A Simple Robust Model Selection Test for Large Models

Zhipeng Liao (UCLA)

Xiaoxia Shi (UW - Madison)

February 19, 2014

Model Selection Test

- Economists often have multiple models for one economic phenomenon.
  - Myopic Markov transition model versus lifetime welfare maximization model for schooling choice. (Cameron and Heckman, 1998, *JPE*).
  - Rule-bound group utility maximization model versus individual utility maximization model for voter turnout. (Coate and Conlin, 2004, *AER*).
  - Belief heterogeneity model versus preference-based models for explaining Long-Shot Bias in asset pricing. (Gandhi and Serrano-Padial, 2012).
  - Regressor is exogenous or endogenous in a nonparametric IV model.
- One economic theory may generate multiple empirical models:
  - Function form specification and variable selection in semi/nonparametric models;

Model Selection Test

- Model selection test is a natural way to statistically compare the fit of these models.
- With two models, $\mathcal{M}_1$ and $\mathcal{M}_2$, the null hypothesis of the model selection test is:

$$H_0 : F(\mathcal{M}_1, \mu) = F(\mathcal{M}_2, \mu)$$

  where $F(\cdot, \cdot)$ is a measure of the fit of the model to the true data distribution μ.
- Examples of the measure of the fit include:
  - likelihood ratio (LR): $F(\mathcal{M}_j, \mu) = E_\mu \left[\log f_j(Z, \theta_j^* / f_0(Z) \right]$;
  - regression SSE: $F(\mathcal{M}_j, \mu) = E_\mu \left[- \left| Y - g_j(Z, \theta_j^*) \right|^2 \right]$;
  - LR or SSE with AIC or BIC penalization.

Model Selection Test for Large Model

- We consider the general form

$$F(\mathcal{M}_j, \mu) = \max_{\theta_j \in \Theta_j} E_\mu \left[m_j(Z, \theta_j) \right]$$

  for a known function $m_j(\cdot, \cdot)$ and a set of parameters Θ_j with $j = 1, 2$.
- Large Model: we assume θ_j is "large/high dimensional" in the sense that:
  - θ_j contains unknown function or;
  - θ_j's dimension increases with sample size n.
- The first case includes all semi/non-parametric likelihood or regression models.
- The second case approximates parametric models with many parameters (Cameron-Heckman: > 70, Coate-Conlin:> 40)
- We only need one of the two models to be "large".

Literature

- Model selection test for "small models": Vuong (1989), Kitamura (2000), Rivers and Vuong (2002), Marmer and Otsu (2012), Schennach and Wilhelm (2013), Shi (2013).
- Model selection test for "large models": Lavergne and Vuong (1996), Chen, Hong and Shum (2007).
- The theory of these papers are based on first-order asymptotic approximation (except Shi (2013)).
- The first-order asymptotic approximation is not good in most large model applications. Tests derived therefrom can be severely size-distorted. (Shi, 2013).

The Contribution of This Paper

- We develop asymptotic theory for the natural statistics that takes into account the second order terms.
- The new asymptotic approximation motivates a new robust test that has correct asymptotic size.
- It has correct asymptotic size no matter the compared models are nested or non nested, strictly or weakly non nested, far or close to each other.
- The models we deal with include those in Lavergne and Vuong (1996), and Chen, Hong and Shum (2007) as special cases.
- Comparing to the robust test for small models in Shi (2013), the new robust test for large models is simpler to use.

Nonparametric Tools

- For the j-th model, θ_j is estimated via sieve M estimation:

$$\widehat{\theta}_{j,n} = \arg \max_{\theta_j \in \Theta_{j,n}} \frac{1}{n} \sum_{i=1}^{n} m_j(Z_i, \theta_j)$$

  where $\Theta_{j,n}$ is either a sieve space that becomes dense in Θ_j as $n \to \infty$, or a finite dimensional parameter space with increasing dimension.
- Some properties of the sieve M estimator are known: e.g., consistency in White and Wooldridge (1991), rate of convergence in Shen and Wong (1994), root-n normality in Shen (1997), Chen and Shen (1998), and pointwise normality in Chen, Liao and Sun (2012).
- This paper contributes to the sieve M estimation literature by providing the asymptotic distribution of a quadratic form of the score function.

Quasi Likelihood Ratio Statistic

- Recall the null hypothesis:

$$H_0 : \max_{\theta_1 \in \Theta_1} E_\mu \left[m_1(Z_i, \theta_1) \right] = \max_{\theta_2 \in \Theta_2} E_\mu \left[m_2(Z_i, \theta_2) \right].$$

- θ_j^*: pseudo true value, i.e $\theta_j^* = \arg\max_{\theta_j \in \Theta_j} E_\mu \left[m_j(Z_i, \theta_j) \right]$. Then

$$H_0 : E_\mu \left[m_1(Z_i, \theta_1^*) \right] = E_\mu \left[m_2(Z_i, \theta_2^*) \right].$$

- Consider the natural quasi-likelihood ratio statistic:

$$\widehat{QLR}_n = \frac{1}{n} \sum_{i=1}^{n} \left[m_1(Z_i, \widehat{\theta}_{1,n}) - m_2(Z_i, \widehat{\theta}_{2,n}) \right].$$

- Need to derive the asymptotic distribution of $\widehat{QLR}_n$ to conduct inference on the null hypothesis H_0.

QLR – First Order Asymptotic Distribution

- We can decompose the $\widehat{QLR}_n$ as

$$
\begin{aligned}
\widehat{QLR}_n &= n^{-1} \sum_{i=1}^{n} \left[m_1(Z_i, \widehat{\theta}_{1,n}) - m_2(Z_i, \widehat{\theta}_{2,n}) \right] \\
&= n^{-1} \sum_{i=1}^{n} \left[m_1(Z_i, \theta_1^*) - m_2(Z_i, \theta_2^*) \right] \\
&\quad + v_n \left[m_1(Z, \widehat{\theta}_{1,n}) - m_2(Z, \widehat{\theta}_{2,n}) + m_2(Z, \theta_2^*) - m_1(Z, \theta_1^*) \right] \\
&\quad - E_\mu \left[m_1(Z, \widehat{\theta}_{1,n}) - m_2(Z, \widehat{\theta}_{2,n}) - m_1(Z, \theta_1^*) + m_2(Z, \theta_2^*) \right]
\end{aligned}
$$

where $v_n \left[f(Z_i, \theta) \right] = \frac{1}{n} \sum_{i=1}^{n} \left[f(Z_i, \theta) - E \left[f(Z_i, \theta) \right] \right]$.

QLR – First Order Asymptotic Distribution

- By the stochastic equicontinuity, we have

$$\begin{aligned} & v_n\left[m_1(Z,\widehat{\theta}_{1,n}) - m_2(Z,\widehat{\theta}_{2,n}) + m_2(Z,\theta_2^*) - m_1(Z,\theta_1^*)\right] \\ = \; & v_n\left[m_1(Z,\widehat{\theta}_{1,n}) - m_1(Z,\theta_1^*)\right] - v_n\left[m_2(Z,\widehat{\theta}_{2,n}) - m_2(Z,\theta_2^*)\right] \\ = \; & o_p(n^{-\frac{1}{2}}). \end{aligned}$$

- Under some smoothness conditions,

$$\begin{aligned} & \left|E_\mu\left[m_1(Z,\widehat{\theta}_{1,n}) - m_2(Z,\widehat{\theta}_{2,n}) - m_1(Z,\theta_1^*) + m_2(Z,\theta_2^*)\right]\right| \\ & \leq \left|E_\mu\left[m_1(Z,\widehat{\theta}_{1,n}) - m_1(Z,\theta_1^*)\right]\right| + \left|E_\mu\left[m_2(Z,\widehat{\theta}_{2,n}) - m_2(Z,\theta_2^*)\right]\right| \\ & \leq const. \times \left[\left\|\widehat{\theta}_{1,n} - \theta_1^*\right\|_S^2 + \left\|\widehat{\theta}_{2,n} - \theta_2^*\right\|_S^2\right] = o_p(n^{-\frac{1}{2}}) \end{aligned}$$

where $\|\cdot\|_S$ denotes some metric defined on Θ_j.

