Efficient Inference in Econometric Models When Identification Can Be Weak

Vadim Marmer Zhengfei Yu

University of British Columbia

Introduction: Literature review

- Staiger and Stock (1997): Weak IVs, suggested to use Anderson-Rubin (AR) statistic, as its null distribution does not depend on the strength of IVs.
  - AR-statistic is based on null-restricted residuals (their sample covariance with IVs).
- AR-statistic is inefficient in overidentified models (due to overidentified restrictions).
- Kleibergen's (2002) K-statistic and Moreira's (2003) CLR-statistic: Transform AR's overidentified inference problem into exactly identified.
- Andrews, Moreira, and Stock (2006): Normal model, single endogenous regressor, homoskedastic errors:
  - K- and CLR- statistics are efficient when IVs are strong.
  - There is no efficient test when IVs are weak.
  - CLR-based test is (numerically) nearly efficient. They do not recommend using K-statistic.

Introduction: Literature review

- Cattaneo, Crump, and Jansson (2012): Relax the normality assumption of Andrews, Moreira, and Stock (2006) by considering an asymptotic framework.
- The null-restricted residuals-based approach cannot be used in certain nonlinear models: e.g. models with latent dependent variables.
- Magnusson (2010): AR-, K-, and CLR-statistics for models with latent dependent variables and minimum distance estimation (MDE).
    - Tests are formed through identifying equations that relate structural and reduced-form parameters.

Introduction: Contribution

- We study the efficiency of robust tests in nonlinear models (in the MDE framework).
  - Semiparametric efficiency: Choose your favorite estimators for reduced-form parameters, what is an efficient test?
- Asymptotic experiments approach as in Cattaneo, Crump, and Jansson (2012).
- K and CLR tests are asymptotically efficient when identification is strong, single endogenous parameter.
- Asymptotic efficiency arguments in the multiple endogenous parameters case.
- Loss of efficiency when identification is weak.
  - Analytical power analysis of K-test.
  - An alternative test: Conditional LM (CLM) statistic.
- There is no dominating test, but CLM test can outperform CLR and K in models with general covariance structure.

Minimum Distance (MD) framework

- Let $h : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^k$ be a known function. The framework is defined by

$$h(\pi, \gamma) = 0,$$

where

  - π is a vector of reduced-form parameters. We assume that there is a consistent and asymptotically normal estimator:

$$\sqrt{n}(\hat{\pi}_n - \pi) \to_d N(0, \Omega).$$

    We also assume that Ω can be estimated consistently:

$$\hat{\Omega}_n \to_p \Omega.$$

  - γ is a structural parameter.

Minimum Distance (MD) framework

- In this framework, γ is (usually) estimated by

$$\hat{\gamma}_n = \arg\min_{x \in \mathbb{R}^m} h(\hat{\pi}_n, x)' A(\hat{\pi}_n, x) h(\hat{\pi}_n, x).$$

- Inference on γ is based on the asymptotic approximation:

$$\begin{aligned} \sqrt{n}(\hat{\gamma}_n - \gamma) &= \left(\frac{\partial h(\hat{\pi}_n, \hat{\gamma}_n)'}{\partial \gamma} A(\hat{\gamma}_n) \frac{\partial h(\hat{\pi}_n, \hat{\gamma}_n)}{\partial \gamma'} \right)^{-1} \\ &\quad \times \frac{\partial h(\hat{\pi}_n, \hat{\gamma}_n)}{\partial \pi} \sqrt{n}(\hat{\pi}_n - \pi) + o_p(1). \end{aligned}$$

- The quality of the approximation can be very poor if the rank condition

$$\text{rank} \left(\frac{\partial h(\pi, \gamma)'}{\partial \gamma} A(\gamma) \frac{\partial h(\pi, \gamma)}{\partial \gamma'} \right) = m$$

fails locally: if the smallest eigenvalue of the above matrix is *local-to-zero.*

Example: Horowit'z (1996) transformation model with endogenous regressors

- Structural equation: $\Lambda(Y_{1i}) = \theta Y_{2i} + X_i'\beta + U_i$, where $\Lambda(\cdot)$ is an unknown function. Suppose that $\gamma \in \mathbb{R}$.
- First stage: $Y_{2i} = \Pi_1' X_i + \Pi_2' Z_i + V_i$, where Z_i is a vector of instruments, $\Pi_2 \in \mathbb{R}^k$.
- Null-restricted residuals-based approach cannot be used since $\Lambda(\cdot)$ is unknown.
- The marginal structural effect of Y_{2i}: $\partial Y_{1i}/\partial Y_{2i} = \theta/\Lambda'(Y_{1i})$.
- Identification:

$$\frac{\theta}{\Lambda'(y)}\Pi_{2,j} = -\frac{\partial \Pr\left(Y_{1i} \leq y \mid Z_i = z\right)/\partial z_j}{\partial \Pr\left(Y_{1i} \leq y \mid Z_i = z\right)/\partial y} \text{ for } j = 1, \ldots k.$$

Example: Horowit'z (1996) transformation model with endogenous regressors

- Define: $\gamma = \theta / \Lambda'(y)$.
- Let $w : \mathbb{R}^k \to [0,1]$ be a weight function. Define:

$$\pi_{2,j} = -\int \frac{\partial \Pr\left(Y_{1i} \leq y \mid Z_i = z\right) / \partial z_j}{\partial \Pr\left(Y_{1i} \leq y \mid Z_i = z\right) / \partial y} w(z) dz \text{ for } j = 1, \ldots k.$$

- We have:

$$\begin{aligned} h\left(\left(\pi_2', \Pi_2'\right)', \gamma\right) &= \gamma \Pi_2 - \pi_2. \\ \frac{\partial h}{\partial \gamma} &= \Pi_2. \end{aligned}$$

- Π_2 can be *local-to-zero* (weak IVs).

