Problem Set 1 Solution

April 15th, 2009 by Yang

1. Monte Carlo Simulation: Weak Instruments Problem in the Linear IV Model

First, let us talk about how to generate a normal random vector with correlation. When $w_i \sim N(0, I_K)$, we have $z_i := Dw_i \sim N(0, DD')$. We usually generate a $n \times K$ matrix W as

$$W = \begin{pmatrix} w_1' \\ \vdots \\ w_n' \end{pmatrix}$$

and thus to generate a $n \times K$ matrix Z whose each row z_i' is a vector with variance DD',

$$Z := WD' = \begin{pmatrix} w_1'D' \\ \vdots \\ w_n'D' \end{pmatrix} = \begin{pmatrix} (Dw_1)' \\ \vdots \\ (Dw_n)' \end{pmatrix}$$

Therefore, the MATLAB code would be as follows

```
n = 100;
K = 10;
D = rand(K);
W = randn(n,K);
Z = W*D';
```

The function `mvnrnd()` also generates such a matrix. We will get the same Z by the following.

```
mu = zeros(1,K);
Z = mvnrnd(mu,D*D',n);
```

The former method is more flexible in the sense that once we generate W, we can use the same W to form Z for any η. Now turning to generating $s_i' = (\varepsilon_i, u_i)'$, we can use a similar method. But in this case, we have to find some matrix like D above. The function `chol()` does this. $T = chol(\Omega)$ gives T such that $T'T = \Omega$. Note that the order of multiplication is a little different. So the following gives S whose each row s_i' is distributed normally with variance Ω.

```
rho = 0.5;
Omega = [1,rho;rho,1];
T = chol(Omega);
V = randn(n,2);
S = V*T;
```

Of course, we can use the following instead to get the same S. But the former method enables us to use the same random numbers for any ρ within each iteration.

```
mu = [0,0];
S = mvnrnd(mu,Omega,n);
```

Now look at the estimators. With A_n as a weighting matrix, GMM is equivalent to solving

$$\min_{\beta} \frac{1}{2}(Y - X\beta)' Z A_n' A_n Z' (Y - X\beta)$$

which gives the following first order condition.

$$-X' Z A_n' A_n Z' (Y - X\widehat{\beta}) = 0$$

So,

$$\widehat{\beta} = \left(X' Z A_n' A_n Z' X \right)^{-1} X' Z A_n' A_n Z' Y$$

The first estimator uses $A_n := I_K$, and the second estimator uses A_n such that $A_n' A_n = (Z'Z/n)^{-1}$, so

$$\widehat{\beta}_1 = \left(X' Z Z' X \right)^{-1} X' Z Z' Y$$

$$\widehat{\beta}_2 = \left(X' Z (Z'Z)^{-1} Z' X \right)^{-1} X' Z (Z'Z)^{-1} Z' Y \qquad \text{(2SLS estimator)}$$

We get 1000 estimates for each combination of K, η, and ρ. To get mean bias, we calcualte $E[\widehat{\beta}] - \beta$ by approximation. Also median bias, standard deviation, and root mean squared error can be obtained in a similar way. See the MATLAB code posted with this solution for details. (You are free to use other programs such as GAUSS, R, and SAS.)

Results

	K=1				K=10			
(eta / rho)	(.05/0)	(.05/.5)	(1/0)	(1/.5)	(.05/0)	(.05/.5)	(1/0)	(1/.5)
GMM1 mean bias	1.0310	-0.8982	-0.0006	-0.0058	0.0054	0.0229	0.0002	0.0003
TSLS mean bias	1.0310	-0.8982	-0.0006	-0.0058	0.0021	0.0415	0.0001	0.0002
GMM1 med bias	0.0424	0.4266	-0.0008	-0.0008	0.0064	0.0264	0.0004	0.0004
TSLS med bias	0.0424	0.4266	-0.0008	-0.0008	0.0067	0.0467	0.0002	0.0003
GMM1 st. dev.	38.0180	45.7800	0.1022	0.1034	0.1097	0.1086	0.0055	0.0055
TSLS st. dev.	38.0180	45.7800	0.1022	0.1034	0.0979	0.0954	0.0050	0.0050
GMM1 rms error	38.0130	45.7660	0.1022	0.1035	0.1098	0.1109	0.0055	0.0055
TSLS rms error	38.0130	45.7660	0.1022	0.1035	0.0978	0.1040	0.0050	0.0050

Interpretation

1. When $\eta = 0.05$, there is weak relation between x_i and z_i. On the other hand, $\eta = 1$ says that z_i explains x_i a lot compared to the error u_i, which stands for strong instruments case. In the simulation result, the case where $\eta = 1$ finds estimators much better than the other case.

