Problem Set 2 Solution

April 22nd, 2009 by Yang

1. More Generalized Slutsky Theorem

[Simple and Abstract]

$$\begin{aligned}
\sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\beta_0)| &= \sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\widehat{\beta}_n)+L(\gamma,\widehat{\beta}_n)-L(\gamma,\beta_0)| \\
&\leq \sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\widehat{\beta}_n)|+\sup_{\gamma\in\Gamma}|L(\gamma,\widehat{\beta}_n)-L(\gamma,\beta_0)| \qquad (1)\\
&\leq \sup_{\gamma\in\Gamma}\sup_{\beta\in B(\beta_0,\varepsilon)}|L_n(\gamma,\beta)-L(\gamma,\beta)|+\sup_{\gamma\in\Gamma}|L(\gamma,\widehat{\beta}_n)-L(\gamma,\beta_0)| \\
&= o_p(1)+o_p(1)=o_p(1)
\end{aligned}$$

where the third line can be obtained with probability approaching 1, and the fourth line follows from the assumptions (i), (ii) and (iii).

[Formal and Detailed]

Take any $\xi>0$. Note that by (1)

$$\begin{aligned}
&\lim_{n\to\infty}\Pr\left(\sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\beta_0)|>\xi\right) \\
&\quad\leq \lim_{n\to\infty}\Pr\left(\sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\widehat{\beta}_n)|+\sup_{\gamma\in\Gamma}|L(\gamma,\widehat{\beta}_n)-L(\gamma,\beta_0)|>\xi\right) \\
&\quad\leq \underbrace{\lim_{n\to\infty}\Pr\left(\sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\widehat{\beta}_n)|>\frac{\xi}{2}\right)}_{=:A_1}+\underbrace{\lim_{n\to\infty}\Pr\left(\sup_{\gamma\in\Gamma}|L(\gamma,\widehat{\beta}_n)-L(\gamma,\beta_0)|>\frac{\xi}{2}\right)}_{=:A_2}
\end{aligned}$$

Partitioning the event in $A1$ yields

$$\begin{aligned}
A_1 &= \lim_{n\to\infty}\Pr\left(\sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\widehat{\beta}_n)|>\frac{\xi}{2}\text{ and }\|\widehat{\beta}_n-\beta_0\|>\varepsilon\right) \\
&\quad+\lim_{n\to\infty}\Pr\left(\sup_{\gamma\in\Gamma}|L_n(\gamma,\widehat{\beta}_n)-L(\gamma,\widehat{\beta}_n)|>\frac{\xi}{2}\text{ and }\|\widehat{\beta}_n-\beta_0\|\leq\varepsilon\right) \\
&\leq \lim_{n\to\infty}\Pr(\|\widehat{\beta}_n-\beta_0\|>\varepsilon)+\lim_{n\to\infty}\Pr\left(\sup_{\gamma\in\Gamma}\sup_{\beta\in B(\beta_0,\varepsilon)}|L_n(\gamma,\beta)-L(\gamma,\beta)|>\frac{\xi}{2}\right)=0
\end{aligned}$$

where the last equality follows from the assumptions (i) and (ii). Also by the assumption (iii), there exists $\delta>0$ such that for any $\beta\in B(\beta_0,\delta)$

$$\sup_{\gamma\in\Gamma}|L(\gamma,\beta)-L(\gamma,\beta_0)|\leq\frac{\xi}{2}$$

This implies, by contrapositive,

$$A_2 \leq \lim_{n\to\infty} \Pr(\|\widehat{\beta}_n - \beta_0\| > \delta) = 0$$

where the last equality follows from the assumption (i). Therefore, we have

$$\lim_{n\to\infty} \Pr\left(\sup_{\gamma\in\Gamma} |L_n(\gamma, \widehat{\beta}_n) - L(\gamma, \beta_0)| > \xi\right) = 0$$

for any $\xi > 0$, which completes the proof.

