Problem Set 3 Solution

April 29th, 2009 by Yang

1. Covariance of MD Estimator

(i) Recall that the asymptotic covariance matrix of the MD estimator is given by

$$(\Gamma_0' A' A \Gamma_0)^{-1} \Gamma_0' A' A V_0 A' A \Gamma_0 (\Gamma_0' A' A \Gamma_0)^{-1}$$

where $\Gamma_0 = \frac{\partial}{\partial \theta'} g(\theta_0)$ and V_0 is the asymptotic covariance matrix of $\widehat{\pi}_n$. By the assumption given in the question, $\widehat{V}_n \xrightarrow{p} V_0$. Note also that $A_n \xrightarrow{p} A$. Under the assumption that $g \in C^1(\Theta_0)$ for some neighborhood $\Theta_0 \subset \Theta$ of θ_0 (with probability one), $\widehat{\theta}_n \xrightarrow{p} \theta_0$ implies

$$\widehat{\Gamma}_n := \frac{\partial}{\partial \theta'} g(\widehat{\theta}_n) \xrightarrow{p} \frac{\partial}{\partial \theta'} g(\theta_0) = \Gamma_0$$

by the continuous mapping theorem. Note that the assumption is implied by the ones assumed to derive the asymptotic normaility. So we can consistently estimate the asymptotic covariance matrix by

$$(\widehat{\Gamma}_n' A_n' A_n \widehat{\Gamma}_n)^{-1} \widehat{\Gamma}_n' A_n' A_n \widehat{V}_n A_n' A_n \widehat{\Gamma}_n (\widehat{\Gamma}_n' A_n' A_n \widehat{\Gamma}_n)^{-1}$$

(ii) Note that the choice of A such that $A'A = V_0^{-1}$ minimizes the asymptotic covariance, that is for any A,

$$(\Gamma_0' V_0^{-1} \Gamma_0)^{-1} \leq (\Gamma_0' A' A \Gamma_0)^{-1} \Gamma_0' A' A V_0 A' A \Gamma_0 (\Gamma_0' A' A \Gamma_0)^{-1}$$

Refer to the lecture note for its proof, since the proof coincides exactly with that in the GMM case. So an optimal weight matrix is A_n such that $A_n' A_n = \widehat{V}_n^{-1}$. In practice, we estimate $\widehat{\pi}_n$ and $\widehat{V}_n$ first, and use them to estimate $\widehat{\theta}_n$ that minimizes

$$Q_n(\theta) = \|\widehat{V}_n^{-1/2}(\widehat{\pi}_n - g(\theta))\|^2/2$$

2. Monte Carlo Simulation: Wald Test of 2SLS Estimator

Use the assumption of conditional homoskedasticity to derive the correct formula of the Wald statistic in this model. In the linear IV model, the asymptotic covariance of the 2SLS estimator is given by

$$\sigma^2 \left[E x_i z_i' (E z_i z_i')^{-1} E z_i x_i' \right]^{-1}$$

and this can be estimated consistently by

$$\widehat{\sigma}_n^2 \left[\frac{1}{n} \sum_{i=1}^n x_i z_i' \left(\frac{1}{n} \sum_{i=1}^n z_i z_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_i x_i' \right]^{-1}$$

where $\widehat{\sigma}_n^2$ is a consistent estimator of σ^2 with

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_n)^2$$

Under $H_0 : \beta_0 = 0$,

$$\begin{aligned}
W_{2SLS} &= n(\widehat{\beta}_n - \beta_0)' \left(\widehat{\sigma}_n^2 \left[\frac{1}{n}\sum_{i=1}^{n} x_i z_i' \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i' \right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i x_i' \right]^{-1} \right)^{-1} (\widehat{\beta}_n - \beta_0) \\
&= \frac{n}{\widehat{\sigma}_n^2}\widehat{\beta}_n' \left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i' \right) \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i' \right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \right) \widehat{\beta}_n \\
&= \frac{\widehat{\beta}_n' X' Z (Z'Z)^{-1} Z' X \widehat{\beta}_n}{\widehat{\sigma}_n^2} \xrightarrow{d} \chi_1^2
\end{aligned}$$

Simulation results and discussion are provided in the next qeustion.

