Course materials subject to Copyright

Statement on class recording

University of New South Wales, School of Economics
Financial Econometrics
Tutorial 4 solutions

1. Estimating MA

Consider an invertible MA(1) model: $y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$, $\varepsilon_t \sim iid \text{ WN}(0, \sigma^2)$.

Using the MA(1) equation for $t = 1,2,3$, we find

$\varepsilon_1 = y_1 - \mu - \theta_1 \varepsilon_0$,

$\varepsilon_2 = y_2 - \mu - \theta_1 \varepsilon_1 = y_2 - \mu - \theta_1 (y_1 - \mu - \theta_1 \varepsilon_0)$,

$\varepsilon_3 = y_3 - \mu - \theta_1 \varepsilon_2 = y_3 - \mu - \theta_1 [y_2 - \mu - \theta_1 (y_1 - \mu - \theta_1 \varepsilon_0)]$,

which are *non-linear* functions of the parameters (μ, θ_1). Obviously, the same can be done for $t = 4,5, \ldots$. The shock ε_t is interpreted as the 1-step ahead forecast error. Hence, the "least squares" principle here is to choose the values of (μ, θ_1) to minimise the sum of squared 1-step ahead forecast errors: $\text{SSE}(\mu, \theta_1) = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2$. Note that the SSE is a *non-linear* function of (μ, θ_1) and its minimisation is carried out by numerical search algorithms in practice. For a long series of y_t, say $T = 1000$, the above applies straightforwardly. We note that the expression of ε_3 involves the term $\theta_1^3 \varepsilon_0$. In general, the expression of ε_t involves the term $\theta_1^t \varepsilon_0$. For large T, whether or not ε_0 is known is unimportant, because the term $\theta_1^t \varepsilon_0$ converges to zero quickly as t increases as long as the MA is invertible (where $|\theta_1| < 1$). However that, when the MA(1) is not invertible (implies $|\theta_1| \geq 1$), the term $\theta_1^t \varepsilon_0$ cannot be ignored.

MLE is based on maximizing the likelihood of observing y_t. An important technical caveat here: the probability to observe the exact values y_t is zero (separately of jointly) because we deal with continuous random variable. The likelihood is proportional to the joint probability that in each time the random variable will fall within the range $y_t \pm \delta$, where δ is a very (infinitesimally) small positive number, the joint probability is

$$P(y_1 - \delta < Y_1 < y_1 + \delta \text{ and } y_2 - \delta < Y_2 < y_2 + \delta \text{ and} \ldots \; y_T - \delta < Y_T < y_T + \delta) = CDF(y_1 + \delta, y_2 + \delta, \ldots, y_T + \delta) - CDF(y_1 - \delta, y_2 - \delta, \ldots, y_T - \delta)$$

The same small δ is used in the definition of the derivative:

$$\lim_{\delta \to 0} \frac{CDF(y_1 + \delta, y_2 + \delta, \ldots, y_T + \delta) - CDF(y_1 - \delta, y_2 - \delta, \ldots, y_T - \delta)}{2\delta} = f(y_1, y_2, \ldots, y_T),$$

where $f(y_1, y_2, \dots, y_T)$ is the pdf, derivative of the CDF.

Hence for small positive δ:

$$P(y_1 - \delta < Y_1 < y_1 + \delta, y_2 - \delta < Y_2 < y_2 + \delta, \dots, \ y_T - \delta < Y_T < y_T + \delta) \approx 2\delta f(y_1, y_2, \dots, y_T)$$

Therefore, maximizing $P(\,)$ is equivalent to maximizing $f(\,)$ because $\delta > 0$.

Obviously you do not have to know how to derive the part above, but hopefully it helps with understanding. So, intuitively, we want to maximise the probability that any random sample (a collection of random variables) falls as close as possible to the observed realization of the random sample by choosing the parameters which characterise the specific family of distributions (we typically use Normal with just two parameters). The parameters of the distribution are related to the parameters of the model for a given model. If there is no such relation the parameters of the model are not identified.

Coming back from this illuminating digression the joint pdf is given by $f(y_3, y_2, y_1)$. We can factorise it:

$$f(y_3, y_2, y_1) = f(y_3, y_2|\Omega_1)f(y_1) = f(y_3|\Omega_2)f(y_2|\Omega_1)f(y_1)$$

Note: that *f* is positive, it can be zero in regions where the random variable does not have a support, but this is not the case for normal distribution; *[more exotically *f* may be allowed to turn negative as an explanation for *negative energy* and *dark matter*, but we will leave all the fun with this to physicists]. Natural log *ln* transformation can be used for simplicity because *ln* is an increasing function. Namely, maximizing $f(y_3, y_2, y_1)$ is equivalent to maximizingln $f(y_3, y_2, y_1)$.