2. Eicker-White Robust Covariance Estimation

(a) By $\sqrt{T}(\widehat{\beta}_{LS} - \beta) \xrightarrow{d} Z \sim N(0, \Omega)$, we have $\widehat{\beta}_{LS} \xrightarrow{p} \beta$.[1] This implies, by continuous mapping theorem,

$$\widehat{U}_t = Y_t - X_t'\widehat{\beta}_{LS} \xrightarrow{p} Y_t - X_t'\beta = U_t$$

This implies in turn, by continuous mapping theorem,

$$E\widehat{U}_t^2 X_t X_t' \xrightarrow{p} EU_t^2 X_t X_t'$$

Note that the LHS term is a random variable, while the RHS term is a scalar. Applying LLN,

$$\frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t' \xrightarrow{p} E\widehat{U}_t^2 X_t X_t'$$

under the following assumptions.[2]

- $EU_t^2 < \infty$[3]
- $E|U_t^2 X_{ti} X_{tj}| < \infty$ for all $i, j = 1, \cdots, K$ where $K = dim(X_t)$.
- $E|X_{ti}^2 X_{tj} X_{tk}| < \infty$ for all $i, j, k = 1, \cdots, K$.

Combining these two,

$$\frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t' \xrightarrow{p} EU_t^2 X_t X_t'$$

Also without any further assumption, by applying LLN,

$$\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \xrightarrow{p} EX_t X_t'$$

[1] To see this, note that $\widehat{\beta}_{LS} - \beta = O_p(T^{-1/2})$ as we will prove in problem set 3. So $\widehat{\beta}_{LS} - \beta = o_p(1)$.

[2] We may not apply iid version of LLN directly to $\frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t'$. This is because $\{\widehat{U}_t : 1 \leq t \leq T\}$ are not independent. Instead, we apply uniform LLN to $\frac{1}{T}\sum_{t=1}^{T} (Y_t - X_t'\theta)^2 X_t X_t'$ for $\theta \in B(\beta, \varepsilon)$. Since $\widehat{\beta}_{LS}$ is consistent for β, we can get the desired result under the described assumptions. See White (*Econometrica*, 1980) for details.

[3] Note this is not implied by $E(U_t^2|X_t) < \infty$.

Therefore, by multilication rule,

$$\widehat{\Omega} := \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t' \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \xrightarrow{p} (EX_tX_t')^{-1} EU_t^2 X_t X_t' (EX_tX_t')^{-1} = \Omega$$

under the additional assumption

- EX_tX_t' is nonsingular

(b) Suppose the assumption of conditional homoskedasticity holds so that $E(U_t^2|X_t) = \sigma^2$. Then the variance of $\widehat{\beta}_{LS}$ simplifies as follows.

$$\Omega = (EX_tX_t')^{-1} E\sigma^2 X_t X_t' (EX_tX_t')^{-1} = \sigma^2 (EX_tX_t')^{-1}$$

So if we believe this assumption, we may use the following formula to estimate a covariance matrix.

$$\widehat{\Omega}_{CH} = \widehat{\sigma}^2 \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}$$

where CH denotes "conditional homoskedasticity," and $\widehat{\sigma}^2$ can be estimated by

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \widehat{U}_T^2 = \frac{1}{T}\sum_{t=1}^{T} (Y_t - X_t'\widehat{\beta}_{LS})^2$$

If we are more dubious of the assumption, we can use the Eicker-White robust covariance estimator.

$$\widehat{\Omega}_{EW} = \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t' \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}$$

Let us talk a little on how to program this. Denote by X, the matrix which stacks X_t' by row. Define Y, U and $\widehat{U}$ in the same way. Then, $\sum_{t=1}^{T} X_t X_t' = X'X$, and $\sum_{t=1}^{T} \widehat{U}_T^2 = \widehat{U}'\widehat{U}$. Therfore, to estimate $\widehat{\Omega}_{CH}$, we can use `OmegaCH = (Uhat' * Uhat / T) * inv(X' * X / T)`, or simply drop both T's. However, $\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t'$ cannot be simplified using matrix notation.[4] The most straightforward way to calculate it is using a loop. But there is a simpler way to do this in MATLAB, using element-wise multiplication. Define $W_t := \widehat{U}_t X_t$, and denote by W, the matrix which stacks W_t' by row. Then,

$$W'W = \sum_{t=1}^{T} W_t W_t' = \sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t'$$

Note that W is the matrix obtained by multiplying $\widehat{U}_t$ to each element in t'th row of X. This is equivalent to element-wise multiplication of $\widehat{U}L'$ and X, where L' is a $1 \times k$ row vector of 1's. So by doing `W = (Uhat * ones(1,k)) .* X`, we get such a matrix. Therefore $\widehat{\Omega}_{EW}$ can be calculated by `OmegaEW = inv(X' * X / T) * W' * W / T * inv(X' * X / T)`.

