Econometrics 2b: Handout #7

Matthew J. Lindquist*

April 2009

1 Estimating ARIMA Models

The goal of ARIMA analysis is a parsimonious representation of the process governing the residual. You should use only enough AR and MA terms to fit the properties of the residuals correctly.

1.1 The Box-Jenkins Methodology

This methodology can be applied to stationary series only. So, you must first deal with unit roots and stochastic seasons.

1. Identification
   (a) Address seasonality, s
   (b) Determine order of integration, d
   (c) Find appropriate values of p and q
2. Estimation
   (a) Pure $AR(p)$ models can be estimated (consistently) using OLS, non-linear OLS or maximum likelihood.
   (b) Pure $MA(q)$ models can be estimated (consistently) using non-linear OLS or maximum likelihood.
   (c) ARIMA models, $AR(p)$ models and $MA(q)$ models can all be estimated (consistently) using non-linear OLS and maximum likelihood.
      i. The **arima** function in $\mathbb{R}$ uses maximum likelihood (see **?arima** for more info).

*Department of Economics, Stockholm University. Matthew.Lindquist@ne.su.se

note: Consistency is an asymptotic property (i.e., it hold for large samples). This does not guarantee unbiasedness in small samples when estimating ARIMA models! We will discuss this topic in more detail later.

1. Diagnostics

   (a) model specification
   (b) autocorrelations
   (c) arch
   (d) normality

- **Example 1** *US GDP*
- In an earlier example I identified $s = 0$, $d = 1$, $p = 1$ and $q = 0$.
- Estimate an ARIMA(1,1,0) model for US GDP:

```
> arima(output, order = c(1, 1, 0))
Call:
arima(x = output, order = c(1, 1, 0))
Coefficients:
 ar1
 0.7079
s.e. 0.1292
sigma^2 estimated as 15959: log likelihood = -181.82, aic = 367.65
```

- Alternatively, we could have estimated an AR(1) process for $\Delta output$ using OLS:

```
> ar1 <- lm(diff(output)[2:29] ~ diff(output)[1:28] -1)
> summary(ar1)
—
Coefficients:         Estimate   Std. Error   t value   Pr(>|t|)
diff(output)[1:28]    0.7140     0.1374       5.198     1.79e-05 ***
—
```

- Note how close the estimates of the AR(1) component are to each other; 0.7079 $\approx$ 0.7140.
- Estimate the ARIMA(1,1,0) $\times$ SAR(4) model for Swedish quarterly GDP:

> **gdp <− ts(gdp, start = 1993, frequency = 4)**
> **arima(gdp, order = c(1, 1, 0), seasonal = list(order = c(0, 1, 0)))**

```
Coefficients:    ar1
               -0.4860
               s.e. 0.1453
sigma^2 estimated as 4.361e-32: log likelihood = 1179.19, aic = -2354.38
```

- Alternatively we could estimate an AR(1) process for $\Delta\Delta_4 output$ using OLS:

> **ar1s4 <− lm(diff(diff(gdp, 4))[2:34] ~ diff(diff(gdp,4))[1:33] -1)**

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)
diff(diff(gdp, 4))[1:33]	-0.3446	0.1641	-2.1	0.0437 *

—

1.1.1 Diagnostics

1. Are the residuals from our ARIMA model uncorrelated? Use the Ljung-Box Portmannteau test.

   (a) In $\mathbb{R}$, use **Box.test()** in the package **stats**.

2. Are the residuals normally distributed? Use the Jarque-Bera test or the Shapiro-Wilk test.

   (a) In $\mathbb{R}$, use **jarque.bera.test()** in the package **tseries** or use **shapiro.test** in the package **stats**.

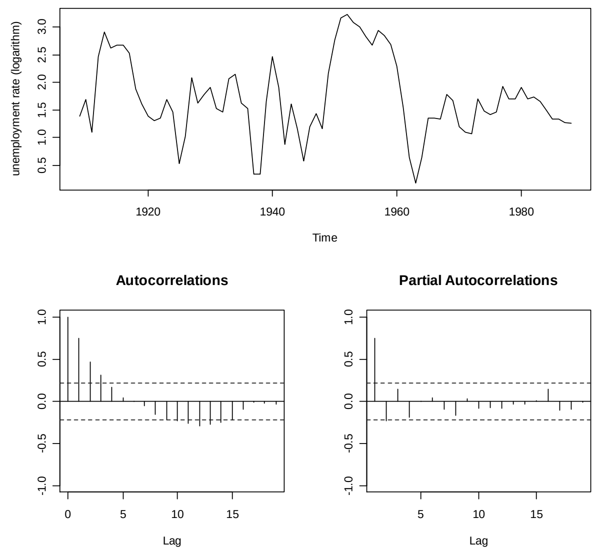
3. Test for ARCH and/or GARCH. More on this below

4. Over- and under-fitting a model. Use the AIC or BIC (reported by the **arima** command) to test if the model would be better off with more or less AR lags or with more or less MA lags.

