Problem Set 6 Solution

May 21st, 2009 by Yang

1. Causal Expression of AR(2)

Let $\phi(z) := (1-\alpha z)(1-\beta z)$. Zeros of $\phi(\cdot)$ are $\frac{1}{\alpha}$ and $\frac{1}{\beta}$, both of which are greater than 1 in absolute value by the assumption in the question. By the theorem mentioned in the lecture, we can write y_t as causal expression

$$y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} = \frac{1}{\phi(L)} \varepsilon_t$$

Note that for $|z| < 1$,

$$\frac{1}{\phi(z)} = \frac{1}{1-\alpha z}\frac{1}{1-\beta z} = \left(1 + \alpha z + \alpha^2 z^2 + \cdots\right)\left(1 + \beta z + \beta^2 z^2 + \cdots\right)$$

and thus

$$\begin{aligned} y_t &= \left(1 + \alpha L + \alpha^2 L^2 + \cdots\right)\left(1 + \beta L + \beta^2 L^2 + \cdots\right)\varepsilon_t \\ &= \left(1 + (\alpha+\beta)L + (\alpha^2 + \alpha\beta + \beta^2)L^2 + \cdots\right)\varepsilon_t \\ &= \left(\sum_{j=0}^{\infty}\left(\sum_{k=0}^{j}\alpha^k\beta^{j-k}\right)L^j\right)\varepsilon_t \\ &= \sum_{j=0}^{\infty}\sum_{k=0}^{j}\alpha^k\beta^{j-k}\varepsilon_{t-j} \end{aligned}$$

Therefore,

$$c_j = \sum_{k=0}^{j}\alpha^k\beta^{j-k}$$

2. Estimation of AR(1)

We have observations (y_1, y_2). We need to obtain the distribution of y_1 to do the MLE. Let us guess $y_1 \sim N(\mu_y, \sigma_y^2)$. Since $y_2 = \alpha y_1 + \varepsilon_t$,

$$y_2 \sim N\left(\alpha\mu_y, \alpha^2\sigma_y^2 + \sigma^2\right)$$

The AR(1) model is stationary, so every observation has the same mean and variance. This implies that the unconditional mean and variance of y_2 is the same with those of y_1. Therefore,

$$\begin{aligned} \alpha\mu_y &= \mu_y \\ \alpha^2\sigma_y^2 + \sigma^2 &= \sigma_y^2 \end{aligned}$$

which implies that

$$y_1 \sim N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$$

The likelihood function is now

$$\begin{aligned} L(\alpha, \sigma^2) &= f(y_1; \alpha, \sigma^2) f(y_2|y_1; \alpha, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\alpha^2)}} \exp\left(-\frac{y_1^2}{2\sigma^2/(1-\alpha^2)}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_2-\phi y_1)^2}{2\sigma^2}\right) \end{aligned}$$

and thus

$$\begin{aligned} \mathcal{L}(\alpha, \sigma^2) &= \log f(y_0; \alpha, \sigma^2) + \log f(y_2|y_1; \alpha, \sigma^2) \\ &= \left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\frac{\sigma^2}{1-\alpha^2} - \frac{y_1^2}{2\sigma^2/(1-\alpha^2)}\right) + \left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{(y_2-\alpha y_1)^2}{2\sigma^2}\right) \\ &= -\log 2\pi - \log\sigma^2 + \frac{1}{2}\log(1-\alpha^2) - \frac{y_1^2(1-\alpha^2)}{2\sigma^2} - \frac{(y_2-\alpha y_1)^2}{2\sigma^2} \end{aligned}$$

Obtain the FOC's as follows.

$$\begin{aligned} -\frac{\alpha}{1-\alpha^2} + \frac{y_1^2\alpha}{\sigma^2} + \frac{(y_2-\alpha y_1)y_1}{\sigma^2} &= 0 \\ -\frac{1}{\sigma^2} + \frac{y_1^2(1-\alpha^2)}{2\sigma^4} + \frac{(y_2-\alpha y_1)^2}{2\sigma^4} &= 0 \end{aligned}$$

