Jackknife, small bandwidth and high-dimensional asymptotics

Yukitoshi Matsushita & Taisuke Otsu

Hitotsubashi University & London School of Economics

May 2018

This paper

- Focus: Analyze **non-standard asymptotic problems** from jackknife perspective, especially
    - **Small bandwidths asymptotics** for semiparametric estimator
    - **Many-weak IV asymptotics** for IV regression
    - **Many covariates asymptotics** for high-dimensional regression
    - **Network asymptotics**, **Infill asymptotics** (hopefully)
- Propose (modified) jackknife empirical likelihood (JEL) method for those problems

1. Background: JEL for semiparametric models under standard asymptotics

(but also new)

Background: Jackknife method

- Useful to estimate bias and variance of estimator
- For statistic S, **Jackknife pseudo-value** is

$$V_i = nS - (n-1)S^{(-i)}$$

  where $S^{(-i)}$ is leave-i-out version of S

- Jackknife bias-corrected estimator is $\bar{V}$ (if S is estimator)
- Jackknife variance estimator of S is

$$\widehat{Var(S)} = \frac{1}{n(n-1)} \sum_{i=1}^{n} (V_i - \bar{V})^2$$

- Idea: **Treat** $\{V_i\}$ **like sample** (Tukey, 1958)

Contribution to literature

- Quenouille (1949), Tukey (1958): Introduce jackknife
- Miller (1964), Arvesen (1969): Consistency of $\widehat{Var(S)}$ for smooth objects
- Shao & Wu (1989): Inconsistency of $\widehat{Var(S)}$ for non-smooth & delete-d jackknife
- Jing, Yuan & Zhou (2009): Introduce JEL method to U-statistic (e.g. ROC curve, $S = n^{-2} \sum_{i,j} \mathbb{I}\{Y_j > X_i\}$)
- **This paper:** Extend JEL to semiparametric problems and provide unified framework for non-standard asymptotics, e.g.
    - Small bandwidth asymptotics
    - Many/weak IV asymptotics
    - Many covariates asymptotics

Semiparametric model

- As a benchmark, first consider inference on semiparametric model under standard asymptotics (but this is also new)
- Moment condition model

$$E[g(Z, \theta, \mu(X))] = 0$$

where

  - θ is parameter of interest
  - $\mu(X)$ is unknown nuisance function (e.g. $\mu(X) = E[Y|X]$)

Example 1: Average treatment effect

- $Y(0), Y(1)$ are potential outcomes for treatment $D = 0$ or 1
- Observe $Z = (Y, X, D)$ where $Y = DY(1) + (1 - D)Y(0)$ and X are covariates
- Under some conditions, ATE is

$$\begin{aligned} \theta &= E[Y(1) - Y(0)] \\ &= E[\mu_1(X) - \mu_0(X)] \end{aligned}$$

where $\mu_d(X) = E[Y|X, D = d]$

- Set

$$g(Z, \theta, \mu(X)) = \mu_1(X) - \mu_0(X) - \theta$$

Example 2: Weighted average derivatives

- Weighted average derivative of regression function $m(X) = E[Y|X]$ is

$$\theta = E\left[w(X)\frac{\partial m(X)}{\partial X}\right]$$

where $w(\cdot)$ is known weight function

- Used for estimation of single index model (e.g. $P(Y=1|X) = F(X'\theta)$)

- Set

$$g(Z,\theta,\mu(X)) = w(X)\mu(X) - \theta$$

with $\mu(X) = \frac{\partial m(X)}{\partial X}$

Semiparametric estimator

- Based on some preliminary estimator $\hat{\mu}$ for μ, θ can be estimated by solving

$$\frac{1}{n}\sum_{i=1}^{n} g(Z_i, \hat{\theta}, \hat{\mu}(X_i)) = 0$$

- By Newey (1994), influence function for $\hat{\theta}$ is

$$\psi(Z, X) = -M_\theta^{-1}\{g(Z, \theta, \mu(X)) + E[g_\mu(Z, \theta, \mu(X))|X]\{Y - \mu(X)\}$$

where $M_\theta = E\left[\frac{\partial g(Z,\theta,\mu(X))}{\partial\theta}\right]$

- Computation of asymptotic variance $Var(\psi(Z, X))$ is complicated

Problems with Wald-type inference (estimate$\pm 2 \cdot$se)