Example 2 *Applying the Box-Jenkins methodology to the US unemployment rate, 1909-1988.*

```
## Pfaff's (2008) example starting on p. 16.
## Rcode 1.3
library(urca)
data(npext)
y <- ts(na.omit(npext$unemploy), start = 1909, end = 1988, frequency = 1)
op <- par(no.readonly = TRUE)
layout(matrix(c(1, 1, 2, 3), 2, 2, byrow = TRUE))
```

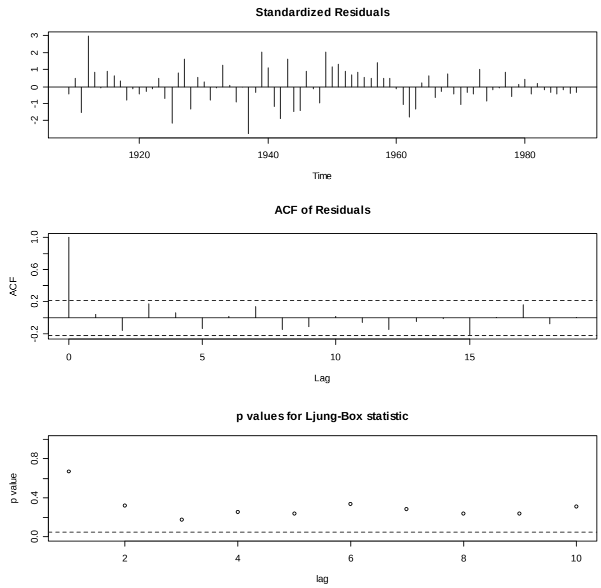
plot(y, ylab = "unemployment rate (logarithm)")
acf(y, main = "Autocorrelations", ylab = "", ylim = c(-1, 1))
pacf(y, main = "Partial Autocorrelations", ylab = "", ylim = c(-1, 1))
par(op)



Tentative ARMA(2,0)
arma20 <- arima(y, order = c(2,0,0))
arma20

c	ar1	ar2
1.6988	0.9297	-0.2356
(0.1079)	(0.1077)	(0.1586)
aic = 105.18		

res20 <- residuals(arma20)
Diagnostics
tsdiag(arma20)



Uncorrelatedness (at lag 20)
Box.test(res20, lag = 20, type = "Ljung-Box")
Normality
shapiro.test(res20)
Underfitting
arma10 <- arima(y, order = c(1, 0,0))
arma10

c	ar1
1.6885	0.7500
(0.1965)	(0.0719)
aic = 107.81	

Overfitting
arma30 <- arima(y, order =c(3, 0, 0))
arma30

c	ar1	ar2	ar3
1.6863	0.9727	-0.3949	0.1669
(0.1851)	(0.1101)	(0.1495)	(0.1103)
aic = 104.93			

- Syntax for an $ARIMA\,(p, d, q)$ x $SAR(s)$ model: **arima.season <- arima(x, order = c(p, d, q), seasonal = list(order = c(p, s, q))**

1.1.2 Using ARIMA Models to Make Forecasts

- One of the main goals of building ARIMA models is to forecast, or, predict future values of an economic variable. These forecasts are often used as benchmarks for comparison when constructing new, more complicated forecasting models.

