Ch 9: Forecasting

- Finally, let's discuss how to forecast/predict future outcomes for a time series of interest!
- Goal: Given $Y_1, Y_2, \ldots, Y_t$ (e.g., $t = n$), forecast Y_{t+g} for $g > 0$.
- To start with:
  - Let's define what "best predictor" to use,
  - Let's project future outcomes for an AR(1) model.
- Material: Ch 9.1 to 9.5.

Forecasting

- Given $Y_1, Y_2, \ldots, Y_t$, forecast Y_{t+g}:
  - t is called the forecast origin,
  - g is called the lead time (referred to as l in the book, I use g to improve readability),
  - the "best" forecast that we will use is denoted by $\hat{Y}_t(g)$.
- The "best" forecast $\hat{Y}_t(g)$ is the function $h(Y_1, Y_2, \ldots, Y_t)$ which minimizes:

$$E\left[(Y_{t+g} - h(Y_1, Y_2, \ldots, Y_t))^2\right],$$

  i.e., we minimize the expected squared difference between the yet unknown Y_{t+g} and our predicted value for it, using the observed $Y_1, \ldots, Y_t$.
- You can show (see App. F), that this predictor is given by:

$$\hat{Y}_t(g) = E(Y_{t+g}|Y_1, Y_2, \ldots, Y_t),$$

  that is, the expected value of Y_{t+g} conditional on (knowing) $Y_1, \ldots, Y_t$.
- What does this expression boil down to for ARMA models?

AR(1) forecasting

- For the AR(1) model, $Y_t = \mu + \phi(Y_{t-1} - \mu) + e_t$, we can forecast one time step as follows:

$$\begin{aligned} Y_{t+1} &= \mu + \phi(Y_t - \mu) + e_{t+1}, \\ \hat{Y}_t(1) &= E(Y_{t+1}|Y_1, \ldots, Y_t) \\ &= \mu + \phi(E(Y_t|Y_1, \ldots, Y_t) - \mu) + (E(e_{t+1}|Y_1, \ldots, Y_t), \\ &= \mu + \phi(Y_t - \mu) + 0. \end{aligned}$$

- For lead time g:

$$\begin{aligned} Y_{t+g} &= \mu + \phi(Y_{t+g-1} - \mu) + e_{t+g}, \\ \hat{Y}_t(g) &= \mu + \phi(E(Y_{t+g-1}|Y_1, \ldots, Y_t) - \mu) + (E(e_{t+g}|Y_1, \ldots, Y_t), \\ &= \mu + \phi(\hat{Y}_t(g-1) - \mu) + 0, \end{aligned}$$

  ah, so we can obtain the forecast for lead time g from the forecast at lead time $g-1$.
  - The equation that relates $\hat{Y}_t(g)$ to $\hat{Y}_t(g-1)$ is called the difference equation form of the forecasts.

AR(1) forecasting

- We can also find an explicit expression for $Y_t(g)$ in terms of $Y_1, \ldots, Y_t$ directly:

$$\begin{aligned}
\hat{Y}_t(g) &= \mu + \phi(\hat{Y}_t(g-1) - \mu), \\
&= \mu + \phi\{\phi(\hat{Y}_t(g-2) - \mu)\}, \\
&\vdots \\
&= \mu + \phi^{g-1}(\hat{Y}_t(1) - \mu), \\
&= \mu + \phi^{g}(Y_t - \mu).
\end{aligned}$$

- What happens for $|\phi| < 1$ and $g \to \infty$?

