Further Topics on ARMA Models

In what follows we assume that the UDC holds. Furthermore let $0<\sigma^2:=\mathbb{E}(\varepsilon_0^2)$.

A. Further Example - ARMA(1,1)

Suppose that $\phi(L)=1-B_1L$ and $\Theta(L)=1-\theta_1L$. As we have already seen the UDC is equivalent to $|B_1|<1$, whence we obtain that the relevant ARMA(1,1) solution is defined by, $\forall t\in\mathbb{Z}$, $y_t=\sum_{j=0}^{\infty}B_1^jL^j(1-\theta_1L)\varepsilon_t=\sum_{j=0}^{\infty}B_1^j\varepsilon_{t-j}-\theta_1\sum_{j=0}^{\infty}B_1^j\varepsilon_{t-j-1}$

$$\overset{s=j+1}{=}\sum_{j=0}^{\infty}B_1^j\varepsilon_{t-j}-\theta_1\sum_{s=1}^{\infty}B_1^{s-1}\varepsilon_{t-s}\overset{s=j}{=}\varepsilon_t+(B_1-\theta_1)\sum_{j=1}^{\infty}B_1^{j-1}\varepsilon_{t-j}.$$

Notice that when $B_1=\theta_1$, i.e. $\phi(L)=\Theta(L)$ [equivalent in this case to that they have the same root], thes as a relevant previous exerise prescribes, the process is actually an ARMA(0,0), i.e. a white noise one. Furthermore we have that for any $k\geq 0$, $\gamma_k=(\sum_{i=0}^{\infty}B_i^*B_{i+k}^*)\sigma^2$, whence,

$B_i^*=\begin{cases}1, & i=0\\ (B_1-\theta_1)B_1^{i-1}, & i>0\end{cases}$ and therefore $\gamma_0=(1+(B_1-\theta_1)^2\sum_{i=1}^{\infty}B_1^{2(i-1)})\sigma^2$

$\overset{p=i-1}{=}(1+(B_1-\theta_1)^2\sum_{p=0}^{\infty}B_1^{2p})\sigma^2=(1+\frac{(B_1-\theta_1)^2}{1-B_1^2})\sigma^2$. Furthermore, for $k>0$,

$$\gamma_k=((B_1-\theta_1)B_1^{k-1}+(B_1-\theta_1)^2\sum_{i=1}^{\infty}B_1^{i-1}B_1^{i+k-1})\sigma^2=((B_1-\theta_1)B_1^{k-1}+(B_1-\theta_1)^2B_1^k$$

$$\sum_{i=1}^{\infty}B_1^{2(i-1)})\sigma^2\overset{p=i-1}{=}((B_1-\theta_1)B_1^{k-1}+(B_1-\theta_1)^2B_1^k\sum_{p=0}^{\infty}B_1^{2p})\sigma^2=$$

$$=((B_1-\theta_1)B_1^{k-1}+(B_1-\theta_1)^2\frac{B_1^k}{1-B_1^2})\sigma^2=(B_1-\theta_1)B_1^{k-1}(1+(B_1-\theta_1)\frac{B_1}{1-B_1^2})\sigma^2.$$

[When $B_1=0$, or/and $\theta_1=0$ we obtain the relevant formulae for the AR(1), MA(1) and white noise case respectively].

B. Invertibility

Definition. The ARMA(p,q) process is called invertible iff

$\forall t \in \mathbb{Z}$, $\varepsilon_t = \sum_{j=0}^{\infty} \rho_j y_{t-j}$ for $(\rho_j)_{j \in \mathbb{N}}$ absolutely summable. $\square$

Remark. It is obvious that the invertibility property implies that $(\varepsilon_t)_{t \in \mathbb{Z}}$ is adapted to $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ with $\mathcal{F}_t := \sigma(y_{t-i}, i \geq 0)$. $\square$

The relevant lemma that introduced the UDC directly implies the following result.

Lemma. If the roots of $\Theta(L)$ satisfy the UDC then the process is invertible.

Remark. Remember the convention on the case where the polynomial in question is constant, that is that the UDC trivially holds. This implies that when $q=0$, i.e. we have an AR(p) process, this is in a trivial manner invertible since $\varepsilon_t = y_t - \sum_{j=1}^{p} B_i y_{t-i}$, $\forall t \in \mathbb{Z}$.

C. Common Roots and Statistical Identification: Parsimony

As a previous relevant exercise established (see also the relevant case in the ARMA(1,1) example above) the existence of k common roots between the Φ and Θ polynomials ($k \leq \min(p,q)$) implies that the solution is actually equal to an ARMA(p-k, q-k) process, something that directly implies that the two relevant statistical models are statistically indistinguishable. The solution to this identification issue is via parsimony, i.e. the lower (p-k, q-k)-order is chosen as a statistical model, something that is consistent to what is called the Box-Jenkins methodology to time series modelling.

