universität innsbruck

Time Series Analysis

Forecasting

Forecasting

Contents

Contents

Contents

Assumption: For the results in this chapter it is assumed that the model is known *exactly* (both order and all parameters).

Justification: Never true in practice but approximately so in large samples when model selection error is small enough.

Otherwise: Uncertainty in model parameters would have to be considered as an additional source of variation.

Forecasting

Minimum Mean Square Error Forecasting

Minimum mean square error forecasting

Goal: Based on the history up to time t, i.e., $Y_1, Y_2, \ldots, Y_{t-1}, Y_t$, *forecast* (or *predict*) $Y_{t+\ell}$.

Jargon: Time t is the *forecast origin* and ℓ the *lead time*.

Notation: $\widehat{Y}_t(\ell)$.

Question: What is the “best” forecast $\widehat{Y}_t(\ell)$?

Answer: W.r.t. mean square forecast errors, the conditional expectation is optimal (i.e., minimizes the mean square forecast errors).

$$\widehat{Y}_t(\ell) \;=\; \mathsf{E}(Y_{t+\ell} \mid Y_1, Y_2, \ldots, Y_{t-1}, Y_t)$$

Minimum mean square error forecasting

Excursion: Conditional expectations. It can be shown that if $g(x) = \mathrm{E}(Y \mid X = x)$ the following holds for $g(X)$

$$\mathrm{E}(g(X)) \;=\; \mathrm{E}(Y)$$

which is often abbreviated to $\mathrm{E}\{\mathrm{E}(Y \mid X)\} = \mathrm{E}(Y)$.

Mean square error prediction: Minimize for a constant c

$$\begin{aligned} g(c) &= \mathrm{E}\{(Y-c)^2\} \\ &= \mathrm{E}(Y^2) \;-\; 2c\mathrm{E}(Y) \;+\; c^2 \end{aligned}$$

Solve analytically

$$\begin{aligned} g'(c) &= 0 \\ -2\mathrm{E}(Y) \;+\; 2c &= 0 \\ c &= \mathrm{E}(y) \;=\; \mu \end{aligned}$$

which also yields $\min_c g(c) = \mathrm{E}\{(Y-\mu)^2\} = \sigma^2$.

Minimum mean square error forecasting

Regression: Use predictor $h(X)$. Minimize

$$\mathrm{E}\{[Y - h(X)]^2\} \;=\; \mathrm{E}(\mathrm{E}\{[Y - h(X)]^2 \mid X\})$$

For the inner expectation

$$\mathrm{E}\{[Y - h(x)]^2 \mid X = x\}$$

the function $h(x)$ is just a constant for each x.

The best choice of $h(x)$ is thus

$$h(x) \;=\; \mathrm{E}(Y \mid X = x)$$

Forecasting

Deterministic Trends

Deterministic trends

Reconsider: Deterministic trend model from Chapter 3.

$$Y_t \;=\; \mu_t \;+\; X_t$$

Here: X_t white noise with zero mean and variance γ_0.

Thus:

$$\begin{aligned}
\widehat{Y}_t(\ell) &= \mathrm{E}(\mu_{t+\ell} + X_{t+\ell} \mid Y_1, \ldots, Y_t) \\
&= \mathrm{E}(\mu_{t+\ell} \mid Y_1, \ldots, Y_t) \;+\; \mathrm{E}(X_{t+\ell} \mid Y_1, \ldots, Y_t) \\
&= \mu_{t+\ell} \;+\; \mathrm{E}(X_{t+\ell}) \\
&= \mu_{t+\ell}
\end{aligned}$$

Example: Linear trend $\mu_t = \beta_0 + \beta_1 \cdot t$. Then

$$\widehat{Y}_t(\ell) \;=\; \beta_0 + \beta_1 \cdot (t + \ell)$$

Deterministic trends

Remark: Lack of dependence (or usage thereof) prevents better forecast than $\mu_{t+\ell}$.

Forecast error:

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \widehat{Y}_t(\ell) \\ &= \mu_{t+\ell} + X_{t+\ell} - \mu_{t+\ell} \\ &= X_{t+\ell} \end{aligned}$$

Hence: All properties are inherited from error term. In particular, forecasts are *unbiased* and error variance identical for all lead times ℓ.

$$\begin{aligned} \mathrm{E}(e_t(\ell)) &= \mathrm{E}(X_{t+\ell}) = 0 \\ \mathrm{Var}(e_t(\ell)) &= \mathrm{Var}(X_{t+\ell}) = \gamma_0 \end{aligned}$$

Deterministic trends

Illustration: Deterministic trend season effect for log number of airline passengers.