Example AR(1) forecasts: color data

Exhibit 1.3 Time Series Plot of Color Property from a Chemical Process

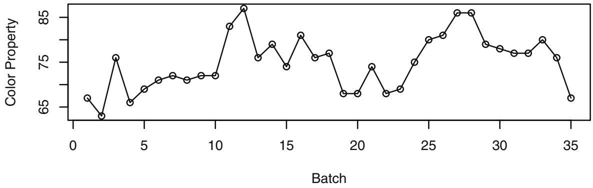


Exhibit 9.1 Maximum Likelihood Estimation of an AR(1) Model for Color

Coefficients:	ar1	intercept†
	0.5705	74.3293
s.e.	0.1435	1.9151

sigma^2 estimated as 24.8: log-likelihood = −106.07, AIC = 216.15

†Remember that the intercept here is the estimate of the process mean μ—not θ_0.

Example AR(1) forecasts: color data

- How to get $\hat{Y}_t(1)$?
- Use $\hat{Y}_t(1) = \mu + \phi(Y_t - \mu)$, and set $\phi = \hat{\phi}$ and $\mu = \hat{\mu}$.
  - Note: In the forecasts, we will only account for uncertainty from unknown error terms and not in the uncertainty in the parameter estimates.
    For large sample sizes this does not seriously effect the results.
- Result:

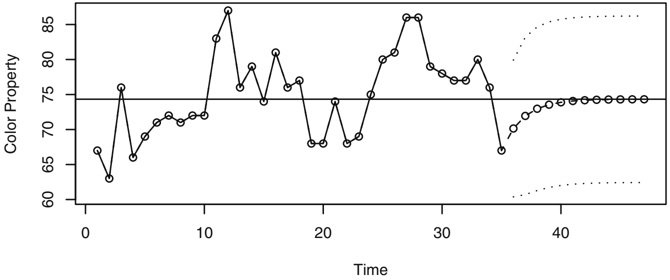
  The last observed value of the color property is 67, so we would forecast one time period ahead as†

$$\begin{aligned}\hat{Y}_t(1) &= 74.3293 + (0.5705)(67 - 74.3293)\\ &= 74.3293 - 4.181366\\ &= 70.14793\end{aligned}$$

- Etc. for $\hat{Y}_t(2), \ldots$

Example AR(1) forecasts: color data

Exhibit 9.3 Forecasts and Forecast Limits for the AR(1) Model for Color



- Note that after observing Y_t, $\hat{Y}_t(g)$ refers to a point prediction, the estimate of the expected value of Y_{t+g} conditional on having observed Y_t, a fixed value (as opposed to a random variable).
- How to get those prediction intervals (dashed lines)?

What about the uncertainty in the forecast?

- The forecast error $e_t(g)$ is given by:

$$e_t(g) \quad = \quad Y_{t+g} - \hat{Y}_t(g).$$

- To examine the properties of $e_t(g)$ it is convenient to use the general linear model notation:

$$Y_t - \mu = \sum_{i=0}^{\infty} \psi_i e_{t-i}$$

  where $\psi_0 = 1$.
- For the AR(1) model, with $\psi_i = \phi^i$:

$$Y_t - \mu = \sum_{i=0}^{\infty} \phi^i e_{t-i}$$

  (Note that in Eq. 9.3.12 in the book, μ is missing).
- The forecast error $e_t(g)$ is given by:...

Forecast errors for an AR(1) model

- For the AR(1) model: $Y_t - \mu = \sum_{i=0}^{\infty} \phi^i e_{t-i}$.
- The forecast error $e_t(g)$ is given by:

$$\begin{aligned}
e_t(g) &= Y_{t+g} - \hat{Y}_t(g), \\
&= Y_{t+g} - (\mu + \phi^g (Y_t - \mu)), \\
&= \sum_{i=0}^{\infty} \phi^i e_{t+g-i} - \phi^g \sum_{i=0}^{\infty} \phi^i e_{t-i}, \\
&= \sum_{i=0}^{g-1} \phi^i e_{t+g-i} + \sum_{i=g}^{\infty} \phi^i e_{t+g-i} - \phi^g \sum_{i=0}^{\infty} \phi^i e_{t-i}, \\
\text{(set } j = i - g) \quad &= \sum_{i=0}^{g-1} \phi^i e_{t+g-i} + \sum_{j=0}^{\infty} \phi^{j+g} e_{t-j} - \phi^g \sum_{i=0}^{\infty} \phi^i e_{t-i}, \\
&= \sum_{i=0}^{g-1} \phi^i e_{t+g-i}.
\end{aligned}$$