- Given the available information set $I_T = \{Y_{-\infty}, ..., Y_T\}$, the optimal predictor, $Y_{T+h|T} \equiv E\left\{Y_{T+h|T} \mid I_T\right\}$, will be chosen so that it minimizes the expected quadratic prediction error

$$\min E\left\{\left(Y_{T+h} - \widehat{Y}_{T+h|T}\right)^2 \mid I_T\right\}.$$

- Derive the optimal predictor for the following $AR(1)$ process

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

- Given this process, $Y_{T+1} = \theta Y_T + \varepsilon_{T+1}$ and the optimal, one-step-ahead predictor is given by

$$\begin{aligned} Y_{T+1|T} &= E\{Y_{T+1} \mid Y_T, ..., Y_{T-\infty}\} \\ &= E\{(\theta Y_T + \varepsilon_{T+1}) \mid Y_T, ..., Y_{T-\infty}\} \\ &= \theta Y_T + E\{\varepsilon_{T+1} \mid Y_T, ..., Y_{T-\infty}\} \\ &= \theta Y_T. \end{aligned}$$

- Given this process, $Y_{T+2} = \theta Y_{T+1} + \varepsilon_{T+2}$ and the optimal, two-step-ahead predictor is given by

$$\begin{aligned} Y_{T+2|T} &= E\{Y_{T+2} \mid Y_T, ..., Y_{T-\infty}\} \\ &= E\{\theta Y_{T+1} + \varepsilon_{T+2} \mid Y_T, ..., Y_{T-\infty}\} \\ &= \theta E\{Y_{T+1} \mid Y_T, ..., Y_{T-\infty}\} \\ &= \theta E\{\theta Y_T + \varepsilon_{T+1} \mid Y_T, ..., Y_{T-\infty}\} \\ &= \theta * \theta Y_T = \theta^2 Y_T. \end{aligned}$$

- In general, the optimal predictor of an $AR(1)$ process is

$$Y_{T+h|T} = \theta^h Y_T.$$

- When the mean of the $AR(1)$ process is not zero, i.e. when

$$Y_t = \mu + \theta Y_{t-1} + \varepsilon_t,$$

  the optimal predictor is

$$Y_{T+h|T} = \mu + \theta^h (Y_T - \mu).$$

- Derive the optimal predictor for the following $MA(1)$ process

$$Y_t = \varepsilon_t + \alpha\varepsilon_{t-1}.$$

- Given this process, $Y_{T+1} = \varepsilon_{T+1} + \alpha\varepsilon_T$ and the optimal one-step-ahead predictor is given by

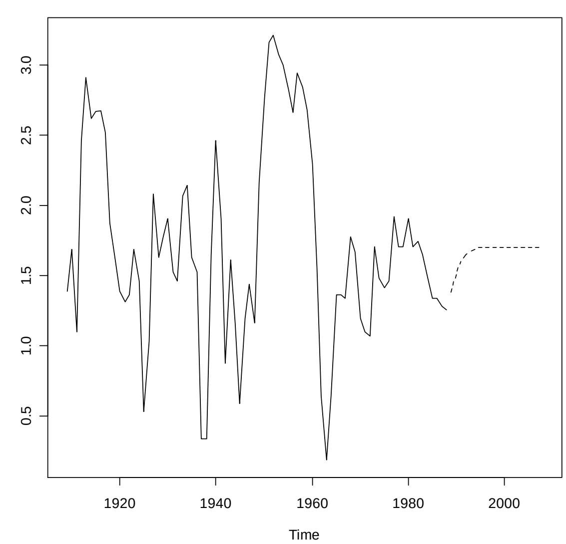
$$\begin{aligned} Y_{T+1|T} &= E\{Y_{T+1} \mid Y_T, ..., Y_{T-\infty}\} \\ &= E\{\varepsilon_{T+1} + \alpha\varepsilon_T \mid Y_T, ..., Y_{T-\infty}\} \\ &= \alpha E\{\varepsilon_T \mid Y_T, ..., Y_{T-\infty}\} \\ &= \alpha\varepsilon_T. \end{aligned}$$

- Given this process, $Y_{T+2} = \varepsilon_{T+2} + \alpha\varepsilon_{T+1}$ and the optimal one-step-ahead predictor is given by

$$\begin{aligned} Y_{T+2|T} &= E\{Y_{T+2} \mid Y_T, ..., Y_{T-\infty}\} \\ &= E\{\varepsilon_{T+2} + \alpha\varepsilon_{T+1} \mid Y_T, ..., Y_{T-\infty}\} \\ &= \alpha E\{\varepsilon_{T+1} \mid Y_T, ..., Y_{T-\infty}\} = 0. \end{aligned}$$