So $y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ and $\varepsilon_t \sim iid$ WN $N(0, \sigma^2)$. Hence

The conditional (and unconditional) distribution of y_t is Normal because y_t is just a linear combination of $\varepsilon_t, \varepsilon_{t-1}$ which are Normally distributed by assumption. Find the mean and the variance of y_t-s and write down the pdfs.

$$y_1 \sim N(\mu, (1+\theta_1^2)\sigma^2); \ f(y_1) = (2\pi(1+\theta_1^2)\sigma^2)^{-1/2} \, exp\{-\frac{(y_1 - \mu)^2}{2(1+\theta_1^2)\sigma^2}\}$$

$$y_2|\Omega_1 \sim N(\mu + \theta_1\varepsilon_1, \sigma^2); \ f(y_2|\Omega_1) = (2\pi\sigma^2)^{-1/2} \, exp\{-\frac{(y_2 - \mu - \theta_1\varepsilon_1)^2}{2\sigma^2}\}$$

$$y_3|\Omega_2 \sim N(\mu + \theta_1\varepsilon_2, \sigma^2); \ f(y_3|\Omega_2) = (2\pi\sigma^2)^{-\frac{1}{2}} \, exp\{-\frac{(y_3 - \mu - \theta_1\varepsilon_2)^2}{2\sigma^2}\}$$

Note that when computing y_1 we treated ε_0 as a random variable. In fact this causes a problem, because when we move one period from *t=1* ahead we cannot distinguish between ε_0 and ε_1 (each of them cannot be separately identified from the data). Therefore it is important to condition on a specific value of ε_0 *[it can be integrated out of the likelihood if we want to assume just a distribution of ε_0, but the integration is not simple].

So, we will maximize $\ln f(y_3, y_2, y_1|\varepsilon_0)$ and

$$y_1|\varepsilon_0 \sim N(\mu + \theta_1\varepsilon_0, \sigma^2);\ f(y_1|\varepsilon_0) = (2\pi\sigma^2)^{-1/2}\, exp\{-\frac{(y_1 - \mu - \theta_1\varepsilon_0)^2}{2\sigma^2}\}$$

Now as we move forward ε_1 can be identified. In fact, we can rewrite all three densities as:

$$\begin{aligned}\ln f(y_3, y_2, y_1|\varepsilon_0) &= \ln[\, f(y_3|y_2, y_1, \varepsilon_0) f(y_2|y_1, \varepsilon_0) f(y_1|\varepsilon_0)] \\ &= \ln f(y_3|y_2, y_1, \varepsilon_0) + \ln f(y_2|y_1, \varepsilon_0) + \ln f(y_1|\varepsilon_0) \\ &= 3\ln\left[(2\pi\sigma^2)^{\frac{1}{2}}\right] \\ &\quad - \frac{1}{2\sigma^2}\left((y_1 - \mu - \theta_1\varepsilon_0)^2 + (y_2 - \mu - \theta_1\varepsilon_1)^2 + (y_3 - \mu - \theta_1\varepsilon_2)^2\right) \\ &= 3(2\pi\sigma^2)^{\frac{1}{2}} - \frac{1}{2\sigma^2}(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2)\end{aligned}$$

Note that in this case in $f(y_2|y_1, \varepsilon_0)$ we do not have to condition on ε_1, it can be discovered if we know y_1, ε_0 (using the equations in the beginning of this question). Similarly ε_2^2 and ε_3^2, are defined already.

Note that in terms of estimating (μ, θ_1) minimizing *SSE* is equivalent to maximizing ln likelihood (conditional on ε_0).

3. Find the unconditional variance of ARMA(1,1) model

$$y_t = \mu + \theta_1\varepsilon_{t-1} + \varphi y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iid\ \mathrm{WN}(0, \sigma^2)$$

Using stationarity and taking into account dependence between ε_{t-1} and y_{t-1}

$$var(y_t) = \theta_1^2\sigma^2 + \varphi^2 var(y_t) + 2\theta_1\varphi cov(\varepsilon_t, y_t) + \sigma^2$$

$cov(\varepsilon_t, y_t) = \sigma^2$ which can be found from $cov(\varepsilon_t, \mu + \theta_1\varepsilon_{t-1} + \varphi y_{t-1} + \varepsilon_t) =$
$E(\varepsilon_t(\mu + \theta_1\varepsilon_{t-1} + \varphi y_{t-1} + \varepsilon_t)) - E(\varepsilon_t)E(y_t) = 0 + 0 + 0 + \sigma^2 - 0 = \sigma^2$

The zeros are due to the fact that $E(\varepsilon_t) = 0$, $E(\varepsilon_t\varepsilon_{t-1})$=0 (white noise property), $E(\varepsilon_t y_{t-1})$=0 (white noise property) and $E(\varepsilon_t\varepsilon_t)$=$\sigma^2$.