Example: Linear structural models with latent dependent variables

- Data: $\{(y_i, Y_i', X_i', Z_i') : i = 1, \ldots, n\}$, where:
- $y_i = G(y_i^*)$, where
  - y_i is observable limited dependent variable,
  - y_i^* is latent dependent variable,
  - $G(\cdot)$ is a known function. Example: $G(u) = \max\{u, 0\}$ (censored regression).
- Structural equation: $y_i^* = \gamma' Y_i + \beta' X_i + u_i$, where
  - Y_i is the $m \times 1$ endogenous regressor,
  - X_i is the $l \times 1$ exogenous regressor.

Censored regression

- First-stage equation: $Y_i = \Pi_1' X_i + \Pi_{2,n}' Z_i + V_i$, where
  - Z_i is the $k \times 1$ vector of IVs ($k \geq m$),
  - Strong IVs: $\Pi_{2,n} = \Pi_2$ (fixed $k \times m$ matrix), or
  - Weak IVs: $\Pi_{2,n} = \Pi_2 / \sqrt{n}$.
- Reduced-from equation: $y_i = \max\left\{\pi_1' X_i + \pi_{2,n}' Z_i + v_i, 0\right\}$, where
  - $\pi_1 = \Pi_1 \gamma + \beta$,
  - $\pi_{2,n} = \Pi_{2,n} \gamma$.

Censored regression

- In this example:

$$h\left((\pi_1', \pi_{2,n}', \text{vec}(\Pi_1)', \text{vec}(\Pi_{2,n})')', (\gamma', \beta')'\right) = \begin{pmatrix} \pi_1 - \Pi_1\gamma - \beta \\ \pi_{2,n} - \Pi_{2,n}\gamma \end{pmatrix}$$

$$\frac{\partial h}{\partial \gamma'} = \Pi_{2,n}.$$

- Since the first-stage: $Y_i = \Pi_1' X_i + \Pi_{2,n}' Z_i + V_i$, the rank condition will fail locally if Z_i is weak ($\Pi_{2,n} = \Pi_2/\sqrt{n}$).

Censored regression: Estimation

- Start from reduced-form estimates:

$$n^{1/2}\begin{bmatrix} \hat{\pi}_{1,n}-\pi_1 \\ \hat{\pi}_{2,n}-\pi_{2,n} \\ \text{vec}\left(\hat{\Pi}'_{1,n}-\Pi'_1\right) \\ \text{vec}\left(\hat{\Pi}'_{2,n}-\Pi'_{2,n}\right) \end{bmatrix} \to_d N\left(0_{((m+1)l+(m+1)k)\times 1}, \Omega\right),$$

  and assume that there is $\hat{\Omega}_n \to_p \Omega$.

- Π_1 and $\Pi_{2,n}$ can be estimated from $Y_i = \Pi_1' X_i + \Pi_{2,n}' Z_i + V_i$ (parametrically or semi-parametrically).

Censored regression: Estimation

- π_1 and π_2 can be estimated from

$$y_i = \max\left\{\pi_1' X_i + \pi_{2,n}' Z_i + v_i, 0\right\}$$

  - Parametrically assuming parametric distribution for v_i (a joint parametric specification for u_i and V_i).
  - semi-parametrically as in Powell (1984; 1986) since the τ-th conditional quantile of y_i is

$$\mathrm{const}_\tau + \pi_1' X_i + \pi_{2,n}' Z_i$$

  for sufficiently large quantile value τ's.

Estimation (Amemiya, 1978)

Write

$$\hat{\pi}_{1,n} = \hat{\Pi}_{1,n}\gamma + \beta + (\hat{\pi}_{1,n} - \pi_1) - \left(\hat{\Pi}_{1,n} - \Pi_1\right)\gamma,$$

$$\hat{\pi}_{2,n} = \hat{\Pi}_{2,n}\gamma + (\hat{\pi}_{2,n} - \pi_{2,n}) - \left(\hat{\Pi}_{2,n} - \Pi_{2,n}\right)\gamma,$$

or

$$\hat{\pi}_n = \hat{H}_n \begin{bmatrix} \gamma \\ \beta \end{bmatrix} + \begin{bmatrix} \eta_{1,n} \\ \eta_{2,n} \end{bmatrix}, \text{ where}$$

$$\hat{\pi}_n = \begin{bmatrix} \hat{\pi}_{1,n} \\ \hat{\pi}_{2,n} \end{bmatrix}, \ \hat{H}_n = \begin{bmatrix} \hat{\Pi}_{1,n} & I_l \\ \hat{\Pi}_{2,n} & 0_{k\times l} \end{bmatrix},$$

$$\eta_{1,n} = \begin{bmatrix} I_l & -(I_l \otimes \gamma') \end{bmatrix} \begin{bmatrix} \hat{\pi}_{1,n} - \pi_1 \\ \operatorname{vec}\left(\hat{\Pi}'_{1,n} - \Pi'_1\right) \end{bmatrix},$$

$$\eta_{2,n} = \begin{bmatrix} I_k & -(I_k \otimes \gamma') \end{bmatrix} \begin{bmatrix} \hat{\pi}_{2,n} - \pi_{2,n} \\ \operatorname{vec}\left(\hat{\Pi}'_{2,n} - \Pi'_{2,n}\right) \end{bmatrix},$$