1.1.3 Forecasting With $\mathbb{R}$

- Return to the example above in which we examine the US unemployment rate, 1909 - 1988. Assume that our preferred model is **arma20 <- arima(y, order = c(2,0,0))**.
- Now use this model to predict the unemployment rate 20 years forward in time: **arma20.predict <- predict(arma20, n.ahead = 20)**.
- Plot the result: **ts.plot(y, arma20.predict$pred, lty = 1:2)**



- Now, let's re-estimate the model and hold back the last 4 observations. Then make a forecast for the last 4 years and compare the forecast to the actual data.

```
y2 <- ts(na.omit(npext$unemploy), start = 1909, end = 1984, frequency = 1)
arma20b <- arima(y2, order = c(2,0,0))
arma20b.predict <- predict(arma20b, n.ahead = 4)
ts.plot(y, arma20b.predict$pred, lty = 1:2, lwd = 3)
```

1.1.4 Forecast Error

- Let us examine the forecast error associated with a one-step ahead predictor of Y_t from the following , rather general model

$$\mathbf{Y}_T = \boldsymbol{\beta}'\mathbf{X}_T + \varepsilon_T$$

where $T = 1, 2, 3,...t$ and where

$$\varepsilon_T \sim IID\ N(0, \sigma^2).$$

- The one-step ahead predictor of Y_t is

$$\mathbf{Y}_{t+1|T} = \boldsymbol{\beta}'\mathbf{X}_{t+1} + \varepsilon_{t+1}$$

and our estimate of the one-step ahead predictor is

$$\widehat{\mathbf{Y}}_{t+1|T} = \mathbf{b}'\mathbf{X}_{t+1}.$$

- The one-step ahead prediction error $e_{t+1|T}$ from this model can be calculated as

$$e_{t+1|T} = \mathbf{Y}_{t+1} - \widehat{\mathbf{Y}}_{t+1|T} = (\boldsymbol{\beta} - \mathbf{b})' \, \mathbf{X}_{t+1} + \varepsilon_{t+1}.$$

- The variance of the one-step ahead prediction error is

$$\begin{aligned} Var\left[e_{t+1|T}\right] &= Var\left[(\boldsymbol{\beta} - \mathbf{b})' \, \mathbf{X}_{t+1}\right] + Var\left[\varepsilon_{t+1}\right] \\ &= Var \quad \underbrace{\left[(\boldsymbol{\beta} - \mathbf{b})' \, \mathbf{X}_{t+1}\right]}_{\text{parameter estimate uncertainty}} \quad + \quad \underbrace{\sigma^2}_{\text{error uncertainty}} \\ &= \mathbf{X}_{t+1}' \left[\sigma^2 (\mathbf{X}_T' \mathbf{X}_T)^{-1}\right] \mathbf{X}_{t+1} + \sigma^2. \end{aligned}$$

- The variance of the h-step ahead prediction error is

$$Var\left[e_{t+h|T}\right] = \mathbf{X}_{t+h}' \left[\sigma^2 (\mathbf{X}_T' \mathbf{X}_T)^{-1}\right] \mathbf{X}_{t+h} + \sigma^2.$$

- In practice, one can compute

$$s^2 \left[1 + \mathbf{X}_{t+h}' \left[(\mathbf{X}_T' \mathbf{X}_T)^{-1}\right] \mathbf{X}_{t+h}\right].$$

- A third source of uncertainty arises when computing out of sample, dynamic forecasts, since we are using the predicted values of $\mathbf{X}_{t+h-1}, ..., \mathbf{X}_{t+1}$in order to predict $\mathbf{X}_{t+h}$. This type of uncertainty is usually to complicated to be dealt with and disappears asymptotically.

- The forecast interval for $\widehat{\mathbf{Y}}_{t+h|T}$ is given by $\widehat{\mathbf{Y}}_{t+h|T} \pm t_{\lambda/2} se(e_{t+h|T})$ where λ is the chosen level of significance.