QLR – First Order Asymptotic Distribution

- Under H_0, easy to derive the first-order asymptotic dist'n

$$
\begin{aligned}
n^{\frac{1}{2}}\widehat{QLR}_n &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\left[m_1(Z_i,\theta_1^*) - m_2(Z_i,\theta_2^*)\right] + o_p(1) \\
&\to_d \;\; N(0,\omega_A^2), \text{ where} \\
\omega_A^2 &= \lim_{n\to\infty} \mathsf{Var}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n}\left[m_1(Z_i,\theta_1^*) - m_2(Z_i,\theta_2^*)\right]\right].
\end{aligned}
$$

- First-order asymptotic test:

$$
\varphi_n^{F.O.}(\alpha) = 1\left\{\frac{\left|n^{\frac{1}{2}}\widehat{QLR}_n\right|}{\widehat{\omega}_A} > z_{\alpha/2}\right\}
$$

where $\widehat{\omega}_A^2 = \widehat{\mathsf{Var}}_n\left[m_1(Z,\widehat{\theta}_{1,n}) - m_2(Z,\widehat{\theta}_{2,n})\right]$.

- The first order asymptotic theory is fragile in large models.

MC Illustration

- Suppose that the data are from

$$Y = X_1\theta_1 + X_2\theta_2 + \theta_3\left(X_{3,1} + \ldots + X_{3,K}\right) + u$$

  where X_1, X_2, $X_{3,j}$ and u are independent standard normal;
- Consider the following two models

$$\begin{aligned} \mathcal{M}_1 &: \quad Y_1 = X_1\beta_1 + X_2\beta_2 + u_1; \\ \mathcal{M}_2 &: \quad Y_2 = X_1\beta_1 + X_{3,1}\beta_{3,1} + \ldots + X_{3,K}\beta_{3,K} + u_2; \end{aligned}$$

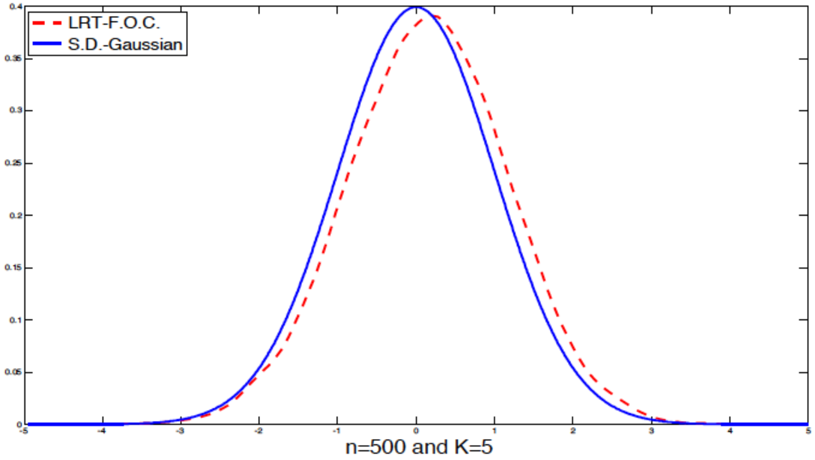
- We set $\theta_1 = 0.5$ and $\theta_3 = \theta_2 K^{-\frac{1}{2}}$ such that the null: H_0 :

$$E\left[\left(Y_1 - X_1\beta_1 - X_2\beta_2\right)^2\right] = E\left[\left(Y_1 - X_1\beta_1 - \sum_{k=1}^{K} X_{3,k}\beta_{3,k}\right)^2\right]$$

  holds for any K and any θ_2.
- We investigate the finite sample densities of the first order LR statistic and our robust LR statistic with 10,000 replications.

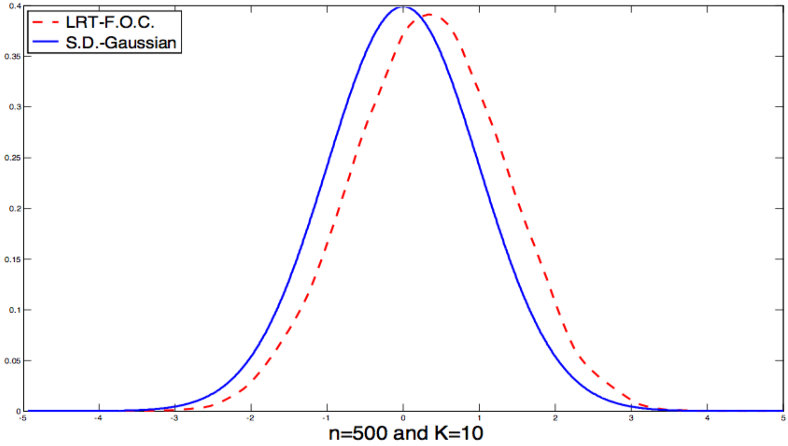
MC Illustration

Figure 1. Density of the LR-test Statistic ($\theta_2 = 0.5$)



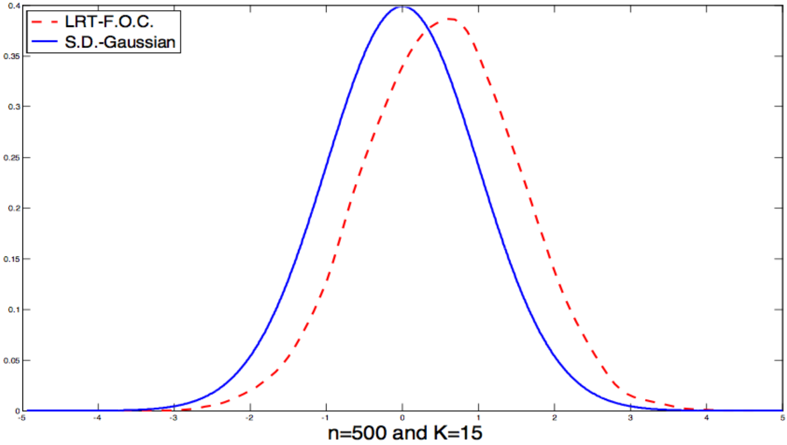
MC Illustration

Figure 2. Density of the LR-test Statistic ($\theta_2 = 0.5$)



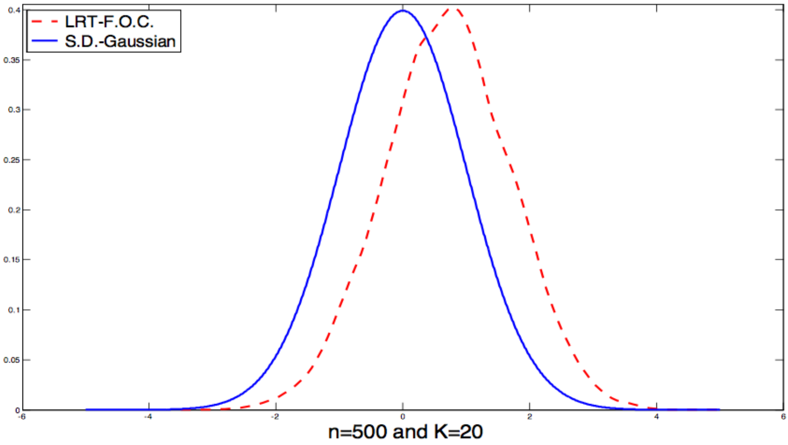
MC Illustration

Figure 3. Density of the LR-test Statistic ($\theta_2 = 0.5$)



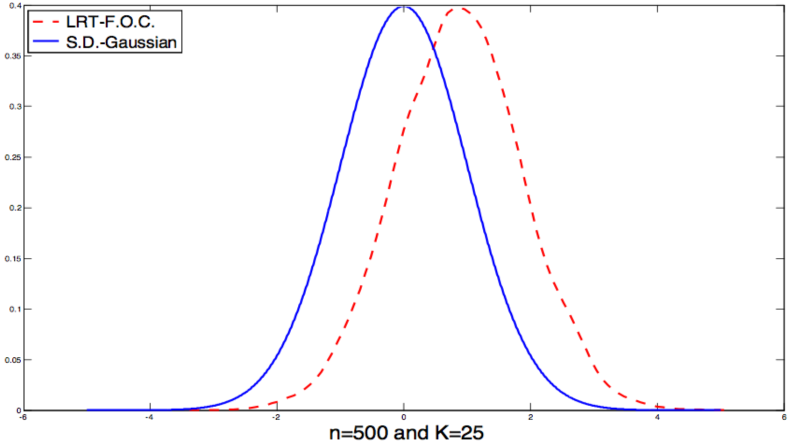
MC Illustration

Figure 4. Density of the LR-test Statistic ($\theta_2 = 0.5$)