3. LM Statistic of Continuously Updating Estimator

(i) The proof is given in the opposite order. Let us first prove $LM_{CUE}(\beta_0) \xrightarrow{p} \chi_d^2$ where $d = \dim\beta_0$, with the lemmas that $\sqrt{n}\widehat{g}(\beta_0) \xrightarrow{d} N(0, \Omega_0)$, $\widehat{\Omega}(\beta_0) \xrightarrow{p} \Omega_0$ and $D(\beta_0) \xrightarrow{p} D_0$, where Ω_0 is nonsingular and D_0 is full rank, and prove these lemmas later. By the lemmas and the Slutsky theorem,

$$D(\beta_0)'\widehat{\Omega}(\beta_0)^{-1}\sqrt{n}\widehat{g}(\beta_0) \xrightarrow{d} N\left(0, D_0'\Omega_0^{-1}D_0\right)$$

Since the LHS is a $d \times 1$ vector where $d = \dim(x_i) = \dim\beta_0$,

$$C_n := \left(D(\beta_0)'\widehat{\Omega}(\beta_0)^{-1}D(\beta_0) \right)^{-1/2} D(\beta_0)'\widehat{\Omega}(\beta_0)^{-1}\sqrt{n}\widehat{g}(\beta_0) \xrightarrow{d} N(0, I_d)$$

and thus,

$$LM_{CUE}(\beta_0) = C_n' C_n \xrightarrow{d} \chi_d^2$$

Now turn to the proof of the lemmas. Under the assumption that

- $E\|\varepsilon_i^2 z_i z_i'\| < \infty$

we can apply the CLT to get the following result.

$$\sqrt{n}\widehat{g}(\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i\varepsilon_i \xrightarrow{d} N(0, E\varepsilon_i^2 z_i z_i')$$

Define $\Omega_0 := E\varepsilon_i^2 z_i z_i'$. Under the same assumption, we apply the LLN to get

$$\widehat{\Omega}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta_0) g_i(\beta_0)' = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 z_i z_i' \xrightarrow{p} \Omega_0$$

Finally, to show $D(\beta_0) \xrightarrow{p} D_0$ for some D_0, note first

$$D(\beta_0) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \widehat{g}(\beta_0)'\widehat{\Omega}(\beta_0)^{-1} g_i(\beta_0) G_i}_{=:D_1} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} G_i}_{=:D_2}$$

It is easy to see that by LLN,

$$D_2 = -\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \xrightarrow{p} -E z_i x_i'$$

under the assumption

- $E\|z_i x_i'\| < \infty$ (usually assumed in the model)

Define a $K \times 1$ vector $h := \widehat{\Omega}(\beta_0)^{-1}\widehat{g}(\beta_0)$, then $h \xrightarrow{p} 0$ since $\widehat{g}(\beta_0) \xrightarrow{p} 0$ and $\widehat{\Omega}(\beta_0) = O_p(1)$. This implies $h_j \xrightarrow{p} 0$ for any $j = 1, \cdots, K$. So

$$D_1 = \frac{1}{n}\sum_{i=1}^{n} h' z_i \varepsilon_i (-z_i x_i') = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{K} h_j z_{ij}\right)\varepsilon_i(-z_i x_i') = -\sum_{j=1}^{K} h_j \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i z_{ij} z_i x_i' \xrightarrow{p} 0$$

since $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i z_{ij} z_i x_i' \xrightarrow{p} E\varepsilon_i z_{ij} z_i x_i'$ by LLN, under the assumption

- $E\|\varepsilon_i z_{ij} z_i x_i'\| < \infty$ for any $j = 1, \cdots, K$.

Therefore,

$$D(\beta_0) \xrightarrow{p} E z_i x_i' =: D_0$$

which is full rank by the model assumption. This completes the proof.

(ii) The results are given in the following pages. Two different choices for $var(z_i)$ are used. Apparently, the Wald test does not control the size at 0.05. When $\beta_0 = 0$ is used to generate data, it rejects more than 5% in many cases. This becomes more clear especially in the cases where instruments are weak and endogeneity is strong. On the contrary, the LM_{CUE} test keeps the size at (less than) 0.05. It is not a coincidence that we observe the same size of LM_{CUE} for different values of η and ρ when $K = 1$. This is because the same random numbers are repeatedly used to perform the test for fixed K. When $K = 1$, $g_i(\beta_0)$, $\widehat{g}(\beta_0)$, and $\widehat{\Omega}(\beta_0)$ are numerically the same for different values of η and ρ, and $D(\beta_0)$ is a scalar, so does not affect the LM_{CUE} statistic. When $\beta_0 \neq 0$ is used to generate data, the Wald test rejects H_0 more often than the LM_{CUE} does, but increasing power in exchange of high size is not desirable. The LM_{CUE} test has weak power in general compared to the Wald test. The exception is the case where $K = 1$ and $\eta = 1$. It is not clear whether the LM_{CUE} test has higher power with weak instruments than with strong instruments.