- Need to derive influence function for each case
- Asymptotic variances of these estimators are rather complicated and contain several unknowns to be estimated
- Also if θ or μ enters to $g(\cdot)$ in nonlinear way, then it may involve numerical derivatives
- Normal approximation often does not work well even when asymptotic variance is known: see Linton (1995), Nishiyama & Robinson (2000, 2005), Ichimura & Linton (2005), Cattaneo, Crump & Jansson (2013, 2014) etc

Construct JEL

- Owen's (1988) original EL for $\theta = E[X]$

$$\ell(\theta) = -2 \sup_{p_1,\ldots,p_n} \sum_{i=1}^{n} \log(np_i)$$

$$\text{s.t. } p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i X_i = \theta$$

i.e. put multinomial weights $\{p_i\}_{i=1}^n$ for $\{X_i\}_{i=1}^n$

- By Lagrange multiplier method, practical dual form of $\ell(\theta)$ is obtained

$$\ell(\theta) = 2 \sup_{\lambda} \sum_{i=1}^{n} \log\left(1 + \lambda'(X_i - \theta)\right) \xrightarrow{d} \chi^2$$

- It is not trivial to extend this idea to parameter $\theta = E[a(X_1, \ldots, X_m)]$ defined by U-statistic

$$U = \binom{n}{m}^{-1} \sum_{i_1 < \cdots < i_m} a(X_{i_1}, \ldots, X_{i_m})$$

where restriction by multinomial weights would be

$$\binom{n}{m}^{-1} \sum_{i_1 < \cdots < i_m} p_{i_1} \cdots p_{i_m} a(X_{i_1}, \ldots, X_{i_m}) = \theta$$

- For this problem, Jing-Yuan-Zhou (2009, JASA) suggested to construct EL based on jackknife pseudo-values

$$V_i = nU - (n-1)U^{(-i)}$$

i.e.

$$\ell(\theta) = -2 \sup_{p_1,\ldots,p_n} \sum_{i=1}^{n} \log(np_i)$$

$$\text{s.t. } p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i V_i = \theta$$

- They showed $\ell(\theta) \xrightarrow{d} \chi^2$ for one and two sample U-statistics with fixed kernel

- Indeed many estimators can be (approximately) written as U-statistic form (or ANOVA-like decomposition)
- We extend JEL approach to
  - Semiparametric estimator (under standard asymptotics)
  - Weighted average derivative estimator (under small bandwidth asymptotics)
  - Jackknife IV estimator (under many-weak IV asymptotics)
  - OLS estimator (under many regressor asymptotics)

JEL for semiparametric model

- Jackknife pseudo-value is

$$V_i = nS - (n-1)S^{(-i)}$$

- For fixed θ, we set

$$\begin{aligned} S(\theta) &= \frac{1}{n}\sum_{j=1}^{n} g(Z_j, \theta, \hat{\mu}(X_j)) \\ S^{(-i)}(\theta) &= \frac{1}{n-1}\sum_{j \neq i} g(Z_j, \theta, \hat{\mu}^{(-i)}(X_j)) \end{aligned}$$

where $\hat{\mu}^{(-i)}$ is leave-i-out version of $\hat{\mu}$

- Based on pseudo-value

$$V_i(\theta) = nS(\theta) - (n-1)S^{(-i)}(\theta)$$

JEL is constructed as

$$\ell(\theta) = -2 \sup_{p_1,\ldots,p_n} \sum_{i=1}^{n} \log(np_i)$$

$$\text{s.t. } p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i V_i(\theta) = 0$$

- By Lagrange multiplier method, dual form of $\ell(\theta)$ is

$$\ell(\theta) = 2 \sup_{\lambda} \sum_{i=1}^{n} \log\left(1 + \lambda' V_i(\theta)\right)$$

(in practice, use this expression)

Theorem: Standard asymptotics

- Under Assumption SP

$$\ell(\theta) \xrightarrow{d} \chi^2(\dim \theta)$$

- Remarks
  - JEL $\ell(\theta)$ is asymptotically pivotal, so no need for variance estimation
  - No need to derive influence function for case-by-case
  - JEL confidence set is obtained by $\{c : \ell(c) \leq \chi^2_\alpha(\dim \theta)\}$