Estimation

$$\hat{\pi}_n = \hat{H}_n \begin{bmatrix} \gamma \\ \beta \end{bmatrix} + \begin{bmatrix} \eta_{1,n} \\ \eta_{2,n} \end{bmatrix}.$$

- Amemiya's OLS: $\begin{bmatrix} \hat{\gamma}_{AOLS,n} \\ \hat{\beta}_{AOLS,n} \end{bmatrix} = \left(\hat{H}_n' \hat{H}_n\right)^{-1} \hat{H}_n' \hat{\pi}_n.$
- Efficient estimation (under strong IVs): Let Σ denote the asymptotic Variance-Covariance of $[\eta_{1,n}', \eta_{2,n}']'$ (depends on γ)
  - Amemiya's GLS: $\begin{bmatrix} \hat{\gamma}_{AGLS,n} \\ \hat{\beta}_{AGLS,n} \end{bmatrix} = \left(\hat{H}_n' \Sigma^{-1} \hat{H}_n\right)^{-1} \hat{H}_n' \Sigma^{-1} \hat{\pi}_n.$
  - Amemiya's feasible GLS: $\begin{bmatrix} \hat{\gamma}_{AFGLS,n} \\ \hat{\beta}_{AFGLS,n} \end{bmatrix} = \left(\hat{H}_n' \hat{\Sigma}_{AOLS,n}^{-1} \hat{H}_n\right)^{-1} \hat{H}_n' \hat{\Sigma}_{AOLS,n}^{-1} \hat{\pi}_n$ where $\hat{\Sigma}_{AOLS,n}$ is a two-step estimator of Σ.

Estimation of γ

- The GLS estimator of γ:

$$\hat{\gamma}_{AGLS,n} = \left(\hat{\Pi}'_{2,n}\Sigma_{22}^{-1}\hat{\Pi}_{2,n}\right)^{-1}\hat{\Pi}'_{2,n}\Sigma_{22}^{-1}\hat{\pi}_{2,n}.$$

- Recall that

$$n^{1/2}\left[\begin{array}{c}\hat{\pi}_{2,n}-\pi_{2,n}\\ \text{vec}\left(\hat{\Pi}'_{2,n}-\Pi'_{2}\right)\end{array}\right] \rightarrow_d \left[\begin{array}{c}w_2\\ \text{vec}(W'_2)\end{array}\right]$$

$$\equiv \quad N\left(0_{(m+1)k\times 1},\left[\begin{array}{cc}\Omega_{22} & \Omega_{24}\\ \ldots & \Omega_{44}\end{array}\right]\right).$$

Estimation of γ under weak IVs

- Amemiya's GLS is inconsistent:

$$
\begin{aligned}
& (\hat{\gamma}_{AGLS,n} - \gamma) \\
= \quad & \left(\left(n^{1/2}(\hat{\Pi}'_{2,n} - \Pi'_{2,n}) + \Pi'_2 \right) \Sigma_{22}^{-1} \left(n^{1/2}(\hat{\Pi}_{2,n} - \Pi_{2,n}) + \Pi_2 \right)' \right)^{-1} \\
& \times \left(n^{1/2}(\hat{\Pi}'_{2,n} - \Pi'_{2,n}) + \Pi'_2 \right) \Sigma_{22}^{-1} \\
& \times \left(n^{1/2}(\hat{\pi}_{2,n} - \pi_{2,n}) - n^{1/2}(\hat{\Pi}_{2,n} - \Pi_{2,n})\gamma \right) \\
\to_d \quad & \left((W_2 + \Pi_2)' \Sigma_{22}^{-1} (W_2 + \Pi_2)' \right)^{-1} (W_2 + \Pi_2)' \Sigma_{22}^{-1} (w_2 - W_2\gamma) .
\end{aligned}
$$

- Amemiya's FGLS is also inconsistent and has a different asymptotic distribution: the OLS-based estimator Σ_{22} converges to a random limit.

Standard hypotheses testing

- $H_0 : \gamma = \gamma_0$ vs $H_1 : \gamma \neq \gamma_0$.
- Let $S_{22,AOLS}$ denote the random limit of the OLS-based estimator for Σ_{22}.
- Non-standard weak-IVs asymptotic null distribution for the Wald statistic:

$$
\begin{aligned}
Wald_{\gamma_0,n} &= n\left(\hat{\gamma}_{AFGLS,n} - \gamma\right)' \left(\hat{\Pi}_{2,n}' \hat{\Sigma}_{22,AOLS,n}^{-1} \hat{\Pi}_{2,n}\right) \\
&\quad \times \left(\hat{\gamma}_{AFGLS,n} - \gamma\right) \\
&\to_d \left\| \left(\left(W_2 + \Pi_2\right)' S_{22,AOLS}^{-1} \left(W_2 + \Pi_2\right) \right)^{-1/2} \right. \\
&\quad \left. \times \left(W_2 + \Pi_2\right)' S_{22,AOLS}^{-1} \left(w_2 - W_2 \gamma\right) \right\|^2 .
\end{aligned}
$$

- Size distortions.

