DEPARTMENT OF ECONOMICS, MANAGEMENT AND QUANTITATIVE METHODS

Academic Year 2018-2019

B-74-3-B Time Series Econometics

Fabrizio Iacone

Discussion of Exercise Sheet 3

1.

This is the Conditional Residual Sum of Squares for a Conditional Maximum Likelihood estimate in a MA(1) model assuming that the disturbances ε_t are normally distributed and $\mu = 0$ (and imposing $\varepsilon_0 = 0$ as usual). (this is the answer to part ii.)

For part i.,

Using $\varepsilon_0 = 0$ for any θ, $\varepsilon_t(\theta) = y_t - \varepsilon_{t-1}(\theta)$

$y_1 = -0.4, \quad y_2 = 0.8, \quad y_3 = 0.6, \quad y_4 = -0.2$

and assuming $\varepsilon_0 = 0$ for the values $\theta = 0.5$, $\theta = -0.5$, $\theta = 0$.

$\varepsilon_t(\theta)$	$t = 1$	$t = 2$
$\theta = 1/2$	$-0.4 - 1/2 * 0 = -0.4$	$0.8 - 1/2 * (-0.4) = 1.0$
$\theta = 0$	$-0.4 - 0 * 0 = -0.4$	$0.8 - 0 * (-0.4) = 0.8$
$\theta = -1/2$	$-0.4 + 1/2 * 0 = -0.4$	$0.8 + 1/2 * (-0.4) = 0.6$

$\varepsilon_t(\theta)$	$t = 3$	$t = 4$
$\theta = 1/2$	$0.6 - 1/2 * 1 = 0.1$	$-0.2 - 1/2 * 0.1 = -0.25$
$\theta = 0$	$0.6 - 0 * 0.8 = 0.6$	$-0.2 - 0 * 0.6 = -0.2$
$\theta = -1/2$	$0.6 + 1/2 * 0.6 = 0.9$	$-0.2 + 1/2 * 0.9 = 0.25$

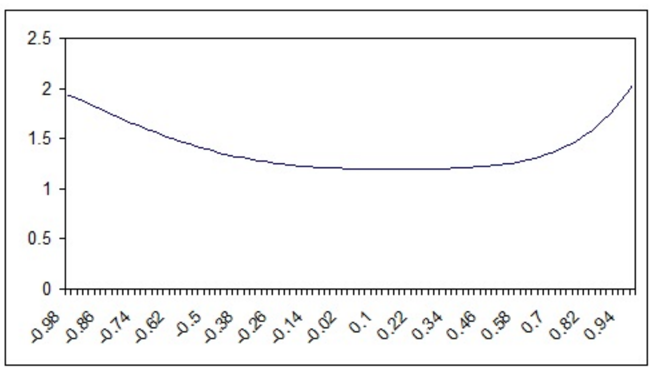
$\varepsilon_t^2(\theta)$	$t=1$	$t=2$	$t=3$	$t=4$
$\theta = 1/2$	$(-0.4)^2 = 0.16$	$1^2 = 1$	$0.1^2 = 0.01$	$(-0.25)^2 = 0.0625$
$\theta = 0$	$(-0.4)^2 = 0.16$	$0.8^2 = 0.64$	$0.6^2 = 0.36$	$(-0.2)^2 = 0.04$
$\theta = -1/2$	$(-0.4)^2 = 0.16$	$0.6^2 = 0.36$	$0.9^2 = 0.81$	$0.25^2 = 0.0625$

	$\sum_{t=1}^T \varepsilon_t^2(\theta)$
$\theta = 1/2$	$0.16 + 1 + 0.01 + 0.0625 = 1.2325$
$\theta = 0$	$0.16 + 0.64 + 0.36 + 0.04 = 1.2$
$\theta = -1/2$	$0.16 + 0.36 + 0.81 + 0.0625 = 1.3925$

iii.

This means that if we were to pick a conditional maximum likelihood estimate $\widehat{\theta}$ between the three candidates $1/2, 0, -1/2$, we would pick $\widehat{\theta} = 0$.

