CHAPTER 2: STATIONARY TIME-SERIES MODELS

Stochastic Difference Equation Models

- Say we have sequence of observations of a variable y over time

$$\{y_0, y_1, y_2, \ldots, y_t\}$$

- A stochastic process describes the probability structure of these observations
- The elements of an observed time series are realizations of this stochastic process
- y is a deterministic variable if there is some r for which $p(y_t = r) = 1$

Stationarity

- A stochastic process is called weakly (covariance) stationary when the mean, the variance and the covariance structure of the process is time independent and finite, that is

$$E(y_t) = \mu < \infty$$

$$var(y_t) = \gamma_0 < \infty$$

$$cov(y_t, y_{t-s}) = E(y_t - \mu)E(y_{t-s} - \mu) = \gamma_{|t-s|} \quad \forall t \neq s$$

- The last condition states that the covariance between y_t and y_s depends only on the displacement $|t - s| = j$
- The set of autocovariances γ_j, $j = 0, \pm 1, \pm 2, ...$ is called the autocovariance function of a stationary process

A White-Noise Process

- It is the simplest stationary process. It has a mean of zero, a constant variance, and is uncorrelated with all other realizations. Formally,

$$E(\varepsilon_t) = 0$$

$$Var(\varepsilon_t) = \gamma_0 = \sigma_\varepsilon^2$$

$$E(\varepsilon_t \varepsilon_{t-j}) = \gamma_j = 0 \text{ for } j = 1, 2, ...$$

ARMA Models

- For a stochastic process, if stability conditions hold, stationarity conditions are satisfied
- *Wold's Decomposition Theorem:* Any discrete stationary covariance time series process $\{y_t\}$ can be expressed as the sum of two uncorrelated processes

$$y_t = d_t + u_t$$

where d_t is purely deterministic and u_t is a purely indeterministic process:

$$u_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i}$$

where $\sum_{i=0}^{\infty} (\beta_i)^2 < \infty$ is necessary for stationarity and ε_t is a white-noise process (it is conventional to define $\beta_0 = 1$)

- Taking d_t as a constant, reparametrizing the infinite order indeterministic process to a finite one (below we discuss the way doing it)

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}. \tag{5}$$

  If all the characteristic roots of (5) are all in the unit circle, $\{y_t\}$ is called an ARMA model for $\{y_t\}$

- The autoregressive part:

$$\sum_{i=1}^{p} a_i y_{t-i}$$

- The moving average part:

$$\sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$$

- If the homogeneous part of the difference equation contains p lags and the moving average part contains q lags, the model is called an $ARMA(p,q)$ model

- Using the lag operator, (5) can be written as

$$(1 - a_1 L - ... - a_p L^p) y_t = a_0 + (1 + \beta_1 L + ... + \beta_p L^p) \varepsilon_t$$

  or

$$A(L) y_t = a_0 + \beta(L) \varepsilon_t$$

  where $A(L)$ and $\beta(L)$ are called the AR polynomial and the MA polynomial, respectively

- Moreover, stability ensures that

$$y_t = \frac{a_0}{A(L)} + \frac{\beta(L)}{A(L)} \varepsilon_t = A^{-1}(L) a_0 + C(L) \varepsilon_t$$

  Hence, Wold decomposition implies that stationary series have an infinite moving average representation

Why use ARMA Models?

- Suppose the true data generating process is as follows

$$y_t = c + x_t + \varepsilon_t + \varepsilon_{t-1}$$

  where x_t exogenous regressors and ε_t is a white noise process

- The above process suggests that shocks to y_t lasts for two periods

- If you estimate the above model through a regression

$$y_t = c + x_t + e_t$$

  then there is a serial correlation in the error terms,

$$cov(e_t, e_{t-1}) = cov(\varepsilon_t + \varepsilon_{t-1}, \varepsilon_{t-1} + \varepsilon_{t-2}) = var(\varepsilon_{t-1}) \neq 0$$