Results using L=diag(1:K)

[Wald test power function = Prob(reject the null | beta)]

(eta / rho)	K=1 (.05/.5)	K=1 (.05/.99)	K=1 (1/.5)	K=1 (1/.99)	K=10 (.05/.5)	K=10 (.05/.99)	K=10 (1/.5)	K=10 (1/.99)
beta = -0.8	0.0130	0.0350	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
beta = -0.6	0.0020	0.0920	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
beta = -0.4	0.0010	0.1840	0.9940	1.0000	0.9720	0.9960	1.0000	1.0000
beta = -0.2	0.0030	0.2750	0.5430	0.5390	0.3400	0.1580	1.0000	1.0000
beta = 0.0	0.0170	0.3300	0.0490	0.0550	0.0860	0.1960	0.0440	0.0440
beta = 0.2	0.0490	0.3770	0.5080	0.5040	0.7030	0.8280	1.0000	1.0000
beta = 0.4	0.1060	0.4150	0.9320	0.8960	0.9810	0.9790	1.0000	1.0000
beta = 0.6	0.1570	0.4490	0.9950	0.9830	1.0000	0.9980	1.0000	1.0000
beta = 0.8	0.2080	0.4850	0.9990	0.9980	1.0000	1.0000	1.0000	1.0000

[LMcue test power function = Prob(reject the null | beta)]

(eta / rho)	K=1 (.05/.5)	K=1 (.05/.99)	K=1 (1/.5)	K=1 (1/.99)	K=10 (.05/.5)	K=10 (.05/.99)	K=10 (1/.5)	K=10 (1/.99)
beta = -0.8	0.0680	0.3700	1.0000	1.0000	0.5370	0.2380	0.0950	0.0930
beta = -0.6	0.0590	0.1040	1.0000	1.0000	0.6870	0.6950	0.1000	0.0970
beta = -0.4	0.0580	0.0620	0.9960	1.0000	0.6850	0.8920	0.1160	0.1060
beta = -0.2	0.0510	0.0500	0.5510	0.6590	0.2550	0.3680	0.2100	0.1850
beta = 0.0	0.0540	0.0540	0.0540	0.0540	0.0270	0.0220	0.0220	0.0220
beta = 0.2	0.0550	0.0550	0.3860	0.3560	0.1950	0.1810	0.2510	0.2840
beta = 0.4	0.0560	0.0530	0.8630	0.7870	0.5160	0.4930	0.1370	0.1650
beta = 0.6	0.0560	0.0550	0.9790	0.9470	0.6710	0.6770	0.1030	0.1260
beta = 0.8	0.0590	0.0580	0.9960	0.9880	0.7150	0.7770	0.0950	0.1100

Results using L=eye(K)

[Wald test power function = Prob(reject the null | beta)]

(eta / rho)	K=1 (.05/.5)	K=1 (.05/.99)	K=1 (1/.5)	K=1 (1/.99)	K=10 (.05/.5)	K=10 (.05/.99)	K=10 (1/.5)	K=10 (1/.99)
beta = -0.8	0.0130	0.0350	1.0000	1.0000	0.3500	0.1520	1.0000	1.0000
beta = -0.6	0.0020	0.0920	1.0000	1.0000	0.0980	0.5490	1.0000	1.0000
beta = -0.4	0.0010	0.1840	0.9940	1.0000	0.0310	0.8650	1.0000	1.0000
beta = -0.2	0.0030	0.2750	0.5430	0.5390	0.1030	0.9630	1.0000	1.0000
beta = 0.0	0.0170	0.3300	0.0490	0.0550	0.3590	0.9920	0.0600	0.0780
beta = 0.2	0.0490	0.3770	0.5080	0.5040	0.6370	0.9960	1.0000	1.0000
beta = 0.4	0.1060	0.4150	0.9320	0.8960	0.7950	0.9980	1.0000	1.0000
beta = 0.6	0.1570	0.4490	0.9950	0.9830	0.8940	0.9980	1.0000	1.0000
beta = 0.8	0.2080	0.4850	0.9990	0.9980	0.9420	0.9980	1.0000	1.0000