[4] Note that $\widehat{U}'\widehat{U}X'X = (\sum_{t=1}^{T} \widehat{U}_t^2)(\sum_{t=1}^{T} X_t X_t')$ is totally different from $\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t'$.

See the MATLAB code posted for details. It would be good to run the code with different choices of R, T, k, and most importantly, σ^2. The following shows the result with two sets of parameters with $R = 1000$, $T = 100$. On the left, $k = 1$ and $\sigma^2 = 9$, and on the right, $k = 3$ and $\sigma^2 = 25$.[5]

```
  True  beta      1.0000                                          1.0000    2.0000    3.0000
------------------------                                       ------------------------------
              (estimates of beta)
     Mean         0.9967                                          0.9833    2.0103    2.9852
 Standard Dev     0.2930                                          0.5068    0.5054    0.5046
------------------------                                       ------------------------------
  True Omega      9.0000                                         25.0000    0.0000    0.0000
                                                                  0.0000   25.0000    0.0000
                                                                  0.0000    0.0000   25.0000
------------------------                                       ------------------------------
  Mean  Bias  (estimates of Omega1: correct formula)
                  0.0964                                          0.3992   -0.0154   -0.0291
                                                                 -0.0154    0.6921    0.0654
                                                                 -0.0291    0.0654    0.5944
              (estimates of Omega2: White formula)
                 -0.0527                                         -0.0192    0.0564   -0.1876
                                                                  0.0564    0.3143    0.0031
                                                                 -0.1876    0.0031    0.4074
------------------------                                       ------------------------------
 Standard Dev (estimates of Omega1: correct formula)
                  1.8753                                          5.1120    2.6270    2.5793
                                                                  2.6270    5.3957    2.6308
                                                                  2.5793    2.6308    5.3192
              (estimates of Omega2: White formula)
                  2.5961                                          6.8917    4.3187    4.2576
                                                                  4.3187    7.1830    4.3945
                                                                  4.2576    4.3945    7.2081
------------------------                                       ------------------------------
 R M S Error  (estimates of Omega1: correct formula)
                  1.8769                                          5.1250    2.6257    2.5781
                                                                  2.6257    5.4372    2.6303
                                                                  2.5781    2.6303    5.3497
              (estimates of Omega2: White formula)
                  2.5953                                          6.8883    4.3169    4.2596
                                                                  4.3169    7.1863    4.3923
                                                                  4.2596    4.3923    7.2160
------------------------                                       ------------------------------

  Norm Stats      Omega1    Omega2                                          Omega1    Omega2
----------------------------------                                       --------------------
     Mean         1.4679    1.9923                                         10.3988   15.1711
 Standard Dev     1.1702    1.6641                                          4.1532    5.7600
   Minimum        0.0002    0.0094                                          2.7274    4.2677
   Median         1.2196    1.6877                                          9.6479   14.2907
   Maximum        7.2610   15.2920                                         33.2757   48.9044
```

[5] You may try another set of assumptions on β, EX_tX_t', the distribution of X_t, and that of U_t. But when you change the distributions of X_t and U_t, make sure that the correct variances of X_t and U_t are used in the code.

To summarize the result, the standard deviation of each element of Eicker-White estimator is 30-40% bigger than that of the estimator using the correct formula. So is the root mean squared error. But looking at the mean bias, Eicker-White estimator is less biased (by mean) in many cases, especially on diagonal elements. These properties hold for many choices of R, T, k, and σ^2 that have ever been tried. In the table, "Norm stats" are also reported. The norm is the distance between true Ω and estimates $\widehat{\Omega}$, calcualted by taking square root of the summed squares of difference in each element. The distributions of 4 sets of 1000 norms are summarized as seen. This implies that the Eicker-White estimator is bigger in general, but not that bad. Moreover it is robust to heteroskedasticity of errors, so it would be better to use this formula if we are not sure of conditional homoskedasticity.

