_t] = \widehat{Y}_t(l-1).\end{aligned}$$

- $W_t = \nabla Y_t = Y_t - Y_{t-1}$ follows an MA(1) model; i.e.,

$$W_t = e_t - \theta e_{t-1} \quad \text{with } \mu = 0.$$

9.5.1. Nonstationary model predictions (con't)

- We have already shown that for an MA(1) process with $\mu = 0$,

$$\widehat{W}_t(l) = \begin{cases} -\theta e_t, & l = 1 \\ 0, & l > 1. \end{cases}$$

- When $l = 1$, note that

$$\widehat{W}_t(1) = -\theta e_t = Y_t - \theta e_t - Y_t = \widehat{Y}(1) - Y_t,$$

and, when $l > 1$, note that

$$\widehat{W}_t(l) = 0 = \widehat{Y}_t(l) - \widehat{Y}_t(l-1).$$

9.5.1. Nonstationary model predictions (con't)

- We have shown that
  - forecasting the original (nonstationary) series Y_t
  - forecasting the stationary differenced series W_t and then summing to obtain the forecast in original terms

  are identical procedures in the IMA(1,1) case.

- This finding extends to any ARIMA(p, d, q) model.

9.5.1. Nonstationary model predictions (con't)

- The analyst can calculate predictions from the nonstationary model for Y_t or the stationary model $W_t = \nabla^d Y_t$.
- The predictions in both cases will be equal (hence, the resulting standard errors will be the same too).

Example 9.1(a). Forecasting for the global temperature data

Example

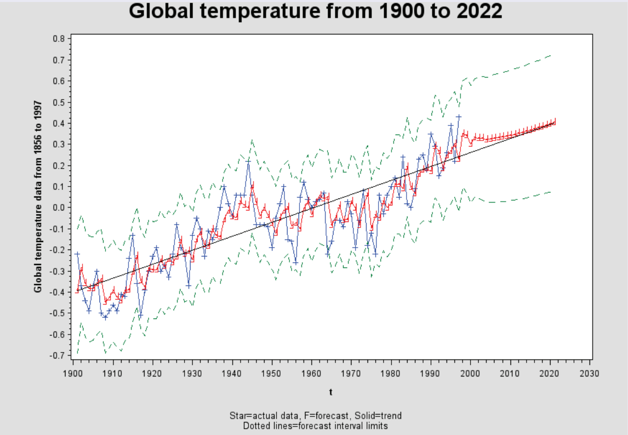
We refit the global temperature deviation data (1901-1997). The fitted model via proc autoreg procedure via ML is

$$\widehat{Y}_t = -12.9731 + 0.006617t + X_t,$$
$$X_t = e_t + 0.4921X_{t-1} + 0.2342X_{t-4}$$

where $t = 1901, \cdots, 1997$ and $\hat{\sigma}^2 = 0.01095$.

For the global temperature data, please see the forecasts in sas output in sas programm Example91_temperature.

Figure 9.2. Figure of temperature data



Example 9.2. The Lake Huron elevation data

Example

In Example 7.2, we examined the Lake Huron elevation data (from 1875- 1972) and we used an AR(2) process with a linear trend to model them. The fit using maximum likelihood was

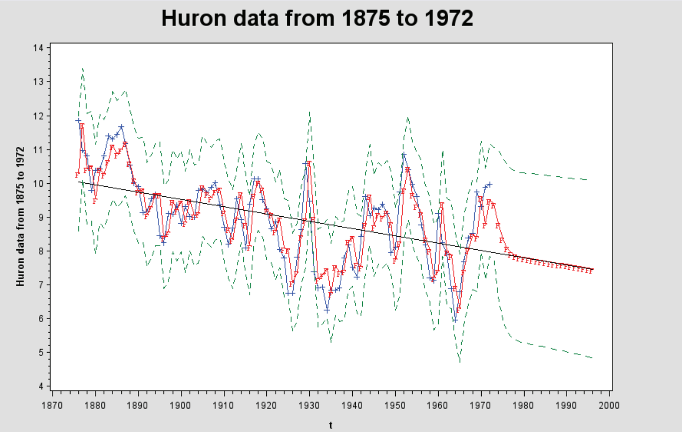
$$Y_t = 50.5109 - 0.0216t + X_t, \quad X_t = 1.0048X_{t-1} - 0.2913X_{t-2} + e_t,$$

so that $\hat{\sigma}_e^2 = 0.47605.$

Example 9.2. The Lake Huron elevation data (con't)

- SAS can provide forecasts and estimated standard errors of the prediction error for any ARIMA(p, d, q) model fit.
- For the Lake Huron data, please see the forecasts in sas output in sas programm Example92_LakeHuron.

Figure 9.3.