D. Topics in Semi-Parametric Statistical Inference on ARMA models [Assume that σ^2 is known and $\sigma^2=1$. The following can be easily extended to the case where σ^2 is unknown]

In what follows we assume the existence of a sample, i.e. a random element $(y_t)_{t=-p+1,-p+2,\ldots,1,\ldots,T}$, $T>0$, from an ARMA(p,q) process, where the actual coefficient $\varphi_0 := (B_{1_0},\ldots B_{p_0}, \theta_{1_0},\ldots,\theta_{q_0})' \in \Theta \subseteq \mathbb{R}^{p+q}$ is unknown, while it could also be the case that the actual values of p and or q are unknown. We will initially assume that the actual (p,q) is known, while we are interested in topics on the issue of semi-parametric inference about φ_0, i.e. inference without parametric assumptions on the fidis of $(\varepsilon_t)_{t\in\mathbb{Z}}$.

1. $q=0$ (i.e. AR(p) model) and the OLSE.

If $q=0$ whence the statistical model can be specified as a linear one, i.e.

$$Y = X\varphi + u, \text{ where } Y = (y_t)_{t=1,\ldots,T},$$

$$X = \begin{pmatrix} y_0 & y_{-1} & \cdots & y_{1-p} \\ y_1 & y_0 & \cdots & y_{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-1} & y_{T-2} & & y_{T-p} \end{pmatrix} = (X_1,\ldots,X_p) \text{ with } X_i = (y_{t-i})_{t=1,\ldots,T}$$

$i=1,\ldots,p$ and $u = (\varepsilon_t)_{t=1,\ldots,}$, $\varphi \in \Theta \subseteq \mathbb{R}^p$. When $\Theta = \mathbb{R}^p$, the OLSE is* $\varphi_T = (X'X)^{-1}X'u$ and properties of which have already been examined in cases where $p=1$.

Notice that analogously to the examined cases if $(\varepsilon_t)_{t\in\mathbb{Z}}$ is stationary and ergodic then due to

* Given that rank$[X]=p$.

Birkhoff's LLN, $\frac{x'u}{T} = \left(\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t y_{t-i}\right)_{i=1,\dots,p}$

$\mathbb{P}$ a.s. converges to zero (provide the full details)

while in an analogous manner $\frac{x'x}{T} = \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-i} x_{t-j}\right)_{i,j=1,\dots,p}$

$\mathbb{P}$ a.s. converges to the matrix $\left(\mathbb{E}(x_{-i}x_{-j})\right)_{i,j=1,\dots,p}$

$= \left(\gamma_{|i-j|}\right)_{i,j=1,\dots,p}$ (provide the full details). Since $\gamma_0 > 0$

and $\gamma_k \to 0$ as $k \to +\infty$, it is possible to prove that the matrix $\left(\gamma_{|i-j|}\right)_{i,j=1,\dots,p}$ is invertible, and thereby using

analogous arguments to the case $p=1$ (provide the details) we can prove that $\varphi_n \to \varphi_0$ $\mathbb{P}$ a.s.

Using finally extensions of the relevant conditions used in the case where $p=1$ it is possible to derive the $\sqrt{T}$ rate and asymptotic normality for the $\sqrt{T}(\varphi_n - \varphi_0)$ (via e.g., among others, the use of multivariate extension of the CLT we have been studying), and so on.

2. $q>0$ and the Gaussian Quasi-MLE.

When $q>0$ the OLSE is infeasible due to the fact that in the relevant extension of the statistical model, the regressor matrix would contain components which are latent, i.e. unobservable, e.g. $(\varepsilon_{t-1})_{t=1,\dots,T}$. Consider following criterion:

$$Q_T(\varphi) := -\frac{1}{2T}\sum_{t=1}^{T}\left(y_t - \sum_{j=1}^{p} \beta_j y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}\right)^2$$

and notice that this is the average log-likelihood for an in the case where $\varepsilon_t \overset{\text{iid}}{\sim} N(0,1)$ (up to a constant) conditionally on $y_0, \dots, y_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q}$, via a decomposition of the latter by consecutive appropriate conditional densities (derive it!)

In the case that the aforementioned parametric assumption on $(\varepsilon_t)_{t \in \mathbb{Z}}$ is valid, then this is termed as conditional likelihood function. In the general case, that allows for the possibility that the assumption above is not true, then Q_T is termed as Gaussian (Conditional) Quasi-Likelihood function.