Fit model with linear trend and season pattern:

```
R> data("AirPassengers", package = "datasets")
R> ap <- log(AirPassengers)
R> ap_lm <- dynlm(ap ~ trend(ap) + season(ap))
```

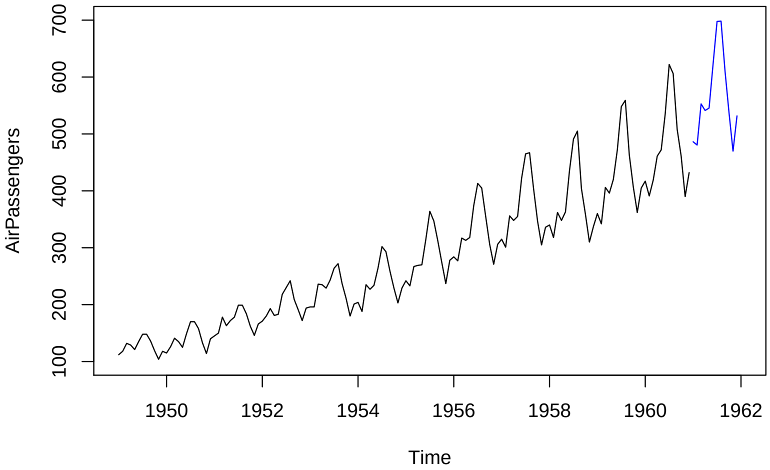
Compute predictions "by hand" because no nice `predict()` function is available for this model class.

```
R> cf <- coef(ap_lm)
R> ap_pred <- cf[1] + cf[2] * (length(ap) + 1:12)/12 + c(0, cf[3:13])
R> ap_pred <- ts(exp(ap_pred), start = 1961, freq = 12)
```

Visualize

```
R> plot(AirPassengers, xlim = c(1949, 1962), ylim = c(100, 700))
R> lines(ap_pred, col = 4)
```

Deterministic trends



Forecasting

ARIMA Forecasting

ARIMA forecasting: AR(1)

Special case: AR(1) with (known) mean μ and autocorrelation coefficient ϕ.

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

Prediction: One-step-ahead prediction of Y_{t+1} given $Y_1, \dots, Y_t$.

$$\begin{aligned}
\widehat{Y}_t(1) &= \mathsf{E}(Y_{t+1} \mid Y_1, \dots, Y_t) \\
&= \mathsf{E}(\mu + \phi(Y_t - \mu) + e_{t+1} \mid Y_1, \dots, Y_t) \\
&= \mu \ + \ \mathsf{E}(\phi(Y_t - \mu) \mid Y_1, \dots, Y_t) \ + \ \mathsf{E}(e_{t+1} \mid Y_1, \dots, Y_t) \\
&= \mu \ + \ \phi(\mathsf{E}(Y_t \mid Y_1, \dots, Y_t) - \mu) \ + \ \mathsf{E}(e_{t+1}) \\
&= \mu \ + \ \phi(Y_t - \mu)
\end{aligned}$$

Thus: The prediction is not only the mean μ but additionally some proportion of the previous deviation from the mean is predicted to occur in the next step.

ARIMA forecasting: AR(1)

Analogously: For a general lead time $\ell \geq 1$.

$$\widehat{Y}_t(\ell) \;=\; \mu \;+\; \phi \left\{ \widehat{Y}_t(\ell - 1) - \mu \right\}$$

Thus: Forecasts can be built recursively. Equation also known as *difference equation form* of forecasts.

Furthermore: Recursion can be solved explicitly.

$$\begin{aligned}
\widehat{Y}_t(\ell) \;&=\; \mu \;+\; \phi \left\{ \widehat{Y}_t(\ell - 1) - \mu \right\} \\
&=\; \mu \;+\; \phi \left[\phi \left\{ \widehat{Y}_t(\ell - 2) - \mu \right\} \right] \\
&\;\;\vdots \\
&=\; \mu \;+\; \phi^{\ell - 1} \left\{ \widehat{Y}_t(1) - \mu \right\} \\
&=\; \mu \;+\; \phi^{\ell} \left\{ Y_t - \mu \right\}
\end{aligned}$$

ARIMA forecasting: US GDP

Illustration: AR(1) model for growth of US gross domestic product (GDP, in billion USD), quarterly series from 1950 Q1 to 2000 Q4 taken from Greene (2003).

```
R> data("USMacroG", package = "AER")
R> gdp <- USMacroG[, "gdp"]
```

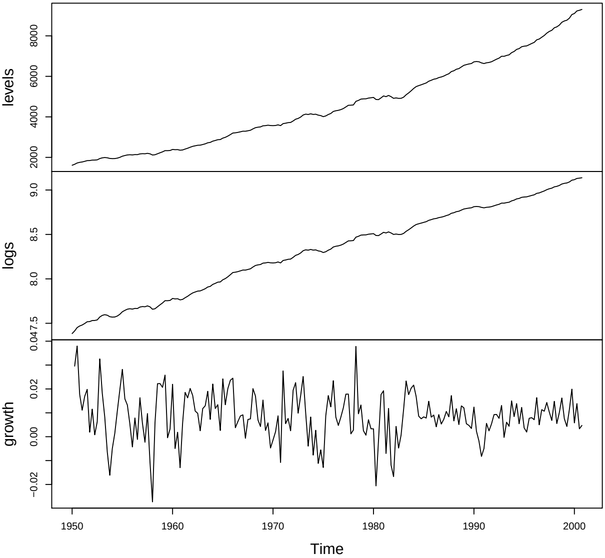
Visualization:

```
R> plot(ts.union(levels = gdp, logs = log(gdp),
+    growth = diff(log(gdp))), main = "")
```