Huron data from 1875 to 1972
Huron data from 1875 to 1972
14
13
12
11
10
9
8
7
6
5
4
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
t

9.5.2. Forecasting log transformed series

- In Chapter 5, we discussed the Box-Cox family of transformations

$$T(Y_t) = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(Y_t), & \lambda = 0, \end{cases}$$

  where λ is called the transformation parameter.
- Many time series data Y_t exhibit nonconstant variability which can be stabilized by taking logarithms. We focus on this transformation explicitly.

Forecasting log transformed series

- The log function $T(x) = \log x$ is not a linear function, so transformations on the log-scale can not simply be "undone" as easily as with differenced series (differencing is a linear transformation).

- For notational purposes, set $Z_t = \log Y_t$, and suppose that the MMSE forecast for Z_{t+l} is denoted by $\widehat{Z}_t(l)$.

Forecasting log transformed series

- The argument on pp 210 (CC) shows that the corresponding MMSE forecast for Y_{t+l} is

$$\widehat{y}_t(l) = \exp\left\{\widehat{Z}_t(l) + \frac{1}{2}\text{var}[e_t(l)]\right\},$$

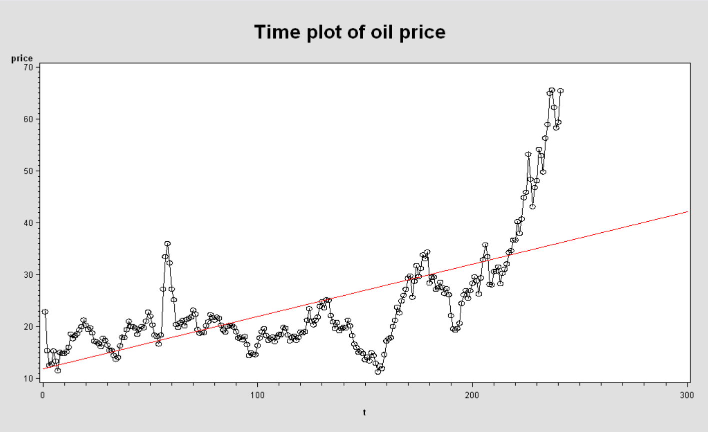
where $\text{var}[e_t(l)]$ is the variance of the l-steps ahead forecast error $e_t(l) = Z_{t+l} - \widehat{Z}_t(l)$.

Example 9.3. Crude oil price data

Example

The data in Figure 9.3 are monthly spot prices for crude oil (measured in U.S. dollars per barrel) from Cushing, OK, from $1/86$ to $1/06$, along with the MMSE forecasts and predictions limits for $l = 1, 2, \cdots, 12$ (i.e., for $2/06$ to $1/07$), based on an IMA(1,1) fit for $\log Y_t$. This model was fit in Example 7.4.

Figure 9.4. Crude oil price data



Example 9.3. Crude oil price forecasting (con't)

- The first fitted model is as follows

$$logY_t = logY_{t-1} + e_t + 0.29372e_{t-1},$$

- The forecasts and estimated standard errors (on the log scale) are given in the predict output below:

	Jan	Feb	Mar	Apr	May	Jun
forecast	4.1817	4.2108	4.2152	4.2195	4.2239	4.2283
se		0.0819	0.1339	0.1708	0.2010	0.2272

Figure 9.5. Crude oil logprice forecasting via 1st fitted model

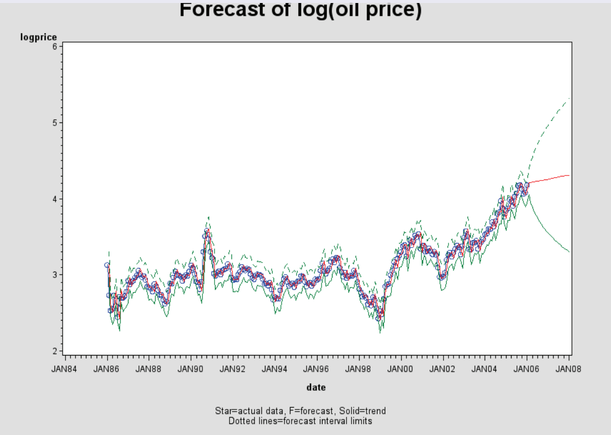
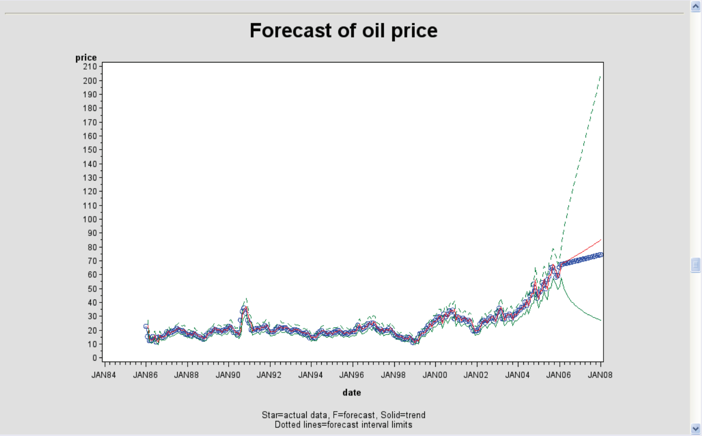