If we used the whole $[-0.98, 0.98]$ the estimate $\widehat{\theta}$ would be 0.14. The function $\sum_{t=1}^T \varepsilon_t^2(\theta)$ is



2. The estimation output is

Dependent Variable: Y
Sample: 1 99
Included observations: 99
Convergence achieved after 33 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.068594	0.157992	0.434161	0.6651
MA(1)	0.730583	0.084889	8.606327	0.0000
SIGMASQ	0.833208	0.096710	8.615568	0.0000

R-squared	0.390882	Mean dependent var	0.074318
Adjusted R-squared	0.378192	S.D. dependent var	1.175521
S.E. of regression	0.926955	Akaike info criterion	2.723719
Sum squared resid	82.48761	Schwarz criterion	2.802359
Log likelihood	-131.8241	Hannan-Quinn criter.	2.755537
F-statistic	30.80244	Durbin-Watson stat	1.840167
Prob(F-statistic)	0.000000		

Inverted MA Roots	-.73

We know that, for MA(1) model, the LS/ML estimate is such that

$$\sqrt{T}(\widehat{\theta} - \theta) \rightarrow_d N(0, 1 - \theta^2)$$

Under H_0, one feasible test statistic is therefore

$$\sqrt{T}\frac{(\widehat{\theta} - \theta)}{\sqrt{1 - \theta^2}} \rightarrow_d N(0, 1)$$

so, under $H_0 := \{\theta = 0.9\}$, this takes value

$$\sqrt{99}\frac{(0.730583 - 0.9)}{\sqrt{1 - 0.9^2}} = -3.867$$

and $| - 3.867| > 1.96$ so the null hypothesis is not rejected at the standard 5 % significance level. Alternatively, we can use a consistent estimate to estimate θ in the variance. In this case, a feasible test statistic is

$$\sqrt{T}\frac{(\widehat{\theta}-\theta)}{\sqrt{1-\widehat{\theta}^2}} \rightarrow_d N(0,1)$$

and

$$\sqrt{99}\frac{(0.730583-0.9)}{\sqrt{1-0.730583^2}} = -2.46869$$

As another alternative still, using the regression output, we could have calculated

$$\frac{(0.730583-0.9)}{0.084889} = -1.99$$

which again is a non-rejection of H_0.

Finally, we could have run the test directly using e-views: the Wald test gives

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	-1.995747	96	0.0488
F-statistic	3.983004	(1, 96)	0.0488
Chi-square	3.983004	1	0.0460

Null Hypothesis: C(2)-0.9=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-0.9 + C(2)	-0.169417	0.084889

Restrictions are linear in coefficients.

As we repeat the exercise for a MA(2) model, we estimated

Dependent Variable: Y
Method: ARMA Maximum Likelihood (OPG - BHHH)
Sample: 1 99
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.066712	0.172953	0.385720	0.7006
MA(1)	0.799120	0.121864	6.557489	0.0000
MA(2)	0.089314	0.108816	0.820784	0.4138
SIGMASQ	0.826064	0.113130	7.301906	0.0000
R-squared	0.396105	Mean dependent var		0.074318
Adjusted R-squared	0.377034	S.D. dependent var		1.175521
S.E. of regression	0.927818	Akaike info criterion		2.735568
Sum squared resid	81.78034	Schwarz criterion		2.840421
Log likelihood	-131.4106	Hannan-Quinn criter.		2.777991
Inverted MA Roots	-.13	-.66		

and the outcome of the test is

Wald Test:

Test Statistic	Value	df	Probability
F-statistic	2.693180	(2, 95)	0.0728
Chi-square	5.386361	2	0.0677

Null Hypothesis: C(2)=0.9, C(3)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-0.9 + C(2)	-0.100880	0.121864
C(3)	0.089314	0.108816

Restrictions are linear in coefficients.

Therefore, the null hypothesis is not rejected in this case.

This difference in outcomes seems puzzling: when we assumed $\theta_2 = 0$ and tested $\theta_1 = 0.9$ we rejected the hypothesis that a MA(1) model with $\theta_1 = 0.9$ was appropriate. However, when we estimated the MA(2), we did not rejected the hypothesis that a MA(1) model with $\theta_1 = 0.9$ was appropriate.

To understand why, check the estimated standard errors for $\widehat{\theta}_1$ in both the estimates: this is 0.084 for the MA(1) model, and 1.12 for the MA(2). This is not surprising: from the formulae for the variance - covariance matrix of the estimates, we know that the variance of $\widehat{\theta}_1$ should be $(1-\theta_1^2)/T$ when the MA(1) is estimated but $(1-\theta_2^2)/T$ when the MA(2) is estimated. If $\theta_2 = 0$ then the variance in the MA(2) model is much larger. This reflects the fact that information is used to estimate θ_2 as well, so we are less confident about θ_1 and this additional uncertainty makes us not reject the null hypothesis.

This example shows that we should estimate parsimonious models, as we gain less information from non-parsimonious models.

Finally, a note regarding the estimated standard errors from eviews. We know that for a MA(1) (for example) the asymptotic variance is $(1-\theta_2^2)/T$: eviews however does not use this bit of information, and estimates the variance as outer product of gradients. This is because obviously eviews whould have to change the formula for the asymptotic variance any time we change model, and therefore should have the formula (in terms of θ and ϕ for any possible ARMA modeland this is impossible. Using the outer product of the gradients gives a consistent estimate of the variance for any model, thus avoiding the problem.