Remark. When $q=0$ then Q_T is proportional to minus the sum of squares criterion. □

Definition. $\varphi_T \in \arg\max_{\varphi \in \Theta} Q_T(\varphi)$ is called Quasi-MLE (QMLE) [conditional MLE when the assumption is valid]

Remark. When $q=0$ and if $\operatorname{rank} X = p$ then QMLE=OLSE. □

When $q>0$ the QMLE is infeasible due to the same reason that results to the infeasibility of the OLSE. However in this case, and for any value of the sample, the Q_T can be approximated by a filtering device for the latent components, as follows.

Filtering Algorithm

For a given value of φ:

1. Set $e_{1-q} = e_{2-q} = \dots = e_0 = 0$,

2. Set $e_t(\varphi) = y_t - \sum_{j=0}^{q} \beta_j y_{t-j} + \sum_{j=1}^{p} \theta_j e_{t-j}(\varphi)$, $t = 1, \dots, T$,

3. Evaluate $Q_T^*(\varphi) = -\frac{1}{2T} \sum_{t=1}^{T} e_t^2(\varphi)$

The filtering algorithm above implies that $Q_T^*(\varphi)$ is "computable" for any value of φ, and thereby can be maximized w.r.t. $\varphi \in \Theta$, in order to obtain an estimator for φ_0.

Definition. Term also $\varphi_T \in \underset{\varphi \in \Theta}{\arg\max}\, Q_T^*(\varphi)$ as the (**feasible**) QMLE for φ_0.

Remark. The optimization above is generally not analytically feasible, hence performed via a numerical optimization procedure (e.g. Newton-Raphson algorithm) which generally works as follows:

1. Initiate $\varphi \in \Theta$,
2. Given φ, use the filtering algorithm to evaluate $Q_T^*(\varphi)$,
3. Change the value of φ according to some procedure (e.g. use information from the derivatives of Q_T^*)
4. Go to 2 and repeat until a prescribed set of conditions is met,
5. φ_T is the selection of φ in the final execution of step 2.

[Given φ_T, several diagnostics can be evaluated, e.g. $ASSR_T := -2\, Q_T^*(\varphi_T)$]

Remark. Numerical procedures involve optimization errors hence φ_T should be more generally defined as an approximate maximiser of Q_T^*. We do not pursue this path for reasons of simplicity. □

Example. Suppose that $p=0, q=1$, i.e. we have the MACD model and that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is also stationary and ergodic.

$$Q_t^*(\theta_1) = -\frac{1}{2T}\sum_{t=1}^{T} e_t^2(\theta_1) \quad \text{where}$$

$$e_t(\theta_1) = \begin{cases} 0, & t=0 \\ y_t + \theta_1 \varepsilon_{t-1}(\theta_1), & t>0 \end{cases}.$$

Leaving aside the details of issues of appropriate approximation, and if Θ is a nonempty compact subset of $(-1,1)$, $\theta_{1_0} \in \Theta$ (i.e. the MA(1) model is invertible), it is possible to show that "asymptotically" as $T \to +\infty$ φ_T "maximizes" $\mathbb{P}$ a.s. w.r.t. $\theta_1 \in \Theta$

$$\frac{1}{T}\sum_{t=1}^{T} e_t^{*2}(\theta_1) \quad \text{where}$$

$(e_t^*)_{t\in\mathbb{Z}}$ is defined by

$$e_t^*(\theta_1) = y_t + \theta_1 e_{t-1}^*(\theta_1) \ \forall t \in \mathbb{Z}$$

$$\Leftrightarrow \ e_t^*(\theta_1) = \theta_1 e_{t-1}^*(\theta) + \varepsilon_t - \theta_{1_0}\varepsilon_{t-1} \ \forall t \in \mathbb{Z}$$

i.e. the process $(e_t^*(\theta_1))$ is a stationary and ergodic invertible ARMA(1,1) process $\forall \theta_1 \in \Theta$, whereas when $\theta_1 = \theta_{1_0}$ the common root discussion above implies that $e_t^*(\theta_{1_0}) = \varepsilon_t \ \forall t \in \mathbb{Z}$. An argument based on an application of Birkhoff's ULLN, implies that $\frac{1}{2T}\sum_{t=1}^{T} e_t^*(\theta_1)$ converges uniformly over Θ, $\mathbb{P}$ a.s. to $Q : \Theta \to \mathbb{R}$, with $Q(\theta_1) := \mathbb{E}[e_0^*(\theta_1)]^2 \underset{\substack{\text{see} \\ \text{the ARMA(1,1)} \\ \text{above}}}{=} -\frac{1}{2}\left[1 + \frac{(\theta_1 - \theta_{1_0})^2}{1-\theta_1^2}\right]$ which is continuous w.r.t θ_1 and uniquely maximized at $\theta_1 = \theta_{1_0}$. Hence the "standard" asymptotic theory for M-estimators implies that φ_T is strongly consistent. □

In the general case it is possible to establish relevant regularity conditions, under which φ_T is consistent, with $\sqrt{T}$ rate and asymptotically normality holds. [*]

3. (p,q) unknown.