- Serial correlation violates the standard assumption of regression theory that error terms are uncorrelated

  - Reported standard errors and t-statistics are invalid

- Things can even get worse: Suppose the true data generating process is as follows

$$y_t = c + y_{t-1} + x_t + \varepsilon_t + \varepsilon_{t-1}$$

  and if you estimate it by using the following model

$$y_t = c + y_{t-1} + x_t + e_t$$

  regressors and the error terms become correlated

$$cov(y_{t-1}, e_t) = cov(y_{t-1}, \varepsilon_t + \varepsilon_{t-1}) \neq 0$$

  - In this case, OLS estimates are biased and inconsistent

E-views Application

wfcreate (wf=income process) u 1000

series y=0

series e=nrnd

smpl @first+1 @last

y=0.7*y(-1)+e+0.5*e(-1)

ls y y(-1)

Dependent Variable: Y
Method: Least Squares
Date: 01/08/16 Time: 00:25
Sample: 2 1000
Included observations: 999

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.865888	0.015835	54.68208	0.0000

R-squared	0.747475	Mean dependent var	0.200655
Adjusted R-squared	0.747475	S.D. dependent var	2.102149
S.E. of regression	1.056369	Akaike info criterion	2.948553
Sum squared resid	1113.684	Schwarz criterion	2.953464
Log likelihood	-1471.802	Hannan-Quinn criter.	2.950420
Durbin-Watson stat	1.428442		

- Serial correlation is a common occurrence in time series data because the data is ordered (over time) so that the effect of shocks could easily last for more then one period. Moreover, each new arriving observation is stochastically depending on the previously observed

- If accounted for, this time dependence is useful. It allow us topredict future values of series
- ARMA models accounts for this time dependence so that the model captures all of the relevant structure
- In what follows:

  1. We will use stationary data
     * So that the mean, variance, and autocorrelations of the series can be obtained based on the single set of realizations
  2. We will identify the data generating process (type of ARMA model)
     * For this, we will use autocorrelation and partial autocorrelation functions

- There are methods to make the nonstationary data stationary, such as differencing, detrending, and filtering. We defer this discussion to the next chapters

The Autocorrelation (ACF) and Partial Autocorrelation (PACF) Functions

- In the case of stationary processes, the *autocorrelation coefficient* at lag j, denoted by ρ_j, is defined as the correlation between y_t and y_{t-j}:

$$\rho_j = \frac{cov(y_t, y_{t-s})}{\sqrt{var(y_t)}\sqrt{var(y_{t-j})}} = \frac{\gamma_j}{\gamma_0}, \qquad j = 0, \pm 1, \pm 2, ...$$

- The plot of ρ_j against j (for $j \geq 1$) is called correlogram
  - The properties of autocorrelation function (ACF) are:

$$\begin{aligned} \rho_0 &= 1 \\ |\rho_j| &\leq 1 \end{aligned}$$

- The *partial autocorrelation coefficient*, on the other hand, measures the linear association between y_t and y_{t-j} adjusted for the effects of the intermediate values $y_{t-1}, ..., y_{t-j+1}$
- Therefore, it is the coefficient a_j in the linear regression model:

$$y_t = a_0 + a_1 y_{t-1} + ... + a_j y_{t-j} + e_t$$

- We will examine the ACF and PACF functions of three extreme cases of $ARMA(p,q)$ model

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \qquad (5)$$

  1. White Noise Process ($a_0 = p = q = 0$)
  2. Pure Autoregressive Process ($q = 0$)
  3. Pure Moving Average Process ($p = 0$)

1- White Noise Process

- The simplest ARMA model is a white noise process

$$y_t = \varepsilon_t$$

which has no memory

- We can easily show that this is a stationary process

$$E(y_t) = 0 < \infty$$

$$E(y_t - \mu)^2 = \sigma_\varepsilon^2 < \infty$$

$$E(y_t - \mu)E(y_{t-j} - \mu) = 0 \;\; \text{for } j = 1, 2, ...$$

all of which are time independent and finite terms

- Its autocorrelation (ACF) and partial autocorrelation (PACF) functions are:

$$\rho_j = 0 \quad if \quad j \neq 0$$

$$a_j = 0 \quad if \quad j \neq 0$$

———————————-

E-views Application

wfcreate (wf=income process) u 5000

series e=nrnd
series y=e
graph aa y
show aa
y.correl(10)