Robust inference

- The usual approach to testing $H_0 : \gamma = \gamma_0$ is to use the null-restricted errors $y_i^* - \gamma_0' Y_i$.
- The usual approach cannot be used since y_i^* is unobservable.
- Magnusson (2010) proposes robust tests (Anderson-Rubin-, Kleibergen-, and Moreira-type) that are based on: $\hat{\pi}_{2,n} - \hat{\Pi}_{2,n}\gamma_0$.
- Our focus is on the efficiency.

Weak-IVs-robust inference: an asymptotic experiment

- As in Cattaneo, Crump, and Jansson (2012), focus on an asymptotic experiment with a single normally distributed observation.
- $m = 1$ (single endogenous regressor) and $k > 1$ (multiple IVs).
- What is the optimal test of $H_0 : \gamma_n = \gamma_0$ against $H_1 : \gamma_n = \gamma_0 + \delta_n$? Here:
  - $\delta_n = \delta/\sqrt{n}$ when $\Pi_{2,n} = \Pi_2$ (strong IVs).
  - $\delta_n = \delta$ when $\Pi_{2,n} = \Pi_2/\sqrt{n}$ (weak IVs).
- Choose your favorite estimators of the reduced-form and first-stage:

$$\begin{bmatrix} \hat{\pi}_{2,n} \\ \hat{\Pi}_{2,n} \end{bmatrix} \overset{a}{\sim} N\left(\begin{bmatrix} \Pi_{2,n}\gamma_n \\ \Pi_{2,n} \end{bmatrix}, n^{-1} \begin{pmatrix} \Omega_{22} & \Omega_{24} \\ \Omega_{42} & \Omega_{44} \end{pmatrix} \right),$$

  where we used the model's restriction $\pi_{2,n} = \Pi_{2,n}\gamma_n$.
- Ω's can be treated as known.
- $\Pi_{2,n}$ is the only remaining unknown nuisance parameter.

Asymptotic experiment

- An equivalent asymptotic specification:

$$\begin{bmatrix} S_{n,\gamma_0} \\ T_{n,\gamma_0} \end{bmatrix} \overset{a}{\sim}$$

$$N\left(\begin{bmatrix} \delta_n \Pi_{2,n} \\ (I_k - \delta_n \Lambda_{\gamma_0+\delta_n})\Pi_{2,n} \end{bmatrix}, n^{-1} \begin{pmatrix} \Sigma_{22,\gamma_0+\delta_n} & 0 \\ 0 & \Sigma_{44,\gamma_0+\delta_n} \end{pmatrix} \right),$$

where

$$\begin{aligned} S_{n,\gamma} &\equiv \hat{\pi}_{2,n} - \hat{\Pi}_{2,n}\gamma, \\ T_{n,\gamma} &\equiv \hat{\Pi}_{2,n} - \Lambda_\gamma(\hat{\pi}_{2,n} - \hat{\Pi}_{2,n}\gamma), \\ \Sigma_{22,\gamma} &\equiv \Omega_{22} - \gamma\Omega_{24} - \gamma\Omega_{42} + \gamma^2\Omega_{44}, \\ \Sigma_{24,\gamma} &\equiv \Omega_{24} - \gamma\Omega_{44}, \\ \Sigma_{44,\gamma_n} &\equiv \Omega_{44} - \Sigma_{42,\gamma}\Sigma_{22,\gamma}^{-1}\Sigma_{24,\gamma}, \\ \Lambda_\gamma &\equiv \Sigma_{42,\gamma}\Sigma_{22,\gamma}^{-1}. \end{aligned}$$

Asymptotic experiment: Strong IVs

- When IVs are strong, the asymptotic experiment is

$$\begin{bmatrix} \sqrt{n} S_{n,\gamma_0} \\ \sqrt{n}\,(T_{n,\gamma_0} - \Pi_2) \end{bmatrix} \overset{a}{\sim} N\left(\begin{bmatrix} \delta \Pi_2 \\ -\delta \Lambda_{\gamma_0} \Pi_2 \end{bmatrix}, \begin{pmatrix} \Sigma_{22,\gamma_0} & 0 \\ 0 & \Sigma_{44,\gamma_0} \end{pmatrix} \right).$$

- $H_0 : \delta = 0$ vs. $H_1 : \delta \neq 0$.
- Π_2 is an unknown nuisance parameter, however: $T_{n,\gamma_0} \to_p \Pi_2$ under H_0 or H_1.
- Optimal infeasible test: By the Neyman-Pearson lemma, when Π_2 is known, the Asymptotically Uniformly Most Powerful (AUMP) unbiased test is to reject H_0 when $\left| LR^*_{n,\gamma_0} \right| > c$, where

$$LR^*_{n,\gamma_0} = \Pi_2' \Sigma_{22,\gamma_0}^{-1} \sqrt{n} S_{n,\gamma_0} - \Pi_2' \Lambda_{\gamma_0}' \Sigma_{44,\gamma_0}^{-1} \sqrt{n} (T_{n,\gamma_0} - \Pi_2).$$

