Problem Set 4 Solution

May 7th, 2009 by Yang

1. Asymptotic Distribution of J-statistic

(i) There are lots of matrices C that satisfy $CC' = \Omega^{-1}$. For example, there are lower triangular matrices and upper triangular matrices. There are also many symmetric matrices C such that $CC' = \Omega^{-1}$. It can be obtained by decomposing Ω^{-1} into a form of $Q\Lambda Q'$, where Λ is a diagonal matrix with eigenvalues, and Q is a matrix of corresponding orthonormal eigenvectors. Then $C = Q\Lambda^{1/2}Q'$. Note that all of such matrices are invertible.[1] So C satisfies $(C')^{-1}C^{-1} = \Omega$.

(ii)

$$n\left[C'\overline{g}_n(\widehat{\beta})\right]'(C'\widehat{\Omega}^{-1}C)^{-1}\left[C'\overline{g}_n(\widehat{\beta})\right] = n\overline{g}_n(\widehat{\beta})'CC^{-1}\widehat{\Omega}^{-1}(C')^{-1}C'\overline{g}_n(\widehat{\beta}) = J_n$$

(iii) Note that

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

and also

$$\overline{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)z_i = \frac{1}{n}Z'(Y - X\beta)$$

Therefore,

$$\begin{aligned}
D_nC'\overline{g}_n(\beta_0) &= C'\overline{g}_n(\beta_0) - C'\left(\frac{1}{n}Z'X\right)\left(\frac{1}{n}X'Z\widehat{\Omega}^{-1}\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{n}X'Z\widehat{\Omega}^{-1}\right)\overline{g}_n(\beta_0)\\
&= C'\overline{g}_n(\beta_0) - C'\left(\frac{1}{n}Z'X\right)\left(\frac{1}{n}X'Z\widehat{\Omega}^{-1}\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{n}X'Z\widehat{\Omega}^{-1}\frac{1}{n}Z'Y\right)\\
&\qquad + C'\left(\frac{1}{n}Z'X\right)\left(\frac{1}{n}X'Z\widehat{\Omega}^{-1}\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{n}X'Z\widehat{\Omega}^{-1}\frac{1}{n}Z'X\right)\beta_0\\
&= C'\overline{g}_n(\beta_0) + C'\left(\frac{1}{n}Z'X\right)(\beta_0 - \widehat{\beta})\\
&= C'\cdot\frac{1}{n}Z'(Y - X\widehat{\beta}) = C'\overline{g}_n(\widehat{\beta})
\end{aligned}$$

(iv) Since $\widehat{\Omega}^{-1} \xrightarrow{p} \Omega^{-1} = CC'$ and $\frac{1}{n}Z'X \xrightarrow{p} Ez_ix_i'$,

$$\begin{aligned}
D_n &\xrightarrow{p} I_k - C'Ez_ix_i'(Ex_iz_i'\Omega^{-1}Ez_ix_i')^{-1}Ex_iz_i'\Omega^{-1}(C')^{-1}\\
&= I_k - C'Ez_ix_i'(Ex_iz_i'CC'Ez_ix_i')^{-1}Ex_iz_i'C\\
&= I_k - R(R'R)^{-1}R'
\end{aligned}$$

[1] Its proof is simple. Suppose not. Then, C has a rank less than its dimension. By the property of rank, $CC' = \Omega^{-1}$ has a rank less than its dimension. This contradicts to invertible Ω^{-1}.

(v) Note first that

$$n^{1/2}\overline{g}_n(\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i z_i \xrightarrow{d} N(0, \underbrace{E z_i z_i' U_i^2}_{=\Omega})$$

Therefore, by Slutsky theorem,

$$n^{1/2}C'\overline{g}_n(\beta_0) \xrightarrow{d} N \sim N(0, C'\Omega C) \stackrel{d}{=} N(0, I_k)$$

(vi) Apply Slutsky theorem again, then it follows from (iv) and (v) that

$$n^{1/2}D_n C'\overline{g}_n(\beta_0) \xrightarrow{d} (I_k - R(R'R)^{-1}R')N$$

Since $C'\widehat{\Omega}C \xrightarrow{p} I_k$, (iii) implies

$$H_n := n^{1/2}(C'\widehat{\Omega}C)^{-1/2}C'\overline{g}_n(\widehat{\beta}) = n^{1/2}(C'\widehat{\Omega}C)^{-1/2}D_n C'\overline{g}_n(\beta_0) \xrightarrow{d} (I_k - R(R'R)^{-1}R')N$$

Therefore, (ii) implies

$$J_n = H_n' H_n \xrightarrow{d} N'(I_k - R(R'R)^{-1}R')N$$

since $(I_k - R(R'R)^{-1}R')$ is a projection matrix.