Date: 01/08/16 Time: 09:25
Sample: 1 5000
Included observations: 5000

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.022	-0.022	2.3303	0.127
		2	-0.025	-0.025	5.3546	0.069
		3	0.010	0.009	5.8307	0.120
		4	0.013	0.013	6.7007	0.153
		5	-0.023	-0.022	9.2899	0.098
		6	0.013	0.013	10.202	0.116
		7	0.020	0.019	12.221	0.094
		8	-0.024	-0.022	15.134	0.057
		9	0.001	0.001	15.141	0.087
		10	0.001	-0.001	15.144	0.127

Hence, a white noise process in an ARMA(0,0) model; it shows no history dependence in any form

———————————-

2- Pure Autoregressive Process

- *Example: AR(1) Process*

$$\begin{aligned} y_t &= a_0 + a_1 y_{t-1} + \varepsilon_t \\ (1 - a_1 L) y_t &= a_0 + \varepsilon_t \\ y_t &= \frac{a_0}{1 - a_1 L} + \frac{\varepsilon_t}{1 - a_1 L} \end{aligned}$$

- If $\mid a_1 \mid < 1$, the last equation can be written as

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

- AR process has an infinite memory so that it can be written as a collection of past shocks
- Notice that y_t is a stationary process
  - The mean of the sequence is finite and time invariant

$$E(y_t) = \frac{a_0}{1 - a_1}$$

  - The variance is finite and time independent

$$Var(y_t) = \gamma_0 = E[(\varepsilon_t + a_1 \varepsilon_{t-1} + a_1^2 \varepsilon_{t-2} + ...)^2] = \frac{\sigma_\varepsilon^2}{1 - a_1^2}$$

  - The covariance between y_t and y_{t-s} is constant and time invariant for all t and $t - s$

$$Cov(y_t, y_{t-j}) = \gamma_j = E[(a_1^j y_{t-j} + \sum_{i=0}^{j} a_1^i \varepsilon_{t-i})(y_{t-j})] = a_1^j \gamma_0$$

- Its autocorrelation (ACF) and partial autocorrelation (PACF) functions are as follows:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = a_1^j \qquad a_j = \left\{ \begin{array}{cc} a_1 & j = 1 \\ 0 & j > 1 \end{array} \right\}$$

- Thus, ACF converges to zero geometrically
- PACF is zero at lag 2 and greater

- *Example: AR(p) Process*

$$\begin{aligned} y_t &= a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t \\ A(L) y_t &= a_0 + \varepsilon_t \end{aligned}$$

- The process is stable and stationary if the characteristic roots of the first equation all lie inside the unit circle (the roots of the polynomial $(1 - \sum a_i L^i)$ must lie outside the unit circle)

- Under stationarity, the above process can be written as

$$y_t = \frac{a_0}{1 - \sum_{i=1}^{p} a_i} + \frac{\varepsilon_t}{1 - \sum_{i=1}^{p} a_i L^i}$$
$$= \frac{a_0}{1 - \sum_{i=1}^{p} a_i} + \sum_{i=1}^{\infty} \beta_i \varepsilon_{t-i}$$

- AR(p) process has an infinite memory as well
- You can use Yule-Walker Conditions to calculate variance and autocovariance of a stationary process
- Basically, for a stationary AR(p) process

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + \varepsilon_t \qquad (1^{**})$$

  - by multiplying (1**) with y_t

$$\gamma_0 = a_1\gamma_1 + a_2\gamma_2 + ... + a_p\gamma_p + \sigma_\varepsilon^2$$