The equations can be simplified as follows.

$$\begin{aligned} -\alpha\sigma^2 + (1-\alpha^2)y_1y_2 &= 0 \\ -2\sigma^2 + y_1^2 - 2\alpha y_1y_2 + y_2^2 &= 0 \end{aligned}$$

From the second equation, obtain an expression of σ^2 in terms of y_1, y_2 and α as

$$\sigma^2 = \frac{y_1^2 - 2\alpha y_1y_2 + y_2^2}{2}$$

and plug in into the first equation, then

$$-\alpha\left(\frac{y_1^2 - 2\alpha y_1y_2 + y_2^2}{2}\right) + (1-\alpha^2)y_1y_2 = 0$$

or

$$-\alpha\left(y_1^2 + y_2^2\right) + 2y_1y_2 = 0$$

and thus

$$\widehat{\alpha} = \frac{2y_1y_2}{y_1^2 + y_2^2}$$

Plug this back into the equations, then

$$\widehat{\sigma}^2 = \frac{(y_1^2 - y_2^2)^2}{2(y_1^2 + y_2^2)}$$

3. Size of T-test in AR(1) Model

Let $c = \mu$, then $y_t = c + \alpha y_{t-1} + u_t$. The OLS estimator is given by

$$\begin{pmatrix} \widehat{c} \\ \widehat{\alpha} \end{pmatrix} = \left[\sum_{t=1}^{T} \begin{pmatrix} 1 \\ y_{t-1} \end{pmatrix} \begin{pmatrix} 1 & y_{t-1} \end{pmatrix} \right]^{-1} \sum_{t=1}^{T} \begin{pmatrix} 1 \\ y_{t-1} \end{pmatrix} y_t$$

and the variance is estimated by

$$var(\widehat{\alpha}) = \widehat{\sigma}^2 \left[\sum_{t=1}^{T} \begin{pmatrix} 1 \\ y_{t-1} \end{pmatrix} \begin{pmatrix} 1 & y_{t-1} \end{pmatrix} \right]^{-1}$$

where $\widehat{\sigma}^2 := \frac{1}{T-2} \sum_{t=1}^{T} (y_t - \widehat{c} - \widehat{\alpha} y_{t-1})^2$. The nominal 95% two-sided CI for α is obtained as follows.

$$\mathrm{CI}_{0.95} = \left[\widehat{\alpha} - 1.96\sqrt{var(\widehat{\alpha})}, \widehat{\alpha} + 1.96\sqrt{var(\widehat{\alpha})} \right]$$

Then the empirical coverage probability $\widehat{\Pr}(\alpha \in CI_{0.95}) = \frac{1}{R} \sum_{r=1}^{R} 1(\alpha \in CI_{0.95}^{r})$ is given as follows.

True alpha	0.00	0.30	0.60	0.90	0.99
Mean Bias	-0.0136	-0.0202	-0.0274	-0.0320	-0.0070
Median Bias	-0.0170	-0.0181	-0.0203	-0.0252	-0.0053
St. Dev.	0.1000	0.0920	0.0822	0.0503	0.0134
R.M.S.E.	0.1009	0.0941	0.0866	0.0596	0.0152
Cov. Prob.	0.9495	0.9540	0.9505	0.9290	0.9230

In fact, there seems to exist no problem. Note that, however, the given process is nonstationary, since the unconditional mean and variance of y_t depends on t. Correct this by letting $c = \mu(1 - \alpha)$ so that

$$y_t = \mu(1 - \alpha) + \alpha y_{t-1} + u_t$$

then $y_t \sim N\left(\mu, \frac{\sigma^2}{1-\alpha^2}\right)$ for any t as easily verified. Use this to generate data and estimate α, then we get the following result. Now we can see that when $\alpha \to 1$, the empirical coverage probability is strictly less than 0.95. This means that we obtain too narrow confidence intervals as $\alpha \to 1$. Simply inverting a t-test is thus not a good idea to obtain CI when α is believed to be close to 1.