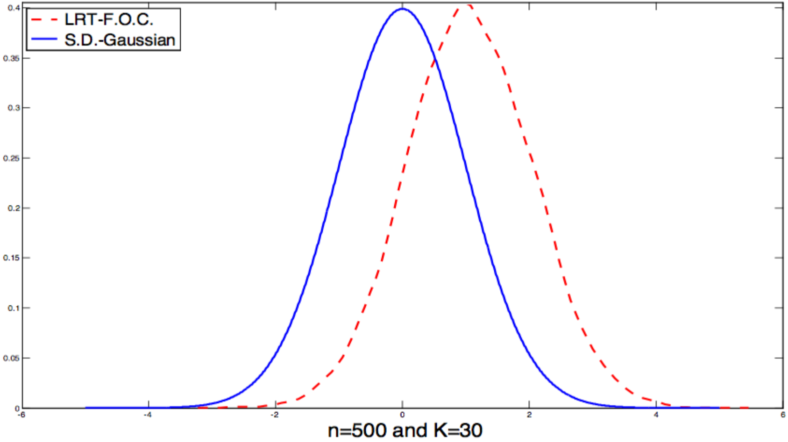
MC Illustration

Figure 5. Density of the LR-test Statistic ($\theta_2 = 0.5$)



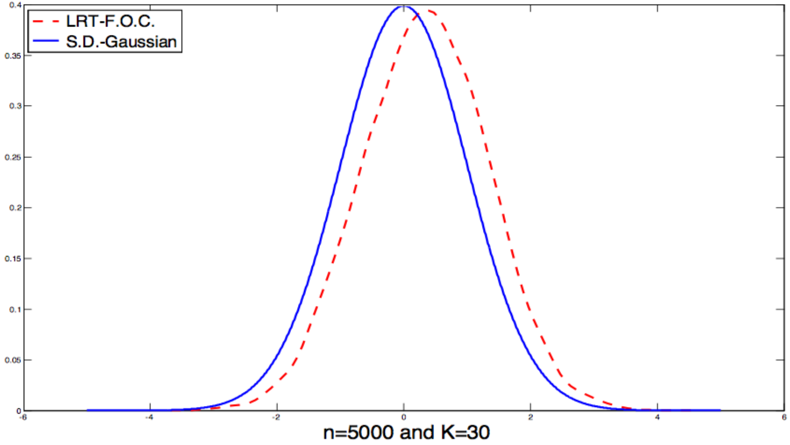
MC Illustration

Figure 6. Density of the LR-test Statistic ($\theta_2 = 0.5$)



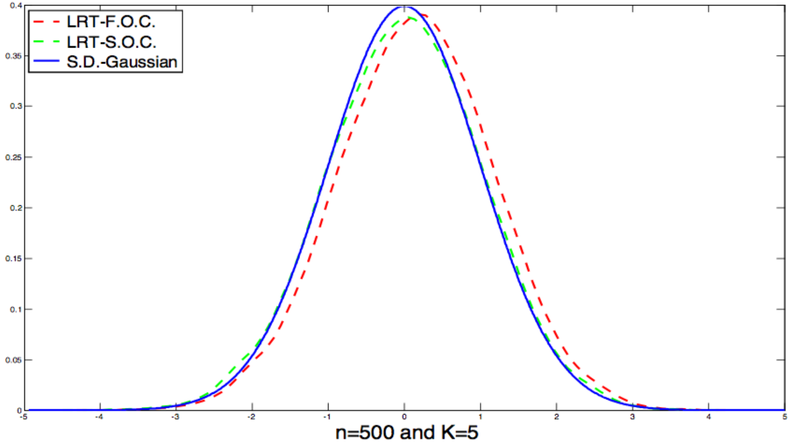
MC Illustration

Figure 7. Density of the LR-test Statistic ($\theta_2 = 0.5$)



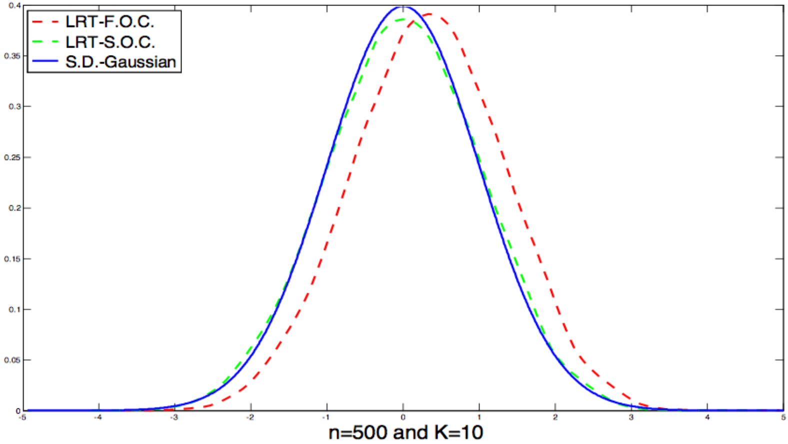
MC Illustration

Figure 8. Densities of the LR and Robust LR Statistics ($\theta_2 = 0.5$)



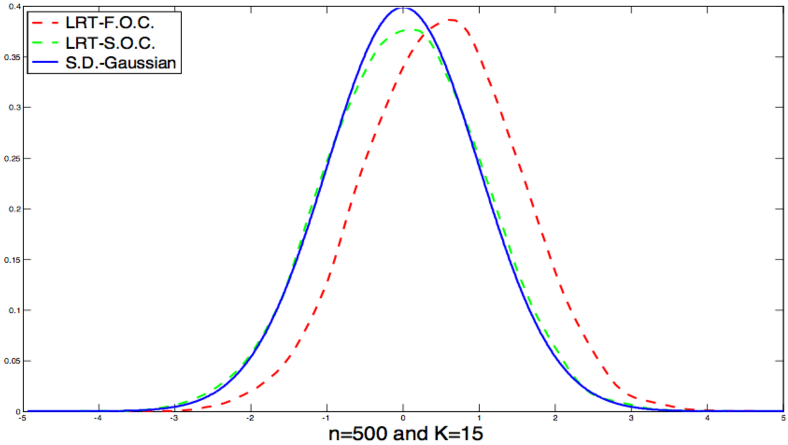
MC Illustration

Figure 9. Densities of the LR and Robust LR Statistics ($\theta_2 = 0.5$)



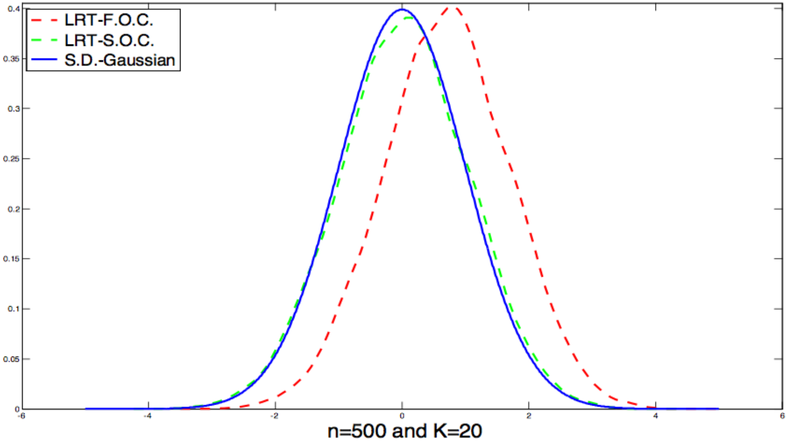
MC Illustration

Figure 10. Densities of the LR and Robust LR Statistics ($\theta_2 = 0.5$)