Asymptotic experiment: Power under strong IVs

- Non-central χ_1^2 distribution:

$$LR_{n,\gamma_0}^{*2}/Var\left(LR_{n,\gamma_0}^{*}\right) \overset{a}{\sim} \chi_1^2\left(\delta^2\left(\Pi_2'\Sigma_{22,\gamma_0}^{-1}\Pi_2 + \Pi_2'\Lambda_{\gamma_0}'\Sigma_{44,\gamma_0}^{-1}\Lambda_{\gamma_0}\Pi_2\right)\right).$$

- Upper bound for the non-centrality parameter:

$$\delta^2\left(\Pi_2'\Sigma_{22,\gamma_0}^{-1}\Pi_2 + \Pi_2'\Lambda_{\gamma_0}'\Sigma_{44,\gamma_0}^{-1}\Lambda_{\gamma_0}\Pi_2\right),$$

  - From the infeasible part $(\sqrt{n}(T_{n,\gamma_0} - \Pi_2))$: $\delta^2\Pi_2'\Lambda_{\gamma_0}'\Sigma_{44,\gamma_0}^{-1}\Lambda_{\gamma_0}\Pi_2$.
  - From the feasible part $(\sqrt{n}S_{n,\gamma_0})$: $\delta^2\Pi_2'\Sigma_{22,\gamma_0}^{-1}\Pi_2$ - **The effective upper bound.**

Efficient feasible statistic under strong IVs

- LR^*_{n,γ_0} requires the knowledge of Π_2 for $\sqrt{n}(T_{n,\gamma_0} - \Pi_2)$. One can relax this following the approach of Choi, Hall, and Schick (1996).
- Consider a local perturbation $\Pi_{2,n} = \Pi_2 + \tau/\sqrt{n}$ and a test of $H_0 : \delta = 0$ and $\tau = 0$.
- The power of the optimal test is described by the non-centrality parameter which depends on τ:

$$\delta^2 \Pi_2' \Sigma_{22,\gamma_0}^{-1} \Pi_2 + (\tau - \delta\Lambda\Pi_2)' \Sigma_{44,\gamma_0}^{-1} (\tau - \delta\Lambda\Pi_2).$$

- The least favorable direction (power minimizing) is $\tau^* = \delta\Lambda_{\gamma_0}\Pi_2$.

Efficient feasible statistic under strong IVs

- The effective power bound is achieved with Kleibergen's LM statistic (Kleibergen (2002); Magnusson (2010)):

$$K_{n,\gamma_0} = \frac{n\left(T'_{n,\gamma_0}\Sigma^{-1}_{22,\gamma_0}S_{n,\gamma_0}\right)^2}{T'_{n,\gamma_0}\Sigma^{-1}_{22,\gamma_0}T_{n,\gamma_0}} \overset{a}{\sim} \chi^2_1\left(\delta^2\Pi'_2\Sigma^{-1}_{22}\Pi_2\right).$$

- The effective power bound is also achieved by Moreira's CLR statistic (Moreira (2003); Andrews, Moreira, Stock (2006); Magnusson (2010)):

$$\begin{aligned}\bar{S}_{n,\gamma_0} &= \Sigma^{-1/2}_{22,\gamma_0}S_{n,\gamma_0},\\ \bar{T}_{n,\gamma_0} &= \Sigma^{-1/2}_{44,\gamma_0}T_{n,\gamma_0},\\ CLR_{n,\gamma_0} &= n\bar{S}'_{n,\gamma_0}\bar{S}_{n,\gamma_0} - n\lambda_{\min}\begin{pmatrix}\bar{S}'_{n,\gamma_0}\bar{S}_{n,\gamma_0} & \bar{S}'_{n,\gamma_0}\bar{T}_{n,\gamma_0}\\ \bar{T}'_{n,\gamma_0}\bar{S}_{n,\gamma_0} & \bar{T}'_{n,\gamma_0}\bar{T}_{n,\gamma_0}\end{pmatrix},\\ &\overset{a}{\sim} \chi^2_1\left(\delta^2\Pi'_2\Sigma^{-1}_{22}\Pi_2\right).\end{aligned}$$

Optimality when $m > 1$ (Strong IVs)

- $m > 1$ (multiple endogenous regressors) and $k > m$ (over-identification).
- For $A = k \times m$ (fixed), consider linear (based on $A' S_{n,\gamma_0}$) statistics:

$$LS_{n,\gamma_0}(A) = n S'_{n,\gamma_0} A \left(A' \Sigma_{22,\gamma_0} A \right)^{-1} A' S_{n,\gamma_0}.$$

- The non-centrality parameter for $LS_{n,\gamma_0}(A)$ (for $\delta = \sqrt{n}(\gamma_n - \gamma_0) \in \mathbb{R}^m$):

$$\delta' \Pi_2' A \left(A' \Sigma_{22,\gamma_0} A \right)^{-1} A' \Pi_2 \delta.$$

- The upper bound for the non-centrality parameter:

$$\begin{aligned} \Pi_2' \Sigma_{22,\gamma_0}^{-1} \Pi_2 - \Pi_2' A \left(A' \Sigma_{22,\gamma_0} A \right)^{-1} A' \Pi_2 &\geq 0, \\ A^* &= \Pi_2' \Sigma_{22,\gamma_0}^{-1}. \end{aligned}$$

Optimality properties when $m > 1$ (Strong IVs)

- Kleibergen's statistic

$$K_{n,\gamma_0} = nS'_{n,\gamma_0}\left(\Sigma_{22,\gamma_0}^{-1} T_{n,\gamma_0}\right)\left(T'_{n,\gamma_0}\Sigma_{22,\gamma_0}^{-1} T_{n,\gamma_0}\right)^{-1}\left(T'_{n,\gamma_0}\Sigma_{22,\gamma_0}^{-1}\right) S_{n,\gamma_0}$$

  - is linear since T_{n,γ_0} is asymptotically independent from S_{n,γ_0},
  - attains the optimal non-centrality parameter $\delta'\Pi_2'\Sigma_{22,\gamma_0}^{-1}\Pi_2\delta$.