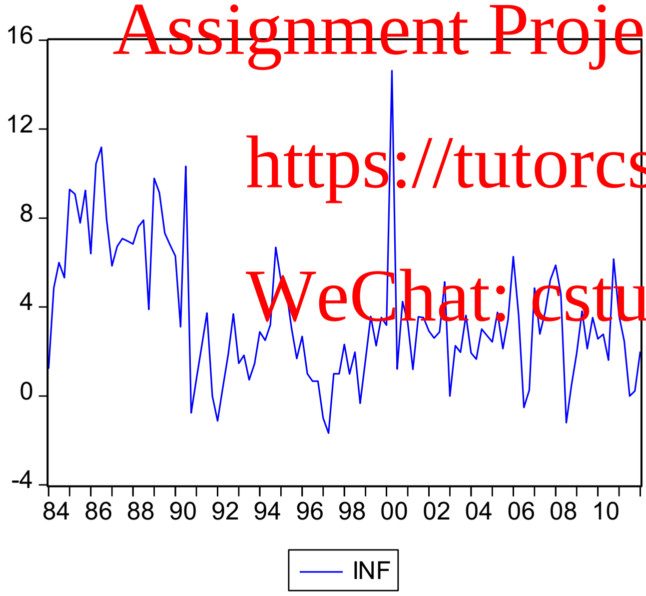
DETOUR: Note that white noise property does NOT imply that $E(\varepsilon_{t-1}y_t)$=0. We can show that $E(\varepsilon_{t-1}y_t) = (\theta_1 + \varphi)\sigma^2$ is this example.

Hence,

$$var(y_t) = \frac{\sigma^2(1+\theta_1^2+2\theta_1\varphi)}{1-\varphi^2} = \frac{\sigma^2(1-\varphi^2+\varphi^2+2\theta_1\varphi+\theta_1^2)}{1-\varphi^2} = \sigma^2\left(1+\frac{(\theta_1+\varphi)^2}{1-\varphi^2}\right)$$

4. [Box-Jenkins methodology]

(a) The time series plot of INF shows that the inflation was on average lower after 1990 than before 1990. You may use the unit-root test with or without a time trend (by selecting or not selecting **Trend and intercept** in the unit root test pop-up menu). From the testing results below, the null of a unit root is rejected (very small p-values). Hence INF can be regarded as a stationary (or trend stationary) time series.



Augmented Dickey-Fuller Unit Root Test on INF

Null Hypothesis: INF has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic based on SIC, MAXLAG=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.910993	0.0027
Test critical values:	1% level	-3.489117	
	5% level	-2.887190	
	10% level	-2.580525	

*MacKinnon (1996) one-sided p-values.

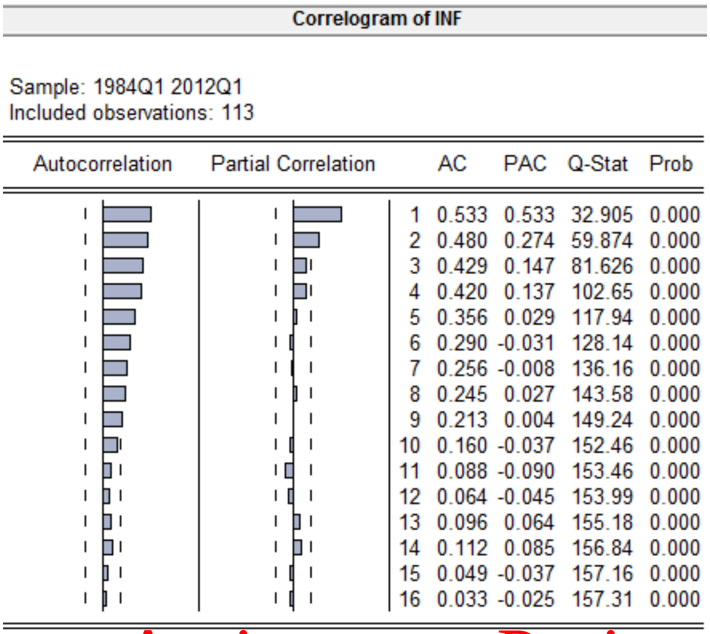
Augmented Dickey-Fuller Unit Root Test on INF

Null Hypothesis: INF has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.674579	0.0013
Test critical values:	1% level	-4.041280	
	5% level	-3.450073	
	10% level	-3.150336	

*MacKinnon (1996) one-sided p-values.