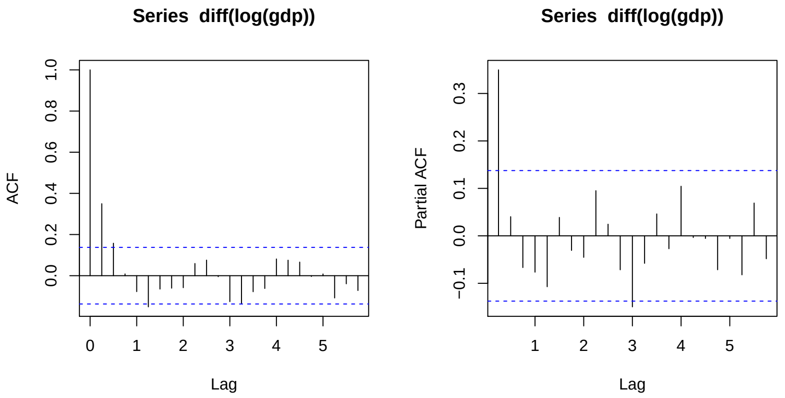
(Partial) autocorrelations:

```
R> acf(diff(log(gdp)))
R> pacf(diff(log(gdp)))
```

ARIMA forecasting: US GDP



ARIMA forecasting: US GDP



ARIMA forecasting: US GDP

AR(1) for growth rates (in %):

```
R> gdp <- 100 * diff(log(gdp))
R> gdp_ar1 <- arima(gdp, c(1, 0, 0))
R> gdp_ar1

Call:
arima(x = gdp, order = c(1, 0, 0))

Coefficients:
        ar1  intercept
      0.356      0.869
s.e.  0.066      0.101

sigma^2 estimated as 0.866:  log likelihood = -273.4,  aic = 552.9
```

ARIMA forecasting: US GDP

Predictions:

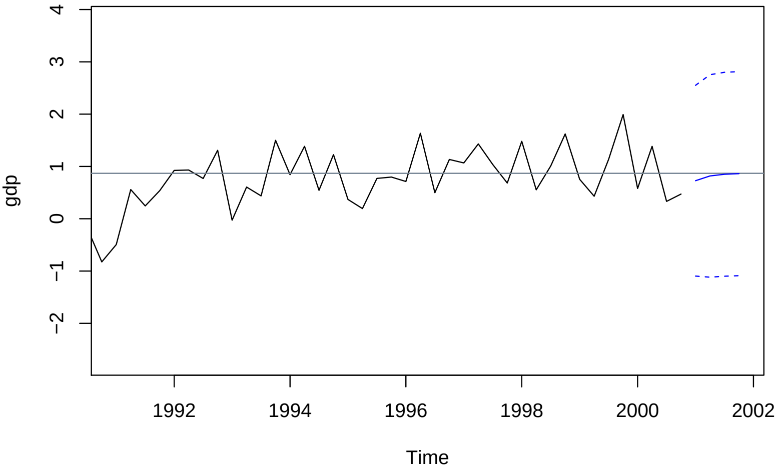
```
R> cf <- coef(gdp_ar1)
R> cf[2] + cf[1]^(1:4) * (gdp[length(gdp)] - cf[2])
[1] 0.7274 0.8183 0.8507 0.8623
R> pred <- predict(gdp_ar1, n.ahead = 4)
R> pred
$pred
       Qtr1   Qtr2   Qtr3   Qtr4
2001 0.7274 0.8183 0.8507 0.8623

$se
       Qtr1   Qtr2   Qtr3   Qtr4
2001 0.9303 0.9875 0.9945 0.9954
```

Visualization:

```
R> plot(gdp, xlim = c(1991, 2001.75))
R> abline(h = cf[2], col = "slategray")
R> lines(pred$pred, col = 4)
R> lines(pred$pred + qnorm(0.025) * pred$se, col = 4, lty = 2)
R> lines(pred$pred + qnorm(0.975) * pred$se, col = 4, lty = 2)
```

ARIMA forecasting: US GDP



ARIMA forecasting: US GDP

Alternatively: Employ generic `forecast()` function from the *forecast* package.