Assumption SP

- (i) $\{Y_i, X_i, Z_i\}_{i=1}^n$ is iid. X is compactly supported and its density f is uniformly bounded from above and away from zero. μ and f are continuously differentiable to order s. $E[|Y - \mu(X)|^{2+\delta}] < \infty$ for some $\delta > 0$, $E[Y^p] < \infty$ for some $p \geq 4$, and $E[Y^p|X = x]f(x)$ is bounded. g has bounded second derivative in μ
- (ii) K is an s-th order kernel function that integrates to 1 in its compact support. As $n \to \infty$, $n^{1/2} h^{\dim X} / \log n \to \infty$ and $nh^{2s} \to 0$

2. JEL under small bandwidth asymptotics

Small bandwidth asymptotics

- Consider density-weighted average derivative

$$\theta = E\left[f(X)\frac{\partial E[Y|X]}{\partial X}\right] = -2E\left[Y\frac{\partial f(X)}{\partial X}\right]$$

"$=$" is by integration-by-parts

- Powell, Stock & Stoker (1989) estimator is

$$\hat{\theta} = -\frac{2}{n}\sum_{j=1}^{n} Y_j \frac{\partial \hat{f}(X_j)}{\partial X}$$

where $\hat{f}$ is leave-one-out kernel density estimator

$$\hat{f}(X_j) = \frac{1}{n-1}\sum_{k\neq j} h^{-d} K\left(\frac{X_j - X_k}{h}\right)$$

for kernel K, bandwidth h and $d = \dim X$

- $\hat{\theta}$ admits U-statistic representation ($\dot{K}$ =derivative of K)

$$\hat{\theta} = \frac{2}{n(n-1)} \sum_{j<k} U_{jk}$$
$$U_{jk} = -h^{-d-1} \dot{K}\left(\frac{X_j - X_k}{h}\right)(Y_j - Y_k)$$

- This is decomposed as

$$\sqrt{n}(\hat{\theta} - \theta) = \underbrace{\sqrt{n}B}_{O(\sqrt{n}h^s)} + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} L_j + \underbrace{\frac{2}{\sqrt{n}(n-1)} \sum_{j<k} W_{jk}}_{O_p\left(1/\sqrt{nh^{d+2}}\right)}$$

where $B = E[\hat{\theta}] - \theta$ (bias), $L_j = 2(E[U_{jk}|Z_j] - E[U_{jk}])$ and $W_{jk} = U_{jk} - (L_j + L_k)/2 - E[U_{jk}]$

Small bandwidth asymptotics for $\hat{\theta}$

- By Cattaneo, Crump & Jansson (2014), under Assumption SB
- **Standard asymptotics:** If $nh^{d+2} \to \infty$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma)$$

  where $\Sigma = E[LL']$

- **Small** h **asymptotics:** If $nh^{d+2} \to \kappa \in (0, \infty)$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \Sigma + \frac{2}{\kappa}\Delta\right)$$

  where $\Delta = 2E[Var(Y|X)f(X)] \int \dot{K}(u)\dot{K}(u)'du$

- **(Very) Small** h **asymptotics:** If $nh^{d+2} \to 0$

$$\sqrt{\binom{n}{2} h^{d+2}}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Delta)$$

Assumption SB

- (i) f is $s+1$ times differentiable, and f and its first $s+1$ derivatives are bounded for some $s \geq 2$. m is twice differentiable, $e = mf$ has the bounded second derivative, $v(x) = E[Y^2|X = x]$ is differentiable, vf has the bounded first derivative, and $\lim_{|x|\to\infty}\{m(x) + |e(x)|\} = 0$. $E[Y^4] < \infty$, $E[Var(Y|X)f(X)] > 0$, and $Var\left(\frac{\partial e(X)}{\partial X} - Y\frac{\partial f(X)}{\partial X}\right)$ is positive definite
- (ii) Same assumptions on K as Cattaneo, Crump and Jansson (2014). $\min\{nh^{d+2}, 1\}nh^{2s} \to 0$ and $n^2h^d \to \infty$

Remark

- Asymptotic distribution of $\hat{\theta}$ is different for each case
- CCJ suggested to apply Wald test based on estimated variance $\hat{V}_{SB} = \hat{\Sigma} + \frac{2}{\kappa}\hat{\Delta}$
- Estimation of $\hat{V}_{SB}$ may involve separate bandwidth selection from that of $\hat{\theta}$ (CCJ argued that same bandwidth for $\hat{\theta}$ and $\hat{V}_{SB}$ may fail to yield positive definite $\hat{V}_{SB}$)