Implications for AR(1) forecasts

- We found $e_t(g) = Y_{t+g} - \hat{Y}_t(g) = \sum_{i=0}^{g-1} \phi^i e_{t+g-i}$, thus

$$\begin{aligned} e_t(g) &\sim N(0, Var(e_t(g))), \text{ where} \\ Var(e_t(g)) &= \sigma_e^2 \frac{1-\phi^{2g}}{1-\phi^2}. \end{aligned}$$

- For $g \to \infty$,

$$Var(e_t(g)) \approx \frac{\sigma_e^2}{1-\phi^2}.$$

  - Wait, doesn't that expression look familiar?
  - Indeed, $Var(e_t(g)) \approx \frac{\sigma_e^2}{1-\phi^2} = Var(Y_t) = \gamma_0$.
- We also found that as $g \to \infty$, $\hat{Y}_t(g) \to \mu$.
- Conclusion: Eventually, the starting point Y_t doesn't matter and the projected distribution for Y_{t+g} is given by its stationary distribution: $N(\mu, \gamma_0)$.

Prediction intervals

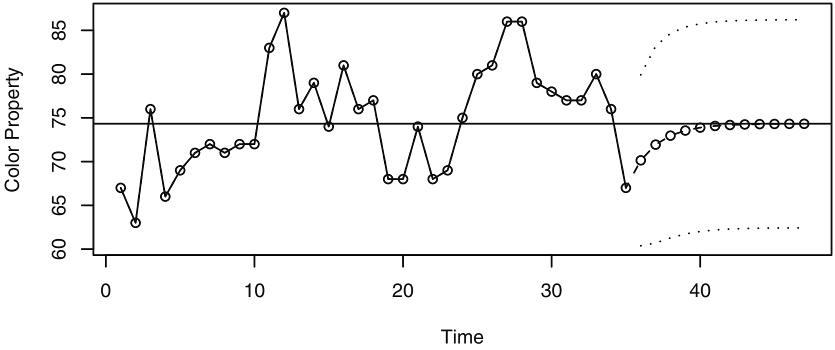
- We can construct prediction intervals (PI) for the future observation Y_{t+g}, using $e_t(g) = Y_{t+g} - \hat{Y}_t(g) \sim N(0, Var(e_t(g)))$ thus

$$P\left(-z_{1-\alpha/2} \leq \frac{Y_{t+g} - \hat{Y}_t(g)}{\sqrt{Var(e_t(g))}} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

- E.g. the 95% PI for Y_{t+g} is given by $\hat{Y}_t(g) \pm 1.96\sqrt{\widehat{Var}(e_t(g))}$ (ignoring additional uncertainty that follows from estimating model parameters).

Example of AR(1) projection with uncertainty

Exhibit 9.3 Forecasts and Forecast Limits for the AR(1) Model for Color



Summary so far

- We discussed the minimum mean square error forecast for Y_{t+g} given $Y_1, \ldots, Y_t$, which is given by $\hat{Y}_t(g) = E(Y_{t+g}|Y_1, \ldots, Y_t)$.
- We also discussed the associated forecast error $e_t(g) = Y_{t+g} - \hat{Y}_t(g)$, and derived its distribution for an AR(1) model.
- Given the point prediction for Y_{t+g} and the distribution of the forecast error, we can construct predictions and prediction intervals.
- Next: forecasting for ARIMA(p, d, q) models.
  - From MA(1) to MA(q) to ARMA(p, q) to ARIMA(p, d, q).
- We assume throughout that models are invertible; we can write $(Y_t - \mu) = e_t + \sum_{i=1}^{\infty} \pi_i (Y_{t-i} - \mu)$.