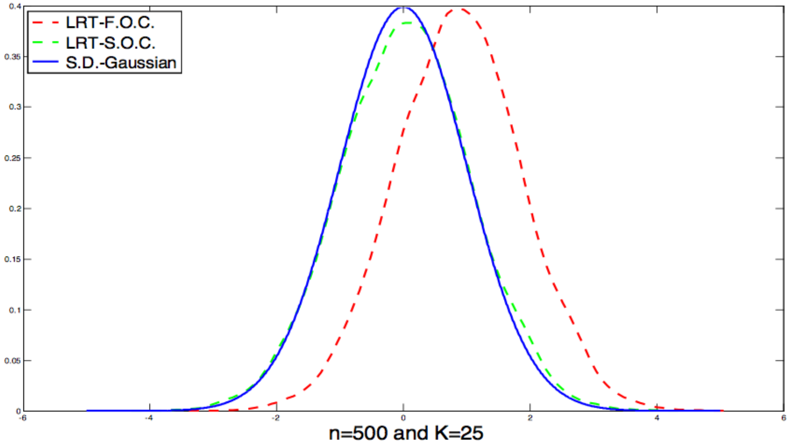
MC Illustration

Figure 11. Densities of the LR and Robust LR Statistics ($\theta_2 = 0.5$)



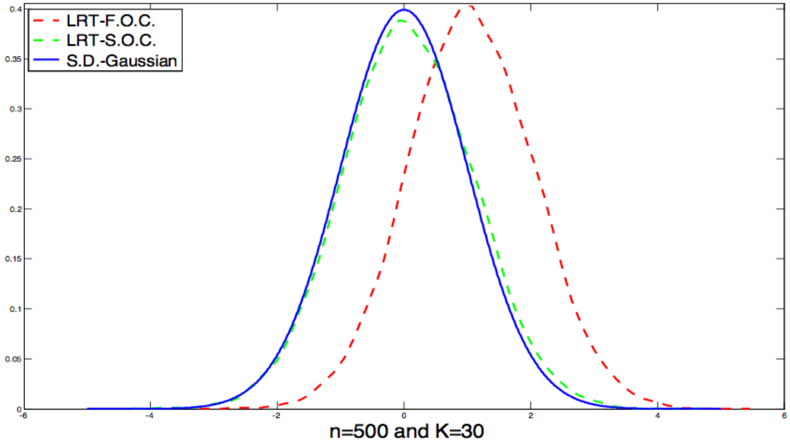
MC Illustration

Figure 12. Densities of the LR and Robust LR Statistics ($\theta_2 = 0.5$)



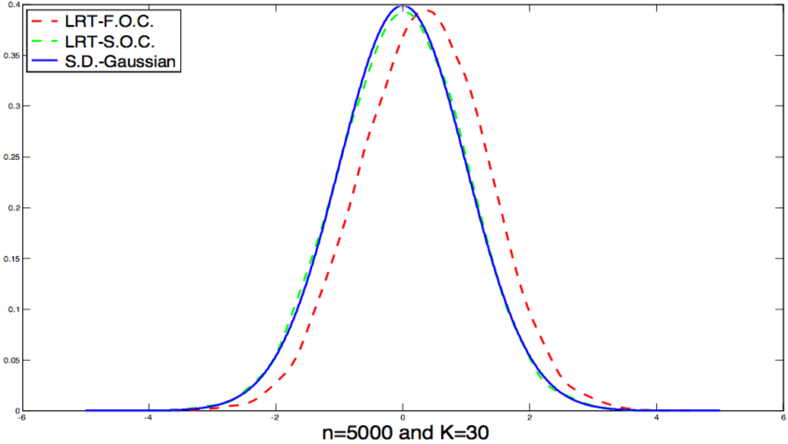
MC Illustration

Figure 13. Densities of the LR and Robust LR Statistics ($\theta_2 = 0.5$)



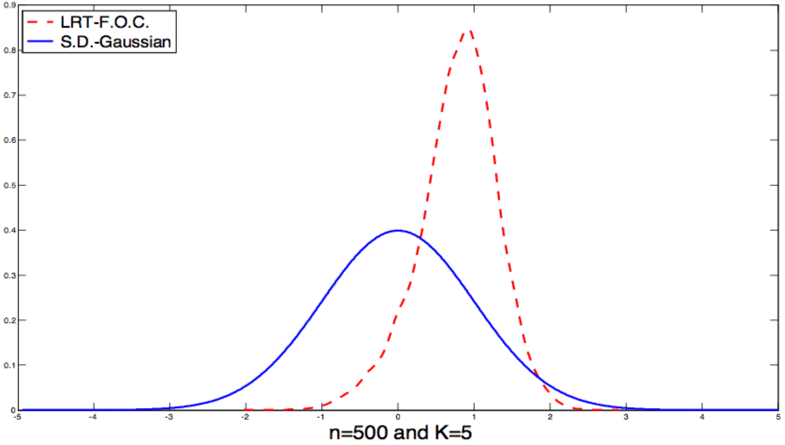
MC Illustration

Figure 14. Densities of the LR and Robust LR Statistics ($\theta_2 = 0.5$)



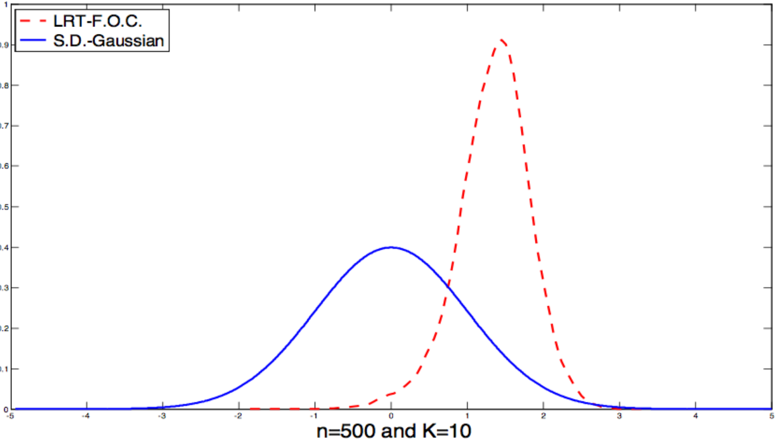
MC Illustration

Figure 15. Densities of the LR test Statistics ($\theta_2 = 0$)



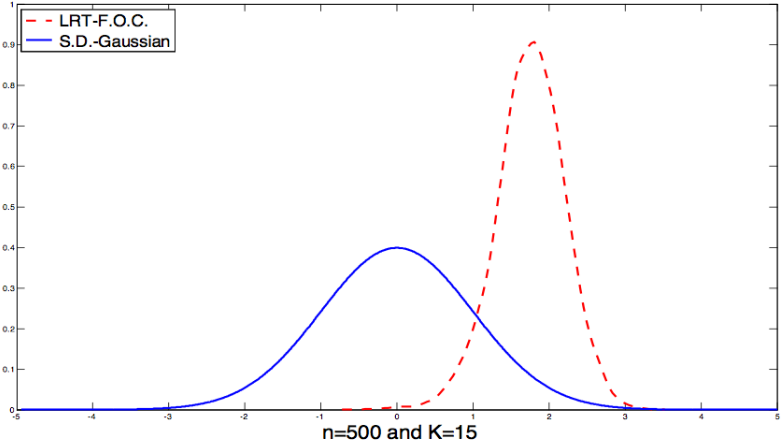
MC Illustration

Figure 16. Densities of the LR test Statistics ($\theta_2 = 0$)



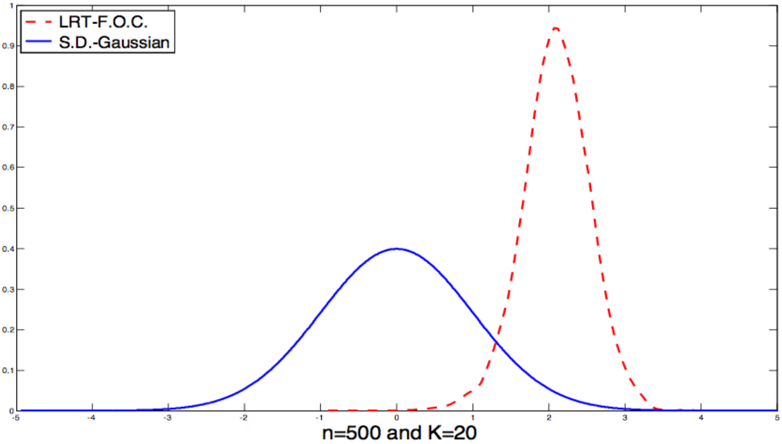
MC Illustration

Figure 17. Densities of the LR test Statistics ($\theta_2 = 0$)



