Time Series Econometrics

Nicky Grant

ECON5221: Problem Set 2 Solutions

Semester 2 2019/2020

Exercise Questions 2

Problem Set 2

1. Consider a bivariate (ie, two variable) $VAR(1)$ system

$$\mathbf{Y}_t = \boldsymbol{\alpha} + \Phi_1 \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\varepsilon}_t$ is vector white noise.

(a) Stationarity of VARs. Write $\Phi(L) = I_2 - \phi_1 L$

With $\phi_{11} = 0.6, \phi_{12} = 0.1, \phi_{21} = -0.6, \phi_{22} = 1.1$:

$$|\Phi(L)| = \begin{vmatrix} 1 - 0.6L & -0.1L \\ 0.6L & 1 - 1.1L \end{vmatrix} = 1 - 1.7L + 0.72L^2$$

and characteristic equation

$$\begin{aligned} \lambda^2 - 1.7\lambda + 0.72 &= (\lambda - 0.9)(\lambda - 0.8) \\ \lambda_1 &= 0.9; \;\; \lambda_2 = 0.8 \end{aligned}$$

Both roots are less than one in absolute value and hence the VAR is stationary.

(b) From the stationary VAR , we have

$$\begin{aligned} \mathbb{E}[\mathbf{Y}_t] &= [\Phi(1)]^{-1}\boldsymbol{\alpha} \\ \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} &= \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.6 & 0.1 \\ -0.6 & 1.1 \end{bmatrix} \right\}^{-1} \begin{bmatrix} -0.2 \\ 0.1 \end{bmatrix} \\ &= \begin{bmatrix} 0.4 & -0.1 \\ 0.6 & -0.1 \end{bmatrix}^{-1} \begin{bmatrix} -0.2 \\ 0.1 \end{bmatrix} \\ &= \frac{1}{0.02} \begin{bmatrix} -0.1 & 0.1 \\ -0.6 & 0.4 \end{bmatrix} \begin{bmatrix} -0.2 \\ 0.1 \end{bmatrix} \\ &= \begin{bmatrix} 1.5 \\ 8.0 \end{bmatrix} \end{aligned}$$

(c) The infinite VMA representation is

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Phi}_1^2 \boldsymbol{\varepsilon}_{t-2} + \boldsymbol{\Phi}_1^3 \boldsymbol{\varepsilon}_{t-3} + ... \\ &= \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Theta}_2 \boldsymbol{\varepsilon}_{t-2} + \boldsymbol{\Theta}_3 \boldsymbol{\varepsilon}_{t-3} + .. \end{aligned}$$

Hence

$$\begin{aligned}
\boldsymbol{\Theta}_0 &= \boldsymbol{I}_2 \\
\boldsymbol{\Theta}_1 &= \boldsymbol{\Phi}_1 = \begin{bmatrix} 0.6 & 0.1 \\ -0.6 & 1.1 \end{bmatrix} \\
\boldsymbol{\Theta}_2 &= \boldsymbol{\Phi}_1^2 = \begin{bmatrix} 0.30 & 0.17 \\ -1.02 & 1.15 \end{bmatrix}
\end{aligned}$$

(d) Impulse response functions (corresponding to $s = 0, 1, 2$) are read from the elements of $\boldsymbol{\Theta}_s$ ($s = 0, 1, 2$)

i. The effect of a unit shock to Y_{1t} on Y_{1t} and Y_{2t} : (namely $\epsilon_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\epsilon_s = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ for all $s \neq t$. In essence we assume the process is in equilibrium and we introduce a one unit shock to Y_{1t} and see the dynamic response to both $Y_{1,t+j}$, $Y_{2,t+j}$ for $j = 0, 1, 2$).

$$\begin{aligned}
\text{Effect on } Y_1 &: \; 1, 0.6, 0.30 \\
\text{Effect on } Y_2 &: \; 0, -0.6, -1.02
\end{aligned}$$

ii. The effect of a unit shock to Y_{2t} on Y_{1t} and Y_{2t}, (namely $\epsilon_t = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\epsilon_s = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ for all $s \neq t$).

$$\begin{aligned}
\text{Effect on } Y_1 &: \; 0, \; 0.1, \; 0.17 \\
\text{Effect on } Y_2 &: \; 1, \; 1.1, \; 1.15
\end{aligned}$$

2. Consider a stationary AR(2) process

$$Y_t = \mu_0 + \phi_{10} Y_{t-1} + \phi_{20} Y_{t-2} + \varepsilon_t \qquad \varepsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{1}$$

where a researcher observes a sample realisation of size $T > 2$, $(y_1, \cdots, y_T)$ from $\{Y_t\}$.