True alpha	0.00	0.30	0.60	0.90	0.99
Mean Bias	-0.0136	-0.0202	-0.0275	-0.0386	-0.0496
Median Bias	-0.0170	-0.0182	-0.0199	-0.0306	-0.0406
St. Dev.	0.1000	0.0921	0.0824	0.0553	0.0439
R.M.S.E.	0.1009	0.0942	0.0869	0.0674	0.0662
Cov. Prob.	0.9495	0.9555	0.9480	0.9225	0.7845

4. Inclusion of Irrelevant Parameter

Note first that in the hints, θ is a row vector, and thus x_t is also a row vector. For completeness, let us begin from the scratch. The conditional likelihood of an MA(q) process is given by

$$\mathcal{L}(\theta) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \sum_{t=1}^{T}\frac{\varepsilon_t(\theta)^2}{2\sigma^2}$$

where

$$\begin{aligned} \varepsilon_t(\theta) &:= y_t - \theta_1\varepsilon_{t-1}(\theta) - \cdots - \theta_1\varepsilon_{t-q}(\theta) \\ \theta &:= (\theta_1, \cdots, \theta_q) \end{aligned}$$

The first order conditions with respect to θ read as

$$0 = \frac{\partial\mathcal{L}(\theta)}{\partial\theta} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\frac{\partial\varepsilon_t(\theta)}{\partial\theta}\varepsilon_t = -\frac{1}{\sigma^2}\sum_{t=1}^{T}x_t\varepsilon_t$$

where $x_t := -\frac{\partial\varepsilon_t(\theta)}{\partial\theta}$. So the information matrix is given by

$$\begin{aligned} E\left[\frac{\partial\mathcal{L}}{\partial\theta'}\frac{\partial\mathcal{L}}{\partial\theta}\right] &= \frac{1}{\sigma^4}E\left[\sum_{t=1}^{T}x_t'\varepsilon_t\sum_{s=1}^{T}x_s\varepsilon_s\right] \\ &= \frac{1}{\sigma^4}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[x_t'x_s\varepsilon_t\varepsilon_s\right] \\ &= \frac{1}{\sigma^4}\sum_{t=1}^{T}E\left[x_t'x_t\right]E\varepsilon_t^2 = \frac{1}{\sigma^2}\sum_{t=1}^{T}Ex_t'x_t \end{aligned}$$

The third equality holds by the following. Note that

$$\begin{aligned} x_t = -\frac{\partial\varepsilon_t(\theta)}{\partial\theta} &= (\varepsilon_{t-1}, \cdots, \varepsilon_{t-q}) + \theta_1\frac{\partial\varepsilon_{t-1}}{\partial\theta} + \cdots + \theta_q\frac{\partial\varepsilon_{t-q}}{\partial\theta} \\ &= (\varepsilon_{t-1}, \cdots, \varepsilon_{t-q}) - \theta_1 x_{t-1} - \cdots - \theta_q x_{t-q} \end{aligned}$$

x_t is an AR(q) process and thus can be expressed as a weigthed sum of past errors under some condition.[1] For $t > s$, x_t, x_s and ε_s are independent of ε_t, so $E[x_t'x_s\varepsilon_t\varepsilon_s] = E[x_t'x_s\varepsilon_s]E\varepsilon_t = 0$. For $t < s$, x_t, x_s and ε_t are independent of ε_s, so $E[x_t'x_s\varepsilon_t\varepsilon_s] = E[x_t'x_s\varepsilon_t]E\varepsilon_s = 0$. Since the asymptotic variance of $\widehat{\theta}_{MLE}$ is the limit of T times the inverse of the information matrix,

$$\begin{aligned} \text{AV}(\widehat{\theta}_{MLE}) &= \lim_{T\to\infty} T\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}Ex_t'x_t\right)^{-1} \\ &= \lim_{T\to\infty}\left(\frac{1}{\sigma^2}E\left[\frac{1}{T}\sum_{t=1}^{T}x_t'x_t\right]\right)^{-1} \\ &= \sigma^2\left(\text{plim}\frac{1}{T}\sum_{t=1}^{T}x_t'x_t\right)^{-1} = \sigma^2(Ex_t'x_t)^{-1} \end{aligned}$$