(b) Visually, the correlogram of INF shows a PAC cutoff at lag $j = 2$ with exponential decaying AC, which fits the cutoff pattern of an AR(2) model. On the other hand, the PAC may also be interpreted as an exponential decay. For this reason, we consider the following models as candidates: AR(2), ARMA(1,1), ARMA(2,1), ARMA(1,2) and ARMA(2,2).



Correlogram of INF

Sample: 1984Q1 2012Q1
Included observations: 113

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.533	0.533	32.905	0.000
		2	0.480	0.274	59.874	0.000
		3	0.429	0.147	81.626	0.000
		4	0.420	0.137	102.65	0.000
		5	0.356	0.029	117.94	0.000
		6	0.290	-0.031	128.14	0.000
		7	0.256	-0.008	136.16	0.000
		8	0.245	0.027	143.58	0.000
		9	0.213	0.004	149.24	0.000
		10	0.160	-0.037	152.46	0.000
		11	0.088	-0.090	153.46	0.000
		12	0.064	-0.045	153.99	0.000
		13	0.096	0.064	155.18	0.000
		14	0.112	0.085	156.84	0.000
		15	0.049	-0.037	157.16	0.000
		16	0.033	-0.025	157.31	0.000

(c) The AIC and SIC values for the models considered in (b) are listed in the table below. Clearly, both AIC and SIC select ARMA(1,1).

ARMA(p,q)	ARMA(2,0)	ARMA(1,1)	ARMA(2,1)	ARMA(1,2)	ARMA(2,2)
AIC	4.675253	4.657635	4.668168	4.673624	4.683174
SIC	4.747661	4.730044	4.764713	4.770169	4.803855

(d) The estimation output of ARMA(1,1) indicates that the roots of the AR polynomial and the MA polynomial are outside the unit circle (note that EViews reports inverted roots). There is no common root. The estimated ARMA coefficients are all statistically and economically significant. The residual-actual-fitted plot shows that the model fits the data quite well (also $\bar{R}^2 = 0.337$), although a few residual points are beyond the 2-standard deviation bands. The key of an ARMA model is to capture the dependence structure in the time series. The residual from an adequate model should not exhibit autocorrelation. The residual correlogram of the ARMA(1,1) confirm that there are no autocorrelations in the residuals (large p-values, which are adjusted for the degrees of freedom lost in estimating the ARMA parameters). Further the histogram of the residual indicates that the normality is strongly rejected, which explains why a few residual points are outside the 2-se bands in the residual plot. Overall, the ARMA(1,1) model is adequate in capturing the dependence structure in INF.

Dependent Variable: INF
Method: Least Squares

Sample: 1984Q1 2012Q1
Included observations: 113
Convergence achieved after 10 iterations
Backcast: 1983Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.600090	0.919616	3.914776	0.0002
AR(1)	0.881293	0.067676	13.02229	0.0000
MA(1)	-0.532092	0.116949	-4.549767	0.0000

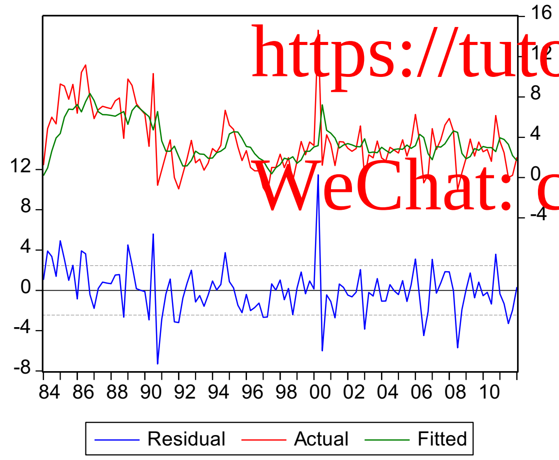
R-squared	0.349057	Mean dependent var	3.602539
Adjusted R-squared	0.337222	S.D. dependent var	3.011510
S.E. of regression	2.451706	Akaike info criterion	4.657635
Sum squared resid	661.1948	Schwarz criterion	4.730044
Log likelihood	-260.1564	F-statistic	29.49284
Durbin-Watson stat	1.987690	Prob(F-statistic)	0.000000