MA(1)

- $Y_t = \mu + e_t - \theta e_{t-1}$, thus we can forecast one time step as follows:

$$\begin{aligned} Y_{t+1} &= \mu + e_{t+1} - \theta e_t, \\ \hat{Y}_t(1) &= \mu + E(e_{t+1}|Y_1, \ldots, Y_t) - \theta E(e_t|Y_1, \ldots, Y_t), \\ &\approx \mu - \theta e_t. \end{aligned}$$

  Why?
  - Independence: $E(e_{t+1}|Y_1, \ldots, Y_t) = E(e_{t+1}) = 0$.
  - $E(e_t|Y_1, \ldots, Y_t) \approx e_t$ because Y_t is invertible thus e_t can be written as $e_t = Y_t - \mu - \sum_{i=1}^{\infty} \pi_i (Y_{t-i} - \mu)$.
  - The approximation becomes very accurate as t increases because π_i decays to zero as $i \to \infty$.
  - The estimate for e_t is obtained from the estimation process.
- For lead time $g > 1$:

$$\begin{aligned} Y_{t+g} &= \mu + e_{t+g} - \theta e_{t+g-1}, \\ \hat{Y}_t(g) &= \mu - E(e_{t+g} - \theta e_{t+g-1}|Y_1, \ldots, Y_t) = \mu. \end{aligned}$$

- Generalize to an MA(q) process...

Forecasting MA(q) model

$$
\begin{aligned}
Y_{t+g} &= \mu + e_{t+g} - \theta_1 e_{t+g-1} - \theta_2 e_{t+g-2} - \ldots - \theta_q e_{t+g-q}, \\
\hat{Y}_t(g) &= E(Y_{t+g} | Y_1, \ldots, Y_t), \\
&= \mu + E(e_{t+g} | Y_1, \ldots, Y_t) - \theta_1 E(e_{t+g-1} | Y_1, \ldots, Y_t) \\
&\quad -\theta_2 E(e_{t+g-2} | Y_1, \ldots, Y_t) - \ldots - \theta_q E(e_{t+g-q} | Y_1, \ldots, Y_t),
\end{aligned}
$$

where

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

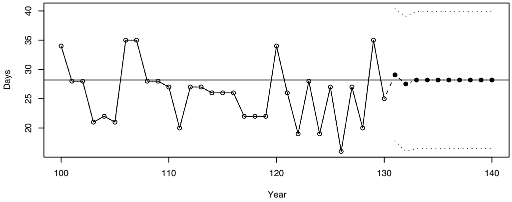
What happens for $g > q$?

Example: Fitting an MA(2) model to "days" (Ex. 9.22)

- Data set "days" contains accounting data from the Winegard Co. of Burlington, Iowa (the number of days until Winegard receives payment for 130 consecutive orders from a particular distributor).
- Forecasting, e.g. for $g = 10$, is easy with built-in functions
  - Check that forecast becomes constant for $g > 2$.

```
> model=arima(daysmod,order=c(0,0,2)); plot(model,n.ahead=10)$pred
```

```
Time Series:
Start = 131
End = 140
Frequency = 1
 [1] 29.07436 27.52056 28.19564 28.19564 28.19564 28.19564 28.19564 28.19564 28.19564
[10] 28.19564
```



