Athens University of Economics and Business
Department of Economics

Postgraduate Program - Master's in Economic Theory
Course: Econometrics II
Prof: Stelios Arvanitis
TA: Alecos Papadopoulos

Semester: Spring 2017-2018

March 29, 2018

Estimating an MA(1) process.

(this document is to be studied together with

"Tutorial 3: a GMM estimator in an MA(1) model.")

We are given a data sample $\{y_0, ..., y_T\}$ and we are told (in all honesty) that it comes from an $\text{MA}(1)$ process without drift,

$$y_t = \theta_0 \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}\left(\sigma_\varepsilon^2 = 1\right), \; \theta_0 \in \mathbb{R} \qquad [1]$$

We are also given that $E\left(\varepsilon^4\right) = \kappa_4 < +\infty$. We are asked to estimate the parameter θ_0. Note that we do *not* exclude the possibility that $\theta_0 = 0$.

I. An attempt at OLS estimation.

Looking at [1], it has all the looks of a regression equation, except for one tiny detail: the "regressor" is ε_{t-1}, on which we do not have data. Still, being hooked on OLS estimation, we want to find a way to estimate θ_0 using OLS. Can we do that?

Given [1], it also holds that

$$y_{t-1} = \theta_0 \varepsilon_{t-2} + \varepsilon_{t-1} \Rightarrow \; \varepsilon_{t-1} = y_{t-1} - \theta_0 \varepsilon_{t-2} \qquad [2]$$

Inserting [2] into [1] we get

$$y_t = \theta_0\left(y_{t-1} - \theta_0\varepsilon_{t-2}\right) + \varepsilon_t \Rightarrow y_t = \theta_0 y_{t-1} + v_t, \quad v_t = \varepsilon_t - \theta_0^2\varepsilon_{t-2} \qquad [3]$$

Now it appears we can apply OLS estimation, and we get

$$\hat{\theta}_{OLS} = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \frac{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}y_t}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2} \qquad [4]$$

We can certainly calculate this magnitude. To examine its properties, we proceed in the usual way,

$$\hat{\theta}_{OLS} = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \frac{\sum_{t=1}^{T} y_{t-1}\left(\theta_0 y_{t-1} + v_t\right)}{\sum_{t=1}^{T} y_{t-1}^2} = \theta_0 + \frac{\sum_{t=1}^{T} y_{t-1}\left(\varepsilon_t - \theta_0^2\varepsilon_{t-2}\right)}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$= \theta_0 + \frac{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2} - \frac{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}\varepsilon_{t-2}}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2}\theta_0^2$$

Substituting for y_{t-1} in the 3d term, we have

$$\hat{\theta}_{OLS} = \theta_0 + \frac{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2} - \frac{\left(1/T\right)\sum_{t=1}^{T}\left(\theta_0\varepsilon_{t-2} + \varepsilon_{t-1}\right)\varepsilon_{t-2}}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2}\theta_0^2$$

$$\Rightarrow \hat{\theta}_{OLS} = \theta_0 + \frac{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2} - \frac{\left(1/T\right)\sum_{t=1}^{T} \varepsilon_{t-1}\varepsilon_{t-2}}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2}\theta_0^2 - \frac{\left(1/T\right)\sum_{t=1}^{T} \varepsilon_{t-2}^2}{\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^2}\theta_0^3 \qquad [5]$$

Under the assumptions made (see "Tutorial3..."), and the fact that the *y*-process is MA(1), we have the following convergence results

$$\left(1/T\right)\sum_{t=1}^{T} y_{t-1}^{2} \xrightarrow{p} E\left(y^{2}\right) = \operatorname{Var}\left(y\right) = 1+\theta_0^2$$

$$\left(1/T\right)\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{p} 0, \qquad \left(1/T\right)\sum_{t=1}^{T} \varepsilon_{t-1}\varepsilon_{t-2} \xrightarrow{p} 0$$

$$\left(1/T\right)\sum_{t=1}^{T} \varepsilon_{t-2}^{2} \xrightarrow{p} \operatorname{Var}\left(\varepsilon\right) = 1$$