3. Asymptotic Theory

(a) Additionally we need to assume

1. [CF i] $\theta_0 \in int(\Theta)$
2. [EE2 ii] $\frac{\partial}{\partial\theta} Q_n(\widehat{\theta}_n) = o_p(n^{-1/2})$
3. $g(\theta) \in C^2(\Theta_0)$ for some neighborhood $\Theta_0 \subset \Theta$ of θ_0 (with probability one)
4. $\frac{\partial}{\partial\theta'} g(\theta_0)$ has full column rank.

Note that we already assumed

5. $\sqrt{n}(\widehat{\pi}_n - \pi_0) \xrightarrow{d} N(0, V_0)$
6. $A_n \xrightarrow{p} A$ where A is nonsingular.
7. There exists a unique value $\theta_0 \in \Theta$ such that $\pi_0 = g(\theta_0)$.
8. Θ is compact.
9. [EE] $\widehat{\theta}_n \in \Theta$ and $Q_n(\widehat{\theta}_n) \le \inf_{\theta\in\Theta} Q_n(\theta) + o_p(1)$

- g is continuous.
- $\widehat{\pi}_n \xrightarrow{p} \pi_0$

Since the last two are implied by 3 and 5 repectively, we need assumptions 1-9. Now let us check how the above assumptions imply CF and EE2. EE2 is implied by 2-3 and 5-9. CF i is assumed in 1. CF ii is implied by 3. CF iii is implied by 3, 5 and 6. To see what Ω_0 is, consider

$$\sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0) = -\frac{\partial}{\partial\theta} g'(\theta_0) A_n' A_n \sqrt{n}(\widehat{\pi}_n - g(\theta_0)) = -\frac{\partial}{\partial\theta} g'(\theta_0) A_n' A_n \sqrt{n}(\widehat{\pi}_n - \pi_0)$$

Define $\Gamma_0 := \frac{\partial}{\partial \theta'} g(\theta_0)$. Then by 5 and 6,

$$\sqrt{n}\frac{\partial}{\partial \theta}Q_n(\theta_0) \xrightarrow{d} N(0, \Gamma_0' A' A V_0 A' A \Gamma_0)$$

Let us check that CF iv is implied by 3-7. The candidate for $B(\theta)$ can be obatined by 3, 5 and 6.

$$[B(\theta)]_{mj} = \text{plim}\frac{\partial^2}{\partial \theta_m \partial \theta_j}Q_n(\theta) = \frac{\partial}{\partial \theta_m}g'(\theta)A'A\frac{\partial}{\partial \theta_j}g(\theta) - \frac{\partial^2}{\partial \theta_m \partial \theta_j}g'(\theta)A'A(\pi_0 - g(\theta))$$

By 3, this is continuous in θ. Moreover, $\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta)$ is also continuous in θ, and by choice, Θ_0 is compact. So uniform convergence follows from pointwise convergence.

$$\sup_{\theta \in \Theta_0}\left\| \frac{\partial^2}{\partial \theta \partial \theta'}Q_n(\theta) - B(\theta)\right\| \xrightarrow{p} 0$$

Finally, $B(\theta_0)$ is nonsingular by 4 and 6. When evaluated at θ_0, the second summand of $[B(\theta_0)]_{mj}$ is 0, so

$$B_0 := B(\theta_0) = \frac{\partial}{\partial \theta}g'(\theta_0)A'A\frac{\partial}{\partial \theta'}g(\theta_0) = \Gamma_0' A' A \Gamma_0$$

So the asymptotic normality follows.

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N\Big(0, (\Gamma_0'A'A\Gamma_0)^{-1}\Gamma_0'A'AV_0A'A\Gamma_0(\Gamma_0'A'A\Gamma_0)^{-1}\Big)$$

(b) We need to assume the following.

1. [EE] $\widehat{\theta}_n \in \Theta$ and $Q_n(\widehat{\theta}_n) \leq \inf_{\theta \in \Theta} Q_n(\theta) + o_p(1)$
2. [ID1 i] Θ is compact.
3. $G(\theta, \tau)$ is continuous in θ at $\tau = \tau_0$.
4. There exists a unique value $\theta_0 \in \Theta$ such that $G(\theta_0, \tau_0) = 0$.
5. $A_n \xrightarrow{p} A$ where A is nonsingular.
6. $\widehat{\tau}_n \xrightarrow{p} \tau_0$
7. $\sup_{\theta \in \Theta}\sup_{\tau \in B(\tau_0, \varepsilon)}|G_n(\theta, \tau) - G(\theta, \tau)| \xrightarrow{p} 0$ for some $\varepsilon > 0$
8. $G(\theta, \tau)$ is continuous in τ at τ_0 uniformly over $\theta \in \Theta$.