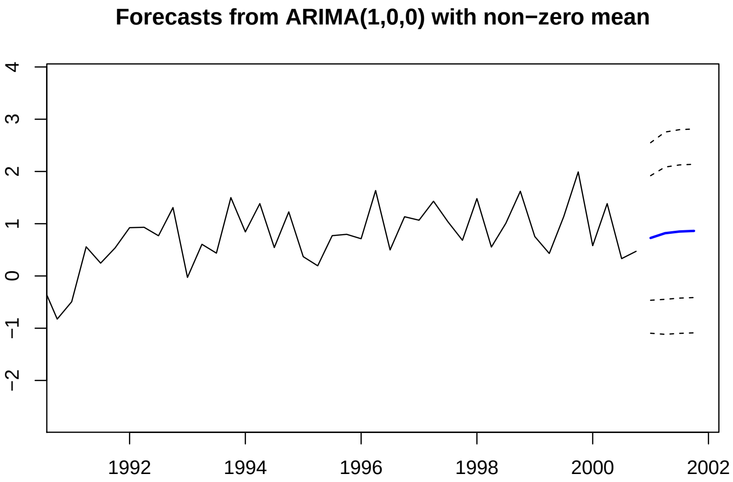
```
R> library("forecast")
R> gdp_fc <- forecast(gdp_ar1, h = 4)
R> gdp_fc
```

```
        Point Forecast   Lo 80 Hi 80  Lo 95 Hi 95
2001 Q1         0.7274 -0.4649 1.920 -1.096 2.551
2001 Q2         0.8183 -0.4472 2.084 -1.117 2.754
2001 Q3         0.8507 -0.4238 2.125 -1.099 2.800
2001 Q4         0.8623 -0.4134 2.138 -1.089 2.813
```

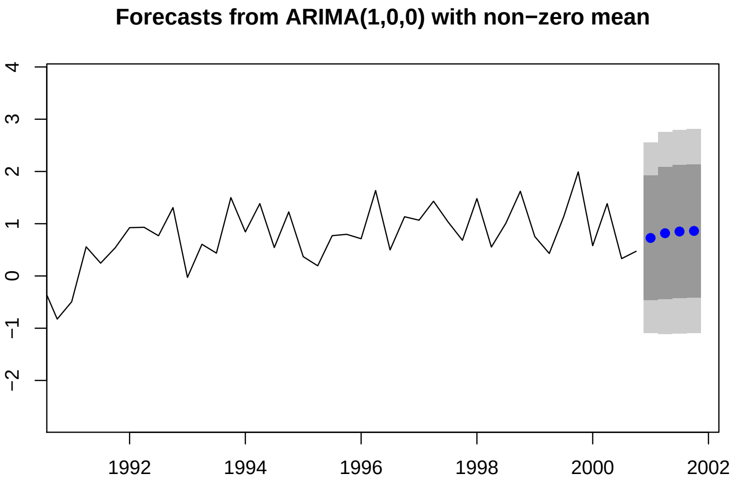
This has also a convenient `plot()` method.

```
R> plot(gdp_fc, shaded = FALSE)
R> plot(gdp_fc, shadecols = gray(c(0.8, 0.6)))
```

ARIMA forecasting: US GDP



ARIMA forecasting: US GDP



ARIMA forecasting: AR(1)

Forecast error: One step ahead.

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

Remarks:

- The white noise process $\{e_t\}$ can be interpreted as a sequence of one-step-ahead forecast errors.
- The same holds for general ARMA processes.
- Implies that the forecast error $e_t(1)$ is independent of the complete history $Y_1, \ldots, Y_t$ up to t.
- If this were not the case, dependence could be exploited for improving the forecasts.

ARIMA forecasting: AR(1)

Furthermore: For ℓ-step-ahead error, rewrite $\{Y_t\}$ in general linear process form, i.e., MA(∞).

$$Y_t - \mu = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \dots$$

Then, the forecast error can be written as:

$$\begin{aligned}
e_t(\ell) &= Y_{t+\ell} - \widehat{Y}_t(\ell) \\
&= Y_{t+\ell} - \mu - \phi^\ell (Y_t - \mu) \\
&= e_{t+\ell} + \phi e_{t+\ell-1} + \dots + \phi^{\ell-1} e_{t+1} + \phi^\ell e_t + \phi^{\ell+1} e_{t-1} + \dots \\
&\quad - \phi^\ell \left(e_t + \phi e_{t-1} + \dots \right) \\
&= e_{t+\ell} + \phi e_{t+\ell-1} + \dots + \phi^{\ell-1} e_{t+1} \\
&= e_{t+\ell} + \psi_1 e_{t+\ell-1} + \dots + \psi_{\ell-1} e_{t+1}
\end{aligned}$$

which can also be shown to hold for general ARMA models.

ARIMA forecasting: AR(1)

Properties: Forecasts are unbiased and error variance increases with lead times ℓ.

$$\begin{aligned} \mathrm{E}(e_t(\ell)) &= 0 \\ \mathrm{Var}(e_t(\ell)) &= \sigma_e^2 \left(1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2\right) \end{aligned}$$

Specifically: For $\ell = 1$ and "large" ℓ, respectively.

$$\begin{aligned} \mathrm{Var}(e_t(1)) &= \sigma_e^2 \\ \mathrm{Var}(e_t(\ell)) &\approx \mathrm{Var}(Y_t) = \gamma_0 \quad \text{for large } \ell \end{aligned}$$

For AR(1).

$$\begin{aligned} \mathrm{Var}(e_t(\ell)) &= \sigma_e^2 \, \frac{1 - \phi^{2\ell}}{1 - \phi^2} \\ &\approx \sigma_e^2 \, \frac{1}{1 - \phi^2} \quad \text{for large } \ell \end{aligned}$$

ARIMA forecasting: MA(1)

Special case: MA(1) with (known) mean μ and MA coefficient θ.