[LMcue test power function = Prob(reject the null | beta)]

(eta / rho)	K=1 (.05/.5)	K=1 (.05/.99)	K=1 (1/.5)	K=1 (1/.99)	K=10 (.05/.5)	K=10 (.05/.99)	K=10 (1/.5)	K=10 (1/.99)
beta = -0.8	0.0680	0.3700	1.0000	1.0000	0.0560	0.7220	0.2080	0.1130
beta = -0.6	0.0590	0.1040	1.0000	1.0000	0.0480	0.2870	0.3630	0.2010
beta = -0.4	0.0580	0.0620	0.9960	1.0000	0.0430	0.0780	0.6380	0.5160
beta = -0.2	0.0510	0.0500	0.5510	0.6590	0.0490	0.0390	0.8870	0.8990
beta = 0.0	0.0540	0.0540	0.0540	0.0540	0.0350	0.0410	0.0350	0.0350
beta = 0.2	0.0550	0.0550	0.3860	0.3560	0.0330	0.0350	0.8590	0.8530
beta = 0.4	0.0560	0.0530	0.8630	0.7870	0.0430	0.0370	0.7950	0.8600
beta = 0.6	0.0560	0.0550	0.9790	0.9470	0.0420	0.0460	0.6530	0.7820
beta = 0.8	0.0590	0.0580	0.9960	0.9880	0.0380	0.0520	0.5470	0.7080

FIGURE1. Rejection Probabilities of Wald and LM_{CUE} when $var(z_{ij}) = j^2$, $j = 1, \cdots, K$