Example: Derive a 95% confidence interval for the 3-step ahead predictor of the following $AR(1)$ process

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

where $\theta < |1|$ and where

$$\varepsilon_t \sim IID \ N(0, \sigma^2).$$

Ignore all parameter estimate uncertainty!

step 1: Derive the optimal 3-step ahead predictor. This was derived above. It is equal to $\theta^3 Y_t$.

step 2: The 3-step ahead prediction error, $e_{t+3|t}$, is defined as

$$e_{t+3|t} = Y_{t+3} - \widehat{Y}_{t+3|t}.$$

step 3: The variance of $e_{t+3|t}$ is equal to

$$\begin{aligned}
V\left\{e_{t+3|t}\right\} &= V\left\{Y_{t+3}-\widehat{Y}_{t+3|t}\right\} = V\left\{\theta Y_{t+2}+\varepsilon_{t+3}-\theta^3 Y_t\right\} \\
&= V\left\{\theta\left(\theta Y_{t+1}+\varepsilon_{t+2}\right)+\varepsilon_{t+3}-\theta^3 Y_t\right\} \\
&= V\left\{\theta\left(\theta\left(\theta Y_t+\varepsilon_{t+1}\right)+\varepsilon_{t+2}\right)+\varepsilon_{t+3}-\theta^3 Y_t\right\} \\
&= V\left\{\theta^3 Y_t+\theta^2\varepsilon_{t+1}+\theta\varepsilon_{t+2}+\varepsilon_{t+3}-\theta^3 Y_t\right\} \\
&= V\left\{\theta^2\varepsilon_{t+1}+\theta\varepsilon_{t+2}+\varepsilon_{t+3}\right\} = V\left\{\theta^2\varepsilon_{t+1}\right\}+V\left\{\theta\varepsilon_{t+2}\right\}+\left\{V\varepsilon_{t+3}\right\} \\
&= \left(1+\theta^2+\theta^4\right)\sigma^2.
\end{aligned}$$

step 4: The forecast interval for $\widehat{\mathbf{Y}}_{t+3|t}$ is given by $\widehat{\mathbf{Y}}_{t+3|t} \pm t_{2.5}\sqrt{\left(1+\theta^2+\theta^4\right)\sigma^2}$.

- Let's now place a confidence interval around our forecast of US unemployment:

ts.plot(y, arma20b.predict$pred, arma20b.predict$pred
+ 1.96*arma20b.predict$se, arma20b.predict$pred
- 1.96*arma20b.predict$se, lty = 1:4, lwd = 3)

1.1.5 Evaluating forecasts

Let's compare the forecast from model A to that from model B.

- Regression method. Let $T = 150$.

1. Estimate each model using the first 100 observations only.
2. Calculate a one-step ahead forecast for each model, $\widehat{Y}_{101|100,i}$ where $i \in \{A, B\}$.
3. Estimate each model using the first 101 observations.
4. Calculate a one-step-ahead forecast for each model, $\widehat{Y}_{102|101,i}$ where $i \in \{A, B\}$..
5. Continue this until you have two series of 50 one-step-ahead forecasts, $\left\{\widehat{Y}_A\right\}_{101}^{150}$ and $\left\{\widehat{Y}_B\right\}_{101}^{150}$.
6. Run the following test regressions

$$\{Y_A\}_{101}^{150} = \alpha_A + \beta_A \left\{\widehat{Y}_A\right\}_{101}^{150} + \{\varepsilon\}_1^{50}$$

$$\{Y_B\}_{101}^{150} = \alpha_B + \beta_B \left\{\widehat{Y}_B\right\}_{101}^{150} + \{\varepsilon\}_1^{50}.$$

7. If these forecasts are unbiased, we should be able to impose the restrictions that $\alpha = 0$ and $\beta = 1$.
8. Use an F-test to see if these restrictions hold. Compare the significance level of these two F-tests.

- Calculate the mean squared prediction errors. Let $T = 150$.

   1. Calculate a series of one-step-ahead prediction errors, $\{e\}_{101}^{150} = \{Y\}_{101}^{150} - \left\{\widehat{Y}\right\}_{101}^{150}$.
   2. $MSPE = 1/50 \sum \{e^2\}_{101}^{150}$

- Calculate the root mean squared (prediction) error: RMSE $= \sqrt{1/50 \sum \{e^2\}_{101}^{150}}$
- Calculate the mean absolute error: $MAE = 1/50 \sum \left|\{e\}_{101}^{150}\right|$
- Maybe you are super risk averse and want to minimize the largest possible loss?
- Model accuracy can differ at different horizons. Some models may be better at longer horizons. Some may be better at shorter horizons. Thus, depending on our end purpose, we may also want to compare the t-step-ahead prediction values of different models.
- Their are number of alternative tests including the Granger-Newbold test, the Diebold-Mariano test and Thiel's U-statistic. Consult a text on forecasting for more info.