In the case where the order of the process is unknown, a variety of procedures can be specified for the estimation of the true value of the order vector, some of which also involve the aforementioned procedures. E.g. if it is known that $p \le p_{max}$ and $q \le q_{max}$ then φ_T from ARMA(p_{max}, q_{max}) can be used for tests of statistical significance. [Δ]

In the same respect, given the previous inequalities, a statistical information criterion can be used for the estimation of (p,q). An example is the Bayesian Information Criterion (bic) defined as: for (p^*, q^*), $p^* = 0, 1, \dots, p_{max}$, $q^* = 0, 1, \dots, q_{max}$,

$$\mathrm{bic}_T(p^*, q^*) := \ln\left[\mathrm{ASSR}_T(p^*, q^*)\right] + (p^* + q^*)\frac{\ln(T)}{T},$$

where $\mathrm{ASSR}_T(p^*, q^*)$ is the ASSR_T from above when estimating an ARMA(p^*, q^*) model. Then the estimator for (p,q) is defined by:

$$(p_T, q_T) = \underset{\substack{(p^*, q^*) \\ p^* = 0, \dots, p_{max} \\ q^* = 0, \dots, q_{max}}}{\operatorname{argmin}} \mathrm{bic}_T(p^*, q^*).$$

E.g. when $q_{max} = 0$, $(\varepsilon_t)_{t \in \mathbb{Z}}$ is iid and $\mathbb{E}(\varepsilon_0^4) < \infty$ it can be proven that p_T is a consistent estimator of p.

4. A simple example of an indirect estimator in the context of MA(1).

Suppose that $p = 0$, $q = 1$, known, the process is known to satisfy $|\theta_{1_0}| < 1$, hence invertibility is the case. The previous imply that Θ can be chosen as $(-1, 1)$ [or some relevant subset if more

[Δ] Notice that for fixed T, the values p_{max}, q_{max} are somehow restricted by the sample size [explain the details].

information on θ_{1_0} is available]*. The sample is $(y_t)_{t=0,\ldots,T}$.

Consider $B_T := \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$, which is the OLSE in the context of an AR(1) model and notice that such a model would be clearly misspecified in this context, except for the case where $\theta_{1_0}=0$. Is it possible to employ the OLSE of the misspecified model for the estimation of θ_{1_0}?

Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is stationary and ergodic. For arbitrary $\theta_1\in\Theta$, and if $(y_t)_{t=0,\ldots,T}$ is a part from an MA(1) process with parameter θ_1, then due to Birkhoff's LLN

$$\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} \\ \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix} \longrightarrow \begin{pmatrix} -\theta_1 \\ (1-\theta_1^2) \end{pmatrix} \text{ a.s.}$$

Thereby, due to the CMT, under the MA(1) process corresponding to $\theta_1\in\Theta$, $B_T \to b(\theta_1) := -\frac{\theta_1}{1+\theta_1^2}$. $b:\Theta\to(-1/2, 1/2)$

is termed as the binding function (notice that the sample is actually a part of the MA(1) process corresponding to θ_{1_0}, hence it actually holds that $B_n \to b(\theta_{1_0}) = -\frac{\theta_{1_0}}{1+\theta_{1_0}^2}$.) Since B_n is

a stochastic approximation of $b(\theta_{1_0})$, an estimator for θ_{1_0} can be defined by minimizing some notion of distance between B_T and $b(\theta)$ w.r.t. $\theta_1\in\Theta$. We completely specify the previous in the following procedure.

* The result on the asymptotic normality for the estimator below, would also hold if $\Theta\subset(-1,1)$ and $\theta_{1_0}\in \text{Interior}(\Theta)$.

Indirect Estimation

1. Employ the possibly misspecified auxiliary AR(1) model to obtain the OLSE, B_T.
2. $\theta_{1_T} := \underset{\theta_1 \in \Theta}{\arg\min} \left(B_T + \frac{\theta_1}{1+\theta_1^2} \right)^2$

θ_{1_T} is called an indirect estimator. Some comments on its definition are the following:

1. It is defined in two optimization steps. In this example each optimization is (at least asymptotically - see below) analytically derivable. Hence its derivation is less computationally costly compared to the Gaussian QMLE.