(vii) Note that R is $k \times d$ matrix.

$$tr(I_k - R(R'R)^{-1}R') = tr(I_k) - tr(R(R'R)^{-1}R') = k - tr((R'R)^{-1}R'R) = k - d$$

Since $(I_k - R(R'R)^{-1}R')$ is symmetric and idempotent, we have

$$N'(I_k - R(R'R)^{-1}R')N \sim \chi^2_{k-d}$$

2. Finite Sample Rejection Probability of J-test

(i) The J-statistic is calculated using the following formula, even though the conditional homoskedasticity assumption holds in the model, which can be used to simplify the formula further.

$$J_n = n \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_n)z_i' \left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_n)^2 z_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\widehat{\beta}_n)$$

For more details in calculating the J-statistic, see TA note 5. The results are provided in the table on the next page. In the given model, the size of the J-test seems to stay at the level of 5%. High η is expected to detect the failure of the moment condition more accurately, which is true in case 2, but turns out to be false in case 1. This is due to the weak power property of the J-test against local failures of the moment condition. In other words, when some specific type of the alternative hypothesis regarding the moment condition is true, the J-statistic tends to have a similar asymptotic property as in the case where the null hypothesis holds. Case 1 is such an example. See TA note 5 for more details.

When $n = 100$, K does not seem to affect the results in a very significant way. We can observe some cases where bigger K decreases the probability to reject the false null hypothesis. It is partly because in case 2, the moment condition fails to hold only in 1 out of 10 equations, which does not lead the test to reject against the other powerful true moment conditions. This happens less when $n = 1000$. It seems that more instruments help reject the false null hypothesis asymptotically.

TABLE1. The Null Rejection Probability of the Pretest (J-test) in the First Stage

Case: n=100

	CASE1				CASE2			
(K / eta)	(2/.05)	(2/1)	(10/.05)	(10/1)	(2/.05)	(2/1)	(10/.05)	(10/1)
delta = -0.30	0.2920	0.0400	0.6880	0.0430	0.2600	0.5140	0.2850	0.3050
delta = -0.25	0.2710	0.0420	0.6820	0.0420	0.2240	0.3820	0.1950	0.2010
delta = -0.20	0.2340	0.0440	0.6220	0.0410	0.1660	0.2670	0.1110	0.1340
delta = -0.15	0.1720	0.0450	0.5070	0.0420	0.1020	0.1480	0.0630	0.0800
delta = -0.10	0.1000	0.0400	0.2720	0.0390	0.0570	0.0830	0.0420	0.0630
delta = -0.05	0.0420	0.0400	0.0880	0.0390	0.0330	0.0520	0.0280	0.0430
delta = -0.03	0.0370	0.0410	0.0440	0.0390	0.0310	0.0460	0.0280	0.0420
delta = -0.02	0.0310	0.0400	0.0390	0.0390	0.0280	0.0450	0.0280	0.0390
delta = -0.01	0.0280	0.0400	0.0300	0.0390	0.0250	0.0410	0.0270	0.0390
delta = 0.00	0.0260	0.0400	0.0280	0.0380	0.0260	0.0400	0.0280	0.0380
delta = 0.01	0.0270	0.0420	0.0230	0.0380	0.0290	0.0480	0.0260	0.0380
delta = 0.02	0.0270	0.0430	0.0190	0.0380	0.0290	0.0460	0.0270	0.0400
delta = 0.03	0.0310	0.0430	0.0180	0.0370	0.0270	0.0480	0.0250	0.0400
delta = 0.05	0.0350	0.0430	0.0300	0.0390	0.0340	0.0540	0.0230	0.0410
delta = 0.10	0.0690	0.0430	0.1480	0.0360	0.0580	0.0940	0.0330	0.0550
delta = 0.15	0.1240	0.0400	0.3880	0.0340	0.0910	0.1690	0.0470	0.0730
delta = 0.20	0.1900	0.0380	0.6000	0.0350	0.1430	0.2990	0.0780	0.1220
delta = 0.25	0.2480	0.0400	0.6870	0.0350	0.1980	0.4360	0.1380	0.1870
delta = 0.30	0.2840	0.0460	0.6910	0.0290	0.2350	0.6000	0.2240	0.3050