JEL under small bandwidth

- Let's first apply directly JEL to $\hat{\theta}$. Again jackknife pseudo-value is

$$V_i(\theta) = nS(\theta) - (n-1)S^{(-i)}(\theta)$$

- For fixed θ, we set

$$\begin{aligned} S(\theta) &= \hat{\theta} - \theta \\ S^{(-i)}(\theta) &= \hat{\theta}^{(-i)} - \theta \end{aligned}$$

where $\hat{\theta}^{(-i)}$ is leave-i-out version of $\hat{\theta}$

- JEL is defined in the same manner

$$\ell(\theta) = 2 \sup_{\lambda} \sum_{i=1}^{n} \log\left(1 + \lambda' V_i(\theta)\right)$$

JEL under small bandwidth asymptotics

- **Standard asymptotics:** If $nh^{d+2} \to \infty$

$$\ell(\theta) \xrightarrow{d} \chi^2(d)$$

- **Small h asymptotics:** If $nh^{d+2} \to \kappa \in (0, \infty)$

$$\ell(\theta) \xrightarrow{d} Z' V_J^{-1} Z$$

where $Z \sim N(0, \Sigma + 2\kappa^{-1}\Delta)$ and $V_J = \Sigma + 4\kappa^{-1}\Delta$

- **(Very) Small h asymptotics:** If $nh^{d+2} \to 0$

$$\ell(\theta) \xrightarrow{d} \frac{1}{2}\chi^2(d)$$

Remark

- Under small h asymptotics, JEL is not asymptotically pivotal
- Under very small h asymptotics, JEL recovers asymptotic pivotalness ("$\frac{1}{2}$" appears by setting $\Sigma = 0$ for small h case)
- We modify JEL to have same $\chi^2(d)$ limiting distribution **for all cases**
- Idea: Reconcile discrepancy in $Z \sim N(0, \Sigma + 2\kappa^{-1}\Delta)$ and $V_J = \Sigma + 4\kappa^{-1}\Delta$

Key: Efron-Stein inequality

- Efron & Stein (1981) showed jackknife variance estimate tends to be **upward biased**. Suppose

$$S_n = \theta + \frac{1}{n}\sum_{i=1}^{n} L(X_i) + \frac{1}{n^2}\sum_{i<j} W(X_i, X_j)$$

Efron-Stein inequality showed

$$\begin{aligned} Var(S_n) &= \frac{\sigma_L^2}{n} + \frac{1}{2}\frac{(n-1)\sigma_W^2}{n^3} \\ &< \frac{\sigma_L^2}{n} + \frac{(n-2)\sigma_W^2}{n(n-1)^2} = E[\widehat{Var(S_n)}] \end{aligned}$$

where $\sigma_L^2 = Var(L(X_i))$, $\sigma_W^2 = Var(W(X_i, X_j))$ and $\widehat{Var(S_n)}$ is jackknife variance estimate

- Derivation

$$
\begin{aligned}
E[\widehat{Var(S_n)}] &= \frac{n-1}{n} E\left[\sum_{i=1}^{n} (S^{(-i)} - S^{(\cdot)})^2\right] \\
&= \frac{n-1}{n} E\left[\frac{1}{n} \sum_{i<i'} (S^{(-i)} - S^{(-i')})^2\right] \\
&= \frac{n-1}{n} E\left[\frac{1}{n} \sum_{i<i'} \left\{ \begin{array}{c} \frac{1}{n-1}(L_{i'} - L_i) \\ +\frac{1}{(n-1)^2} \sum_{j \neq i,i'} (W_{i'j} - W_{ij}) \end{array} \right\}^2\right] \\
&= \frac{\sigma_L^2}{n} + \frac{(n-2)\sigma_W^2}{n(n-1)^2}
\end{aligned}
$$