2. When $\rho = 0$, there is no endogeneity issue. Using weak instruments ($K = 1, \eta = 0.05$) in this case is as bad as in the case where there is indeed an endogeneity problem, while using better instruments ($K = 10$ or $\eta = 1$) work quite well. However, note that the OLS would be the best linear estimator when $\rho = 0$, so the IV estimator will always be inefficient. When $\rho = 0.5$, we have endogeneity problem. IV estimators are generally good, except for the case of weak instruments ($K = 1, \eta = 0.05$). Especially, when $K = 10$ and $\eta = 1$, IV estimators are excellent, doing as good as in the case where there is no endogeneity problem.

3. When $K = 1$, $(Z'Z)^{-1}$ is a scalar, so $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are numerically the same in any case as we can see. There seems to be greater biases, variances, and mean squared errors compared to those of corresponding cases in $K = 10$. More instruments work better in any case.

4. While $n = 100$ is not sufficient for the large sample, the Monte Carlo simulation finds that 2SLS estimator has less variance than the other estimator in most of the cases. This is consistent with the theory that 2SLS estimator has the least asymptotic variance when errors are conditionally homoskedastic. While 2SLS estimator has greater biases when $K = 10$ and $\eta = 0.05$, mean squared errors are still less. Note again that when $K = 1$, 2SLS and the other estimators are numerically the same, so there cannot be any comparison.

2. Some Proofs: (i) (a) While we can use even the definition of limit in proving these claims, this approach is not taken here, and it is assumed that everybody is familiar with the definition of limit. Let X_n and Y_n be $o_p(1)$. First we want to prove $X_n + Y_n = o_p(1)$. Take any $\varepsilon > 0$. By definition, we have

$$\lim_{n\to\infty} \Pr\left(|X_n| > \frac{\varepsilon}{2}\right) = 0$$
$$\lim_{n\to\infty} \Pr\left(|Y_n| > \frac{\varepsilon}{2}\right) = 0$$

Note that the event $|X_n + Y_n| > \varepsilon$ is a subset of the event $|X_n| + |Y_n| > \varepsilon$, which is a subset of the event that $|X_n| > \frac{\varepsilon}{2}$ or $|Y_n| > \frac{\varepsilon}{2}$. So

$$\begin{aligned}
\lim_{n\to\infty} \Pr(|X_n + Y_n| > \varepsilon) &\leq \lim_{n\to\infty} \Pr(|X_n| + |Y_n| > \varepsilon) \\
&\leq \lim_{n\to\infty} \Pr\left(|X_n| > \frac{\varepsilon}{2} \text{ or } |Y_n| > \frac{\varepsilon}{2}\right) \\
&\leq \lim_{n\to\infty} \left[\Pr\left(|X_n| > \frac{\varepsilon}{2}\right) + \Pr\left(|Y_n| > \frac{\varepsilon}{2}\right)\right] \\
&= \lim_{n\to\infty} \Pr\left(|X_n| > \frac{\varepsilon}{2}\right) + \lim_{n\to\infty} \Pr\left(|Y_n| > \frac{\varepsilon}{2}\right) = 0
\end{aligned}$$

A probability is always nonnegative, so $\lim_{n\to\infty} \Pr(|X_n + Y_n| > \varepsilon) = 0$. Next we want to prove $X_n Y_n = o_p(1)$. Take any $\varepsilon > 0$. By definition, we have

$$\lim_{n\to\infty} \Pr\left(|X_n| > \sqrt{\varepsilon}\right) = 0$$
$$\lim_{n\to\infty} \Pr\left(|Y_n| > \sqrt{\varepsilon}\right) = 0$$

In a similar way,

$$\begin{aligned}
\lim_{n\to\infty} \Pr(|X_n Y_n| > \varepsilon) &\leq \lim_{n\to\infty} \Pr\left(|X_n| > \sqrt{\varepsilon} \text{ or } |Y_n| > \sqrt{\varepsilon}\right) \\
&\leq \lim_{n\to\infty} \left[\Pr\left(|X_n| > \sqrt{\varepsilon}\right) + \Pr\left(|Y_n| > \sqrt{\varepsilon}\right)\right] \\
&= \lim_{n\to\infty} \Pr\left(|X_n| > \sqrt{\varepsilon}\right) + \lim_{n\to\infty} \Pr\left(|Y_n| > \sqrt{\varepsilon}\right) = 0
\end{aligned}$$

The choice of X_n, Y_n, and ε are arbitrary, so we are done.