Case: n=1000

	CASE1				CASE2			
(K / eta)	(2/.05)	(2/1)	(10/.05)	(10/1)	(2/.05)	(2/1)	(10/.05)	(10/1)
delta = -0.30	0.1280	0.0560	0.2390	0.0400	0.5460	1.0000	0.9950	1.0000
delta = -0.25	0.1270	0.0540	0.2550	0.0430	0.5330	1.0000	0.9780	1.0000
delta = -0.20	0.1240	0.0540	0.2480	0.0440	0.5020	0.9880	0.9450	0.9930
delta = -0.15	0.1190	0.0550	0.2390	0.0410	0.4520	0.9070	0.8170	0.8940
delta = -0.10	0.1170	0.0560	0.2090	0.0420	0.3480	0.5690	0.4730	0.5010
delta = -0.05	0.0940	0.0590	0.1390	0.0430	0.1600	0.1920	0.1280	0.1250
delta = -0.03	0.0780	0.0590	0.0980	0.0440	0.0960	0.1060	0.0740	0.0660
delta = -0.02	0.0670	0.0590	0.0740	0.0440	0.0760	0.0800	0.0530	0.0520
delta = -0.01	0.0520	0.0560	0.0540	0.0420	0.0610	0.0610	0.0460	0.0430
delta = 0.00	0.0540	0.0590	0.0430	0.0430	0.0540	0.0590	0.0430	0.0430
delta = 0.01	0.0510	0.0600	0.0440	0.0440	0.0590	0.0620	0.0420	0.0450
delta = 0.02	0.0480	0.0600	0.0460	0.0440	0.0680	0.0780	0.0540	0.0560
delta = 0.03	0.0520	0.0600	0.0610	0.0440	0.0910	0.1100	0.0710	0.0750
delta = 0.05	0.0710	0.0580	0.0970	0.0440	0.1760	0.2250	0.1320	0.1350
delta = 0.10	0.1230	0.0580	0.1870	0.0450	0.4180	0.6440	0.4970	0.5210
delta = 0.15	0.1290	0.0590	0.2310	0.0490	0.5730	0.9280	0.8960	0.9040
delta = 0.20	0.1290	0.0600	0.2510	0.0490	0.6160	0.9950	0.9890	0.9970
delta = 0.25	0.1270	0.0610	0.2570	0.0480	0.6300	1.0000	0.9970	1.0000
delta = 0.30	0.1280	0.0600	0.2390	0.0520	0.6270	1.0000	0.9990	1.0000

(ii) Note that the true β is 0 in any of the cases, so the Wald test should not reject the null hypothesis more than 5%. But we can find that for any K, η, and specification of c, the size is not controlled at the level of 5%. We can choose δ that makes the size arbitrarily close to 1. This shows an example that the size of the two stage test may be distorted, even when instruments are strong.

TABLE2. The Conditional Null Rejection Probability of the Wald Test in the Second Stage