Forecasting ARMA(p, q) model

$$
\begin{aligned}
Y_{t+g} - \mu &= \phi_1(Y_{t+g-1} - \mu) + \phi_2(Y_{t+g-2} - \mu) + \dots \\
&\quad +\phi_p(Y_{t+g-p} - \mu) + e_{t+g} \\
&\quad -\theta_1 e_{t+g-1} - \theta_2 e_{t+g-2} - \dots - \theta_q e_{t+g-q}, \\
\hat{Y}_t(g) - \mu &= E(Y_{t+g} - \mu | Y_1, \dots, Y_t), \\
&= \phi_1(\hat{Y}_t(g-1) - \mu) + \phi_2(\hat{Y}_t(g-2) - \mu) + \dots \\
&\quad +\phi_p(\hat{Y}_t(g-p) - \mu) + 0 \\
&\quad -\theta_1 E(e_{t+g-1} | Y_1, \dots, Y_t) - \theta_2 E(e_{t+g-2} | Y_1, \dots, Y_t) \\
&\quad - \dots - \theta_q E(e_{t+g-q} | Y_1, \dots, Y_t),
\end{aligned}
$$

where

$$
\hat{Y}_t(j) = \begin{cases} Y_{t+j} & \text{for } j \leq 0, \\ \hat{Y}_t(j) \text{ (true forecast)} & \text{for } j > 0, \end{cases}
$$

and

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

What happens for $g > q$?

ARMA(p,q) forecasting, for $g > q$

- Difference equation in terms of deviations from μ:

$$\begin{aligned}\hat{Y}_t(g) - \mu &= \phi_1(\hat{Y}_t(g-1) - \mu) + \phi_2(\hat{Y}_t(g-2-\mu) + \ldots \\ &\quad + \phi_p(\hat{Y}_t(g-p) - \mu), \text{ for } g > q,\end{aligned}$$

  only the AR-part is left!
- What happens when $g \to \infty$ for a stationary ARMA model?
  - Note that $(\hat{Y}_t(g) - \mu)$ satisfies the YW equation (set $\rho_g = (\hat{Y}_t(g) - \mu)$):

$$\rho_g = \phi_1\rho_{g-1} + \phi_2\rho_{g-2} + \phi_3\rho_{g-3} + \ldots + \phi_p\rho_{g-p}.$$

  - Remember that $\rho_g \to 0$ as g increases in an ARMA process...
  - So $Y_t(g) \to \mu$ as g increases in an ARMA process.
- In summary:
  - For $g = 1, \ldots, q$ the point prediction is determined by AR and MA terms (past Y_t's and past white noise),
  - For $g = q+1, q+2, \ldots$ the point prediction is determined by AR terms (past Y_t's) but their influence decreases over time as the forecast gets closer to μ.

What about the forecast errors?

- Helpful to use yet another expression for the invertible ARIMA model (derivation in App G outside class material):
  The truncated linear process form is given by
  $$Y_{t+g} = C_t(g) + I_t(g),$$
  for $g > 1$ where
  - $C_t(g)$ is a function of $Y_t, Y_{t-1}, \ldots, Y_1$ (if t is reasonably large),
  - $I_t(g) = e_{t+g} + \psi_1 e_{t+g-1} + \psi_2 e_{t+g-2} + \ldots + \psi_{g-1} e_{t+1}$, where ψ_j's are the coefficients in $Y_t - \mu = \sum_{j=0}^{\infty} \psi_j e_{t-j}$.
- Why is this useful to get the forecast errors?

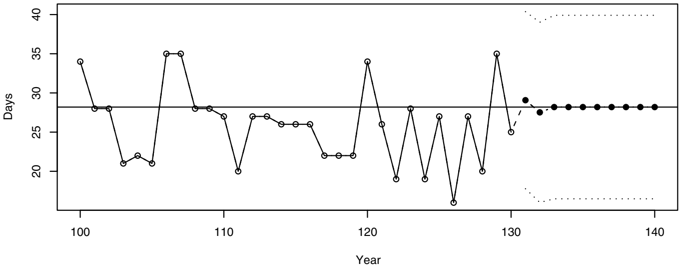
$$\begin{aligned}
e_t(g) &= Y_{t+g} - \hat{Y}_t(g), \\
&= C_t(g) + I_t(g) - E(C_t(g) + I_t(g) | Y_1, \ldots, Y_t), \\
&= C_t(g) + I_t(g) - C_t(g), \\
&= I_t(g), \\
&= e_{t+g} + \psi_1 e_{t+g-1} + \psi_2 e_{t+g-2} + \ldots + \psi_{g-1} e_{t+1},
\end{aligned}$$