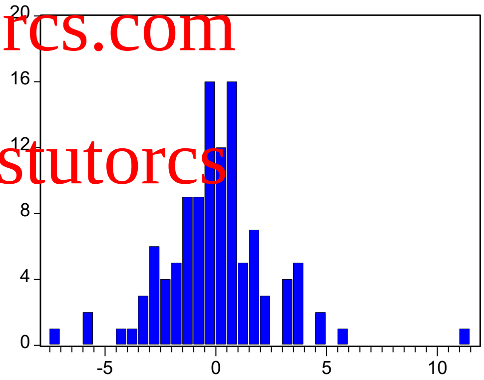
Inverted AR Roots	.88
Inverted MA Roots	.53

Correlogram of Residuals

Sample: 1984Q1 2012Q1
Included observations: 113
Q-statistic probabilities adjusted for 2 ARMA term(s)

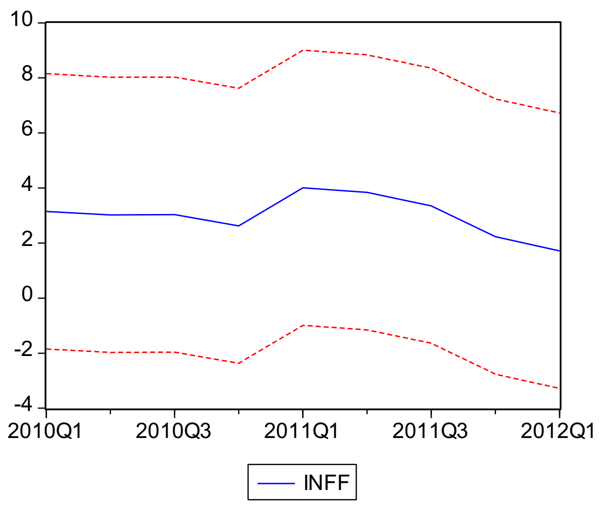
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.005	0.005	0.0030	
		2	0.036	0.036	0.1567	
		3	0.025	0.025	0.2300	0.632
		4	0.113	0.111	1.7414	0.419
		5	0.057	0.056	2.1365	0.545
		6	-0.003	-0.012	2.1379	0.710
		7	0.002	-0.007	2.1383	0.830
		8	0.061	0.047	2.5916	0.858
		9	0.075	0.065	3.2959	0.856
		10	0.021	0.018	3.3531	0.910
		11	-0.093	-0.101	4.4637	0.878
		12	-0.088	-0.109	5.4610	0.858
		13	0.011	-0.006	5.4772	0.906
		14	0.079	0.085	6.2890	0.901
		15	-0.037	-0.009	6.4715	0.927
		16	-0.044	-0.026	6.7356	0.944

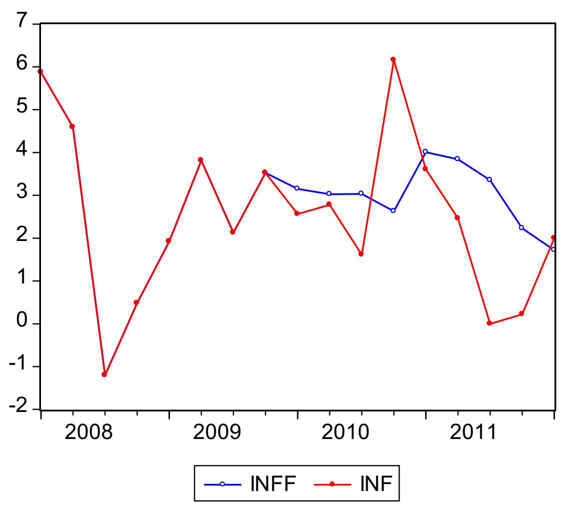




Series: Residuals	
Sample 1984Q1 2012Q1	
Observations 113	
Mean	0.035138
Median	-0.034075
Maximum	11.46643
Minimum	-7.301654
Std. Dev.	2.429461
Skewness	0.660686
Kurtosis	7.072407
Jarque-Bera	86.30622
Probability	0.000000

(e) For the forecast exercise, you should be able to see the graphs below, where INFF (blue curve) stands for the 1-step ahead forecasts from the ARMA(1,1) model and INF (red curve) is the actual. The second graph is from the command "**plot inff inf**". The first graph is from the **Forecast** menu, which includes the 2-se bands (dashed red lines). By comparison, the interval forecasts (2-se bands) in fact covered the actuals.


10
8
6
4
2
0
-2
-4
2010Q1
2010Q3
2011Q1
2011Q3
2012Q1
INFF


7
6
5
4
3
2
1
0
-1
-2
2008
2009
2010
2011
INFF
INF