  - by multiplying (1**) with y_{t-1}

$$\gamma_1 = a_1\gamma_0 + a_2\gamma_1 + ... + a_p\gamma_{p-1}$$

  and so no. Finally, by multiplying (1**) with y_{t-p}

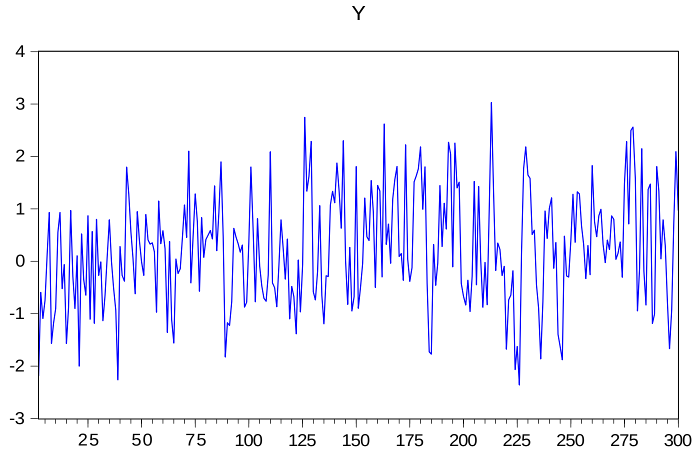
$$\gamma_p = a_1\gamma_{p-1} + a_2\gamma_{p-2} + ... + a_p\gamma_0$$

- *Results:*
- Autoregressive processes have an exponentially declining ACF, whether they are AR(1), AR(2), etc. (nonstationary series also have an ACF that remains significant for some lags)
- The partial autocorrelation of an AR(p) process is zero at lag p+1 and greater. Hence, partial autocorrelations are useful in identifying the order of an autoregressive model (p) from the data

---

E-views Application

```
wfcreate (wf=income process) u 300
series e=nrnd
series y=0
smpl @first+1 @last
y=0.5*y(-1)+e
graph aa y
show aa
```



y.correl(10)

Date: 01/08/16 Time: 09:28
Sample: 2 300
Included observations: 299

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.346	0.346	36.168	0.000
		2	0.133	0.015	41.550	0.000
		3	0.090	0.044	44.002	0.000
		4	-0.034	-0.091	44.358	0.000
		5	-0.092	-0.067	46.933	0.000
		6	-0.070	-0.016	48.448	0.000
		7	-0.069	-0.026	49.908	0.000
		8	0.001	0.052	49.909	0.000
		9	0.000	-0.016	49.909	0.000
		10	0.023	0.023	50.079	0.000

```
ls y c ar(1)
series res=resid
res.correl(10)
```

Date: 01/08/16 Time: 09:28
Sample: 2 300
Included observations: 299

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.0166	0.897
		2	-0.001	-0.002	0.0173	0.991
		3	0.076	0.076	1.7719	0.621
		4	-0.043	-0.042	2.3402	0.673
		5	-0.077	-0.078	4.1481	0.528
		6	-0.027	-0.035	4.3793	0.625
		7	-0.061	-0.056	5.5228	0.596
		8	0.029	0.039	5.7894	0.671
		9	-0.010	-0.011	5.8183	0.758
		10	-0.009	-0.009	5.8415	0.828

3- Pure Moving Average Process

$$y_t = a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} = a_0 + \beta(L)\varepsilon_t$$

- Stationarity:

$$E(y_t) = a_0$$

$$Var(y_t) = E[(\beta_0\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + ...)^2] = \sigma^2 \sum_{i=0}^{q} \beta_i^2$$

$$cov(y_t, y_{t-s}) = \sigma^2(\beta_s + \beta_{s+1}\beta_1 + \beta_{s+2}\beta_2 + ...)$$

all of which are time independent and also finite terms as long as $\sum \beta_i^2$ is finite