Figure 9.6. Oil price forecasting via 1st model with a revision



Example 9.3. Crude oil price forecasting via the second fitted model

- The second fitted model is as follows

$$logY_t = 3.1029 - 0.006567t + 0.0000418t^2 + X_t$$

with

$$X_t = e_t + 1.1832X_{t-1} - 0.2768X_{t-2}.$$

Figure 9.7. Oil logprice forecasting via 2nd fitted model

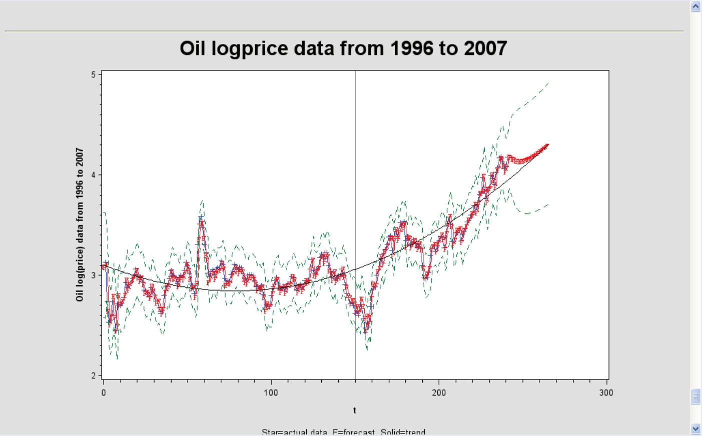
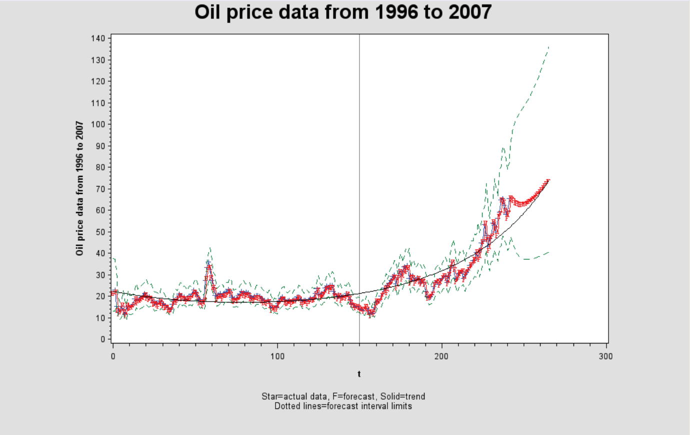


Figure 9.8. Oil price forecasting via 2nd fitted model without revision



Example 9.3. Crude oil price forecasting (con't)

- Figure 9.8 does not use

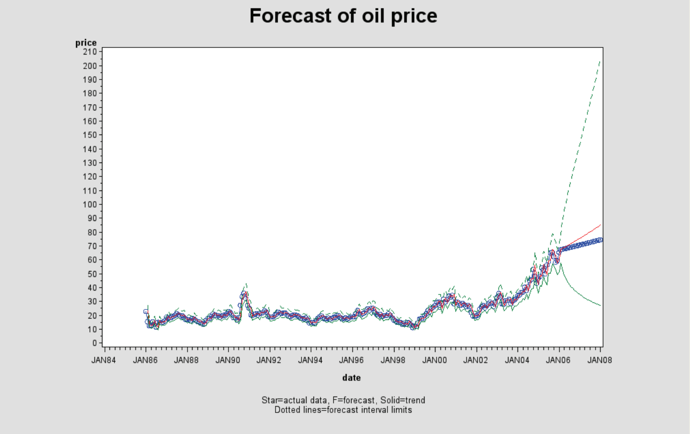
$$\widehat{y}_t(l) = \exp\left\{\widehat{Z}_t(l) + \frac{1}{2}\mathrm{var}[e_t(l)]\right\}.$$

- We can use another way to solve this problem.
  - Consider $X_t = logY_t - 3.1029 + 0.006567t - 0.0000418t^2$.
  - use ARIMA procedure for X_t with $p = 2$ and get its forecast.
  - Use

$$\hat{Y}_t(l) = exp\{3.1029 - 0.006567t + 0.0000418t^2 + \hat{X}_t(l) + std^2/2\}.$$

- See Figure 9.9.

Figure 9.9. Oil price forecasting via 2nd model with a revision



Have a nice day !