2. When $B_T \in (-1/2, 1/2)$, the second optimization problem can be analytically solved to obtain $\theta_{1_T} = \frac{-1+\sqrt{1-4B_T^2}}{2B_T}$ (the definition of θ_{1_T} can be modified so that it can be analytically obtained for any possible value of B_T). Notice that since $B_T \to b(\theta_{1_0}) \in (-1/2, 1/2)$ a.s. as $T \to +\infty$, θ_{1_T} admits asymptotically the previous expression a.s. This expression is the consequence of that b is 1-1 on $(-1,1)$ something that is termed indirect identification.

3. Comment 2 essentially means that as $T \to +\infty$, $\theta_{1_T} = b^{-1}(B_T)$ a.s. Since $b^{-1}(B) := \frac{-1+\sqrt{1-4B^2}}{2B}$, $B \in (-1/2, 1/2)$ is continuous in B, and $B_T \to b(\theta_{1_0})$, a.s. the CMT implies that $\theta_{1_T} \to b^{-1}(b(\theta_{1_0})) = \theta_{1_0}$ a.s. as $T \to \infty$, i.e. it is strongly consistent.

4. It is possible to prove using tedious algebra, several asymptotic considerations, and the CLT for stationary ergodic smd processes that if $(\varepsilon_t)_{t\in\mathbb{Z}}$ is iid, (see lemma 2 in Ref below)

$$\sqrt{T}\left(B_T - b(\theta_{1_0})\right) \xrightarrow{d} z \sim N\left(0, v_{\theta_{1_0}}\right), \text{ where } v_{\theta_{1_0}} = \frac{(1+\theta_{1_0}^4)^2 + \theta_{1_0}^2(1+\theta_{1_0}^2)^2}{(1+\theta_{1_0}^2)^4}.$$

Note that the previous does not need existence of $\mathbb{E}(\varepsilon_0^4)$, as in the application of the Gordin's CLT case in the relevant tutorial.

Now notice that due to the inverse function theorem (why is it applicable?), for any $\theta_1 \in (-1,1)$, $\frac{db^{-1}(b(\theta_1))}{d\beta} = \left(\frac{db(\theta)}{d\theta}\right)^{-1} =$

$$\left[\frac{d\left(\frac{-\theta_1}{1+\theta_1^2}\right)}{d\theta_1}\right]^{-1} = \left[\frac{-(1+\theta_1^2)+\theta_1 \cdot 2\theta_1}{(1+\theta_1^2)^2}\right]^{-1} = \frac{(1+\theta_1^2)^2}{\theta_1^2-1}$$ which is continuous. Hence, the Delta method implies that (explain the details)

$$\sqrt{T}\left(\theta_{1_T}-\theta_{1_0}\right) \xrightarrow{d} z_* \sim N\left(0, V_I(\theta_{1_0})\right) \text{ where}$$

$$V_I(\theta_{1_0}) = \frac{(1+\theta_{1_0}^2)^4}{(1-\theta_{1_0}^2)^2} V_{\theta_{1_0}} = \frac{(1+\theta_{1_0}^4)^2+\theta_{1_0}^2(1+\theta_{1_0}^2)^2}{(1-\theta_{1_0}^2)^2}.$$

Exercise. Derive a consistent estimator for $V_I(\theta_{1_0})$ assuming that κ_4 is known.

Ref.: Arvanitis Stelios, (2014), A simple example of an indirect estimator with discontinuous limit theory in the MA(1) model, Journal of Time Series Analysis, 35, pages 536–557. DOI: 10.1111/jtsa.12080

[*] Further remark on the Gaussian QMLE:

The actual existence, as a measurable function, of the θ_T can also be facilitated (depending on the form of Θ), if the definition is extended so as to allow for approximate maximizers. This is also in accordance with its usual numerical approximation. In the particular MA(1) example above, Θ could be chosen as $(-1,1)$ (as we did in class) and the aforementioned extention would facilitate its existence in this case. Furthermore, if $\Theta=(-1,1)$ then consistency would be implied by that Q_T^* converges P a.s.

to Q locally uniformly (i.e. uniformly w.r.t. any non empty compact subset of $(-1,1)$), by the fact that Q is uniquely maximized at $\theta_{1,0}$, and that the maximizer is "somehow distinguishable."

Exercise. Show that in the ARMA(1,1) example presented above, $\forall k>1$

$$\gamma_k = \beta_1 \gamma_{k-1}.$$

[The notes are in a state of perpetual correction. They do not substitute the lectures. Please report any typos to stelios@aueb.gr or the course's e-class.]