n=100

	CASE1				CASE2			
(K / eta)	(2/.05)	(2/1)	(10/.05)	(10/1)	(2/.05)	(2/1)	(10/.05)	(10/1)
delta = -0.30	0.0537	0.9938	0.9199	1.0000	0.0216	0.5761	0.1524	0.1640
delta = -0.25	0.0302	0.9551	0.8365	1.0000	0.0168	0.4256	0.1267	0.1164
delta = -0.20	0.0157	0.8013	0.6508	1.0000	0.0108	0.2892	0.1215	0.0878
delta = -0.15	0.0097	0.5351	0.2901	0.9969	0.0056	0.1843	0.1142	0.0750
delta = -0.10	0.0089	0.2792	0.0646	0.8835	0.0106	0.1047	0.1127	0.0683
delta = -0.05	0.0073	0.1010	0.0526	0.3163	0.0103	0.0643	0.1255	0.0543
delta = -0.03	0.0093	0.0667	0.0722	0.1509	0.0093	0.0556	0.1317	0.0574
delta = -0.02	0.0083	0.0573	0.0884	0.0905	0.0093	0.0503	0.1337	0.0572
delta = -0.01	0.0082	0.0500	0.1196	0.0656	0.0103	0.0480	0.1387	0.0541
delta = 0.00	0.0103	0.0500	0.1368	0.0541	0.0103	0.0500	0.1368	0.0541
delta = 0.01	0.0113	0.0532	0.1679	0.0665	0.0124	0.0504	0.1427	0.0541
delta = 0.02	0.0134	0.0648	0.2110	0.1237	0.0134	0.0524	0.1470	0.0542
delta = 0.03	0.0134	0.0805	0.2607	0.2004	0.0123	0.0557	0.1477	0.0552
delta = 0.05	0.0238	0.1338	0.3639	0.4256	0.0145	0.0708	0.1546	0.0574
delta = 0.10	0.0548	0.3292	0.6397	0.9191	0.0255	0.1313	0.1830	0.0688
delta = 0.15	0.1005	0.6146	0.8529	0.9990	0.0440	0.2178	0.2120	0.0917
delta = 0.20	0.1543	0.8493	0.9475	1.0000	0.0665	0.3153	0.2495	0.1230
delta = 0.25	0.1875	0.9563	0.9744	1.0000	0.0985	0.4663	0.2842	0.1550
delta = 0.30	0.2067	0.9937	0.9741	1.0000	0.1229	0.6325	0.3247	0.1957

n=1000

	CASE1				CASE2			
(K / eta)	(2/.05)	(2/1)	(10/.05)	(10/1)	(2/.05)	(2/1)	(10/.05)	(10/1)
delta = -0.30	0.6583	1.0000	1.0000	1.0000	0.4824	NaN	1.0000	NaN
delta = -0.25	0.6415	1.0000	1.0000	1.0000	0.4604	NaN	1.0000	NaN
delta = -0.20	0.6199	1.0000	1.0000	1.0000	0.4357	1.0000	0.7636	0.8571
delta = -0.15	0.5834	1.0000	1.0000	1.0000	0.2974	0.9140	0.3060	0.4151
delta = -0.10	0.4009	0.9905	1.0000	1.0000	0.0767	0.5360	0.0626	0.1623
delta = -0.05	0.0475	0.5728	0.9454	0.9990	0.0048	0.1807	0.0493	0.0560
delta = -0.03	0.0087	0.2540	0.4191	0.8243	0.0100	0.1018	0.0605	0.0460
delta = -0.02	0.0075	0.1413	0.1307	0.4906	0.0173	0.0728	0.0686	0.0443
delta = -0.01	0.0148	0.0742	0.0264	0.1461	0.0234	0.0650	0.0839	0.0449
delta = 0.00	0.0317	0.0574	0.0972	0.0460	0.0317	0.0574	0.0972	0.0460
delta = 0.01	0.0685	0.0840	0.3013	0.1904	0.0457	0.0618	0.1054	0.0534
delta = 0.02	0.1229	0.1596	0.6279	0.5460	0.0676	0.0857	0.1173	0.0646
delta = 0.03	0.2089	0.2894	0.8935	0.8912	0.1001	0.1079	0.1356	0.0724
delta = 0.05	0.4306	0.6497	0.9978	0.9990	0.1881	0.2206	0.1740	0.0890
delta = 0.10	0.7275	0.9915	1.0000	1.0000	0.4811	0.6320	0.3121	0.1921
delta = 0.15	0.7440	1.0000	1.0000	1.0000	0.6019	0.8750	0.5192	0.3021
delta = 0.20	0.7428	1.0000	1.0000	1.0000	0.5911	1.0000	0.8182	1.0000
delta = 0.25	0.7434	1.0000	1.0000	1.0000	0.5676	NaN	1.0000	NaN
delta = 0.30	0.7420	1.0000	1.0000	1.0000	0.5657	NaN	1.0000	NaN

FIGURE1. The Size and Power of the Two Tests ($n = 100$)