What about the forecast errors? (ctd)

- We found that $e_t(g) = Y_{t+g} - \hat{Y}_t(g) = e_{t+g} + \psi_1 e_{t+g-1} + \ldots + \psi_{g-1} e_{t+1}$, thus
$$e_t(g) \quad \sim \quad N(0, Var(e_t(g))),$$
where $Var(e_t(g)) = \sigma_e^2 \sum_{j=0}^{g-1} \psi_j^2$.
- The expression for the $(1-\alpha)\cdot$ 100% prediction interval (PI) for Y_{t+g} is again given by $\hat{Y}_t(g) \pm z_{1-\alpha/2}\sqrt{\widehat{Var}(e_t(g))}$.
- What happens as $g \to \infty$?
  - We get $Var(e_t(g)) = \sigma_e^2 \sum_{j=0}^{\infty} \psi_j^2 = \gamma_0$.
  - We also found that as $g \to \infty$, $\hat{Y}_t(g) \to \mu$.
  - Conclusion:
    In predictions using a stationary ARMA model, eventually the past Y_t's don't matter and the predicted distribution for Y_{t+g} is given by its stationary distribution: $N(\mu, \gamma_0)$.
- (Check that the same expression and conclusion were true for the AR(1) model).

Example: prediction intervals for the "days" with an MA(2) model

- Data set "days" contains accounting data from the Winegard Co. of Burlington, Iowa (the number of days until Winegard receives payment for 130 consecutive orders from a particular distributor).
- What are the expressions for the variances of the forecast errors here?



ARIMA(p, d, q) models with $d > 0$

- Approach 1: forecast the differenced series, and then obtain the forecasts for the original series from the forecasted differences.
- Approach 2 (same result, but equations give some additional insights):
  - We can write an ARIMA (p, d, q) with $d > 1$ as non-stationary ARMA$(p + d, q)$ model.
  - Derivations used for the stationary ARMA model can be easily applied (see next slides).
  - We find that the variance of the forecast errors keeps increasing.

ARIMA(p, d, q) models with $d > 0$: more details

- We can write an ARIMA (p, d, q) with $d > 1$ as non-stationary ARMA$(p + d, q)$ model.
  - E.g., for ARIMA(1,1,1) with $W_t = Y_t - Y_{t-1}$ (with zero-mean to simplify equations):

$$\begin{aligned} W_{t+g} &= \phi W_{t+g-1} + e_{t+g} - \theta_1 e_{t+g-1}, \\ Y_{t+g} - Y_{t+g-1} &= \phi(Y_{t+g-1} - Y_{t+g-2}) + e_{t+g} - \theta e_{t+g-1}, \\ Y_{t+g} &= (1+\phi)Y_{t+g-1} - \phi Y_{t+g-2} + e_{t+g} - \theta e_{t+g-1}, \end{aligned}$$

- For point forecast $\hat{Y}_t(g)$, just follow the same derivation as on Slide 17.
  - E.g., for ARIMA(1,1,1):

$$\begin{aligned} \hat{Y}_t(g) &= (1+\phi)\hat{Y}_t(g-1) - \phi\hat{Y}_t(g-2) \\ &\quad -\theta E(e_{t+g-1}|Y_1, \ldots, Y_t). \end{aligned}$$

  - Contrast this prediction with prediction for a stationary ARMA model: For the ARIMA model we do NOT conclude that the influence of past Y_t's decays to zero.