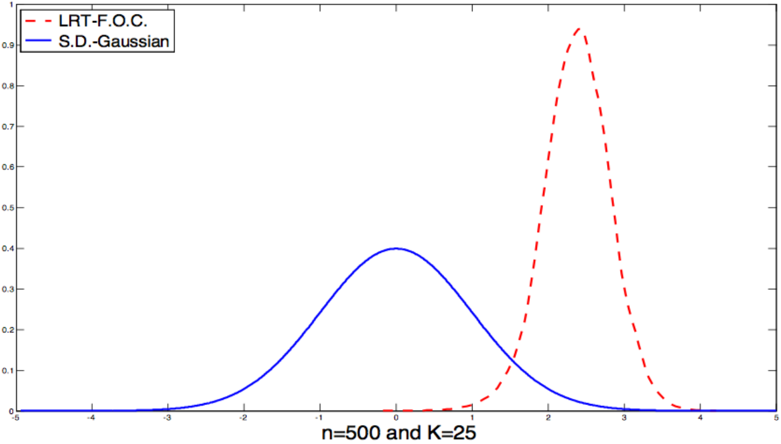
MC Illustration

Figure 18. Densities of the LR test Statistics ($\theta_2 = 0$)



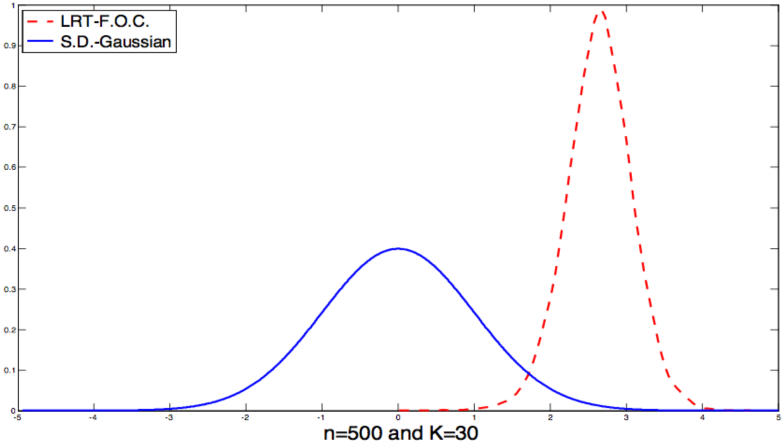
MC Illustration

Figure 19. Densities of the LR test Statistics ($\theta_2 = 0$)



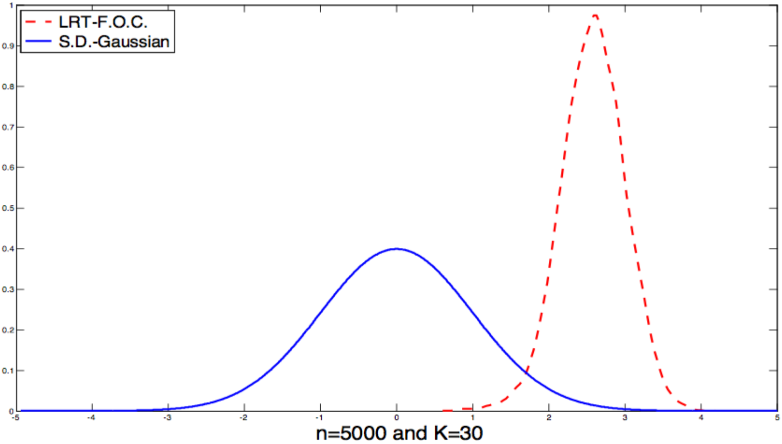
MC Illustration

Figure 20. Densities of the LR test Statistics ($\theta_2 = 0$)



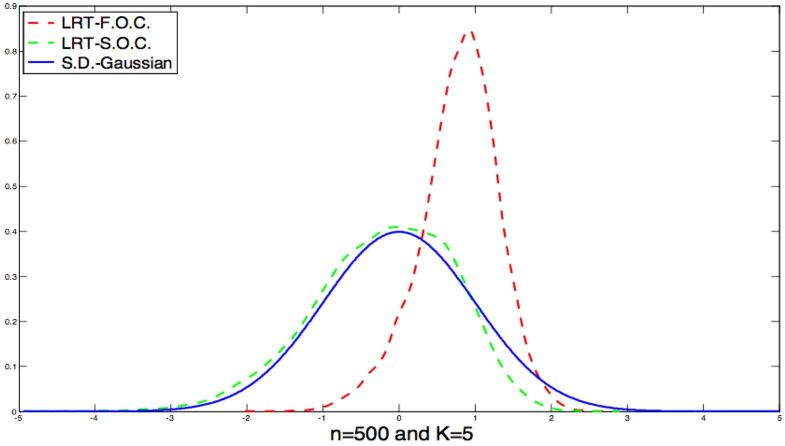
MC Illustration

Figure 21. Densities of the LR test Statistics ($\theta_2 = 0$)