Let us first determine what $Q(\theta)$ is. By 5-7, we have pointwise convergence such that for any $\theta \in \Theta$,

$$Q_n(\theta) \xrightarrow{p} \|AG(\theta, \tau_0)\|^2/2 =: Q(\theta)$$

For this $Q(\theta)$, ID holds by 2-5. In particular, 4 and 5 imply ID iii. To see this, note first that $Q(\theta_0) = 0$ and that this is the minimum since $Q(\theta) \geq 0$ for any θ. Suppose that θ satisfies $Q(\theta) = 0$. Then, by the property of norm operator, $AG(\theta,\tau_0)=0$. By 5, $G(\theta,\tau_0)=0$, and thus by 4, $\theta=\theta_0$, which proves ID iii. Now let us turn to UWCON. 2-3 and 5-8 imply UWCON. First, applying Question 1, 6-8 imply

$$\sup_{\theta\in\Theta}\|G_n(\theta,\widehat{\tau}_n)-G(\theta,\tau_0)\| \xrightarrow{p} 0$$

Using 2, 3, 5, 6 and this,

$$\begin{aligned}\sup_{\theta\in\Theta}|Q_n(\theta)-Q(\theta)| &\le \sup_{\theta\in\Theta}|Q_n(\theta)-G'(\theta,\tau_0)A_n'A_nG(\theta,\tau_0)/2|+\sup_{\theta\in\Theta}|G'(\theta,\tau_0)A_n'A_nG(\theta,\tau_0)/2-Q(\theta)|\\ &\le 2\cdot\frac{1}{2}\sup_{\theta\in\Theta}|G'(\theta,\tau_0)|A_n'A_n\sup_{\theta\in\Theta}|G_n(\theta,\widehat{\tau}_n)-G(\theta,\tau_0)| \qquad \text{(wp approaching 1)}\\ &\qquad+\frac{1}{2}\sup_{\theta\in\Theta}|G'(\theta,\tau_0)||A_n'A_n-A'A|\sup_{\theta\in\Theta}|G(\theta,\tau_0)|\\ &= O_p(1)O_p(1)o_p(1)+O_p(1)o_p(1)O_p(1)=o_p(1)\end{aligned}$$

4. Monte Carlo Simulation: Goodness of Approximation to the Assymptotic Distribution

(i) As we derived before, under conditional homoskedasticity, $\widehat{\beta}_{2SLS}$ has the asymptotic variance of $\sigma^2\left(Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i'\right)^{-1}$. Since $\sigma^2=1$, $Ez_iz_i'=I_K$ and $Ez_iu_i=0$,

$$Ex_iz_i'=E(z_i'\pi z_i'+u_iz_i')=\pi'(Ez_iz_i')+Eu_iz_i'=\pi'$$

and thus we can calculate the asymptotic variance as follows.

$$\sigma^2\left(Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i'\right)^{-1}=(\pi'I_K\pi)^{-1}=\frac{1}{K\eta^2}$$

It depends on K and η but not on ρ.

(ii) Looking at the first figure, we can see many huge outliers of estimates in the case where $K=1$ and $\eta=0.05$. This is consistent with the fact that the 2SLS estimator has no finite first moment when it is exactly identified. Even when $\eta=1$, we can observe some outliers although they are not very big. The 2SLS estimator tends to be biased when $\rho=0.95$ and the instruments are weak. K and η affect the variance of estimates in general.

(iii) In the second figure, the histogram of estimates was normalized so that the area of histogram bars is 1, and the pdf of the distribution approximated by the asymptotic distribution is plotted together.[6] To focus on the shape of the distribution, the outliers are dropped from this figure. When $\eta=1$, the asymptotic distribution approximates the simulated distribution quite well, while it does not if $\eta=0.05$. When $K=1$ and $\rho=0$, there does not seem to be a bias, but there are outliers and estimates which are not outliers are centered more than the approximated distribution. In the thrid figure, the numbers drawn from the approximated distribution are plotted, along with the histogram of estimates without rescaling. It looks similar to the second figure.