3. Recall

Akaike information criterion $AIC = -2 \ln lik\,(p,q) + 2\,(p+q)$

Bayes information criterion $BIC = -2 \ln lik\,(p,q) + \ln T\,(p+q)$.

(p,q)	$\ln lik\,(p,q)$	AIC	BIC
$(1,0)$	-248.6914	499.38	502.68
$(0,1)$	-257.1481	516.30	519.59
$(1,1)$	-248.6750	501.35	507.95
$(0,2)$	-251.3668	506.73	513.33
$(2,0)$	-247.8323	499.66	506.26

So both the AIC and the BIC selected an AR(1) model.

NOTE 1: notice that the highest maximized log-likelihood is for AR(2). As it happens, the AR(2) nests the AR(1) (i.e, we can write the AR(1) as a restriction of the AR(2)), so we can compare them with a likelihood ratio test. If we tested the AR(1) against the AR(2) using a likelihood ratio test, $2\,(+248.6914 - 247.8323) = 1.7182$ so the null hypothesis that the additional parameter is 0 would not be rejected (at 5% size).

NOTE 2: I discussed both AIC and BIC to give an example. However, discussing only one of them would have been sufficient for a complete solution. In fact I do not recommend running both of them, as this could leave to conflicting results: suppose, for example, that AIC selected AR(2) and BIC selected AR(1), which one would you choose? We studied reasons to prefer AIC and reasons to prefer BIC. For example, if I prefer BIC because it gives consisent estimate of (p,q), then I should not use AIC, so it is not necessary to compute it.

NOTE 3: some candidates may note that, given the formula of the Information Criterion and the values of the maximised log-likelihoods, in this case it is clear that the recommended model can only be the AR(1) (best log-likelihood when $p+q=1$) or the AR(2) (best log-likelihood when $p+q=2$). This is very elegant and perfectly acceptable. By the way, at this point, as the AR(1) is nested in the AR(2), of course it is natural to compare them with a likelihood ratio test (although using an information criterion instead is also perfectly acceptable).

4.

When the model is correctly specified, the residuals estimate the original i.i.d. disturbances. The Portmanteau statistic

$$T \sum_{j=1}^{m} r_j^2 \to_d \chi^2_{m-(p+q)}$$

as $T \to \infty$, where p and q are the numbers of AR and MA parameters. in this case the Portmanteau statistic takes value $200\left(0.05^2 + (-0.07)^2 + 0.1^2\right) = 3.48$: we have $m = 3$, $p = 1$, $q = 1$, so the critical distribution is a χ^2_1. Taking the size as 5% as usual, the critical value is 3.84, so the hypothesis is not rejected, and we can conclude that the approximation is satisfactory.

5.

Maximised log-likelihoods are

MA(1): -131.8241

MA(2): -131.4106 Thus the MA(2) has higher maximised likelihood. However, we know that adding parameters always increase the likelihood, so maximising the likelihood does not deliver a consistent estimate. Comparing these with the information criteria,

BIC (Shwarz) are

MA(1): 2.802359

MA(2): 2.840421 so the BIC selected the MA(1) model.

NOTE: Notice that, as MA(1) and MA(2) are nested (i.e., we can write MA(1) as a restriction on parameters of MA(2)) we could compare them also using a Likelihood Ratio tests. Likelihood ratio tests are asymptotically equivalent to Wald tests, so we could use results from exercise 2 to conclude that MA(1) should be preferred. Indeed,using the likelihood ratio test (or the Wald test) would be the best thing to do (because these are statistical tests, and because the likelihood ratio test has nice power properties under some conditions): however, we compared MA(1) vs. MA(2) using the information criterion to familiarize ourselves with it.

The Portmanteau test on the residuals (using up to 10 autocorrelation) yields

Sample: 1 99
Included observations: 99
Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.070	0.070	0.5015	
		2	0.041	0.037	0.6772	0.411
		3	0.018	0.012	0.7097	0.701
		4	0.043	0.040	0.9062	0.824
		5	0.036	0.029	1.0423	0.903
		6	0.109	0.102	2.3120	0.805
		7	-0.035	-0.052	2.4417	0.875
		8	0.124	0.123	4.1390	0.764
		9	0.056	0.038	4.4863	0.811
		10	0.095	0.076	5.4905	0.790

From the P-value of the Q statistic, we conclude that the null hypothesis of no residual autocorrelation is not rejected. Thus, the MA(1) is an acceptable specification.