- Jackknife variance estimate **doubles** second term (due to mismatch of characterizing quadratic term in Hoeffding decomposition)
- This bias is lower order than first term $\sigma_L^2/(n-1)$, so is asymptotically negligible
- Efron-Stein suggested bias correction by estimating σ_W^2

$$\hat{\sigma}_W^2 = \frac{(n-1)^2}{n(n-1)/2-1} \sum_{j<k} Q_{jk}^2$$

where

$$Q_{jk} = n\hat{\theta} - (n-1)(\hat{\theta}^{(-j)} + \hat{\theta}^{(-k)}) + (n-2)\hat{\theta}^{(j,k)}$$

and $\hat{\theta}^{(j,k)}$ is leave-(j,k)-out version of $\hat{\theta}$

Our view

- Under non-standard asymptotics, **Efron-Stein bias emerges in first order**. It applies at least
  - Small bandwidth asymptotics
  - Many/weak IV asymptotics
  - Many covariates asymptotics
- We propose **unified inference approach that works for all cases above** (at least)

Modified JEL

- Modified JEL is defined by

$$\ell^m(\theta) = 2 \sup_{\lambda} \sum_{i=1}^{n} \log\left(1 + \lambda V_i^m(\theta)\right)$$

where

$$V_i^m(\theta) = V_i(\hat{\theta}) - \hat{\Gamma}\tilde{\Gamma}^{-1}\{V_i(\hat{\theta}) - V_i(\theta)\}$$

and $\hat{\Gamma}$ and $\tilde{\Gamma}$ are given by

$$\hat{\Gamma}\hat{\Gamma}' = \frac{1}{n}\sum_{i=1}^{n} V_i(\hat{\theta})V_i(\hat{\theta})'$$

$$\tilde{\Gamma}\tilde{\Gamma}' = \frac{1}{n}\sum_{i=1}^{n} V_i(\hat{\theta})V_i(\hat{\theta})' - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} Q_{ij}Q_{ij}'$$

- This modification internalizes Efron-Stein bias correction into JEL construction
- cf. Chandler-Bate (2007, Biometrika) parametric likelihood correction for clustered data
- Intuitively we change curvature of likelihood function but leaving other aspects of its shape unchanged, so that "Bartlett identity" is asymptotically recovered

Theorem: Modified JEL under small bandwidth asymptotics

- Under Assumption SB

$$\ell^m(\theta) \xrightarrow{d} \chi^2(d)$$

without any condition on nh^{d+2}

- Remarks
  - Modified JEL $\ell^m(\theta)$ is asymptotically pivotal for all standard, small h, and very small h asymptotics
  - Correction term $\frac{1}{n+1}\sum_{j<k} Q_{jk}^2$ is used to eliminate Efron-Stein bias

3. JEL under many-weak IV asymptotics

Many-weak IV

- IV regression (θ is scalar)

$$\begin{aligned} y &= X\theta + U \\ X &= Z'\gamma_n + \epsilon \end{aligned}$$

where $\gamma_n = \frac{1}{\sqrt{n}}\mu_n\pi$ and μ_n is scalar sequence (concentration parameter) and π is K-dimensional constant vector

- **Many-weak IV asymptotics:**
  - (i) K is fixed and $\mu_n = O(n^{1/2})$ (standard case)
  - (ii) $K \to \infty$ and $K/\mu_n^2 \to \alpha < \infty$ (many weak case)
  - (iii) $K \to \infty$ and $K/\mu_n^2 \to \infty$ (many very weak case)
- IV estimator $\hat{\theta}$ has different limiting distributions for above cases

JIVE

- Jackknife IV estimator by Angrist-Imbens-Krueger (1999)

$$\hat{\theta} = \left(\sum_{k=1}^{n} \sum_{l \neq k} X_k P_{kl} X_l \right)^{-1} \sum_{k=1}^{n} \sum_{l \neq k} X_k P_{kl} Y_l$$

where $P_{kl} = Z_k' \left(\sum_{h=1}^{n} Z_h Z_h' \right)^{-1} Z_l$

- JIVE is robust to many instruments in contrast to LIML and 2SLS estimators

JIVE under many-weak IV asymptotics

- Chao et al. (2012) showed that under Assumption MW
- **(i) Standard asymptotics**

$$\mu_n(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1})$$

- **(ii) Many weak case**

$$\mu_n(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1} + \alpha H^{-1}\Psi H^{-1})$$

- **(iii) Many very weak case**

$$\frac{\mu_n^2}{\sqrt{K}}(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}\Psi H^{-1})$$

- Chao et al. (2012) suggested robust inference by estimating unknown components H, Σ and Ψ