- *Example: MA(1) Process*

$$y_t = a_0 + \varepsilon_t + \beta\varepsilon_{t-1}$$

- Stationarity Check:

$$E(y_t) = a_0$$

- Yule-Walker Conditions:

$$\gamma_0 = Var(y_t) = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_t + \beta\varepsilon_{t-1})] = (1+\beta^2)\sigma^2$$

$$\gamma_1 = Cov(y_t, y_{t-1}) = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-1} + \beta\varepsilon_{t-2})] = \beta\sigma^2$$

$$\gamma_2 = Cov(y_t, y_{t-2}) = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-2} + \beta\varepsilon_{t-3})] = 0$$

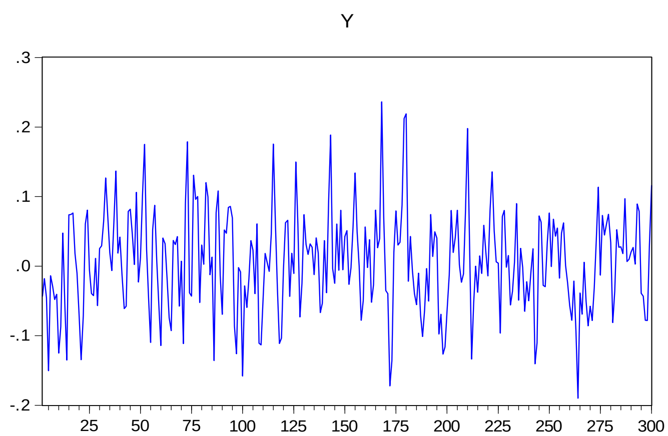
- Hence, the autocorrelation of an MA(q) process is zero at lag q+1 and greater
- As a result, autocorrelations are useful in identifying the order of an autoregressive model

- *Note:* Moving average processes have a geometrically (or oscillatory) declining PACF; hence, they cannot be used in identifying the order of an autoregressive model

---

E-views Application

```
wfcreate (wf=income process) u 300
series e=nrnd/15
series y=0
smpl @first+1 @last
y=e+0.5*e(-1)
graph aa y
show aa
```




```
y.correl(10)
```

Date: 01/08/16 Time: 09:31
Sample: 2 300
Included observations: 299

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.356	0.356	38.333	0.000
		2	-0.036	-0.187	38.725	0.000
		3	-0.042	0.047	39.253	0.000
		4	-0.006	-0.015	39.265	0.000
		5	0.043	0.053	39.839	0.000
		6	0.045	0.008	40.453	0.000
		7	-0.075	-0.105	42.192	0.000
		8	-0.095	-0.019	44.997	0.000
		9	-0.043	-0.016	45.583	0.000
		10	0.012	0.021	45.629	0.000

```
ls y c ma(1)
series res=resid
res.correl(10)
```

Date: 01/08/16 Time: 09:32
Sample: 2 300
Included observations: 299

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.003	-0.003	0.0020	0.964
		2	-0.021	-0.021	0.1412	0.932
		3	-0.036	-0.036	0.5265	0.913
		4	-0.002	-0.003	0.5284	0.971
		5	0.019	0.018	0.6410	0.986
		6	0.064	0.063	1.8870	0.930
		7	-0.075	-0.074	3.6135	0.823
		8	-0.060	-0.057	4.7112	0.788
		9	-0.017	-0.016	4.8033	0.851
		10	-0.013	-0.020	4.8547	0.901

**AN IMPORTANT NOTE

- Both E-VIEWS and STATA models ARMA structure in the disturbances; that is, ARMA(1,1) model is defined as

$$y_t = c + \mu_t$$

where

$$\mu_t = \rho\mu_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

which can be estimated in e-views as **-ls y ar(1) ma(1)-**

- Note that this process can be also be written as autocorrelation in the dependent variable

$$y_t = \beta + \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

where

$$\beta = (1-\rho)c$$

which can be estimated as **-ls y y(-1) ma(1)-**

- In this case, it does not matter how we model the AR structure
- However, things are different for ARMAX models, where

$$y_t = c + \gamma X_t + \mu_t$$

and

$$\mu_t = \rho\mu_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

- These two processes imply that

$$y_t = (1-\rho)c + \rho y_{t-1} + \gamma X_t - \rho\gamma X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

- Hence, it is better if you specify lagged dependent variable separate from the error structure!