Therefore

$$\hat{\theta}_{OLS} \xrightarrow{p} \theta_0 - \frac{\theta_0^3}{1+\theta_0^2} = \frac{\theta_0}{1+\theta_0^2} \qquad [6]$$

Clearly, $\hat{\theta}_{OLS}$ will be consistent *only if* $\theta_0 = 0$ i.e. only in the case where, in reality we have $y_t = \varepsilon_t$. In any other case, $\hat{\theta}_{OLS}$ will be inconsistent. Moreover this is also an example of what is called "attenuation bias": The OLS estimator will tend to underestimate the true value of the parameter, i.e. to give an etimate closer to zero in absolute terms than the true parameter (remember that the true parameter may be negative). In fact, the *larger* θ_0 is in absolute terms, the *smaller and closer to zero* will the OLS estimator tend to be. In other words, either for $\theta_0 \approx 0$ or for $\left|\theta_0\right| \nearrow\nearrow$, the OLS estimator will tend to give us estimates close to zero. Obviously, a very misleading estimator.

Exercises related to the OLS estimation attempt.

1. After equation [4], arrive at equation [6] by using the substitution $y_t = \theta_0\varepsilon_{t-1} + \varepsilon_t$ instead of $y_t = \theta_0 y_{t-1} + v_t$ that was used.

2. Contemplate the following: equation [3] looks like an AR(1) process, but where we have *not* constrained the coefficient on the lagged dependent variable to lie in $\left(-1,1\right)$. In general we know that if the parameter on the lag is outside this interval, the AR process is explosive. But we started by asserting that the process is an ergodic stationary MA(1) process, and [3] is just a different representation of [1] through a simple and straightforward substitution. *What happens here*? (***Hint***: *This is an example and a warning about "artificial" alternative representations of time series that appear to have contradicting properties*).

3. Using an Econometrics software of your choice, (the open source **Gretl** is a user-friendly and quality choice if you are not already using something), generate 50,000 i.i.d observations from a $\varepsilon_t \sim N(0,1)$ random variable, form the MA(1) processes

$$a)\ y_t = \varepsilon_{t-1} + \varepsilon_t$$

$$b)\ y_t = 3\varepsilon_{t-1} + \varepsilon_t$$

$$c)\ y_t = -3\varepsilon_{t-1} + \varepsilon_t$$

and estimate them (separately) by OLS to verify the theoretical result in [6].

4. The transformation we applied to the MA(1) process led essentially (eq. [3]), to an ARMA model. Now check in e-class the document "*An Example of an Inconsistent OLSE in the ARMA Context*". There too, it is shown that the OLS estimator is inconsitent, for the parameter that functions as the autoregressive coefficient. But compare the two and understand why in our case, where the ARMA model was *induced* in order to be able to apply OLS, the estimator performs worse compared to the OLS in the case of an "original" ARMA model (***Hint***: *deduce that in our case the OLS is bounded, irrespective of the value of the unknown parameter. Also, how many distinct parameters are there in the "original" ARMA used in the other document, and how many in our case?)*

5. Try to obtain the limiting distribution and variance of the OLS estimator, meaning a *properly centered* (and scaled) function of it so that the limiting distribution will exist and will have zero-mean. Determine which CLT applies here, if any (of those you know).

II. GMM estimation.

To be accurate, we will be implementing an exactly identified GMM estimator, which then is nothing more than the "traditional" Method of Moments estimator.

Method of Moments estimators are formed by finding "moment equations" that hold with respect to the theoretical/true values, and then invoke and apply the Analogy Principle, by considering their sample analogues. Moments are expected values, and with an ergodic stationary sample and finite moments, sample means are consistent estimators of the corresponding expected values.

The model has a single unknown parameter, so we need just one moment condition. But it has to be a moment condition that allows us to "identify" the unknown parameter. In practice, this means that we need a moment condition where θ_0 appears "outside" the *y*-variable as a stand-alone magnitude (and this is because we will need to use the *y*-variable as is, since it is the only available data we have).