- where

$$
\begin{aligned}
H &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (1 - P_{kk}) \pi' Z_k Z_k' \pi \\
\Sigma &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \sigma_k^2 (1 - P_{kk})^2 \pi' Z_k Z_k' \pi \\
\Psi &= \lim_{n\to\infty} \frac{1}{K} \sum_{k=1}^{n} \sum_{l\neq k} P_{kl}^2 \{ \sigma_k^2 E[\epsilon_l^2] + E[\epsilon_k U_k] E[\epsilon_l U_l] \}
\end{aligned}
$$

Assumption MW

- (i) There are positive constants C and C_1 such that $\max_{1\leq i\leq n} P_{ii} \leq C < 1$ and $C_1^{-1} \leq \pi' \left(\frac{1}{n}\sum_{i=1}^n Z_i Z_i'\right)\pi \leq C_1$ for all n large enough. Also, $n^{-2}\sum_{i=1}^n |\pi' Z_i|^4 \to 0$ as $n \to \infty$
- (ii) $\{U_i, \epsilon_i\}_{i=1}^n$ are independent with $E[U_i] = 0$ and $E[\epsilon_i] = 0$. Also for some positive constant C_2, minimum eigenvalue of $Var(U_i, \epsilon_i)$ is larger than C_2^{-1} and $\max_{1\leq i\leq n}\{E[U_i^2], E[U_i^4], E[\epsilon_i^2], E[\epsilon_i^4]\} < C_2$
- (iii) Σ, Ψ and Ξ exist. Also $\sqrt{K}/\mu_n^2 \to 0$ as $n \to \infty$

JEL

- Again jackknife pseudo-value is

$$V_i(\theta) = nS(\theta) - (n-1)S^{(-i)}(\theta)$$

- Based on estimating equation for JIVE, we set

$$\begin{aligned} S(\theta) &= \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{l \neq k} X_k P_{kl}(Y_l - X_l \theta) \\ S^{(-i)}(\theta) &= \frac{1}{(n-1)(n-2)} \sum_{k \neq i} \sum_{l \neq i,k} X_k P_{kl}(Y_l - X_l \theta) \end{aligned}$$

- JEL is defined in the same manner

$$\ell(\theta) = 2 \sup_{\lambda} \sum_{i=1}^{n} \log(1 + \lambda' V_i(\theta))$$

JEL under many-weak IV asymptotics

- **(i) Standard asymptotics**

$$\ell(\theta) \xrightarrow{d} \chi^2(1)$$

- **(ii) Many weak case**

$$\ell(\theta) \xrightarrow{d} \frac{\xi^2}{\Sigma + \Xi + 2\alpha\Psi}$$

where $\xi \sim N(0, \Sigma + \alpha\Psi)$ and
$\Xi = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{l\neq k} \sigma_l^2 P_{lk}^2 \pi' Z_k Z_k' \pi$

- **(iii) Many very weak case**

$$\ell(\theta) \xrightarrow{d} \frac{1}{2}\chi^2(1)$$

Remark

- Under many-weak IV asymptotics, JEL is not asymptotically pivotal
- For case (ii), discrepancy of $\xi \sim N(0, \Sigma + \alpha\Psi)$ and $2\alpha\Psi$ in denominator is due to Efron-Stein bias
- Additional term Ξ emerges due to the fact that matrix $[P_{kl}]$ is not exactly projection matrix for $S^{(-i)}(\theta)$
- For case (iii), "$\frac{1}{2}$" appears by setting $\Sigma = \Xi = 0$

Modified JEL under many-weak IV asymptotics

- Under Assumption MW, Modified JEL satisfies

$$\ell^m(\theta) \xrightarrow{d} \chi^2(1)$$

for all cases

- Modified JEL $\ell^m(\theta)$ follows χ^2 limiting distribution for all cases without estimating variance components Σ, Ψ and Ξ

4. JEL under many regressor asymptotics

Many regressor asymptotics

- Regression

$$Y = X\theta + Z'\gamma_n + U$$

where X is scalar and Z is K-dimensional

- Two scenarios
- (i) Standard asymptotics: $\frac{K}{n} \to 0$
- (ii) Many regressor asymptotics: $\frac{K}{n} \to \tau \in (0, 1)$
- Cattaneo-Jansson-Newey (2017) imposed $\tau \in \left(0, \frac{1}{2}\right)$ and derived robust standard error for OLS estimator under Case (ii)