[6] Since the asymptotic distribution is obtained by scaling $\widehat{\beta}_{2SLS}$ up by $\sqrt{n}$, we always need to scale the asymptotic variance down by n to get the approximated variance.

FIGURE1. Histogram of Estimates

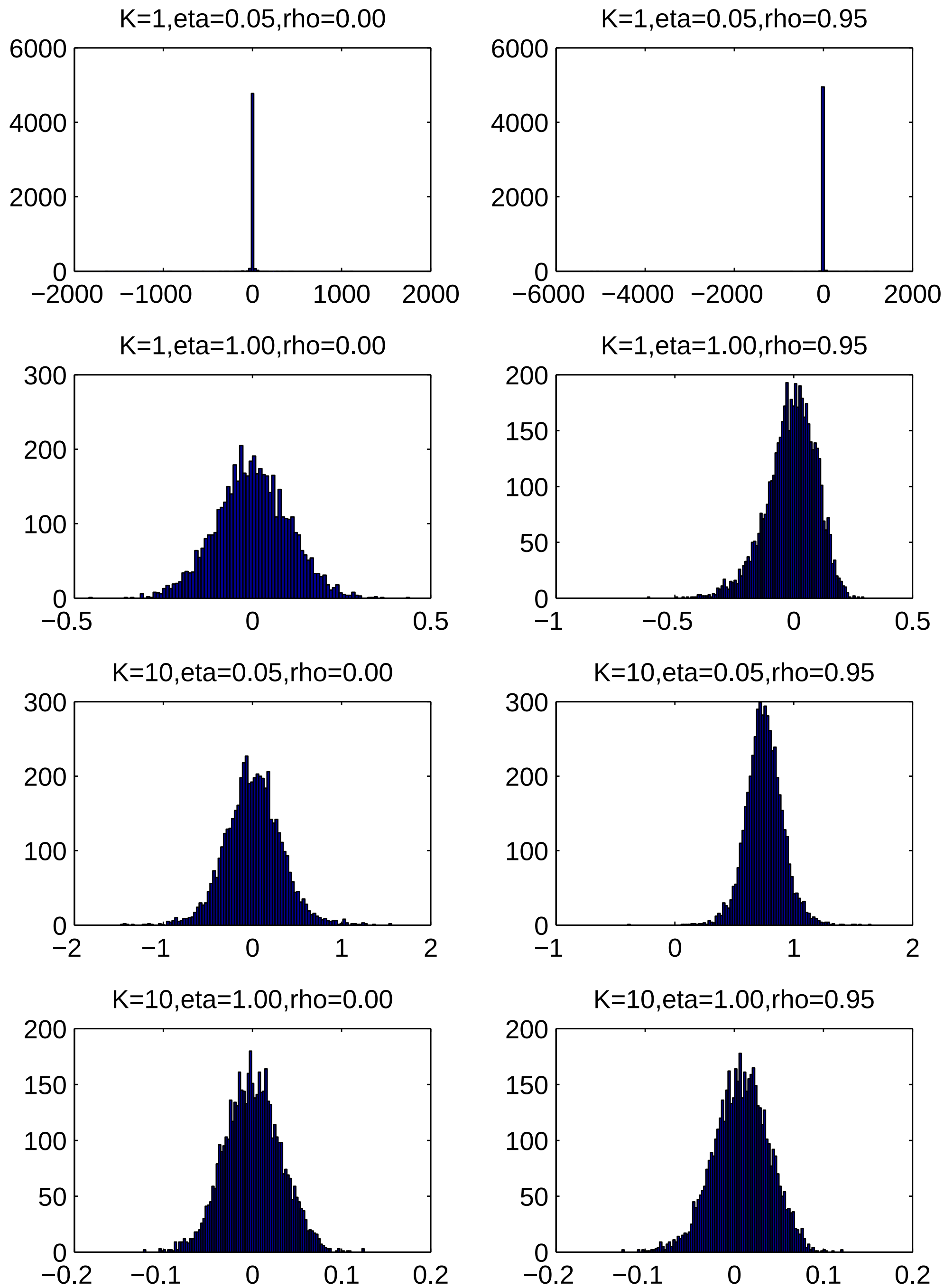

FIGURE2. Histogram and Approximated Distribution

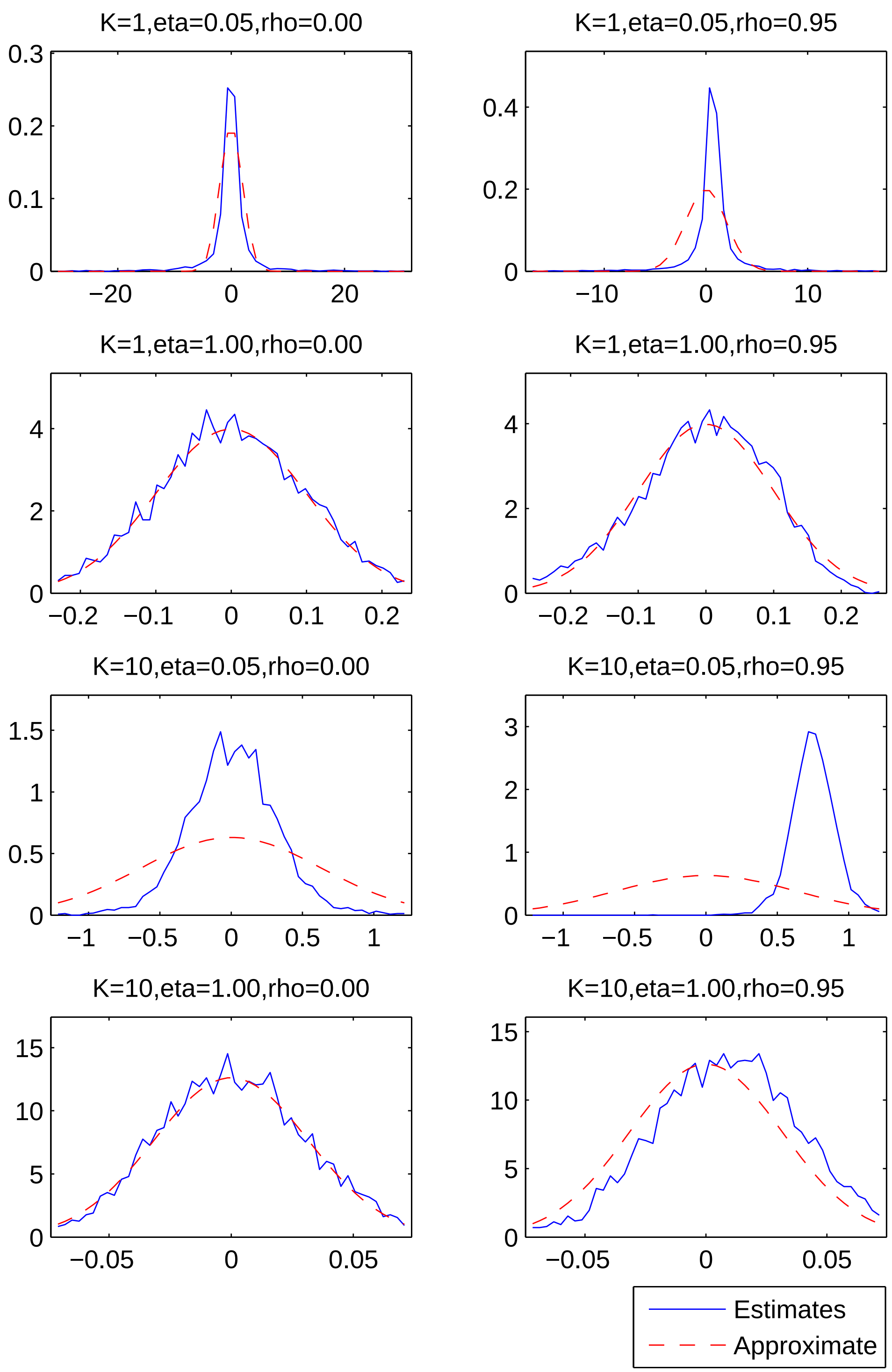

FIGURE3. Histogram and Approximated Distribution (simulated)