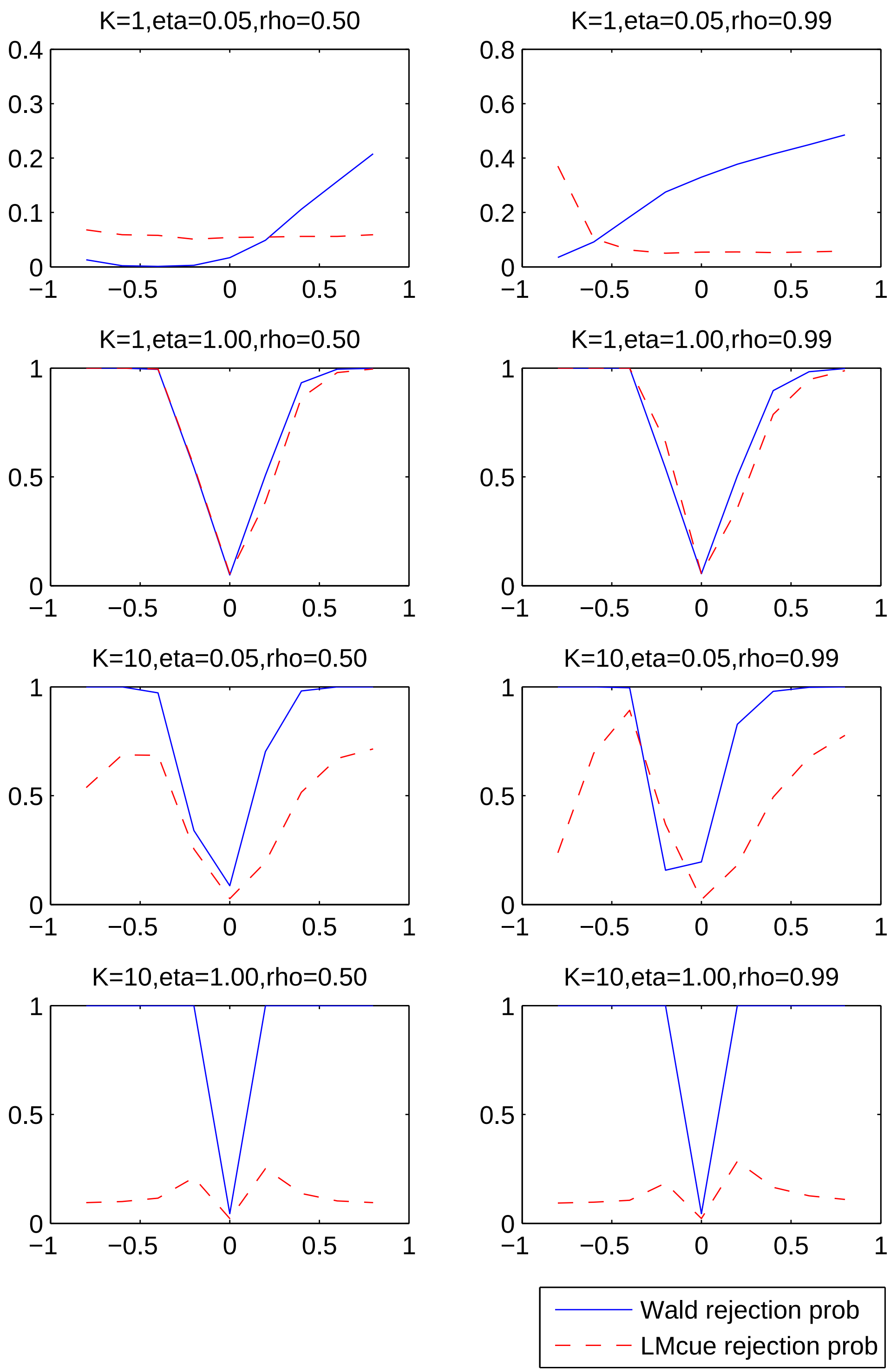
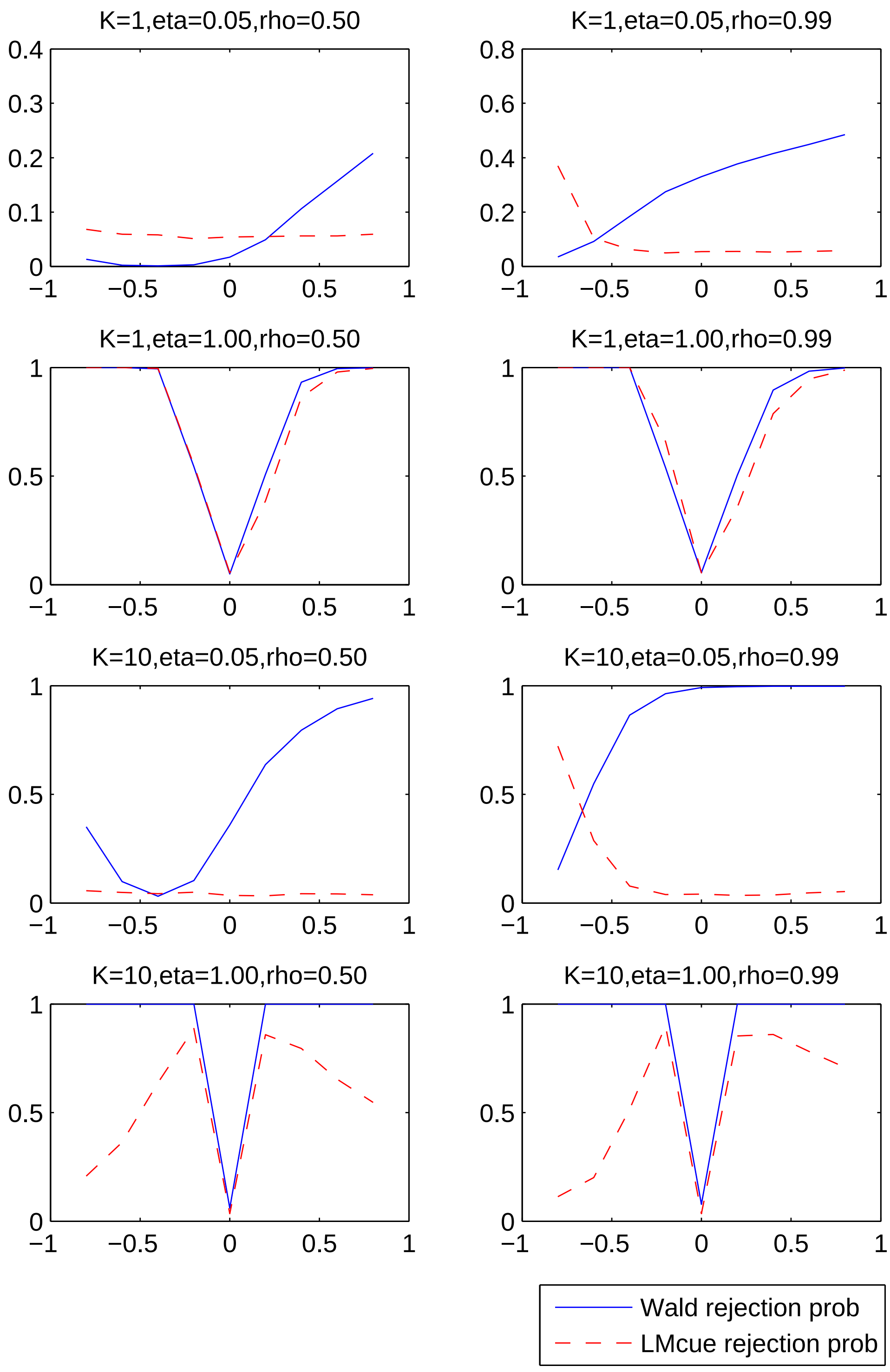