JEL

- Again jackknife pseudo-value is

$$V_i(\theta) = nS(\theta) - (n-1)S^{(-i)}(\theta)$$

- Based on FOC of OLS, we set

$$\begin{aligned} S(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\tilde{X}_k M_{kl}(Y_l - X_l\theta) \\ S^{(-i)}(\theta) &= \frac{1}{n-1}\sum_{k\neq i}\tilde{X}_k M_{kk}(Y_k - X_k\theta) \\ &\quad + \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k}\tilde{X}_k M_{kl}(Y_l - X_l\theta) \end{aligned}$$

where $P_{kl} = Z_k'(\sum_{h=1}^{n} Z_h Z_h')^{-1} Z_l$, $M_{kl} = \mathbb{I}\{k = l\} - P_{kl}$, $\tilde{X}_k = \sum_{l=1}^{n} M_{kl} X_l$

JEL under many regressor asymptotics

- Suppose Assumption MR holds. Then
- **(i) Standard asymptotics**

$$\ell(\theta) \stackrel{d}{\to} \chi^2(1)$$

- **(ii) Many regressor asymptotics**

$$\ell(\theta) \stackrel{d}{\to} \frac{\xi^2}{\Sigma + \Xi + 2\Psi}$$

where $\xi \sim N(0, \Sigma + \Psi)$

- where

$$\begin{aligned}
\Sigma &= p \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} M_{kk}^2 E[\epsilon_k^2 U_k^2 | Z_k] \\
\Psi &= p \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{l\neq k} M_{kl}^2 E[\epsilon_k^2 U_l^2 | Z_k, Z_l] \\
\Xi &= p \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{l\neq k} M_{lk}^2 E[\epsilon_l^2 | Z_l] (Z_k' \gamma)^2
\end{aligned}$$

Assumption MR

- (i) $\{Y_i, X_i, Z_i\}_{i=1}^n$ is independent and identically distributed
- (ii) $\mathrm{rank}(P) = K$ a.s. There are positive constants C, C_1, and C_2 such that $M_{ii} \geq C$ for all $i = 1, \ldots, n$, $E[\epsilon_i^4 | Z_i] \geq C_1$, $E[U_i^4 | X_i, Z_i] \geq C_1$, $E[\epsilon_i^2 | Z_i] \leq C_2$, and $E[U_i^2 | X_i, Z_i] \leq C_2$
- (iii) For some $\alpha_g, \alpha_h > 0$, there is a positive constant C_3 such that

$$\begin{aligned} \min_{\pi_g} E|g(Z_i) - Z_i'\pi_g|^2 &\leq C_3 K^{-2\alpha_g} \\ \min_{\pi_h} E|h(Z_i) - Z_i'\pi_h|^2 &\leq C_3 K^{-2\alpha_h} \end{aligned}$$

where $h(Z_i) = E[X_i | Z_i]$

Remark

- Under many regressor asymptotics, JEL is not asymptotically pivotal
- For case (ii), discrepancy of $\xi \sim N(0, \Sigma + \Psi)$ and 2Ψ in denominator is due to Efron-Stein bias
- Additional term Ξ emerges due to the fact that matrix $[P_{kl}]$ is not exactly projection matrix for $S^{(-i)}(\theta)$

Modified JEL under many-weak IV asymptotics

- Under Assumption MR, Modified JEL satisfies

$$\ell^m(\theta) \xrightarrow{d} \chi^2(1)$$

  for all cases

- Modified JEL $\ell^m(\theta)$ follows χ^2 limiting distribution for all cases without estimating variance components Σ, Ψ and Ξ

- In contrast to Cattaneo-Jansson-Newey (2017), which imposes $\frac{K}{n} \to \tau \in \left(0, \frac{1}{2}\right)$, we only require $\tau \in (0, 1)$. The requirement $\tau < \frac{1}{2}$ is imposed to guarantee consistency of robust standard error (which we circumvent)

5. Simulation

Simulation

- Same design as Cattaneo-Jansson-Newey (2017). Partial linear model

$$\begin{aligned} Y &= \beta X + g(W) + U \\ X &= h(W) + V \end{aligned}$$

$$\begin{aligned} U|X, W &\sim_{\text{iid}} N(0, c_U\{1 + (t(X) + \iota' W)^2\}^{\vartheta}) \\ V|W &\sim_{\text{iid}} N(0, c_V\{1 + (\iota' W)^2\}^{\vartheta}) \\ W &\sim_{\text{iid}} \text{6-dimensional independen } U[-1, 1] \end{aligned}$$