Asymptotic experiment: Weak IVs

- The asymptotic experiment:

$$\begin{bmatrix} \sqrt{n}S_{n,\gamma_0} \\ \sqrt{n}T_{n,\gamma_0} \end{bmatrix} \overset{a}{\sim} N\left(\begin{bmatrix} \delta\Pi_2 \\ (I_k - \delta\Lambda_{\gamma_0+\delta})\,\Pi_2 \end{bmatrix}, \begin{pmatrix} \Sigma_{22,\gamma_0+\delta} & 0 \\ 0 & \Sigma_{44,\gamma_0+\delta} \end{pmatrix}\right)$$

  Note that δ now appears in the variance expression.

- The most powerful test is δ-specific and infeasible since Π_2 is unknown and cannot be estimated consistently:

$$LR_{n,\gamma_0}(\delta) = 2\delta\Pi_2'\Sigma_{22,\gamma_0}^{-1}\sqrt{n}S_{n,\gamma_0} - \left(\sqrt{n}S_{n,\gamma_0} - \delta\Pi_2\right)'\left(\Sigma_{22,\gamma_0+\delta}^{-1} - \Sigma_{22,\gamma_0}^{-1}\right)\left(\sqrt{n}S_{n,\gamma_0} - \delta\Pi_2\right).$$

Asymptotic experiment: Weak IVs, K-statistic

- The non-centrality parameter for K_{n,γ_0}:

$$\delta^2 \frac{\left(\left(\mathcal{N}_{\gamma_0+\delta} + \left(I_k - \delta \Lambda_{\gamma_0+\delta} \right) \Pi_2 \right)' \Sigma_{22,\gamma_0+\delta}^{-1} \Pi_2 \right)^2}{\left(\mathcal{N}_{\gamma_0+\delta} + \left(I_k - \delta \Lambda_{\gamma_0+\delta} \right) \Pi_2 \right) \Sigma_{22,\gamma_0+\delta}^{-1} \left(\mathcal{N}_{\gamma_0+\delta} + \left(I_k - \delta \Lambda_{\gamma_0+\delta} \right) \Pi_2 \right)},$$

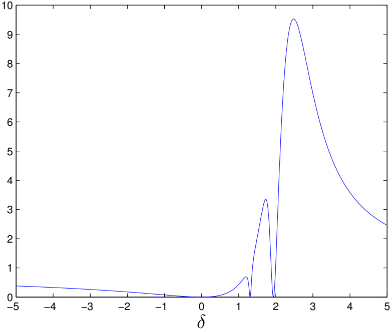
  where $\mathcal{N}_\gamma \overset{a}{\sim} N\left(0, \Sigma_{44,\gamma}\right)$.

- Only trivial power (no power) for

$$\delta = \left((\mathcal{N}_{\gamma_0+\delta} + \Pi_2)' \Sigma_{22,\gamma_0+\delta}^{-1} \Pi_2 \right) / \left(\Pi_2' \Sigma_{22,\gamma_0+\delta}^{-1} \Lambda_{\gamma_0+\delta} \Pi_2 \right).$$

Asymptotic experiment: Weak IVs, K-statistic

- The Non-centrality parameter for K_{n,γ_0} with Weak IVs conditional on $\mathcal{N}_{\gamma_0+\delta} = 0$.



Conditional LM test

- K-test: the reason for loss of power under weak IVs is the transformation that makes $\sqrt{n}S_{n,\gamma_0}$ independent from the estimator of Π_2 (so that χ^2 critical values can be used).
- Instead, consider the following statistic:

$$CLM_{n,\gamma_0} = \frac{n(\hat{\Pi}_{2,n}' \Sigma_{22,\gamma_0}^{-1} S_{n,\gamma_0})^2}{\hat{\Pi}_{2,n}' \Sigma_{22,\gamma_0}^{-1} \hat{\Pi}_{2,n}}.$$

- Critical values can be generated conditionally on T_{n,γ_0} by using the simulated quantiles of

$$\frac{((T_{n,\gamma_0} + \Lambda_{\gamma_0} S_n^*)' \Sigma_{22,\gamma_0}^{-1} \sqrt{n} S_n^*)^2}{(T_{n,\gamma_0} + \Lambda_{\gamma_0} S_n^*)' \Sigma_{22,\gamma_0}^{-1} (T_{n,\gamma_0} + \Lambda_{\gamma_0} S_n^*)},$$

where $\sqrt{n}S_n^* \sim N(0, \Sigma_{22,\gamma_0})$ and independent of T_{n,γ_0}.

  - This is similar to critical values simulations for the CLR statistic (Moreira, 2003).