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

One-step-ahead forecast: Using for $t+1$ instead of t and taking conditional expectations yields

$$\begin{aligned}
\widehat{Y}_t(1) &= \mu + \mathsf{E}(e_{t+1} \mid Y_1, \dots, Y_t) - \theta \mathsf{E}(e_t \mid Y_1, \dots, Y_t) \\
&\approx \mu - \theta e_t \\
e_t &= Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \dots
\end{aligned}$$

for an invertible MA(1) process. In finite samples, this is typically conditioned on $e_t = 0$ for $t \leq 0$ (see Chapter 7). Except for the approximation, this yields again

$$e_t(1) = Y_{t+1} - \widehat{Y}_t(1) = e_{t+1}.$$

ARIMA forecasting: MA(1)

Analogously: For $\ell > 1$.

$$\begin{aligned}
\widehat{Y}_t(\ell) &= \mu + \mathsf{E}(e_{t+\ell} \mid Y_1, \ldots, Y_t) - \theta \mathsf{E}(e_{t+\ell-1} \mid Y_1, \ldots, Y_t) \\
&= \mu \\
e_t(\ell) &= Y_{t+\ell} - \widehat{Y}_t(\ell) \\
&= e_{t+\ell} - \theta e_{t+\ell-1} \\
&= e_{t+\ell} + \psi_1 e_{t+\ell-1} + \ldots + \psi_{\ell-1} e_{t+1}
\end{aligned}$$

because trivially $\psi_1 = -\theta$ and $\psi_j \geq 0$ for all $j > 1$.

Properties: Expectation and variance can be computed using the same formulas as before (based on the ψ_j coefficients).

ARIMA forecasting: Random walk with drift

Special case: ARIMA(0, 1, 0) with intercept θ_0, i.e., random walk with drift.

$$Y_t \;=\; Y_{t-1} \;+\; \theta_0 \;+\; e_t$$

Prediction: As above, one-step-ahead forecasts are obtained by taking conditional expectations. For general lead times $\ell \geq 1$, the the difference equation form or recursively solved equation can be used.

$$\begin{aligned}
\widehat{Y}_t(1) &= \mathsf{E}(Y_t \mid Y_1, \dots, Y_t) \;+\; \theta_0 \;+\; \mathsf{E}(e_{t+1} \mid Y_1, \dots, Y_t) \\
&= Y_t \;+\; \theta_0 \\
\widehat{Y}_t(\ell) &= \widehat{Y}_t(\ell - 1) \;+\; \theta_0 \\
&= Y_t \;+\; \theta_0 \cdot \ell
\end{aligned}$$

Hence, predictions are very different with/without drift, especially for large lead times ℓ.

ARIMA forecasting: Random walk with drift

Forecast error: As before.

$$
\begin{aligned}
e_t(1) &= Y_{t+1} - \widehat{Y}_t(1) = e_{t+1} \\
e_t(\ell) &= Y_{t+\ell} - \widehat{Y}_t(\ell) \\
&= (Y_t + \theta_0 \ell + e_{t+1} + \cdots + e_{t+\ell}) - (Y_t + \theta_0 \ell) \\
&= e_{t+\ell} + e_{t+\ell-1} + \ldots + e_{t+1} \\
&= e_{t+\ell} + \psi_1 e_{t+\ell-1} + \ldots + \psi_{\ell-1} e_{t+1}
\end{aligned}
$$

where the MA(∞) representation has $\psi_j = 1$ for all j.

Properties: Unbiased but variance diverges for increasing ℓ (characteristic for all nonstationary ARIMA processes).

$$
\begin{aligned}
\mathrm{E}(e_t(\ell)) &= 0 \\
\mathrm{Var}(e_t(\ell)) &= \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 = \ell \cdot \sigma_e^2
\end{aligned}
$$

ARIMA forecasting: ARMA(p, q)

Special case: General stationary ARMA(p, q).

$$\begin{aligned}\widehat{Y}_t(\ell) \;&=\; \phi_1 \widehat{Y}_t(\ell-1) \,+\, \ldots \,+\, \phi_p \widehat{Y}_t(\ell-p) \,+\, \theta_0 \,- \\ &\quad\; \theta_1 \mathrm{E}(e_{t+\ell-1} \mid Y_1, \ldots, Y_t) \,-\, \ldots \,-\, \theta_q \mathrm{E}(e_{t+\ell-q} \mid Y_1, \ldots, Y_t)\end{aligned}$$

where the innovations up to time t can be (approximately) computed (for large enough t) from the AR(∞) representation. Expectations of future innovations are always zero.

$$\mathrm{E}(e_{t+j} \mid Y_1, \ldots, Y_t) \,=\, \begin{cases} 0 & \text{for } j > 0 \\ e_{t+j} & \text{for } j \leq 0 \end{cases}$$

ARIMA forecasting: ARMA(p, q)

Specifically: For $\ell > q$, a Yule-Walker type recursion holds.

$$\begin{aligned}
\widehat{Y}_t(\ell) &= \phi_1 \widehat{Y}_t(\ell - 1) + \ldots + \phi_p \widehat{Y}_t(\ell - p) + \theta_0 \\
\widehat{Y}_t(\ell) - \mu &= \phi_1 \left(\widehat{Y}_t(\ell - 1) - \mu \right) + \ldots + \phi_p \left(\widehat{Y}_t(\ell - p) - \mu \right)
\end{aligned}$$

because $\theta_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$.