For example, the following is a valid moment equation for our MA(1) process, together with its sample analogue:

$$E(y_t) = 0 \xrightarrow{\text{Analogy Principle}} \frac{1}{T+1}\sum_{t=0}^{T} y_t = 0 \qquad [7]$$

We could certainly use [7] to *test* whether some of the properties of an MA(1) process hold, but does it allow us to estimate θ_0? No. In order to bring θ_0 into the surface in [7] we would have to use $y_t = \theta_0\varepsilon_{t-1} + \varepsilon_t$ -but then we would not be able to use the *y*-variable as is (i.e. our data) any more.

So where do the *y*'s (which are the available data) and the parameter θ_0 meet in a way that allows us to estimate θ_0? If first-order moment equations are not helpful, we try the second-order moments: The autocovariance function of the process. Since this is an MA(1) we have (for unitary variance of the white noise),

$$E\left(y_t y_{t-k}\right)=\begin{cases}1+\theta_0^2 & k=0\\ \theta_0 & k=1\\ 0 & k>1\end{cases} \Rightarrow \begin{cases}E\left(y_t^2\right)=1+\theta_0^2\\ E\left(y_t y_{t-1}\right)=\theta_0\end{cases} \qquad [8]$$

In principle, we can use either moment condition (or even both and form a true over-identified GMM estimator). Sticking to the exactly identified case, we prefer $E\left(y_t y_{t-1}\right)=\theta_0$ because it is linear in the unknown parameter.

In "Tutorial3... " you can find the (here, ceremonial) steps related to writing the estimation procedure in the GMM way. In practice it is directly before our eyes that our estimator will be

$$\hat{\theta}_{GMM}=\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_t \qquad [9]$$

and we know that it will be consistent (why?). We also have that it is unbiased since

$$E\left(\hat{\theta}_{GMM}\right)=E\left[\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_t\right]=\frac{1}{T}\sum_{t=1}^{T} E\left(y_{t-1}y_t\right)=\frac{1}{T}\sum_{t=1}^{T} E\left[\left(\theta_0\varepsilon_{t-1}+\varepsilon_t\right)\left(\theta_0\varepsilon_{t-2}+\varepsilon_{t-1}\right)\right]$$

$$=\frac{1}{T}\sum_{t=1}^{T} E\left[\theta_0^2\varepsilon_{t-1}\varepsilon_{t-2}+\theta_0\varepsilon_{t-1}^2+\theta_0\varepsilon_t\varepsilon_{t-2}+\varepsilon_t\varepsilon_{t-1}\right]$$

$$=\frac{1}{T}\sum_{t=1}^{T}\left[\theta_0^2 E\left(\varepsilon_{t-1}\varepsilon_{t-2}\right)+\theta_0 E\left(\varepsilon_{t-1}^2\right)+\theta_0 E\left(\varepsilon_t\varepsilon_{t-2}\right)+E\left(\varepsilon_t\varepsilon_{t-1}\right)\right]$$

$$=\frac{1}{T}\sum_{t=1}^{T}\left[\theta_0^2\cdot 0\ +\ \theta_0\cdot 1\ +\ \theta_0\cdot 0+0\right]=\frac{1}{T}T\theta_0=\theta_0$$

But note that unbiasedness rests crucially on the assumption that $E\left(\varepsilon_{t-1}^2\right)=\mathrm{Var}\left(\varepsilon\right)=1$.

Going through the tedious calculations (see "Tutorial3-Corrected") we also find that the *finite sample* variance of the estimator is

$$\text{Var}\left(\hat{\theta}_{GMM}\right) = \frac{1}{T}\left[1+\left(1+\kappa_4\right)\theta_0^2+\theta_0^4\right] + \frac{\left(T-1\right)\left(T+2\right)}{T^2}\theta_0^2 - \theta_0^2$$

$$\Rightarrow \text{Var}\left(\hat{\theta}_{GMM}\right) = \frac{1}{T}\left[1+\left(1+\kappa_4\right)\theta_0^2+\theta_0^4\right] + \frac{T-2}{T^2}\theta_0^2 \qquad [10]$$

It is evident that $\text{Var}\left(\hat{\theta}_{GMM}\right) \xrightarrow{T\to\infty} 0$. This, together with unbiasedness, are sufficient conditions for the consistency of the estimator.