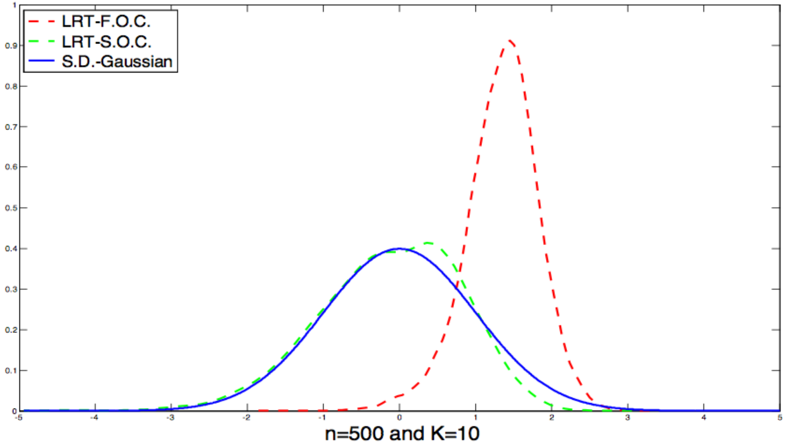
MC Illustration

Figure 22. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



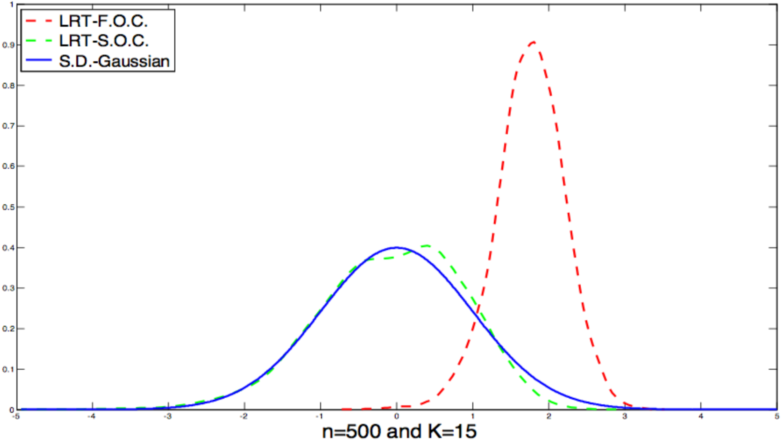
MC Illustration

Figure 23. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



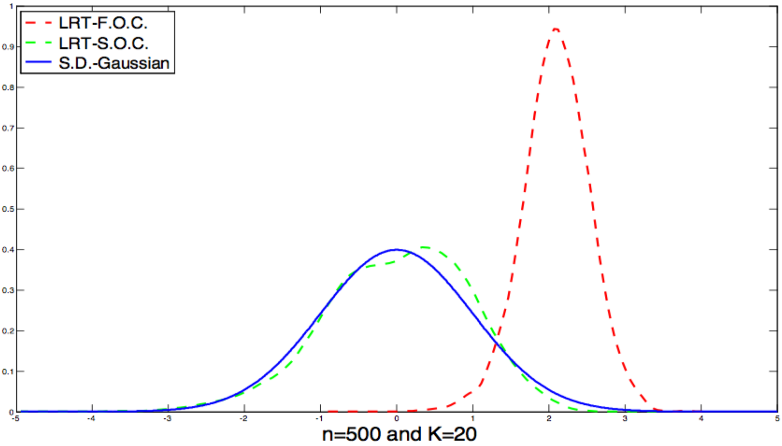
MC Illustration

Figure 24. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



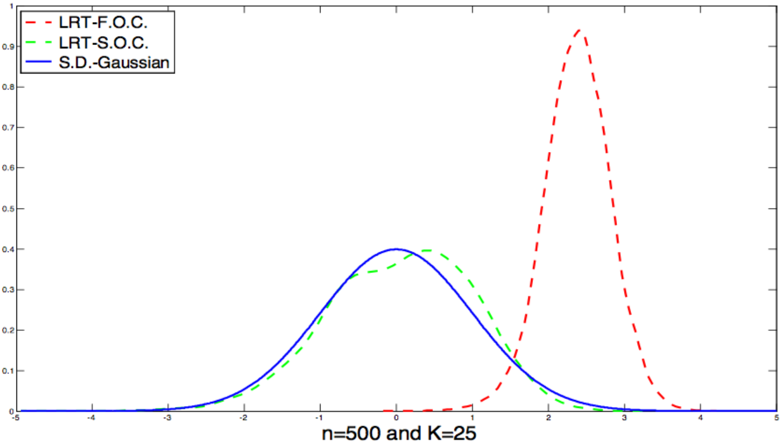
MC Illustration

Figure 25. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



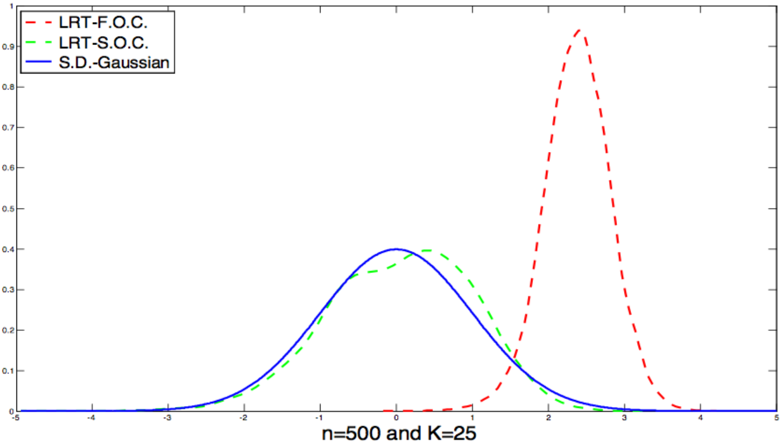
MC Illustration

Figure 26. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



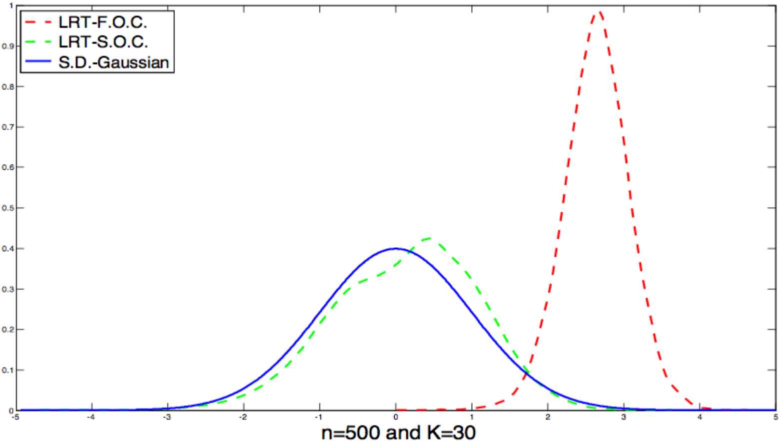
MC Illustration

Figure 27. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



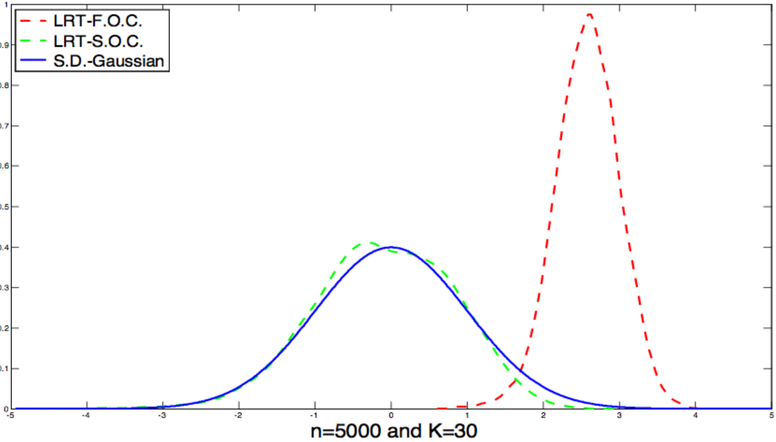
MC Illustration

Figure 27. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



MC Illustration

Figure 28. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



The Second Order Expansion of QLR

- Let $m_i(\theta) = m_1(Z_i, \theta_1) - m_2(Z_i, \theta_2)$, where $\theta' = (\theta_1', \theta_2')$.
- To make the presentation simple, let's focus on nonparametric models: θ_j^* is an unknown function and the sieve space

$$\Theta_{j,n} = \left\{ \theta_j : \theta_j = P_{k_j}' \beta_{k_j,j} \text{ with } \beta_{k_j,j} \in B_{k_j} \right\}$$

  where P_{k_j}: vector of sieve basis functions and k_j: dimension of sieve space.
- Define $\Theta_n = \Theta_{1,n} \times \Theta_{2,n}$, $P_k = \begin{pmatrix} P_{k_1} & 0 \\ 0 & P_{k_2} \end{pmatrix}$ and $\beta_k' = (\beta_{k_1,1}', \beta_{k_2,2}')$. Then for any $\theta \in \Theta_n$, we can write $\theta = P_k' \beta_k$.
- Also let's assume $m_i(P_k' \beta)$ is differentiable in θ a.e.

The Second Order Expansion of QLR

- Under some regularity conditions, we show that

$$
\begin{aligned}
n^{\frac{1}{2}}\widehat{QLR}_n \quad = \quad & \overbrace{n^{-\frac{1}{2}}\sum_{i=1}^{n}\left[m_1(Z_i,\theta_1^*) - m_2(Z_i,\theta_2^*)\right]}^{F.O.} \\
& \overbrace{-2^{-1}n^{\frac{1}{2}}(\widehat{\beta}_{k,n} - \beta_k^*)' H_{n,k}(\widehat{\beta}_{k,n} - \beta_k^*)}^{S.O.} \\
& + o_p(n^{-\frac{1}{2}}(k_1+k_2)^{\frac{1}{2}})
\end{aligned}
$$

where $H_{n,k} = E_\mu\left[\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k' \partial \beta_k}\right]$.

- Var$(F.O.)$ =Var$[m_1(Z_i,\theta_1^*) - m_2(Z_i,\theta_2^*)] = \omega_{F.O.}^2$.
- $S.O. = O_p(n^{\frac{1}{2}})\left\|\widehat{\beta}_{k,n} - \beta_k^*\right\|^2 = O_p((k_1+k_2)n^{-\frac{1}{2}})$.

The Second Order Expansion of QLR

- Var$(F.O.) = \omega^2_{F.O.}$ and $S.O. = O_p((k_1 + k_2)n^{-\frac{1}{2}})$.
- Cases where $\omega^2_{F.O.}$ is too small relative to Var$(S.O.)$ for finite n:
  - Strict degeneracy: $\omega^2_{F.O.} = 0$ $(m_1(Z, \theta_1^*) = m_2(Z, \theta_2^*)$ a.e.)

    e.g. both are likelihood models and correctly specified.

    e.g. both are regression models and share the nonredundant regressors.
  - Near degeneracy in large models: $\omega^2_{F.O.}$ is small and $k_1 + k_2$ is relatively large

    e.g. both likelihood models are mildly misspecified

    e.g. regressors of both regression models overlap and/or are correlated
- The second order term may not be negligible in finite samples even when two models are strictly non-nested.