Remarks:

- Roots of AR characteristic polynomial determine behavior of $\widehat{Y}_t(\ell) - \mu$, i.e., linear combination of exponentially decaying terms and damped sine waves.
- For "large" ℓ $\widehat{Y}_t(\ell) = \mu$.
- Dependence dies out and only "naive" forecasts can be used.

ARIMA forecasting: ARMA(p, q)

Forecast error: It can be shown in general that

$$e_t(\ell) \;=\; e_{t+\ell} \;+\; \psi_1 e_{t+\ell-1} \;+\; \ldots \;+\; \psi_{\ell-1} e_{t+1}$$

Properties:

$$\begin{aligned}
\mathrm{E}(e_t(\ell)) &= 0 \\
\mathrm{Var}(e_t(\ell)) &= \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 \\
&\approx \sigma_e^2 \sum_{j=0}^{\infty} \psi_j^2 \;=\; \gamma_0 \qquad \text{for large } \ell
\end{aligned}$$

ARIMA forecasting: Nonstationary series

Special case: ARIMA(p, 1, q) can always be written as nonstationary ARMA($p+1$, q).

$$\begin{aligned} Y_t &= \varphi_1 Y_{t-1} + \ldots + \varphi_{p+1} Y_{t-p-1} + \theta_0 \\ &\quad + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \end{aligned}$$

with

$$\varphi_j = \begin{cases} 1 + \phi_1 & \text{for } j = 1 \\ \phi_j - \phi_{j-1} & \text{for } j = 2, \ldots, p \\ -\phi_p & \text{for } j = p+1 \end{cases}$$

and then formulas for ARMA(p, q) can be applied.

ARIMA forecasting: Nonstationary series

Forecast errors: Unbiased but variance diverges because ψ_j weights do not decay to zero.

Examples:

- ARIMA(0, 1, 1): $\psi_j = 1 - \theta$.
- ARIMA(1, 1, 0): $\psi_j = (1 - \phi^{j+1})/(1 - \phi)$.

Forecasting

Prediction Limits

Prediction limits

Goal: Complement point predictions with prediction intervals.

Idea: If forecast errors $e_t(\ell)$ can be assumed to be (approximately) normally distributed, then pointwise prediction intervals for each ℓ can be easily computed.

$$\mathsf{P}\left(-z_{1-\alpha/2} \;<\; \frac{Y_{t+\ell} - \widehat{Y}_t(\ell)}{\sqrt{\mathsf{Var}(e_t(\ell))}} \;<\; z_{1-\alpha/2}\right) \;=\; 1 - \alpha$$

where $1 - \alpha$ is the confidence level and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile from the standard normal distribution.

Prediction interval: Or prediction limits, at level $1 - \alpha$.

$$\widehat{Y}_t(\ell) \;\pm\; z_{1-\alpha/2}\sqrt{\mathsf{Var}(e_t(\ell))}$$

Prediction limits

In practice: $\text{Var}(e_t(\ell))$ is typically unknown but has to be estimated from the data.

Example: Deterministic trend $Y_t = \mu_t + X_t$ and $\text{Var}(e_t(\ell)) = \gamma_0$.

Finite samples: Actually, the correct forecast error variance can be shown (as in standard regression models) to be $\gamma_0(1 + 1/n + c_{n,\ell})$ where $c_{n,\ell}$ depends on the type of deterministic trend. E.g., for linear trend $c_{n,\ell} = 3(n + 2\ell - 1)^2/[n(n^2 - 1)] \approx 3/n$ for moderate ℓ and large n.

Similarly: Variance of forecast errors in ARIMA models are typically estimated based on estimated coefficients.