At the same time

$$\text{Var}\left(\sqrt{T}\left(\hat{\theta}_{GMM}-\theta_0\right)\right)=\text{Var}\left(\sqrt{T}\hat{\theta}_{GMM}\right) = T\left[\frac{1}{T}\left[1+\left(1+\kappa_4\right)\theta_0^2+\theta_0^4\right] + \frac{T-2}{T^2}\theta_0^2\right]$$

$$=1+\left(1+\kappa_4\right)\theta_0^2+\theta_0^4 + \frac{T^2-2T}{T^2}\theta_0^2 \xrightarrow{T\to\infty} 1+\left(1+\kappa_4\right)\theta_0^2+\theta_0^4 + \theta_0^2$$

$$\Rightarrow \text{Var}\left(\sqrt{T}\hat{\theta}_{GMM}\right) \xrightarrow{T\to\infty} 1+\left(2+\kappa_4\right)\theta_0^2+\theta_0^4 = \left(1+\theta_0^2\right)^2 + \kappa_4\theta_0^2 \qquad [11]$$

Informally, we expect that [11] will be the variance of the limiting disgtribution of $\sqrt{T}\left(\hat{\theta}_{GMM}-\theta_0\right)$. We say "informally", because the fact that the sequence $\left\{\text{Var}\left(\sqrt{T}\hat{\theta}_{GMM}\right)\right\}$ converges to a finite limit does not automatically guarantee that this limit will be the variance of the limiting distribution. Some additional condition is needed (Hayashi p. 91 Lemma 2.1 is relevant but has unfortunately neglected to include this additional condition. See errata of the book, available on the book's website). See also

https://stats.stackexchange.com/a/88519/28746

Limiting Distribution of the GMM estimator.

Consistency and the previous variance result does suggest that the properly centered and scaled function of the estimator that will have a limiting distribution is $\sqrt{T}\left(\hat{\theta}_{GMM}-\theta_0\right)$. We need to determine which CLT applies here. First, and since the estimator is

$\hat{\theta}_{GMM} = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_t$ and consistent , we examine whether the process $\left(y_{t-1}y_t - \theta_0\right)$ is a Martingale Difference . We have

$$\left(y_{t-1}y_t - \theta_0\right) = \left[\left(\theta_0\varepsilon_{t-2} + \varepsilon_{t-1}\right)\left(\theta_0\varepsilon_{t-1} + \varepsilon_t\right) - \theta_0\right] =$$

$$= \theta_0^2\varepsilon_{t-2}\varepsilon_{t-1} + \theta_0\varepsilon_{t-2}\varepsilon_t + \theta_0\varepsilon_{t-1}^2 + \varepsilon_{t-1}\varepsilon_t - \theta_0$$

Denote $f_{t-i} = \sigma\left(\varepsilon_{t-i}, \varepsilon_{t-i-1,}...\right)$ the filtration containing the history of the white noise process. **Note** that we "went down" to the foundations of the $\left(y_{t-1}y_t - \theta_0\right)$ process, we didn't attempt to examine the martingale-difference property examining $\left(y_{t-1}y_t - \theta_0\right)$ as is and forming filtrations in terms of the *y*-variable (can you guess why?)

Then, for $\left(y_{t-1}y_t - \theta_0\right)$ to be a Martingale Difference we examine whether $E\left[\left(y_{t-1}y_t - \theta_0\right)\middle| f_{t-1}\right]$ is equal to zero for every *t*. We have

$$E\left[y_{t-1}y_t - \theta_0 \,|\, f_{t-1}\right] = E\left[\theta_0^2\varepsilon_{t-2}\varepsilon_{t-1} + \theta_0\varepsilon_{t-2}\varepsilon_t + \theta_0\varepsilon_{t-1}^2 + \varepsilon_{t-1}\varepsilon_t - \theta_0 \,|\, f_{t-1}\right]$$

$$= \theta_0^2\varepsilon_{t-2}\varepsilon_{t-1} + 0 + \theta_0\varepsilon_{t-1}^2 + 0 - \theta_0 \neq 0$$

So the process is *not* a martingale difference, and we need some other CLT.