Conditional LM test

- **Under strong IVs**: $\hat{\Pi}_{2,n} \to_p \Pi_2$, and CLM_{n,γ_0} attains the efficiency bound:

$$CLM_{n,\gamma_0} = \frac{n(\hat{\Pi}_{2,n}' \Sigma_{22,\gamma_0}^{-1} S_{n,\gamma_0})^2}{\hat{\Pi}_{2,n}' \Sigma_{22,\gamma_0}^{-1} \hat{\Pi}_{2,n}} \to_d \chi_1^2 \left(\delta^2 \Pi_2' \Sigma_{22}^{-1} \Pi_2 \right).$$

- The critical values converge to χ_1^2 critical values: $T_{n,\gamma_0} \to_p \Pi_2$, $S_n^* \to_p 0$,

$$\frac{((T_{n,\gamma_0} + \Lambda_{\gamma_0} S_n^*)' \Sigma_{22,\gamma_0}^{-1} \sqrt{n} S_n^*)^2}{(T_{n,\gamma_0} + \Lambda_{\gamma_0} S_n^*)' \Sigma_{22,\gamma_0}^{-1} (T_{n,\gamma_0} + \Lambda_{\gamma_0} S_n^*)} \to_d \chi_1^2(0).$$

- **Under weak IVs:** Nonstandard, nonpivotal distribution.

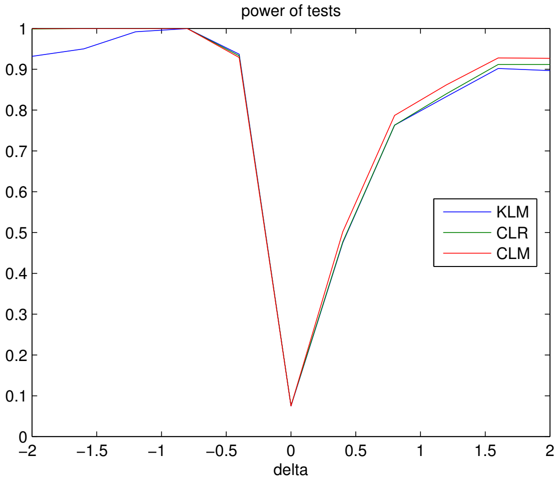
Power comparisons

- Model (censored regression): $m = 1$ and $k = 2$.

$$\begin{aligned} Y_{1i} &= \max\{\beta_0 + (\gamma_0 + \delta)Y_{2i} + U_i, 0\}, \\ Y_{2i} &= \Pi_2' Z_i + V_i, \\ \begin{pmatrix} U_i \\ V_i \end{pmatrix} &\sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \\ Z_i &\sim N(0, I_2). \end{aligned}$$

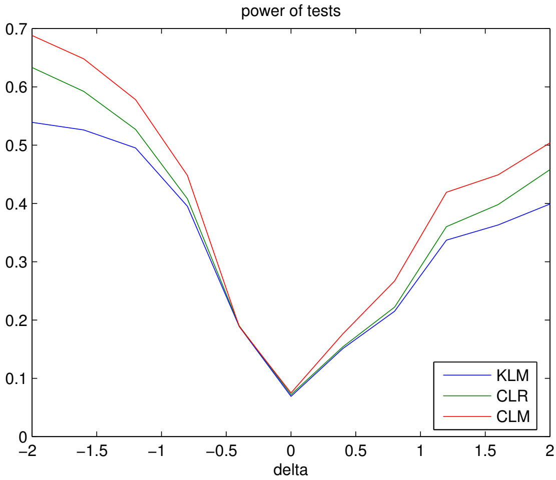
- Degree of endogeneity: ρ
- Concentration parameter (strength of IVs): $\lambda = n\Pi_2'\Pi_2$.
- Reduced-form parameters estimation: quantile regression of Y_{1i} against Z_i and a constant for quantile value $\tau = 0.75$.
- First-stage estimation: OLS.
- We perform CLM, CLR, and K tests of $H_0 : \delta = 0$ for different values of ρ, λ, and δ.

Power comparisons: Strong IVs, high degree of endogeneity



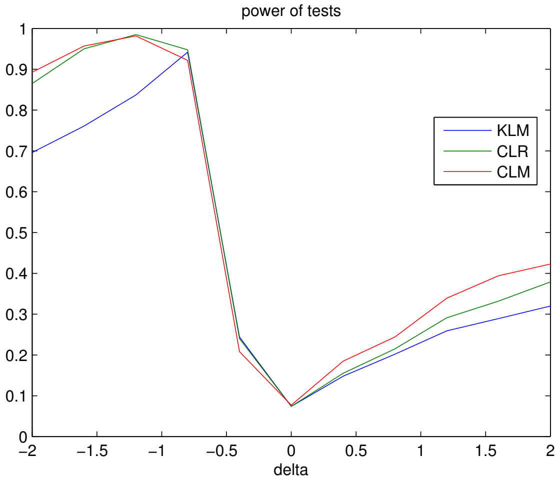
- $\lambda = 22.5,\ \rho = 0.9.$

Power comparisons: Weak IVs, low degree of endogeneity



- $\lambda = 4.9$, $\rho = 0.4$.

Power comparisons: Weak IVs, high degree of endogeneity



- $\lambda = 4.9,\ \rho = 0.9.$

Efficient robust inference in the MDE framework

- $h(\pi, \gamma) = 0$ with an estimator for reduced-form parameters: $\sqrt{n}(\hat{\pi}_n - \pi) \to_d N(0, \Omega)$.
- Consider $H_0 : \gamma = \gamma_0$. We want a test that:
  - Valid if identification is weak.
  - Efficient when identification is strong.
- Assume strong identification and local alternatives $\gamma = \gamma_0 + \delta/\sqrt{n}$.