FIGURE2. Rejection Probabilities of Wald and LM_{CUE} when $var(z_{ij}) = 1,\ j = 1, \cdots, K$



4. Proof and Discussion

(i) Let $X_n = O_p(1)$ and $Y_n = o_p(1)$. Take any $\varepsilon > 0$. For any $\delta > 0$, there exist M_δ and N_δ such that for any $n > N_\delta$,

$$\Pr(\|X_n\| > M_\delta) < \delta$$

By definition, for such a M_δ,

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

Therefore,

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

Since δ was chosen arbitrarily, let $\delta \to 0$, then we are done.

(ii) Take any $\varepsilon > 0$. Find $M < \infty$ such that

$$\begin{aligned}
\Pr(|X| \geq M) &< \frac{\varepsilon}{2} \\
\Pr(X_n \leq M) &\longrightarrow \Pr(X \leq M) \\
\Pr(X_n \leq -M) &\longrightarrow \Pr(X \leq -M)
\end{aligned}$$

This should be possible since X is a random variable, and by definition of convergence in distribution, $\Pr(X_n \leq a) \to \Pr(X \leq a)$ for all continuity points of X. By definition of convergence, there exists N such that for any $n > N$,

$$\begin{aligned}
|\Pr(X_n \leq M) - \Pr(X \leq M)| &< \frac{\varepsilon}{4} \\
|\Pr(X_n \leq -M) - \Pr(X \leq -M)| &< \frac{\varepsilon}{4}
\end{aligned}$$

This implies

$$\begin{aligned}
\Pr(X_n \leq M) &< \Pr(X \leq M) + \frac{\varepsilon}{4} \\
-\Pr(X_n \leq -M) &< -\Pr(X \leq -M) + \frac{\varepsilon}{4}
\end{aligned}$$

Then,

$$\begin{aligned}
\Pr(|X_n| > M) &= \Pr(X_n < -M) + \Pr(X_n > M) \\
&\leq \Pr(X_n \leq -M) + 1 - \Pr(X_n \leq M) \\
&< \Pr(X \leq -M) + \frac{\varepsilon}{4} + 1 - \Pr(X \leq M) + \frac{\varepsilon}{4} \\
&\leq \Pr(X \leq -M) + 1 - \Pr(X < M) + \frac{\varepsilon}{2} \\
&= \Pr(|X| \geq M) + \frac{\varepsilon}{2} < \varepsilon
\end{aligned}$$

(iii) Suppose that H_1 is true. Then, $\widehat{\theta}_n \xrightarrow{p} \theta_1$ with $h(\theta_1) \neq 0$. Under appropriate assumptions, $\widehat{B}_n \xrightarrow{p} B_1$, $\widehat{\Omega}_n \xrightarrow{p} \Omega_1$, and $\widehat{H} \xrightarrow{p} \mathcal{H}_1$, where B_1, Ω_1 and $\mathcal{H}_1$ are finite, so

$$W_n = nh(\widehat{\theta}_n)' \left(\widehat{H} \widehat{B}_n^{-1} \widehat{\Omega}_n \widehat{B}_n^{-1} \widehat{H}' \right)^{-1} h(\widehat{\theta}_n) \xrightarrow{p} \infty$$

If we reject H_0 when $W_n > \chi^2_{r,1-\alpha}$, we would reject with probability 1 asymptotically, but if we reject H_0 when $W_n < \chi^2_{r,\alpha}$, we would reject with probability 0 asymptotically. Accordingly, when the sample size is large, we may reject with very low probability if we perform the second type of test. Given the nominal size fixed at α, this is not the best test we can do. The way we maximize power of the test (in other words, the way we reject false H_0 as often as possible) is to reject H_0 when $W_n > \chi^2_{r,1-\alpha}$.