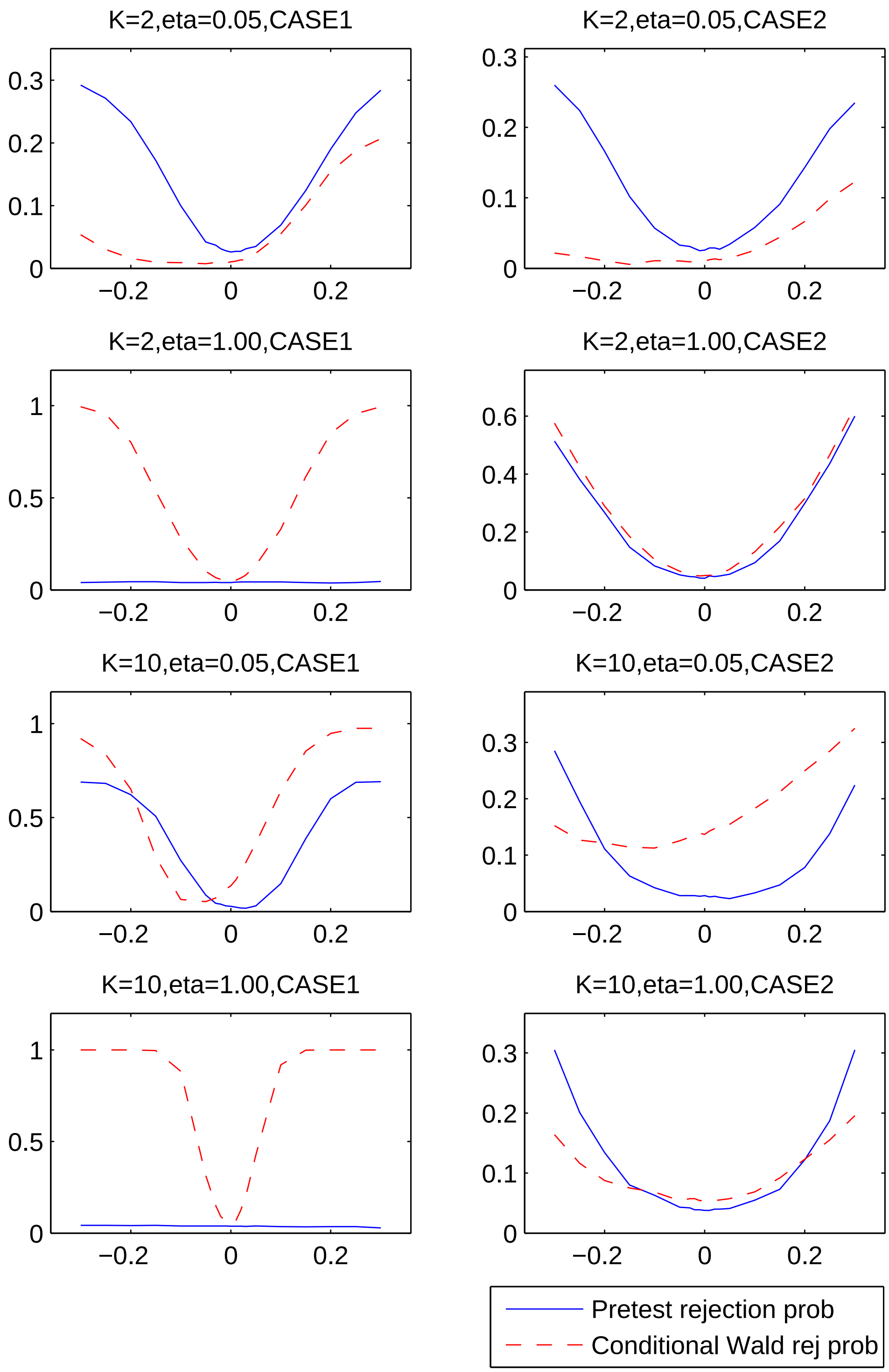