We move on to **Gordin's CLT**. For it to be applicable we require

1) A strictly stationary and ergodic process. We have it.

2) **Gordin's condition.**

GC-1 : Second moment of the process is finite. $E\left(y_{t-1}y_t - \theta_0\right)^2 < +\infty$

It holds (do the calculations).

GC-2 : $E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-i}\right) \xrightarrow{m.s} 0, \quad i \longrightarrow \infty$

We have from before

$$E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-i}\right) = E\left[\theta_0^2 \varepsilon_{t-2}\varepsilon_{t-1} + \theta_0 \varepsilon_{t-2}\varepsilon_t \; + \; \theta_0 \varepsilon_{t-1}^2 \; + \; \varepsilon_{t-1}\varepsilon_t \; - \theta_0 \,|\, f_{t-i}\right]$$

Then

for $i = 1 \Rightarrow E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-i}\right) = \theta_0^2 \varepsilon_{t-2}\varepsilon_{t-1} \; + \theta_0\left(\varepsilon_{t-1}^2 - 1\right)$

for $i > 1 \Rightarrow E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-i}\right) = E\left[\theta_0\left(\varepsilon_{t-1}^2 - 1\right)\right] = \theta_0\left(1 - 1\right) = 0$

So the condition $E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-i}\right) \xrightarrow{m.s} 0, \quad i \longrightarrow \infty$ is obviously satisfied.

GC-3. We need to form the telescoping sum. We have

$$y_{t-1}y_t - \theta_0 = r_{t0} + r_{t1} + \ldots + r_{t,j-1} + E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-j}\right),$$

$$r_{t,m} \equiv E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-m}\right) - E\left(y_{t-1}y_t - \theta_0 \,|\, f_{t-m-1}\right)$$

From immediately before, we have that

$$r_{t,m} = \begin{cases} y_{t-1}y_t - \theta_0 & m = 0 \\ \theta_0^2 \varepsilon_{t-2}\varepsilon_{t-1} \; + \theta_0\left(\varepsilon_{t-1}^2 - 1\right) & m = 1 \\ 0 & m > 1 \end{cases}$$

We are asked to consider the infinite sum of $\sqrt{E\left(r_{t,m}^2\right)}$. Squaring $r_{t,m}$ and taking the expected value will give us no higher than its fourth moment, which has been assumed finite. Then, since the terms of the infinite sum are zero after a point, and the expected values up to that point are finite it follows that

$$\sum_{i=0}^{\infty}\left[E\left(r_{t,i}^2\right)\right]^{1/2} < +\infty, \quad \forall t \qquad \text{holds.}$$

So Gordin's condition is satisfied and we can apply Gordin's CLT that states

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_t - \theta_0\right) = \sqrt{T}\left(\hat{\theta}_{GMM} - \theta_0\right) \xrightarrow{d} N\left(0, \mathrm{V}\right), \quad \mathrm{V} = \gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k$$

where γ_k is the autocovariance function of the process $\left(y_{t-1}y_t - \theta_0\right)$.

Exercises related to GMM estimation.

1. Derive the autocovariance function of the process $\left(y_{t-1}y_t - \theta_0\right)$ and verify that the variance of the limiting distribution of the GMM estimator is equal to expression [11].

2. Compare the expressions for the OLS (eq. [4] and the GMM estimator (eq. [9]). Can you explain in some "intuitive" way, why OLS in this case "falls victim of its own success" in other cases?

3. Use the other available moment condition, and calculate the limiting distribution and variance of the GMM estimator. Compare.

4. Use both moment conditions to form a GMM estimator proper (i.e. over-identified). Determine the optimal weighting matrix.

5. Assume that $\mathrm{Var}\left(\varepsilon_t\right) = \sigma_\varepsilon^2$. Namely, the model has now two unknown parameters.

(i) Write out in detail the objective function that the (once again exactly identified) GMM estimator will minimize. Derive the first-order conditions.

(ii) Calculate the covariance between the two estimators $\hat{\theta}$ and $\hat{\sigma}_\varepsilon^2$.

--