[1] The condition is invertibility of the MA(q) process of y_t, which is equivalent to causality of the AR(q) process of x_t. If x_t is noncausal stationary, we may use a noncausal expression to claim $E[x_t'x_s\varepsilon_t\varepsilon_s] = 0$ for $t \neq s$.

where the third equality holds by the (dependent version of) law of large numbers and the continuous mapping theorem, and the last equality holds by identical distribution of x_t.

(i) MA(1) model

Consider $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$, then $x_t := -\frac{\partial \varepsilon_t}{\partial \theta_1} = -\theta_1 x_{t-1} + \varepsilon_{t-1}$ is an AR(1) process. As is well known, the variance of the AR(1) process is

$$Ex_t^2 = \frac{\sigma^2}{1-\theta_1^2}$$

so the asymptotic variance of $\widehat{\theta}_1$ is

$$\mathrm{AV}(\widehat{\theta}_1) = \sigma^2 (Ex_t^2)^{-1} = 1 - \theta_1^2$$

(ii) MA(2) model with $\theta_2 = 0$

Consider $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$. We estimate both θ_1 and θ_2, although θ_2 is actually 0.

$$x_t := -\frac{\partial \varepsilon_t(\theta)}{\partial \theta} = (\varepsilon_{t-1}, \varepsilon_{t-2}) - \theta_1 x_{t-1} - \theta_2 x_{t-2}$$

is an AR(2) process with $\theta_2 = 0$. This implies that x_t is actually an AR(1) process.[2]

$$x_t = -\theta_1 x_{t-1} + (\varepsilon_{t-1}, \varepsilon_{t-2})$$

So

$$Ex_t' x_t = \theta_1^2 E x_{t-1}' x_{t-1} + E\begin{bmatrix} \varepsilon_{t-1}^2 & \varepsilon_{t-1}\varepsilon_{t-2} \\ \varepsilon_{t-2}\varepsilon_{t-1} & \varepsilon_{t-2}^2 \end{bmatrix} - \theta_1 E\left[x_{t-1}'(\varepsilon_{t-1}, \varepsilon_{t-2})\right] - \theta_1 E\left[\begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} x_{t-1}\right]$$

[2] We may obtain $Ex_t'x_t$ from an AR(2) process, and use $\theta_2 = 0$ later. Since $x_t'x_t = [(\varepsilon_{t-1}, \varepsilon_{t-2})' - \theta_1 x_{t-1}' - \theta_2 x_{t-2}'] \cdot [(\varepsilon_{t-1}, \varepsilon_{t-2}) - \theta_1 x_{t-1} - \theta_2 x_{t-2}]$,

$$Ex_t'x_t = \theta_1^2 Ex_{t-1}'x_{t-1} + \theta_2^2 Ex_{t-2}'x_{t-2} + \theta_1\theta_2 Ex_{t-1}'x_{t-2} + \theta_1\theta_2 Ex_{t-2}'x_{t-1} - \cdots$$

Also since $x_t'x_{t-1} = [(\varepsilon_{t-1}, \varepsilon_{t-2})' - \theta_1 x_{t-1}' - \theta_2 x_{t-2}']\, x_{t-1}$,

$$Ex_t'x_{t-1} = E(\varepsilon_{t-1}, \varepsilon_{t-2})' x_{t-1} - \theta_1 Ex_{t-1}'x_{t-1} - \theta_2 Ex_{t-2}'x_{t-1}$$