Stationarity Restrictions for an ARMA(p,q) Model

- For the general $ARMA(p, q)$ model, the condition for stationary includes stationary conditions for both AR and MA parts of the model
- Using lag operators

$$(1 - \sum_{i=1}^{p} a_i L^i) y_t = a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$$

  - The roots of the polynomial $(1 - \sum a_i L^i)$ must lie outside the unit circle
  - $\sum \beta_i^2$ must be finite

- Mixed (ARMA) processes typically show exponential declines in both the ACF and the PACF
- *Notes:*
- If you are estimating a higher-order AR process, EViews requires you to include all lower-order terms
  - If you simply type ar(3) and omit other terms, this forces the estimate of ar(1) and ar(2) to zero
- After from estimation, you may want to ensure that the residuals from the estimated model mimic a white-noise process

Box-Jenkins Model Selection

- Box–Jenkins (1976) popularized a three-stage method aimed at selecting an appropriate model for the purpose of estimating and forecasting a univariate time series

  1 *Identification:* Examine the time plot of the series, the autocorrelation function, and the partial correlation function

  2 *Estimation*: Fit the model by OLS or any other alternative methods

  3 *Diagnostic Checking:* Ensure that the residuals from the estimated model mimic a white-noise process

Alternative Methods of Checking for Serial Correlation (you are not responsible from this part)

- The last two columns reported in the correlogram are the Ljung-Box Q-statistics and their p-values. The Q-statistic at lag is a test statistic for the null hypothesis that there is no autocorrelation up to order and is computed as

$$Q = T \sum_{k=1}^{s} r_k^2$$

  high sample autocorrelations lead to large values of Q

- Breusch-Godfrey (null hypothesis of no serial correlation)

E-views Application

```
wfcreate (wf=income process) u 300
series e=nrnd
series y=0
smpl @first+1 @last
y=e+0.5*e(-1)
ls y c
y.correl(10)
```

Date: 01/08/16 Time: 09:33
Sample: 2 300
Included observations: 299

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.356	0.356	38.297	0.000
		2	-0.090	-0.248	40.753	0.000
		3	-0.040	0.108	41.234	0.000
		4	0.013	-0.042	41.284	0.000
		5	0.079	0.108	43.212	0.000
		6	0.161	0.109	51.199	0.000
		7	0.055	-0.042	52.140	0.000
		8	-0.053	-0.014	53.004	0.000
		9	-0.079	-0.063	54.955	0.000
		10	0.055	0.118	55.892	0.000

```
auto(1)
```

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	43.18304	Prob. F(1,297)	0.0000
Obs*R-squared	37.95524	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 01/08/16 Time: 09:34
Sample: 2 300
Included observations: 299
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000700	0.056877	-0.012315	0.9902
RESID(-1)	0.356475	0.054247	6.571380	0.0000

R-squared	0.126941	Mean dependent var	6.94E-17
Adjusted R-squared	0.124001	S.D. dependent var	1.050792
S.E. of regression	0.983487	Akaike info criterion	2.811242
Sum squared resid	287.2723	Schwarz criterion	2.835994
Log likelihood	-418.2807	Hannan-Quinn criter.	2.821149
F-statistic	43.18304	Durbin-Watson stat	1.816415
Prob(F-statistic)	0.000000		

```
ls y c ar(1)
series res=resid
```

ls res c
auto(1)

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	2.390362	Prob. F(1,297)	0.1232
Obs*R-squared	2.387245	Prob. Chi-Square(1)	0.1223

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 01/08/16 Time: 09:35
Sample: 2 300
Included observations: 299
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000347	0.056653	-0.006133	0.9951
RESID(-1)	0.089564	0.057930	1.546080	0.1232

