Ch 5+: Review of ARIMA "mean" parameters and log-transformations

Motivating example for review on mean parameters:

- Suppose Y_t follows an ARIMA$(p, 1, q)$ model, and $W_t = Y_t - Y_{t-1}$.
- If $E(W_t) = \mu \neq 0$, what does that imply for Y_t?
- To discuss:
  - Review: How to formulate/interpret/simulate/estimate/forecast ARMA(p, q) models with non-zero mean μ,
  - How to formulate/interpret/simulate/estimate/forecast ARIMA$(p, 1, q)$ models with non-zero mean μ for $(Y_t - Y_{t-1})$.

ARMA(p, q) models constant term θ_0

- A stationary ARMA(p, q) model can be written compactly as

$$\phi(B)Y_t = \theta_0 + \theta(B)e_t,$$

with constant term θ_0 and AR and MA characteristic polynomials

$$\begin{aligned}\phi(x) &= 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p, \\ \theta(x) &= 1 - \theta_1 x - \theta_2 x^2 - \ldots - \theta_q x^q.\end{aligned}$$

- Or equivalently

$$\begin{aligned}Y_t &= \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}.\end{aligned}$$

- With $E(e_t) = 0$, it follows that

$$E(Y_t) = E(\theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p})$$

and because Y_t is stationary, $E(Y_t) = \mu$, a constant given by:

$$\mu = \theta_0/(1 - \phi_1 - \ldots - \phi_p).$$

Rewriting ARMA(p, q) models with constant term θ_0

- Instead of

$$\begin{aligned} Y_t &= \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}, \end{aligned}$$

  we can also write the ARMA model as

$$\begin{aligned} Y_t - \mu &= \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}. \end{aligned}$$

- Two ways to verify that this expression is correct:
  - Just plug in $\theta_0 = \mu(1 - \phi_1 - \ldots - \phi_p)$ in the first expression
  - Start with $X_t \sim ARMA(p, q)$ with $E(X_t) = 0$:

    $$\phi(B)X_t = \theta(B)e_t,$$

    and define $Y_t = X_t + \mu$:

    $$\phi(B)(Y_t - \mu) = \theta(B)e_t.$$

- How to estimate μ for a given time series?

Estimating μ and forecasting Y_t for the ARMA(p, q) model

- If $E(Y_t) = \mu \neq 0$, μ is included in the likelihood function, and we can obtain the MLE for μ (see Ch.7).
- The MLE for μ is used for forecasting Y_{t+g} (see Ch.9).
- A note on reading R output:

```
phis <- 0.5
X.t <- arima.sim(model = list(order = c(1,0,1),
ar = phis, ma = 0.8), n = 500)
mu <- 3
Y.t <- X.t + mu
> arima(Y.t, order = c(1,0,1), method="ML")
         ar1     ma1  intercept
      0.4857  0.8027     3.0354
```

- Does "intercept" refer to μ or θ_0?

```
> theta0 <- mu*(1-sum(phis)); theta0
[1] 1.5
> mu
[1] 3
```

ARIMA(p, d, q) models constant term θ_0

- An ARIMA(p, d, q) model can be written compactly as

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

with constant term θ_0 and AR and MA characteristic polynomials

$$\begin{aligned} \phi(x) &= 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p, \\ \theta(x) &= 1 - \theta_1 x - \theta_2 x^2 - \ldots - \theta_q x^q. \end{aligned}$$

- Or equivalently, for $W_t = (1 - B)^d Y_t$

$$\begin{aligned} W_t &= \theta_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}, \end{aligned}$$

- It follows that

$$E(W_t) = E(\theta_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p})$$

and because W_t is stationary, $E(W_t) = \mu = \theta_0/(1 - \phi_1 - \ldots - \phi_p)$.

- What is $E(Y_t)$ if $E(W_t) = \mu$?