Arranging the above two equations and making use of stationarity of x_t, we have

$$(1 - \theta_1^2 - \theta_2^2) Ex_t'x_t = \theta_1\theta_2 Ex_t'x_{t-1} + \theta_1\theta_2 Ex_{t-1}'x_t + \begin{pmatrix} \sigma^2 & -\theta_1\sigma^2 \\ -\theta_1\sigma^2 & \sigma^2 \end{pmatrix}$$
$$Ex_t'x_{t-1} = -\theta_1 Ex_t'x_t - \theta_2 Ex_{t-1}'x_t + \begin{pmatrix} 0 & 0 \\ \sigma^2 & 0 \end{pmatrix}$$
$$Ex_{t-1}'x_t = -\theta_1 Ex_t'x_t - \theta_2 Ex_t'x_{t-1} + \begin{pmatrix} 0 & \sigma^2 \\ 0 & 0 \end{pmatrix}$$

Solving the system of equations simultaneously,

$$Ex_t'x_t = \frac{\sigma^2}{(1-\theta_2)[(1+\theta_2)^2 - \theta_1^2]} \begin{pmatrix} 1+\theta_2 & -\theta_1 \\ -\theta_1 & 1+\theta_2 \end{pmatrix}$$

Now use $\theta_2 = 0$, then this simplifies to $Ex_t'x_t = \frac{\sigma^2}{1-\theta_1^2}\begin{pmatrix} 1 & -\theta_1 \\ -\theta_1 & 1 \end{pmatrix}$, which is the same as we get using the other method.

By stationarity of x_t, $Ex_t'x_t = Ex_{t-1}'x_{t-1}$. Note also that

$$E\left[\begin{array}{cc} \varepsilon_{t-1}^2 & \varepsilon_{t-1}\varepsilon_{t-2} \\ \varepsilon_{t-2}\varepsilon_{t-1} & \varepsilon_{t-2}^2 \end{array}\right] = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

and that

$$\begin{aligned} E\left[\begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} x_{t-1}\right] &= E\left[\begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} \begin{pmatrix} \varepsilon_{t-2} & \varepsilon_{t-3} \end{pmatrix}\right] - \theta_1 \underbrace{E\left[\begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} x_{t-2}\right]}_{=0 \ \because\ x_{t-2} \perp (\varepsilon_{t-1}, \varepsilon_{t-2})} \\ &= E\left[\begin{array}{cc} \varepsilon_{t-1}\varepsilon_{t-2} & \varepsilon_{t-1}\varepsilon_{t-3} \\ \varepsilon_{t-2}^2 & \varepsilon_{t-2}\varepsilon_{t-3} \end{array}\right] \\ &= \begin{pmatrix} 0 & 0 \\ \sigma^2 & 0 \end{pmatrix} \end{aligned}$$

Therefore

$$(1-\theta_1^2)Ex_t'x_t = \begin{pmatrix} \sigma^2 & -\theta_1\sigma^2 \\ -\theta_1\sigma^2 & \sigma^2 \end{pmatrix}$$

or equivalently,

$$Ex_t'x_t = \frac{\sigma^2}{1-\theta_1^2} \begin{pmatrix} 1 & -\theta_1 \\ -\theta_1 & 1 \end{pmatrix}$$

So the asymptotic variance of $\widehat{\theta}_{MLE}$ is

$$\text{AV}(\widehat{\theta}_{MLE}) = \sigma^2 (Ex_t'x_t)^{-1} = \begin{pmatrix} 1 & \theta_1 \\ \theta_1 & 1 \end{pmatrix}$$

The asymptotic variance of $\widehat{\theta}_1$ is 1.

Asymptotic Relative Efficiency

From the above result, the asymptotic relative efficiency of $\widehat{\theta}_1$ obtained from MA(2) model with $\theta_2 = 0$ to that obtained from MA(1) model is

$$\text{ARE} = \frac{1-\theta_1^2}{1} = 1 - \theta_1^2$$

This is less than 1. Estimating irrelevant parameter θ_2 increases the variance of $\widehat{\theta}_1$, and thus is relatively inefficient.