ARIMA(p, d, q) models with $d > 0$: forecast errors

- For the forecast errors for projecting $\hat{Y}_t(g)$ with an ARIMA(p, d, q) model, the derivation for ARMA(p, q) forecast errors still holds true:

$$
\begin{aligned}
e_t(g) &= Y_{t+g} - \hat{Y}_t(g), \\
&= I_t(g), \\
&= e_{t+g} + \psi_1 e_{t+g-1} + \psi_2 e_{t+g-2} + \ldots + \psi_{g-1} e_{t+1}, \\
e_t(g) &\sim N(0, Var(e_t(g))), \text{ where} \\
Var(e_t(g)) &= \sigma_e^2 \sum_{j=0}^{g-1} \psi_j^2.
\end{aligned}
$$

- However, here the ψ_j's do not decay to zero as g increases, thus prediction intervals will keep getting wider.

Example: IMA(1,1) forecasting

- Robot data (plotted on next slide)
- Code and output below (write out expression for forecast!).
- Note that the width of the PI keeps increasing (although only slightly so in this example).

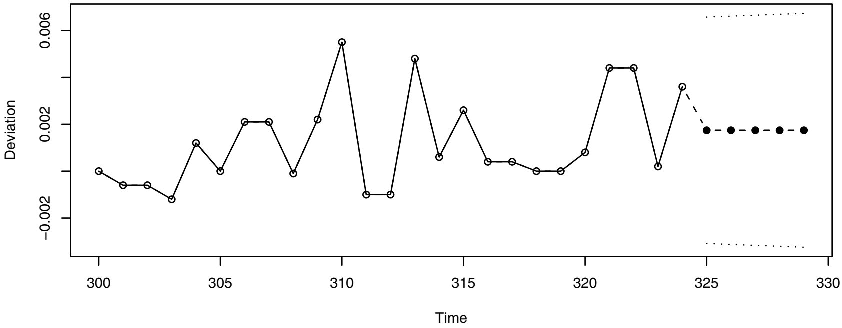
```
> data(robot); model=arima(robot,order=c(0,1,1)); model; plot(model,n.ahead=5)$pred
```

```
Time Series:
Start = 325
End = 329
Frequency = 1
[1] 0.001742672 0.001742672 0.001742672 0.001742672 0.001742672
```

```
> plot(model,n.ahead=5)$upi; plot(model,n.ahead=5)$lpi
```

```
Time Series:
Start = 325
End = 329
Frequency = 1
[1] 0.006669889 0.006710540 0.006750862 0.006790862 0.006830548
Time Series:
Start = 325
End = 329
Frequency = 1
[1] -0.003184545 -0.003225197 -0.003265519 -0.003305518 -0.003345204
```

Example: IMA(1,1) forecasting



Summary Ch. 9

- We discussed how to forecast using any ARIMA(p, d, q) model:
  - the minimum mean square error forecast for Y_{t+g}, given $Y_1, \ldots, Y_t$, is given by $\hat{Y}_t(g) = E(Y_{t+g}|Y_1, \ldots, Y_t)$,
  - we derived the distribution of the forecast error $e_t(g) = Y_{t+g} - \hat{Y}_t(g)$,
  - and the expression for the $(1-\alpha)\cdot$ 100% prediction interval (PI) for Y_{t+g}, which is given by $\hat{Y}_t(g) \pm z_{1-\alpha/2}\sqrt{\widehat{Var}(e_t(g))}$.
- Important findings for stationary ARMA models:
  - For lead time $g = 1, \ldots, q$ the point forecast is determined by AR and MA terms (past Y_t's and past white noise),
  - For $g = q+1, q+2, \ldots$ the point forecast is determined by AR terms (past Y_t's) but their influence decreases over time as the forecast gets closer to μ.
  - Eventually, the past Y_t don't matter and the predicted distribution for Y_{t+g} is given by its stationary distribution: $N(\mu, \gamma_0)$.
- Important findings for ARIMA(p, d, q) models with $d > 1$ (non-stationary models for Y_t):
  - The past Y_t's determine short term as well as long term predictions, the width of prediction intervals keeps increasing with lead time g.