R-squared	0.007984	Mean dependent var	3.12E-17
Adjusted R-squared	0.004644	S.D. dependent var	0.981889
S.E. of regression	0.979607	Akaike info criterion	2.803335
Sum squared resid	285.0098	Schwarz criterion	2.828087
Log likelihood	-417.0986	Hannan-Quinn criter.	2.813242
F-statistic	2.390362	Durbin-Watson stat	1.947693
Prob(F-statistic)	0.123150		

Parsimony

- Incorporating additional coefficients to an ARMA model will necessarily increase fit of the model at a cost of reducing degrees of freedom
- A parsimonious model fits the data well without incorporating any needless coefficients
- Box and Jenkins argue that parsimonious models produce better forecasts than overparameterized models
- If different ARMA models may have similar properties, such as AR(1) and MA(∞), the AR(1) model is the more parsimonious model and is preferred

Model Selection Criteria

- We never know the true data-generating process
- There exist various model selection criteria that trade-off a reduction in the sum of squares of the residuals for a more parsimonious model
- The two most commonly used model selection criteria are the Akaike Information Criterion (AIC) and the Schwartz Bayesian Criterion (SBC)

$$AIC = T * ln(\text{sum of squared residuals}) + 2n$$

$$SBC = T * ln(\text{sum of squared residuals}) + n * ln(T)$$

- where n = number of parameters estimated (p + q + possible constant term)
- T = number of usable observations (notice that when you estimate a model using lagged variables, some observations are lost)

- Notice that increasing the number of regressors increases n but reduces the sum of squared residuals (SSR)
- Ideally, the AIC and SBC will be as small as possible
- EViews and SAS report values for the AIC and SBC using

$$AIC^{*} = -2ln(L)/T + 2n/T$$

$$SBC^{*} = -2ln(L)/T + nln(T)/T$$

where L is the maximized value of the log of the likelihood function

- For a normal distribution, $-2ln(L) = Tln(2\pi) + Tln(\sigma^2) + (1/\sigma^2)(SSR)$

Invertibility

- Invertibility conditions regard the moving average part
- y_t is invertible if it can be represented by a finite-order or convergent autoregressive process
- For instance, MA(1) model:

$$y_t = \varepsilon_t - \beta\varepsilon_{t-1}$$

is invertible if $\mid \beta \mid < 1$

$$\frac{y_t}{(1-\beta L)} = \varepsilon_t$$

or

$$y_t + \beta y_{t-1} + \beta^2 y_{t-2} + \beta^3 y_{t-3} + ... = \varepsilon_t \tag{46}$$

- An AR process is stationary if it is inverted, but not all (stationary) MA process can be invertible. Suppose we have the following model

$$y_t = \varepsilon_t - \varepsilon_{t-1}$$

This process is a stationary one, yet, if $\beta = 1$, (46) finds

$$y_t + y_{t-1} + y_{t-2} + y_{t-3} + ... = \varepsilon_t$$

Clearly, the autocorrelations and partial autocorrelations between y_t and y_{t-j} will never decay

- Moreover, if $\mid \beta \mid > 1$, y_t can be written as

$$\frac{y_t}{-\beta L[1-(\beta L)^{-1}]} = \varepsilon_t$$

or

$$-(\beta L)^{-1}[y_t + \beta^{-1}y_{t+1} + (\beta^{-1})^2 y_{t+2} + (\beta^{-1})^3 y_{t+3} + ...] = \varepsilon_t,$$

or

$$-\beta^{-1}y_{t+1} - (\beta^{-1})^2 y_{t+2} - (\beta^{-1})^2 y_{t+3} - ... = \varepsilon_t,$$

implying that y_t is a function of future values of y_t which is not useful for forecasting

- If invertibility condition is not imposed on a model with an MA component, different sets of MA parameter values give rise to the same autocorrelation function
- Hence, invertibility provides uniqueness of the ACF and PACF functions