Efficient robust inference in the MDE framework

- $\sqrt{n}S_{n,\gamma_0} = \sqrt{n}h\left(\hat{\pi}_n, \gamma_0\right) \overset{a}{\sim} N\left(\frac{\partial h(\pi,\gamma_0)}{\partial \gamma'}\delta, \Sigma_{22,\gamma_0}\right)$, where $\Sigma_{22,\gamma_0} = \frac{\partial h(\pi,\gamma_0)}{\partial \pi'}\Omega\frac{\partial h(\pi,\gamma_0)'}{\partial \pi}$.
- The power of an AUMP is described by non-central χ^2_m distribution with non-centrality parameter

$$\delta' \frac{\partial h(\pi, \gamma_0)}{\partial \gamma'} \Sigma_{22,\gamma_0}^{-1} \frac{\partial h(\pi, \gamma_0)'}{\partial \gamma} \delta.$$

- CLM-statistic:

$$CLM_{n,\gamma_0} = n \left\| \left(\frac{\partial h(\hat{\pi}_n, \gamma_0)'}{\partial \gamma} \hat{\Sigma}_{n,22,\gamma_0}^{-1} \frac{\partial h(\hat{\pi}_n, \gamma_0)}{\partial \gamma'} \right)^{-1/2} \times \frac{\partial h(\hat{\pi}_n, \gamma_0)'}{\partial \gamma} \hat{\Sigma}_{n,22,\gamma_0}^{-1} h\left(\hat{\pi}_n, \gamma_0\right) \right\|^2.$$

Efficient robust inference in the MDE framework

- The critical values for CLM must be simulated.
- Let $T_{n,\gamma_0} = \text{vec}\left(\frac{\partial h(\hat{\pi}_n,\gamma_0)}{\partial \gamma'}\right) - \hat{\Sigma}_{n,42,\gamma_0}\hat{\Sigma}_{n,22,\gamma_0}^{-1} h\left(\hat{\pi}_n, \gamma_0\right)$, where
  - $\Sigma_{44,\gamma} = \frac{\partial}{\partial \pi'} \text{vec}\left(\frac{\partial h(\pi,\gamma)}{\partial \gamma'}\right) \Omega \frac{\partial}{\partial \pi} \text{vec}\left(\frac{\partial h(\pi,\gamma)}{\partial \gamma'}\right)'$.
  - $\Sigma_{42,\gamma} = \frac{\partial}{\partial \pi'} \text{vec}\left(\frac{\partial h(\pi,\gamma)}{\partial \gamma'}\right) \Omega \frac{\partial h(\pi,\gamma)'}{\partial \pi}$.
- Generate $S_{n,\gamma_0}^{*} \sim N(0, \frac{1}{n}\hat{\Sigma}_{n,22,\gamma_0}^{-1})$, and let Π_{n,γ_0}^{*} be such that $\text{vec}(\Pi_{n,\gamma_0}^{*}) = T_{n,\gamma_0} + \hat{\Sigma}_{n,42,\gamma_0}\hat{\Sigma}_{n,22,\gamma_0}^{-1} S_{n,\gamma_0}^{*}$.
- The null distribution of the CLM-statistic is simulated as

$$CLM_{n,\gamma_0}^{*} = n\left\|\left(\Pi_{n,\gamma_0}^{*\prime}\hat{\Sigma}_{n,22,\gamma_0}^{-1}\Pi_{n,\gamma_0}^{*}\right)^{-1/2}\Pi_{n,\gamma_0}^{*\prime}\hat{\Sigma}_{n,22,\gamma_0}^{-1}S_{n,\gamma_0}^{*}\right\|^2.$$

Classical MDE

- Classical MDE: reduced-form parameters (π) and structural parameters (γ) are additively separable:

$$h(\pi, \gamma) = \pi - f(\gamma).$$

- Examples: (i) DSGE models, where models' predictions ($f(\gamma)$) are matched to empirical moments, impulse responses, and etc (π); (ii) Horowitz's transformation model with endo regressors.
- In this case, CLM statistics can be used with χ^2 critical values:

$$CLM_{n,\gamma_0} = n \left\| \left(\frac{\partial f(\gamma_0)'}{\partial \gamma} \hat{\Sigma}^{-1}_{n,22,\gamma_0} \frac{\partial f(\gamma)}{\partial \gamma'} \right)^{-1/2} \times \frac{\partial f(\gamma_0)'}{\partial \gamma} \hat{\Sigma}^{-1}_{n,22,\gamma_0} h\left(\hat{\pi}_n, \gamma_0\right) \right\|^2 .$$

- K-statistic reduces to CLM since $\partial f(\gamma_0)/\partial \gamma'$ is independent from $h(\hat{\pi}_n, \gamma_0)$.

Conclusion

- CLR, K, and CLM are asymptotically efficient (and equivalent) when identification is strong.
- Their behavior differ when identification is weak.
- In general nonlinear models, CLM can dominate CLR and K depending on a model's structure.
- Ongoing work: Understanding the conditions under which this occurs; Comparison of CLM, K, and CLR to the power envelope under weak IVs.