Forecasting

Forecasting Illustrations

Forecasting illustrations

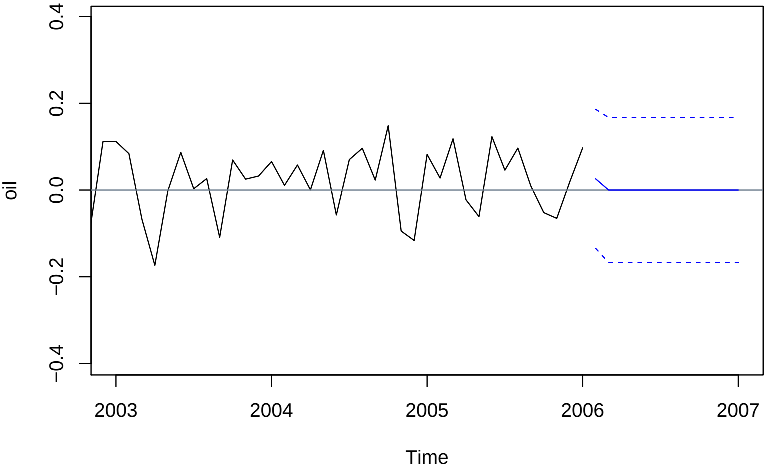
Illustration: MA(1) model without intercept for oil price returns. Estimated parameters are $\hat{\theta} = -0.2956$ and $\widehat{\sigma}_e = 0.0818$.

```
R> data("oil.price", package = "TSA")
R> oil <- diff(log(oil.price))
R> oil_ma1 <- arima(oil, order = c(0, 0, 1), include.mean = FALSE)
R> predict(oil_ma1, n.ahead = 6)
```

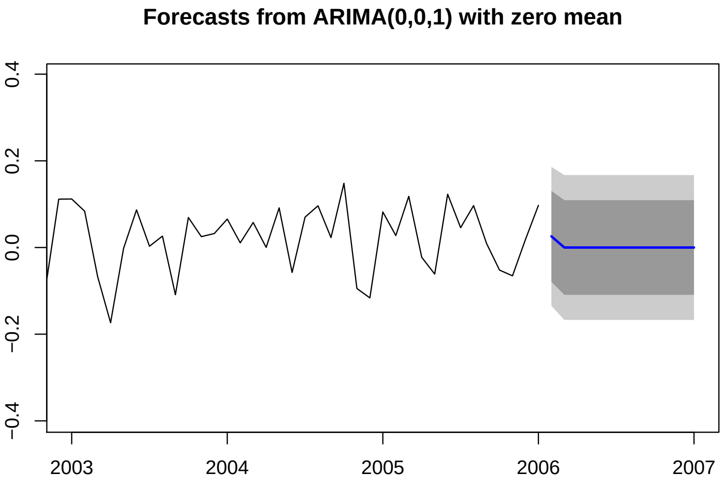
```
$pred
         Feb     Mar     Apr     May     Jun     Jul
2006 0.02581 0.00000 0.00000 0.00000 0.00000 0.00000

$se
         Feb     Mar     Apr     May     Jun     Jul
2006 0.08178 0.08528 0.08528 0.08528 0.08528 0.08528
```

Forecasting illustrations



Forecasting illustrations



Forecasting

Updating ARIMA Forecasts

Updating ARIMA forecasts

Goal: A prediction $\widehat{Y}_t(\ell+1)$ made at time t, should be updated at time $t+1$ when Y_{t+1} was observed.

Solution: Compute $\widehat{Y}_{t+1}(\ell)$, either from scratch or by recursive updating.

Updating equation: One can show that

$$\begin{aligned}\widehat{Y}_{t+1}(\ell) &= \widehat{Y}_t(\ell+1) + \psi_\ell e_{t+1} \\ &= \widehat{Y}_t(\ell+1) + \psi_\ell \left\{ Y_{t+1} - \widehat{Y}_t(1) \right\}\end{aligned}$$

This is also called *adaptive expectations*.

Forecast Weights and Exponentially Weighted Moving Averages

Forecast weights and EWMAs

Motivation:

- Computing forecasts in AR models is straightforward as it involves only $Y_t, \dots, Y_1$.
- In ARIMA models with MA terms the noise terms e_t also appears which is less intuitive.

Idea: Employ AR(∞) representation of the ARIMA model

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \dots + e_t$$

By replacing t with $t+1$ and taking conditional expectations, the one-step-ahead forecasts are

$$\widehat{Y}_t(1) = \pi_1 Y_t + \pi_2 Y_{t-1} + \dots$$

If t is sufficiently large and the weights die out sufficiently quickly, the terms $\pi_t, \pi_{t+1}, \dots$ can be ignored.

Forecast weights and EWMAs

Recursion: For an invertible ARIMA model, the AR(∞) weights are

$$\pi_j = \begin{cases} \displaystyle\sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-1} + \varphi_j & \text{for } 1 \leq j \leq p+d \\ \displaystyle\sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-1} & \text{for } j > p+d \end{cases}$$

with initial value $\pi_0 = -1$.

See also Chapter 4 for MA(∞) and AR(∞) representations of stationary and invertible ARMA models.

Forecast weights and EWMAs

Example: ARIMA(0, 1, 1).