(a) Derive conditional log likelihood function for (1) conditioning on $(Y_1 = y_1, Y_2 = y_2)$.

We derived this for the AR(1) in Lecture 2 (Wk3). It is also covered in Hamilton, the conditional likelihood function for any AR(p) model.

Conditioning on $\theta = (\mu, \phi_1, \phi_2, \sigma^2)$ (being the true parameter) then, $Y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}$ is distributed $N(\mu + \phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma^2)$ since

$\varepsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2)$ hence is independently distributed of Y_{t-2} and Y_{t-1}, both being functions of shocks prior to time period t.

The conditional likelihood function in this case is

$$f_{Y_T,\cdots,Y_3|Y_2,Y_1}(y_T,\cdots,y_1|\theta) = \prod_{t=3}^{T} f_{Y_t|Y_{t-2},Y_{t-1}}(y_t|y_{t-1};\theta) = \prod_{t=3}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t-\mu-\phi_1 y_{t-1}-\phi_2 y_{t-2})^2}{2\sigma^2}\right)$$

where the second inequality holds noting that

$Y_t|Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}$ is distributed $N(\mu + \phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma^2)$ for all $t = 3, \cdots, T$ [We must start at time period 3 as Y_t depends on the outcomes of the last two periods.]

$\ln(f_{Y_T,\cdots,Y_3|Y_2,Y_1}(y_T,\cdots,y_1|\theta) = -\frac{T-2}{2}(\ln(2\pi) + \ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=3}^{T}(y_t - \mu - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2$

(b) The ML estimator maximises the log likelihood function. We can see the maximiser for μ, ϕ_1 is the one which minimises $-\frac{1}{2\sigma^2}\sum_{t=3}^{T}(y_t - \mu - \phi_1 y_{t-1})^2$, hence the same as the maximiser of the sum of squared residuals. Hence the ML estimator is equivalent to OLS in this instance. Normality of errors is not needed to ensure the OLS estimator is consistent, hence establishing that the ML estimator isn't inconsistent in this example.

3. Consider an MA(1) process

$$Y_t = \mu_0 + \theta_{10}\varepsilon_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{2}$$

where a researcher observes a sample realisation of size T, $(y_1, \cdots, y_T)$ from $\{Y_t\}$.

(a) **Conditional likelihood:** $f_{Y_T,\cdots,Y_1|\varepsilon_0=0}(y_T,\cdots,y_1,;\theta) = f_{Y_1|\varepsilon_0=0}(y_1;\theta)\prod_{t=2}^{T} f_{Y_t|Y_{t-1},\cdots,Y_1,\varepsilon_0=0}(y_t|y_{t-1},\cdots,y_1;\theta)$

Conditional on $\varepsilon_0 = 0$ can express realised shocks ε_t as function of $(y_{t-1}, \cdots, y_1)$

$$\varepsilon_1 = y_1 - \alpha \qquad [Y_1 = y_1]$$

$$\varepsilon_2 = y_2 - \alpha - \theta_1\varepsilon_1 \qquad [Y_1 = y_1, Y_2 = y_2] \qquad \vdots$$

$$\varepsilon_T = y_T - \alpha - \theta_1\varepsilon_{T-1} \qquad [Y_1 = y_1, \cdots, Y_T = y_T]$$

$Y_1|\varepsilon_0 = 0 \sim N(\alpha, \sigma^2)$ $(Y_t|Y_{t-1} = y_{t-1}, \cdots Y_1 = y_1, \varepsilon_0 = 0)$ same

distribution as $(Y_t|\varepsilon_{t-1})$ for $t = 2, .., T$

i.e. $(Y_t|Y_{t-1} = y_{t-1}, \cdots, Y_1 = y_1, \varepsilon_0 = 0) \sim N(\alpha + \theta_1\varepsilon_{t-1}, \sigma^2)$

Hence we can put this together to find the likelihood function as $\frac{1}{\sqrt{2\sigma^2}}\exp(-\frac{1}{2\sigma^2}(y_t-\alpha)^2)\prod_{t=2}^{T}\frac{1}{\sqrt{2\sigma^2}}\exp(-\frac{1}{2\sigma^2}(y_t-\alpha-\theta_1\epsilon_{t-1}))$ where ϵ_t for $t = 2, \cdots, T$ are calculated using the recursion above and are a function observed data and the parameters. Then take logarithms to find the log-likelihood. [Check Hamilton, pp. 127-128 who runs through the derivation of the conditional MA(1) log likelihood function]

(b) If the true θ_1 is bigger than 1 (in absolute value) then conditioning on ε_0 is not innocuous (for large sample sizes). Note that we derived ε_t as

a function of the observable data, noting that $\varepsilon_t = y_t - \mu - \theta_1\varepsilon_{t-1}$ and starting at $\varepsilon_0 = 0$. This is an AR(1) process in ε_t, and we know θ_1 here must be less than 1 in absolute value or the impact of past shocks doesn't vanish for large sample sizes.