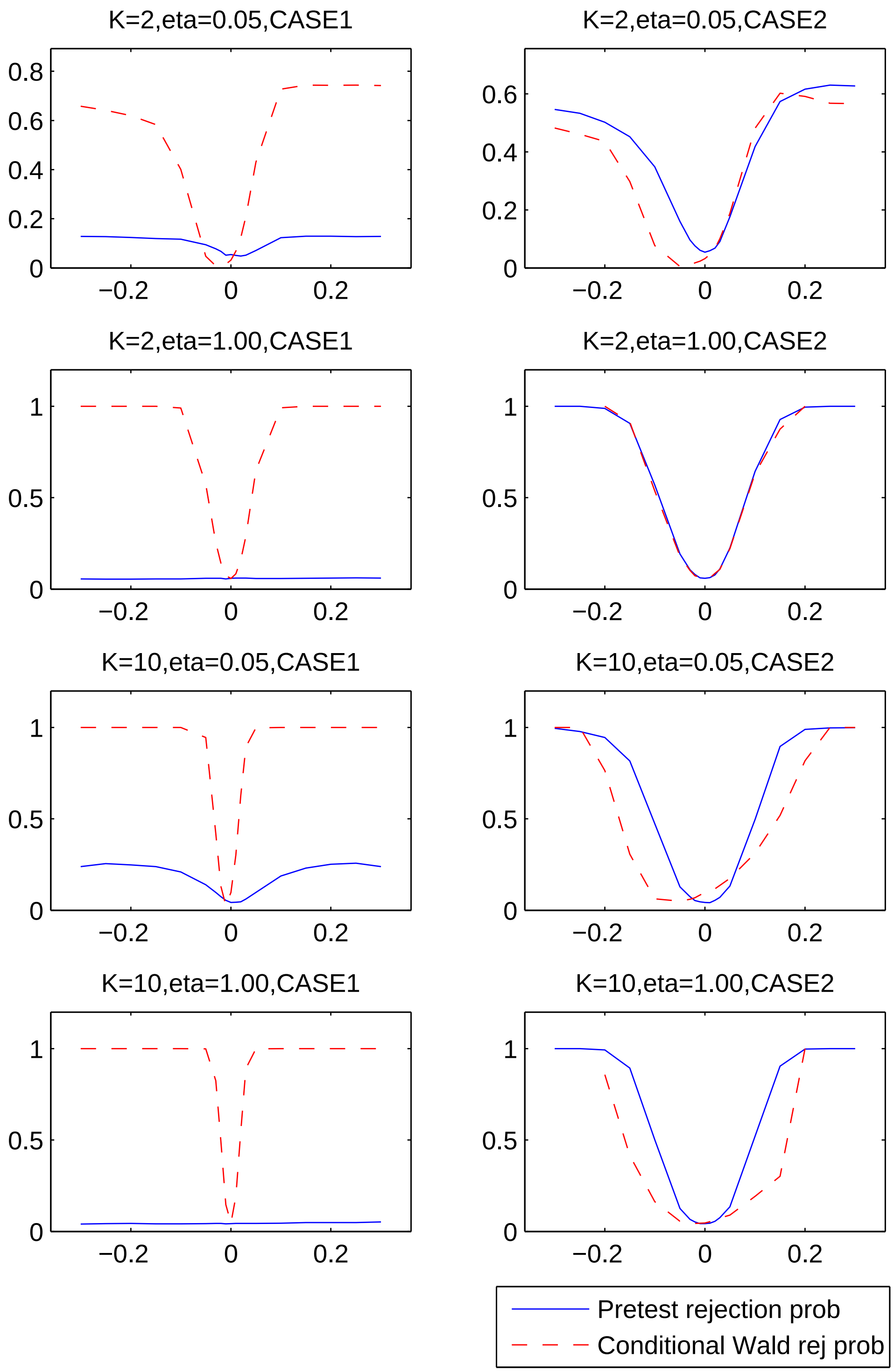