(b) Let X_n be $o_p(1)$. We want to prove that $g(a+X_n)-g(a)$ is $o_p(1)$. Take any $\varepsilon>0$. Because g is continous at a, we can find $\delta>0$ such that $|g(b)-g(a)|<\varepsilon$ for any b with $|b-a|<\delta$. By definition, for such δ

$$\lim_{n\to\infty} \Pr(|X_n|<\delta)=1$$

Now note that, by construction of δ, the event $|g(a+X_n)-g(a)|<\varepsilon$ contains the event $|X_n|<\delta$. Therefore,

$$\lim_{n\to\infty} \Pr\Big(|g(a+X_n)-g(a)|<\varepsilon\Big) \geq \lim_{n\to\infty} \Pr(|X_n|<\delta)=1$$

A probability cannot be greater than 1, so $\lim_{n\to\infty} \Pr\Big(|g(a+X_n)-g(a)|<\varepsilon\Big)=1$, which completes the proof.

(ii) Recall that

$$\limsup_{n\to\infty}(a_n)=\lim_{n\to\infty}\sup_{k\geq n}(a_k)$$

Use the fact that[1]

$$\sup_{k\geq n}(a_k+b_k)\leq \sup_{k\geq n}(a_k)+\sup_{k\geq n}(b_k)$$

Taking limit,

$$\lim_{n\to\infty}\sup_{k\geq n}(a_k+b_k)\leq \lim_{n\to\infty}\sup_{k\geq n}(a_k)+\lim_{n\to\infty}\sup_{k\geq n}(b_k)$$

which is the desired result. The equality does not hold. As a counterexample, consider $a_n:=(-1)^n$ and $b_n:=(-1)^{n+1}$. Then $a_n+b_n=0$, so the LHS is 0, but the RHS would be 2.

3. Minimum Distance Estimator

We have to assume the following.

1. $A_n \xrightarrow{p} A$ and $\widehat{\pi}_n \xrightarrow{p} \pi_0$.
2. There exists a unique value $\theta_0\in\Theta$ such that $\pi_0=g(\theta_0)$.
3. A is nonsingular.
4. g is continuous.
5. Θ is compact.
6. (Assumption EE) $\widehat{\theta}_n\in\Theta$ and $Q_n(\widehat{\theta}_n)\leq \inf_{\theta\in\Theta} Q_n(\theta)+o_p(1)$.

[1] This can be easily proven. Let $c_n=a_n+b_n$. Choose any $m\geq n$. By definition of sup, $a_m\leq \sup_{k\geq n}(a_k)$ and $b_m\leq\sup_{k\geq n}(b_k)$, which implies $c_m\leq \sup_{k\geq n}(a_k)+\sup_{k\geq n}(b_k)$. Since this holds for any $m\geq n$, we have

$$\sup_{k\geq n} c_k\leq \sup_{k\geq n}(a_k)+\sup_{k\geq n}(b_k)$$

as desired. (This proof is taken from Kyoo-il Kim's suggested solution.)

Under these assumptions, we can prove that the MD estimator is consistent. It is easy to see that Q_n converges pointwise in probability to

$$Q := \|A(\pi_0 - g(\theta))\|^2/2$$

In order to prove consistency, it suffices to prove that Assumptions ID and UWCON hold by 1-5. Now check that Assumptions ID1 hold.

(1) Θ is compact by 5.

(2) Q is continuous in θ by 4.

(3) θ_0 uniquely minimizes Q by 2 and 3. To see this, note first that $Q(\theta_0) = 0$ by 2 and that this is the minimum. Let $Q(\theta) = 0$ for some θ. This implies $A(\pi_0 - g(\theta)) = 0$. By 3, $\pi_0 - g(\theta) = 0$, and thus by 2, $\theta = \theta_0$.

Since Assumption ID1 implies Assumption ID as we will prove in the next question, Assumption ID holds. UWCON can be directly proven by 1, 4 and 5.[2] We have to prove that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| > \varepsilon\right) = 0$$

or equivalently, $\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0$. Note that by triangular inequality and Q2 (ii),

$$\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| \le \underbrace{\sup_{\theta\in\Theta} \left|Q_n(\theta) - \|A(\widehat{\pi}_n - g(\theta))\|^2/2\right|}_{(i)} + \underbrace{\sup_{\theta\in\Theta} \left|\|A(\widehat{\pi}_n - g(\theta))\|^2/2 - Q(\theta)\right|}_{(ii)}$$