- Set $n = 100$ and

$$\begin{aligned} \beta &= 1 \quad \text{(parameter of interest)} \\ \vartheta &= \begin{cases} 0 & \text{(homoskedastic)} \\ 1 & \text{(heteroskedastic)} \end{cases} \end{aligned}$$

- Also

$$\begin{aligned} g(W) &= \exp(-|W|^{1/2}), \qquad h(W) = \exp(|W|^{1/2}) \\ t(X) &= X\mathbb{I}\{-2 \leq X \leq 2\} + 2\text{sgn}(X)\{1 - \mathbb{I}\{-2 \leq X \leq 2\}\} \\ c_U, c_V &= \text{constants to normalize } Var(U) = Var(V) = 1 \end{aligned}$$

- Basis functions

K	$p_K(w_i)$
7	$1, w_{1i}, w_{2i}, w_{3i}, w_{4i}, w_{5i}, w_{6i}$
13	$p_7(w_i)', w_{1i}^2, w_{2i}^2, w_{3i}^2, w_{4i}^2, w_{5i}^2, w_{6i}^2$
28	$p_{13}(w_i)$ + first-order interactions
34	$p_{28}(w_i), w_{1i}^3, w_{2i}^3, w_{3i}^3, w_{4i}^3, w_{5i}^3, w_{6i}^3$
84	$p_{34}(w_i)$ + second-order interactions
90	$p_{84}(w_i), w_{1i}^4, w_{2i}^4, w_{3i}^4, w_{4i}^4, w_{5i}^4, w_{6i}^4$

- Compare four confidence intervals
- (i) **Wald-HC0:** Wald by conventional Eicker-White se
- (ii) **Wald-CJN:** Wald by Cattaneo-Jansson-Newey
- (iii) **JEL**
- (iv) **mJEL:** Modified JEL
- Note: Wald-CJN does not cover the case of $\frac{K}{n} > \frac{1}{2}$

Homoskedastic case

- $n = 100$, nominal coverage$= 0.95$

K	Wald-HC0	Wald-CJN	JEL	mJEL
7	0.897	0.909	0.920	0.913
13	0.916	0.934	0.950	0.937
28	0.888	0.922	0.962	0.940
34	0.869	0.930	0.962	0.945
84	0.591	**0.816**	0.967	0.947
90	0.513	**0.578**	0.970	0.959

Heteroskedastic case

- $n = 100$, nominal coverage$= 0.95$

K	Wald-HC0	Wald-CJN	JEL	mJEL
7	0.887	0.897	0.911	0.899
13	0.924	0.947	0.954	0.943
28	0.881	0.928	0.961	0.938
34	0.838	0.924	0.956	0.936
84	0.539	**0.796**	0.963	0.949
90	0.508	**0.620**	0.961	0.946

Conclusion

- Likelihood inference for semiparametric problems and non-standard asymptotics
- JEL inference for semiparametric models
- JEL is not pivotal under non-standard asymptotics (Efron-Stein Bias emerges in first order)
- Modified JEL internalizes Efron-Stein correction into JEL and becomes pivotal for all cases. Examples are
  - Small bandwidth asymptotics
  - Many-weak IV asymptotics
  - Many regressor asymptotics
- Future work: Extend this idea to
  - Network asymptotics
  - Infill asymptotics for high frequency data

Network (ongoing, with Karun Adusumilli)

- Method of moments estimator (Bickel, Chen and Levina, 2011)

$$\hat{P}(R) = \frac{1}{\binom{n}{p}|\mathrm{Iso}(R)|} \sum_{S} \mathbb{I}\{S \sim R\}$$

for subgraph R of G, where n and p are numbers of vertices of G and R, respectively, $\mathrm{Iso}(R)$ is set of subgraphs isomorphic to R

- We find that $\hat{P}(R)$ admits ANOVA-like decomposition

$$\hat{P}(R) - P(R) \approx \frac{1}{n}\sum_i \alpha_i + \frac{1}{n(n-1)\dots(n-p+1)} \sum_{i_1<\dots<i_p} \beta_{i_1 \dots i_p}$$

- BCL considered dense network where first term $\frac{1}{n}\sum_i \alpha_i$ dominates
- Indeed under certain sparse network case ruled out by BCL, both terms are of same order
- Develop sparsity robust jackknife variance estimator and JEL inference