FIGURE2. The Size and Power of the Two Tests ($n = 1000$)



3. Empirical Anaylsis Using Linear IV Model

(i) See the following table.

(ii) There might be some unobserved ability that explains wage well and is correlated with education. *near4* seems to be correlated with education. We may argue that people who live near a 4 year college pay less cost to get higher education. On the other hand, whether people live near a college does not seem to be correlated with their ability. But one may insist that people are affected by the environment and thus get smarter when they live close to a college. See the table for the result of the exact IV estimation.

(iii) The result is closer to that of the OLS estimation, rather than to that of the exact IV estimation. We would argue that parents' education reflects their ability, and genes let their offsprings have similar ability. This explains why parents' education might be correlated with unobserved ability, and thus invalid as instruments.

	OLS	exact IV	2SLS	GMM
beta0	4.7351	2.2659	4.0867	4.0944
(SE)	(0.0796)	(1.0240)	(0.2190)	(0.2230)
(robust)	(0.0837)	(0.9999)	(0.2235)	(0.2230)
beta1	0.0802	0.2226	0.1176	0.1170
(SE)	(0.0041)	(0.0590)	(0.0125)	(0.0128)
(robust)	(0.0043)	(0.0577)	(0.0128)	(0.0128)
beta2	0.0893	0.1507	0.1054	0.1054
(SE)	(0.0081)	(0.0272)	(0.0096)	(0.0098)
(robust)	(0.0081)	(0.0270)	(0.0098)	(0.0098)
beta3	-0.0025	-0.0027	-0.0025	-0.0025
(SE)	(0.0004)	(0.0005)	(0.0004)	(0.0004)
(robust)	(0.0004)	(0.0006)	(0.0004)	(0.0004)
beta4	-0.1359	-0.0856	-0.1227	-0.1234
(SE)	(0.0178)	(0.0302)	(0.0185)	(0.0185)
(robust)	(0.0177)	(0.0294)	(0.0185)	(0.0185)
beta5	-0.1591	-0.0359	-0.1267	-0.1288
(SE)	(0.0238)	(0.0588)	(0.0262)	(0.0261)
(robust)	(0.0238)	(0.0562)	(0.0263)	(0.0261)

Specification	J-statistic	Null rejection
beta_2SLS, first V_hat	16.9332	REJECTED
beta_GMM, first V_hat	16.8708	REJECTED
beta_GMM, second V_hat	16.8884	REJECTED

* critical value = chi2 (quantile .95, df 3) : 7.8147

(iv) The result is provided in the same table, together with the other results. Not surprisingly, it is similar to that of 2SLS estimation using the same moment conditions. This is because the 2SLS estimator is consistent even though the conditional homoskedasticity assumption does not hold. If we iterate on the procedure to get the efficient GMM estimator, we would be able to get the (infeasible) efficient GMM estimator eventually. But iterating only once gives the (feasible) efficient GMM estimator which is asymptotically equivalent to the infeasible one. This can be supported by the fact that the standard errors are numerically very similar, whether they are calculated by approximation to the asymptotic variance, or by robust estimation formula of the variance.

(v) The J-statistic is provided in the above table, using three other methods. The first is evaluated at $\widehat{\beta}_{2SLS}$ and $\widehat{V}_1 := \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_{2SLS})^2 z_i z_i'$. The second is evaluted at $\widehat{\beta}_{EFF}$ and $\widehat{V}_1$. The last is obtained using $\widehat{\beta}_{EFF}$ and $\widehat{V}_2 := \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta}_{EFF})^2 z_i z_i'$. The values are numerically different, but by a small amount, and have the same asymptotic distribution. Whatever method we use, we reject the null hypothesis that all the moment conditions hold at equality. This reinforces the argument made in (iii).

(vi) It seems that the exact IV estimation (ii) gives the most likely result. But we should be careful because the exact IV estimation does not have even a finite first moment. This means that the estimator may suffer from the huge bias, especially when instruments are weak. *Exper* seems exogenous, since it captures age effect, and thus is not likely to be correlated with hidden factors affecting wage.

4. OLS is BLUE

Note that the OLS estimator $\widehat{\beta}_{OLS} = (X'X)^{-1}X'y$. This is linear in y, and also unbiased since

$$E\widehat{\beta}_{OLS} = E\left[(X'X)^{-1}X'(X\beta + \varepsilon)\right] = \beta + (X'X)^{-1}X'E\varepsilon = 0$$

Let Cy be any linear unbised estimator of β. Since

$$ECy = E\left[C(X\beta + \varepsilon)\right] = CX\beta + CE\varepsilon = CX\beta$$

$CX = I_d$ should be satisfied for Cy to be unbiased, where $d = \text{dim}\beta$. Now compare the variances of the two estimators. Note that

$$var(\widehat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$$
$$var(Cy) = var(C\varepsilon) = \sigma^2 CC'$$

Hence $var(Cy) \geq var(\widehat{\beta}_{OLS})$ is equivalent $CC' \geq (X'X)^{-1}$. Use $CX = I_d$ to show

$$CC' - (X'X)^{-1} = CC' - CX(X'X)^{-1}X'C' = C\left[I_n - X(X'X)^{-1}X'\right]C' \geq 0$$

This holds since $I_n - X(X'X)^{-1}X'$ is a projection matrix, and thus it can be written as a quadratic form. Cy is arbitrarily chosen, so the result follows.