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

Thus, $\varphi_1 = 1$. Then,

$$\begin{aligned}
\pi_1 &= \theta\pi_0 + 1 = 1 - \theta \\
\pi_2 &= \theta\pi_1 = \theta(1-\theta) \\
\pi_3 &= \theta\pi_2 = \theta^2(1-\theta) \\
&\vdots \\
\pi_j &= \theta^{j-1}(1-\theta)
\end{aligned}$$

Thus: Weights exponentially decrease and sum to unity.

$$\sum_{j=1}^{\infty}\pi_j = (1-\theta)\sum_{j=1}^{\infty}\theta^{j-1} = (1-\theta)\frac{1}{1-\theta} = 1$$

Forecast weights and EWMAs

Forecast: $\widehat{Y}_t(1)$ is called *exponentially weighted moving average* (EWMA).

$$
\begin{aligned}
\widehat{Y}_t(1) &= (1-\theta)Y_t \;+\; (1-\theta)\theta Y_{t-1} \;+\; (1-\theta)\theta^2 Y_{t-2} \;+\; \dots \\
&= (1-\theta)Y_t \;+\; \theta\widehat{Y}_{t-1}(1) \\
&= \widehat{Y}_{t-1}(1) \;+\; (1-\theta)\left\{Y_t - \widehat{Y}_{t-1}(1)\right\}
\end{aligned}
$$

Remarks:

- Also called *exponential smoothing* with *smoothing constant* θ.
- Often used on an ad-hoc basis in practice (with sometimes rather arbitrary selection of θ).
- Many smoothing methods have ARIMA representation.
- Discussion of exponential smoothing and ARIMA forecasting in Hyndman & Khandakar (*Journal of Statistical Software*, **27**(3)), accompanying *forecast*.

Forecasting

Forecasting Transformed Series

Forecasting transformed series

Question: When ARIMA models are used for a transformed series Z_t obtained from an original series Y_t, how should predictions $\widehat{Y}_t(\ell)$ be computed from $\widehat{Z}_t(\ell)$?

Answer: For all linear transformations, the minimum mean square error forecast can be obtained by the inverse transformation. For nonlinear transformations, the inverse transformation typically does *not* yield the minimum mean square error forecast.

Example: Differencing $Z_t = Y_t - Y_{t-1}$ (i.e., a linear transformation) with ARIMA(0, 1, 1) for Y_t and ARIMA(0, 0, 1) for Z_t.

Forecasting transformed series

The forecasts for Y_t yield

$$\begin{aligned}
\widehat{Y}_t(1) &= Y_t - \theta e_t \\
\widehat{Y}_t(\ell) &= \widehat{Y}_t(\ell-1) \qquad \text{for } \ell > 1
\end{aligned}$$

The forecasts for Z_t yield

$$\begin{aligned}
\widehat{Z}_t(1) &= -\theta e_t \\
\widehat{Z}_t(\ell) &= 0 \qquad \text{for } \ell > 1
\end{aligned}$$

Thus, cumulative sums of $\widehat{Z}_t(\ell)$ yield $\widehat{Y}_t(\ell)$

$$\begin{aligned}
\widehat{Y}_t(1) &= Y_t + \widehat{Z}_t(1) \\
\widehat{Y}_t(\ell) &= \widehat{Y}_t(\ell-1) + \widehat{Z}_t(\ell) \\
&= Y_t + \sum_{j=1}^{\ell} \widehat{Z}_t(j)
\end{aligned}$$

Forecasting transformed series

Example: Logarithms $Z_t = \log(Y_t)$ (i.e., a nonlinear transformation).

It can be shown that

$$\mathsf{E}(Y_{t+\ell} \mid Y_1, \ldots, Y_t) \;\geq\; \exp(\mathsf{E}(Z_{t+\ell} \mid Z_1, \ldots, Z_t))$$

Thus:

- $\exp(\widehat{Z}_t(\ell))$ is *not* the minimum mean square error forecast for $Y_{t+\ell}$ (except in trivial cases).
- However, if the distribution of Z_t can be assumed to be symmetric: $\widehat{Z}_t(\ell)$ is also the minimum *median* square error forecast and so is $\exp(\widehat{Z}_t(\ell))$ due to invariance of median.
- Minimum mean square error forecast can also be computed for certain distributions of Z_t.

Forecasting transformed series

Note: If X has a normal distribution with mean μ and variance σ^2, then

$$\mathrm{E}(\exp(X)) \;=\; \exp\left(\mu + \sigma^2/2\right)$$

If Z_t can be assumed to be normal, then the mean square error forecast for $Y_{t+\ell}$ is

$$\begin{aligned}
& \mathrm{E}(Y_{t+\ell} \mid Y_1, \ldots, Y_t) \\
= \;& \mathrm{E}(\exp(Z_{t+\ell}) \mid Z_1, \ldots, Z_t) \\
= \;& \exp\{\mathrm{E}(Z_{t+\ell} \mid Z_1, \ldots, Z_t) \;+\; \mathrm{Var}(Z_{t+\ell} \mid Z_1, \ldots, Z_t)/2\} \\
= \;& \exp\left\{\widehat{Z}_t(\ell) \;+\; \mathrm{Var}(\widehat{Z}_t(\ell) + e_t(\ell) \mid Z_1, \ldots, Z_t)/2\right\} \\
= \;& \exp\left\{\widehat{Z}_t(\ell) \;+\; \mathrm{Var}(e_t(\ell) \mid Z_1, \ldots, Z_t)/2\right\} \\
= \;& \exp\left\{\widehat{Z}_t(\ell) \;+\; \mathrm{Var}(e_t(\ell))/2\right\}
\end{aligned}$$