Degeneracy

- In these degenerate cases, F.O. cannot dominate S.O..
- Thus, $\frac{n^{\frac{1}{2}}\widehat{QLR}_n}{\widehat{\omega}_A}$ is not close to $N(0,1)$ even in the asymptotic sense.
- Causing size distortion for the first-order asymptotic test.
- There may be (severe) over rejection because

$$E\left[S.O.\right] \approx -\frac{1}{2}n^{-\frac{1}{2}}E\left(\left[\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k'}\right] H_{n,k}^{-1}\left[\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k'}\right]'\right)$$

  may be (very) different from zero.
- The problem actually is more complicated because there is second order term in $\widehat{\omega}_A^2$ as well.

Robust Asymptotic Theory

- The second order term can be further decomposed as

$$
\begin{aligned}
S.O. \quad = \quad & \overbrace{-2^{-1}n^{-\frac{3}{2}}\sum_{i=1}^{n}\left[\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k'}\right] H_{n,k}^{-1}\left[\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k'}\right]'}^{\equiv b_n} \\
& \overbrace{-n^{-\frac{3}{2}}\sum_{i=2}^{n}\left[\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k'}\right] H_{n,k}^{-1}\left[\sum_{j=1}^{i-1}\frac{\partial m_j(P_k'\beta_k^*)}{\partial \beta_k'}\right]}^{\equiv U_n} \\
& +o_p(n^{-\frac{1}{2}}(k_1+k_2)^{\frac{1}{2}});
\end{aligned}
$$

- $\mathrm{Var}(U_n) = \omega_{U_n}^2 = \frac{n-1}{2n^2}\sum_{s=1}^{k_1+k_2}\lambda_s^2$, where λ_s is the eigenvalue of $D_{n,K}^{\frac{1}{2}} H_{n,k}^{-1} D_{n,K}^{\frac{1}{2}}$ and

$$
D_{n,K} = E_\mu\left[\left(\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k'}\right)'\left(\frac{\partial m_i(P_k'\beta_k^*)}{\partial \beta_k'}\right)\right].
$$

Robust Asymptotic Theory

- Now the QLR statistic can be written as

$$n^{\frac{1}{2}}\widehat{QLR}_n = \overbrace{n^{\frac{1}{2}}v_n\left[m_i(\theta^*)\right]}^{F.O.} + \overbrace{b_n + U_n}^{S.O.} + o_p(n^{-\frac{1}{2}}(k_1+k_2)^{\frac{1}{2}}),$$

- That is

$$\frac{n^{\frac{1}{2}}\widehat{QLR}_n - b_n}{\sigma_n} = \frac{n^{\frac{1}{2}}v_n\left[m_i(\theta^*)\right] + U_n}{\sigma_n} + o_p(\sigma_n^{-1}n^{-\frac{1}{2}}(k_1+k_2)^{\frac{1}{2}}),$$

where $\sigma_n^2 = \omega_{F.O.}^2 + \omega_{U_n}^2 =Var[m_i(\theta^*)] + \frac{n-1}{2n^2}\sum_{s=1}^{k_1+k_2}\lambda_s^2$.

Robust Asymptotic Theory

Theorem (Robust Asymptotic Distribution)

Under H_0 and regularity conditions, if $n^{-1}\sigma_n^{-2}(k_1+k_2)=O(1)$ and $\sup_{s,n}\lambda_s<\infty$, then

$$\frac{n^{\frac{1}{2}}\widehat{QLR}_n - b_n}{\sigma_n} = \frac{n^{\frac{1}{2}}\widehat{QLR}_n - b_n}{\sqrt{\omega_{F.O.}^2+\omega_{U_n}^2}} \to_d N(0,1).$$

- The theorem allows $\omega_{F.O.}^2=0$, also allows $\omega_{F.O.}^2$ to depend on n and $\omega_{F.O.}^2\to 0$ at any rate. Thus robust to degeneracy.
- When $\lim\omega_{F.O.}^2=\omega_A^2>0$, b_n and ω_U^2 become smaller order terms. The above theorem is also valid.

Robust Test

- We show that

$$\frac{\widehat{\omega}_A^2 - (\sigma_n^2 + \omega_{U_n}^2)}{\sigma_n^2} \to_p 0$$

where $\widehat{\omega}_A^2 = \widehat{\mathsf{Var}}_n \left[m_1(Z, \widehat{\theta}_{1,n}) - m_2(Z, \widehat{\theta}_{2,n}) \right]$;

- This implies that $\widehat{\omega}_A^2$ is an inconsistent variance estimator.
- Analysis of the F.O. Asy. Theo:

$$\begin{aligned} \frac{n^{\frac{1}{2}} \widehat{QLR}_n}{\widehat{\omega}_A} &= \frac{\sigma_n}{\widehat{\omega}_A} \frac{b_n + n^{\frac{1}{2}} \widehat{QLR}_n - b_n}{\sigma_n} + o_p(1) \\ &= \frac{\sigma_n}{\widehat{\omega}_A} \left[\frac{b_n}{\sigma_n} + \frac{n^{\frac{1}{2}} \widehat{QLR}_n - b_n}{\sigma_n} \right] + o_p(1) \end{aligned}$$

where there are a scaling bias $\frac{\sigma_n}{\widehat{\omega}_A}$ and a additive bias $\frac{b_n}{\sigma_n}$.

Robust Test

- We show the plug-in estimators $\widehat{b}_n$ and $\widehat{\omega}^2_{U_n}$ are consistent:

$$\frac{\widehat{b}_n - b_n}{\sigma_n} = o_p(1) \text{ and } \frac{\widehat{\omega}^2_{U_n} - \omega^2_{U_n}}{\sigma^2_n} = o_p(1),$$

which implies that

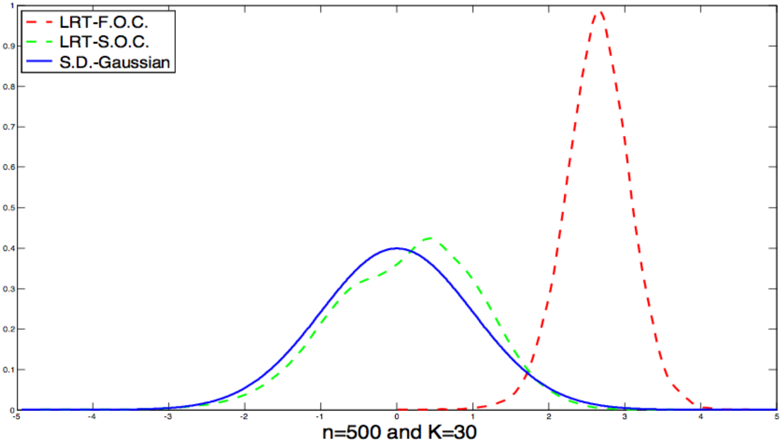
$$\frac{\sqrt{n}\widehat{QLR}_n - \widehat{b}_n}{\sigma_n} = \frac{\sqrt{n}\widehat{QLR}_n - b_n}{\sigma_n} + o_p(1);$$

$$\text{and } \frac{(\widehat{\omega}^2_A - \widehat{\omega}^2_{U_n})}{\sigma^2_n} \to_p 1.$$

- Therefore, $\frac{\sqrt{n}\widehat{QLR}_n - \widehat{b}_n}{\sqrt{\widehat{\omega}^2_A - \widehat{\omega}^2_{U_n}}} \to_d N(0,1)$

MC Illustration

Figure 27. Densities of the LR and Robust LR Statistics ($\theta_2 = 0$)



Robust Test

- Using the results in the previous slide, we get

$$\varphi_n^{Rob-Asy}(\alpha) = 1\left\{ \frac{\left| n^{\frac{1}{2}}\widehat{QLR}_n - \widehat{b}_n \right|}{\widehat{\sigma}_n} > z_{\alpha/2} \right\}$$

- Our general formula of $\widehat{\sigma}_n$:

$$\widehat{\sigma}_n^2 = \max\{\widehat{\omega}_A^2 - \widehat{\omega}_{U_n}^2, \widehat{\omega}_{U_n}^2\}.$$

to ensure that $\widehat{\sigma}_n > 0$.