We can recurse the shocks back t periods and find:
$\varepsilon_t = (y_t-\mu) - \theta(y_{t-1}-\mu) + \theta^2(y_{t-2}-\mu) - \cdots + (-1)^t\theta_1^t\varepsilon_0$

If $|\theta_1| < 1$ then $(-1)^t\theta_1^t\varepsilon_0$ is small for t large, and conditioning on $\varepsilon_0 = 0$ when it may not be true will have no asymptotic (large sample) impact on the approximation of the likelihood function. This is not the case when $|\theta_1| > 1$. In this case we can exploit invertibility of MA models and rewrite the model equivalent with a parameter $1/\theta_1$ and variance of the shocks $\theta_1^2\sigma^2$

4. Derive the autocovaiance function of the VMA(∞) process

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{s=0}^{\infty}\Theta_s\boldsymbol{\varepsilon}_{t-s} \qquad \varepsilon_t \sim WN(\Sigma).$$

where Σ is $k \times k$ var-covariance matrix of $\boldsymbol{\varepsilon}_t$.

Need to find $Cov(Y_t, Y_{t-k})$ for all values of k (the autocovariance function). We can find the covariances for $k = 1, 2, \cdots$ then use the fact $\Gamma(-k) = \Gamma(k)'$ to find the covariance for lags $k = -1, 2....$

We also need to find for $k = 0$, i.e. the variance. Utilising the assumption that $\boldsymbol{\varepsilon}_t \sim$ WN(Σ) it also follows that

$$Var[\mathbf{Y}_t] = \sum_{j=0}^{\infty}\Theta_1 Var(\boldsymbol{\varepsilon}_{t-j})\Theta_1' = \sum_{j=0}^{\infty}\Theta_1\Sigma\Theta_1'$$

Consider the covariance at lag 1, noting that $\mathbb{E}[\mathbf{Y}_t] = \mathbb{E}[\mathbf{Y}_{t-1}] = \mu$ (show this taking expectations of the VMA and using the fact the mean of the shocks are all zero by WN.)

$\Gamma(1) = \mathbb{E}[(\mathbf{Y}_t - \mu)(\mathbf{Y}_{t-1} - \mu)']$ this is equal to

$$\mathbb{E}\left[(\boldsymbol{\varepsilon}_t + \Theta_1\boldsymbol{\varepsilon}_{t-1} + ...)(\boldsymbol{\varepsilon}_{t-1} + \Theta_1\boldsymbol{\varepsilon}_{t-2} + ...)'\right].$$

Arrange this calculation in a **tabular form** (in essence writing to clearly see the common shocks in both $\mathbf{Y}_t$ and $\mathbf{Y}_{t-1}$), placing shocks in a common time in each column (i.e. the only shocks that are correlated due to WN assumption of ε_t), as

$$\begin{array}{lccccc} & \mathbb{E}[(\quad \varepsilon_t & +\Theta_1\varepsilon_{t-1} & & +\Theta_2\varepsilon_{t-2} & +\ldots) \\ & \times (& \varepsilon'_{t-1} & & +\varepsilon'_{t-2}\Theta'_1 & +\ldots)] \\ = & & \Theta_1\mathbb{E}\left[\varepsilon_{t-1}\varepsilon'_{t-1}\right] & +\Theta_2\mathbb{E}\left[\varepsilon_{t-2}\varepsilon'_{t-2}\right]\Theta'_1 + \cdots & & \end{array}$$

The final line follows using the argument that, out of all the products $\varepsilon_{t-i}\varepsilon_{t-j}$ for i,j whose expectation is required, only those for which $i = j$ will have non-zero expected value Σ, because the ε_t in different time periods are uncorrelated by the WN assumption. Hence,

$$\Gamma(1) = Cov\left[\mathbf{Y}_t, \mathbf{Y}_{t-1}\right] = \sum_{j=0}^{\infty} \Theta_{j+1}\Sigma\Theta'_j$$

Notice that the difference of lags between t and $t-1$ (namely 1), is reflected in the differences of the subscripts on the right hand side.

We can perform the prodecure for any $k > 1$, and find

$$\Gamma(k) = Cov\left[\mathbf{Y}_t, \mathbf{Y}_{t-k}\right] = \sum_{j=0}^{\infty} \Theta_{j+k}\Sigma\Theta'_j$$

(work through this to convince yourself, if you have any trouble with this please email to come to my office hour or I can make an online video clip.)