Example: $E(Y_t)$ for IMA(1,1) model with $\theta_0 \neq 0$

- For the IMA(1,1) model, if $E(W_t) = E(Y_t - Y_{t-1}) = \mu$

$$
\begin{aligned}
(1-B)Y_t &= \theta_0 + e_t - \theta e_{t-1}, \\
Y_t - Y_{t-1} &= \mu + e_t - \theta e_{t-1}, \\
Y_t &= \mu + e_t - \theta e_{t-1} + Y_{t-1}.
\end{aligned}
$$

- Substituting the expression for Y_{t-1}, Y_{t-2}, etc we find

$$
\begin{aligned}
Y_t &= \mu + e_t - \theta e_{t-1} + Y_{t-1}, \\
&= \mu + e_t - \theta e_{t-1} + (\mu + e_{t-1} - \theta e_{t-2} + Y_{t-2}), \\
&= 2\mu + e_t + (1-\theta)e_{t-1} - \theta e_{t-2} + Y_{t-2}, \\
&\ldots \\
&= t\mu + e_t + (1-\theta)e_{t-1} + \ldots + (1-\theta)e_1 - \theta e_0 + Y_0.
\end{aligned}
$$

- Suppose $Y_0 = 0$, then $E(Y_t) = t \cdot \mu$.

$E(Y_t)$ in an ARIMA$(p, 1, q)$ model with $\theta_0 \neq 0$

- More generally, for an ARIMA$(p, 1, q)$ with $W_t = Y_t - Y_{t-1}$, with $Y_0 = 0$, we find that if $E(W_t) = \mu$ then:

$$
\begin{aligned}
E(Y_t) &= E(W_t + Y_{t-1}), \\
&= \mu + E(Y_{t-1}), \\
&= \mu + \mu + E(Y_{t-2}), \\
\ldots & \\
&= t \cdot \mu + E(Y_0), \\
&= t \cdot \mu.
\end{aligned}
$$

- Even more generally (Ch. 5), $\theta_0 \neq 0$ in an ARIMA(p, d, q) model results in a mean function for Y_t which is a deterministic polynomial of degree d.

Estimating μ and forecasting Y_t for the ARIMA$(p, 1, q)$ model

- Maximum likelihood estimates for all ARIMA model parameters, including μ can be obtained as usual, based on the likelihood function for W_t.
- The MLE for μ is used for forecasting Y_{t+g} (see Ch.9).
  - E.g., for IMA(1,1) use

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

- You can obtain an estimate for μ in an ARIMA(p,1,q) model using the “arima” function in R but it is less straightforward than doing so using the “Arima” function from the “forecast” package... so let’s check that one out!

Example simulation/estimation/forecast for ARIMA(1,1,1) process

- How to simulate an ARIMA(1,1,1) process with $\theta_0 \neq 0$?
- Steps:
  - Get $X_t \sim ARMA(1,1)$ with mean zero.
  - Get $W_t = X_t + \mu = Y_t - Y_{t-1}$.
  - Fix $Y_0 = 0$ and get $Y_t = W_t + Y_{t-1}$.

R-code for simu and estimation

```
library(forecast)
mu <- 0.5
phis <- -0.8
Wzeromean <- arima.sim(mode = list(ma = -0.5, ar = phis,
 order = c(1,0,1)), n=200)
W.t <- Wzeromean + mu
Y.t <- diffinv(W.t, xi = 0) # xi is starting value Y_0
mod <- Arima(Y.t, order = c(1,1,1), include.drift = TRUE,
 method="ML")
> summary(mod)
          ar1       ma1    drift
      -0.7809   -0.4860   0.4844
```

- The drift term refers to μ (NOT θ_0).

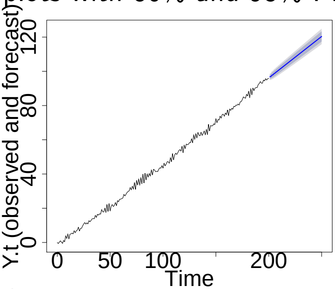
Forecasting in R using "forecast"

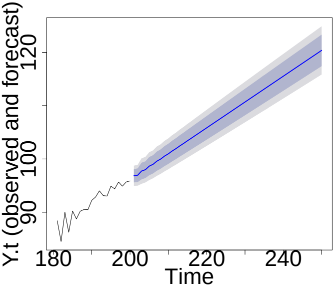
R-code (main points)

```
mod <- Arima(Y.t,
 order = c(1,1,1),
 include.drift = TRUE,
 method="ML")
fcast <- forecast(mod, h=50)
plot(fcast)
plot(fcast, include = 20)

> fcast$mean[50]-fcast$mean[49]
[1] 0.4843754
> coef(mod)['drift']
   drift
0.484378
```