Robust Test – Nested Case

- When the models are known to be nested, we have

$$\frac{n^{\frac{1}{2}}\widehat{QLR}_n - b_n}{\sqrt{\omega_{F.O.}^2 + \omega_{U_n}^2}} = \frac{n^{\frac{1}{2}}\widehat{QLR}_n - b_n}{\omega_{U_n}} \to_d N(0,1);$$

- The inference based on $\widehat{\sigma}_n$ is still asymptotically valid;
- However, more accurate inference of the null hypothesis would be

$$\varphi_n^{Rob-Asy}(\alpha) = 1\left\{\frac{\left|n^{\frac{1}{2}}\widehat{QLR}_n - \widehat{b}_n\right|}{\widehat{\omega}_{U_n}} > z_{\alpha/2}\right\}.$$

Bootstrap

- A bootstrap critical value may be used in place of z_α.
- Let $\widehat{QLR}_n^*$ be computed from the same formula as $\widehat{QLR}_n$ except using the nonparametric i.i.d. bootstrap sample instead of original sample. And let

$$\widehat{\sigma}_n^{2,*} = \widehat{\omega}_A^2 + \widehat{\omega}_{U_n}^2,$$

- Let $cv_n^{BT}(\alpha)$ be the $1-\alpha$ quantile of the conditional (on data) distribution of

$$\frac{n^{\frac{1}{2}}\widehat{QLR}_n^* - n^{\frac{1}{2}}\widehat{QLR}_n - \widehat{b}_n}{\widehat{\sigma}_n^*}$$

- We show that $cv_n^{BT}(\alpha) \to_p z_\alpha$: this bootstrap is consistent.
- Note: both the recentering and the bias correction are needed.

Score Bootstrap

- The bootstrap above requires recomputing $\widehat{\theta}_n$ for every bootstrap sample, which can be demanding in some problems.
- Thus we also propose a score-based bootstrap critical value, which is also consistent.
- Let $\widehat{QLR}_n^{S,*} = \frac{1}{n}\sum_{i=1}^n m(Z_i^*, \widehat{\theta}_n)$ and $S_n^* = \frac{1}{n}\sum_{i=1}^n \frac{\partial m(Z_i^*, P_k'\widehat{\beta}_k)}{\partial \beta_k}$.
- Let the score bootstrap critical value $cv_n^{SBT}(\alpha)$ be the $1-\alpha$ quantile of the conditional (on data) distribution of

$$\frac{n^{\frac{1}{2}}\widehat{QLR}_n^{S,*} - n^{\frac{1}{2}}\widehat{QLR}_n - \frac{\sqrt{n}}{2} S_n^{*\prime} H_{n,k}^{-1} S_n^* - \widehat{b}_n}{\widehat{\sigma}_n^*}$$

Nonparametric Exogeneity Test

- Our test applies to nonparametric exogeneity testing.
- Let Y be a dependent variable, X an explanatory variable, and the structural model says

$$Y = g(X) + \varepsilon$$

  Maintain $E(Z|\varepsilon) = 0$; want to test the exogeneity of X.
- Under exogeneity, we have $E\left[\varepsilon|Z, X\right] = 0$ which implies that

$$E\left[Y|Z, X\right] = E\left[Y|X\right] = g(X).$$

- Thus exogeneity is rejected if the following H_0 is rejected:

$$H_0 : E\left[\left|Y - \theta_1(Z, X)\right|^2\right] = E\left[\left|Y - \theta_2(X)\right|^2\right]$$

  where $\theta_1(Z, X) = E\left[Y|Z, X\right]$ and $\theta_2(X) = E\left[Y|X\right]$.
- This falls into our framework, and two models are nested.

Nonparametric Exogeneity Test

- Difference with Blundell and Horowitz (2007):
  - Our exogeneity is a joint one: $E[\varepsilon|Z, X] = 0$;
  - BH defines exogeneity as separate: $E[\varepsilon|Z] = 0$ and $E[\varepsilon|X] = 0$;
  - Both we and BH maintain: $E[\varepsilon|Z] = 0$.
- Design (same as BH): $X = \Phi(\xi)$, $Z = \Phi(\zeta)$, $\zeta \sim N(0, 1)$, $\xi = \rho\zeta + \sqrt{1 - \rho^2}u$, $u \sim N(0, 1)$,

$$Y = \beta_0 + X\beta_1 + \sigma_\varepsilon \varepsilon$$

  where $\varepsilon = \eta u + \sqrt{1 - \eta^2}v$, $v \sim N(0, 1)$ and $\sigma_\varepsilon = 0.2$.
- X is exogenous if and only if $\eta = 0$. ρ controls the strength of instrument Z.
- Number of Monte Carlo repetitions: $S = 1000$.

Nonparametric Exogeneity Test

Table 1. Rejection Probabilities of Nonpar Exog Test ($\alpha = 0.05$)

n	ρ	η	F.O. Asy.	Rob. Asy.	BT	Score BT	BH*
250	.35	.00	.007	.078	.063	.059	.042
250	.35	.15	.019	.139	.108	.110	.077
250	.35	.25	.043	.249	.218	.216	.119
750	.35	.00	.004	.067	.052	.051	.048
750	.35	.15	.053	.263	.219	.217	.274
750	.35	.25	.209	.576	.523	.510	.596

*: BH simulation results are taken from Blundell and Horowitz (2007).

Non nested Linear Regression

- Regression models:

$$\begin{aligned} \mathcal{M}_1 &: \quad Y = \beta_{0,1} + X_1'\beta_{1,1} + u_1 \text{ with } E\left[u_1|X_1\right] = 0; \\ \mathcal{M}_2 &: \quad Y = \beta_{0,2} + X_2'\beta_{1,2} + u_2 \text{ with } E\left[u_2|X_2\right] = 0; \end{aligned}$$

- Assume that the regressors X_1 and X_2 are not subsets of each other. Thus models are non nested.
- We want to compare the two models by their regression MSE:

$$H_0 : E\left[\left(Y - \beta_{0,1}^* - X_1'\beta_{1,1}^*\right)^2\right] = E\left[\left(Y - \beta_{0,2}^* - X_1'\beta_{1,2}^*\right)^2\right]$$

- DGP: $(X_1', X_2', u) \sim N(0, I_{k_1+k_2-1})$ and

$$Y = 1 + \frac{a_1 X_1' 1_{k_1-1}}{\sqrt{k_1-1}} + \frac{a_2 X_2' 1_{k_2-1}}{\sqrt{k_2-1}} + u.$$

- H_0 holds if and only if $a_1 = a_2$, and $\omega_{F.O.}^2 = 0$ if $a_1 = a_2 = 0$;
- Consider $K_1 = 12$; $K_2 = 3$.

Nonparametric Exogeneity Test

Table 2. Results for the Nonpar Exog Test with $\alpha = 0.1$

n	(a_1, a_2)	F.O. Asy.	Rob. Asy.	BT	Score BT
250	$(.25, .25)$	.248	.120	.094	.128
250	$(.00, .00)$	.467	.145	.103	.154
250	$(.00, .25)$	[.010,.034]*	[.000,.502]	[.000,.453]	[.000,.523]
500	$(.25, .25)$	.169	.117	.107	.117
500	$(.00, .00)$	.444	.142	.113	.154
500	$(.00, .25)$	[.000,.358]	[.000,.850]	[.000,.842]	[.000,.862]

*: The test done is a 2-sided test. The numbers in the brackets are:
[prob. of rejection H_0 and in favor of Model 1,
prob. of rejecting H_0 and in favor of Model 2]

Conclusion

- We developed a robust model selection test for large models.
- The test may use standard normal, bootstrap or score-bootstrap critical value.
- It applies to both nested and non nested models, both strictly nonnested or overlapping models.
- We are extending the test to time-series context. (Models for copula dependence, forecasting quality, etc.)