Now

$$(i) = \frac{1}{2} \sup_{\theta\in\Theta} \left|(\widehat{\pi}_n - g(\theta))'(A_n'A_n - A'A)(\widehat{\pi}_n - g(\theta))\right| = o_p(1)$$

since $A_n'A_n - A'A = o_p(1)$ by 1, and $\widehat{\pi}_n - g(\theta) = O_p(1)$ by compactness of $g(\Theta)$ which is implied by 4 and 5. Also

$$\begin{aligned}
(ii) &= \frac{1}{2} \sup_{\theta\in\Theta} \left|(\widehat{\pi}_n - g(\theta))'A'A(\widehat{\pi}_n - g(\theta)) - (\pi_0 - g(\theta))'A'A(\pi_0 - g(\theta))\right| \\
&= \frac{1}{2} \sup_{\theta\in\Theta} \left|\widehat{\pi}_n'A'A\widehat{\pi}_n - \pi_0'A'A\pi_0 - 2\widehat{\pi}_n'A'Ag(\theta) + 2\pi_0A'Ag(\theta)\right| \\
&\le \underbrace{\frac{1}{2} \left|\widehat{\pi}_n'A'A\widehat{\pi}_n - \pi_0'A'A\pi_0\right|}_{=o_p(1) \text{ by } 1} + \underbrace{\sup_{\theta\in\Theta} \left|(-\widehat{\pi}_n + \pi_0)A'Ag(\theta)\right|}_{=o_p(1) \text{ by 1, compactness of } g(\Theta)} = o_p(1)
\end{aligned}$$

where the last equality holds by Q2 (i) (a). Now that (i) and (ii) imply $\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| \le o_p(1)$, the fact that $|Q_n(\theta) - Q(\theta)| \ge 0$ for any θ and any event further implies $\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| = o_p(1)$, which completes the proof.

[2] This proof is taken from Kyoo-il Kim's suggested solution.

4. ID1 implies ID

Suppose ID1 holds. Take any $\varepsilon > 0$. Since Θ is compact and $B(\theta_0, \varepsilon)$ is open, $\Theta \setminus B(\theta_0, \varepsilon)$ is compact. Since Q is continuous, $Q(\Theta \setminus B(\theta_0, \varepsilon))$ is also compact. Therefore there exists a minimum of $Q(\Theta \setminus B(\theta_0, \varepsilon))$. Define $\theta_\varepsilon \in \Theta \setminus B(\theta_0, \varepsilon)$ as the minimizer of $Q(\Theta \setminus B(\theta_0, \varepsilon))$. Clearly, $\theta_\varepsilon \neq \theta_0$, which is the unique minimizer of Q, and thus $Q(\theta_\varepsilon) > Q(\theta_0)$. So we established that for any $\varepsilon > 0$,

$$\inf_{\theta \in \Theta \setminus B(\theta_0, \varepsilon)} Q(\theta) = \min_{\theta \in \Theta \setminus B(\theta_0, \varepsilon)} Q(\theta) =: Q(\theta_\varepsilon) > Q(\theta_0)$$

Take counterexamples as follows.

1. Θ is not compact.
   Let $\Theta = [0, 2)$, and $Q(\theta) = 1 - (\theta - 1)^2$. $Q(\theta)$ is continuous and is uniquely minimized at $\theta_0 = 0$, but $\inf_{\theta \in \Theta \setminus B(\theta_0, 1)} Q(\theta) = \lim_{\theta \to 2} Q(\theta) = 0 = Q(\theta_0)$.
   Another example would be $\Theta = [0, \infty)$, and $Q(\theta) = \min(\theta, \frac{1}{\theta})$. Again, $Q(\theta)$ is continuous and is uniquely minimized at $\theta_0 = 0$, but $\inf_{\theta \in \Theta \setminus B(\theta_0, 1)} Q(\theta) = \lim_{\theta \to \infty} Q(\theta) = 0 = Q(\theta_0)$.

2. Q is not continuous.
   Let $\Theta = [0, 2]$, and $Q(\theta) = 1 - (\theta - 1)^2$ for $\theta \in [0, 2)$, and $Q(2) = 1$. $Q(\theta)$ is uniquely minimized at $\theta_0 = 0$, but $\inf_{\theta \in \Theta \setminus B(\theta_0, 1)} Q(\theta) = \lim_{\theta \to 2} Q(\theta) = 0 = Q(\theta_0)$.

3. Q is not uniquely minimized at θ_0.
   Let $\Theta = [0, 2]$, and $Q(\theta) = 1 - (\theta - 1)^2$. If $\theta_0 = 0$, $\inf_{\theta \in \Theta \setminus B(\theta_0, 1)} Q(\theta) = Q(2) = 0 = Q(\theta_0)$. If $\theta_0 = 2$, $\inf_{\theta \in \Theta \setminus B(\theta_0, 1)} Q(\theta) = Q(0) = 0 = Q(\theta_0)$. So no true θ_0 satisfies Assumption ID.