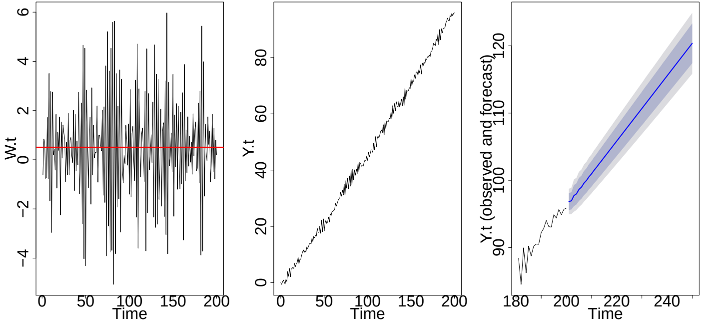
Nice plots with 80% and 95% PIs!





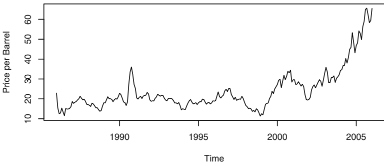
Summary

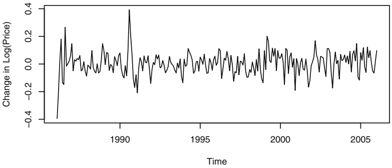
- Suppose Y_t follows an ARIMA$(p, 1, q)$ model, and $W_t = Y_t - Y_{t-1}$.
- If $E(W_t) = \mu \neq 0$, $E(Y_t)$ is a linear function of μ, which we can estimate using ML estimation, and incorporate in the forecast.
- However, do note the implication of including $\mu \neq 0$: decide whether a time trend should be included in the forecast or not!



Log-transformations: Oil price example

- Remember the oil price data: using a log-transform and differencing of Y_t resulted in a "nice looking series".
- Let's discuss some motivation for using a log-transform (to determine when such a transform would be appropriate).
- Assume that Y_t is positive and note that log refers to the natural log-transform unless otherwise stated.
- Motivation 0 to use a log-transform: a log-transform guarantees that forecasts for Y_t will be positive.





Motivation 1 for using log-transformations

- Suppose $E(Y_t) = \mu_t$ and $\sqrt{Var(Y_t)} = \mu_t \sigma$,
  - i.e. the standard deviation of the series is proportional to the level of the series,

  and μ_t changes roughly exponentially ($\mu_t \approx \beta_0 \exp(\beta_1 \cdot t)$).
- Then

$$
\begin{aligned}
\log(Y_t) &\approx \log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t} \text{ (Taylor appr.)}, \\
E(\log(Y_t)) &\approx \log(\mu_t) \approx \log(\beta_0) + \beta_1 \cdot t, \\
Var(\log(Y_t)) &\approx \sigma^2.
\end{aligned}
$$

- Thus a log-transform of Y_t may make the variance of the series constant, and taking differences would solve the non-constant mean problem.

Motivation 2 for using log-transformations

- Suppose you are monitoring Y_t (e.g. sales) that tends to have relatively stable percentage changes, denoted by $X_t \cdot 100\%$:

$$Y_t = (1 + X_t)Y_{t-1}.$$

- Then

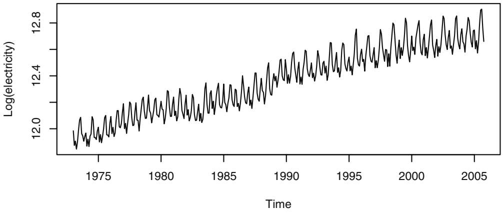
$$\begin{aligned} \nabla \log(Y_t) &= \log(Y_t) - \log(Y_{t-1}), \\ &= log(1 + X_t), \\ &\approx X_t, \end{aligned}$$

  where the approximation is reasonable if the percentage changes are small, e.g. $X_t \cdot 100\%$ less than 20%.
- If X_t is indeed relatively stable, $\nabla \log(Y_t)$ may be well-modeled by a stationary process.

Example: Electricity

- Plot: Electricity generated in the USA (by month).
- Why is this series not stationary?
- Answer: variance seems to increase with the level, and there is a time trend.

Exhibit 5.9 **Time Series Plot of Logarithms of Electricity Values**




```
> plot(log(electricity),ylab='Log(electricity)')
```

Example: Electricity

- Approach: Take log-transform, and difference $\log(Y_t)$.
- Now it may be possible to use an ARMA model to model the series.