Inference in Instrumental Variable Regression Analysis with Heterogeneous Treatment Effects

Kirill Evdokimov   Michal Kolesár

Princeton University

October 16, 2016

Main goal of the paper

Valid inference (standard errors) in linear IV model when

1. Treatment effects (TE) heterogeneous (each pair of instrument values defines potentially different LATE)
   - Need to define estimand (what are we estimating with more than one instrument?)
2. Number of instruments K and covariates L potentially large
   - Many instrument asymptotics under which K and L may both increase with sample size
   - Instruments may be weak: Concentration parameter $\widehat{r}_n \to \infty$ as $n \to \infty$ (rules out Staiger and Stock (1997) asymptotics), maybe more slowly than n

Main goal of the paper

Valid inference (standard errors) in linear IV model when

1. Treatment effects (TE) heterogeneous (each pair of instrument values defines potentially different LATE)
   - Need to define estimand (what are we estimating with more than one instrument?)
2. Number of instruments K and covariates L potentially large
   - Many instrument asymptotics under which K and L may both increase with sample size
   - Instruments may be weak: Concentration parameter $\widehat{r}_n \to \infty$ as $n \to \infty$ (rules out Staiger and Stock (1997) asymptotics), maybe more slowly than n

Empirical motivation

- Most empirical papers careful to interpret IV exercise as estimating TEs for compliers
  - However, reported standard errors typically only valid under constant TE: too small under LATE model
- In addition, in many papers K and L both large, usually when both instruments and covariates are indicators:
  - Angrist and Krueger (1991) interact quarter and state of birth
  - Judge indicators as instruments: in Aizer and Doyle (2015) judges assigned randomly to juvenile offenders in each neighborhood; in Dobbie and Song (2015) to bankruptcy cases in each bankruptcy office
  - Silver (2016): physician's peer group in each work shift random conditional on physician fixed effects
  - Autor and Houseman (2010): Random assignment of social workers to cases

Unconditional inference

In the LATE model, no single β satisfies moment condition $E[(Y_i - X_i\beta - W_i'\delta)Z_i] = 0$: what are we estimating?
Unconditional inference: estimand β_{U} = plim of estimator under standard asymptotics

- For TSLS (and jackknife estimators) β_{U} = weighted average of LATEs (Imbens and Angrist, 1994)
- For LIML no causal interpretation of β_{U} in general (Kolesár, 2013)
- Under standard asymptotics, can use delta method and normality of reduced form coefficients to derive standard errors (appendix in Imbens and Angrist (1994))
- Under many-instrument asymptotics, no results available

Conditional inference

Conditional inference: estimand β_{C} = what we'd estimate if reduced form errors set to 0

- For tsls and some jackknife estimators: β_{C} also weighted average of LATEs, but can only show (so far) weights positive when covariates indicator variables
  - Heuristically, weights reflect strength of instruments based on their variability in sample (rather than population)
- Standard errors are smaller: s.e. for β_{U} needs to also reflect variability of conditional estimand under i.i.d. sampling of instruments and covariates
  - Similar to distinction between conditional vs unconditional best linear predictor in misspecified linear regression (Abadie, Imbens, and Zheng, 2014)

Main results: Many instruments + LATEs

- Focus on jackknife estimators due to problems with LIML interpretation
- Consistency:
  - Bias of JIVE1, original jackknife IV estimator (Phillips and Hale, 1977; Angrist, Imbens, and Krueger, 1999) of order $L/\widehat{r}_n$
  - Bias of IJIVE1, modification by Ackerberg and Devereux (2009), of order $L/\widehat{r}_n \cdot K/n$ if design is "balanced"
- Asymptotic normality:
  - "many instruments term" needs to be added to (un-) conditional variance formula, similar form to constant TE case

Main results: Standard errors

- No consistently estimable "structural error"
  - Harder to estimate s.e.
  - Need to jackknife s.e. formulas
- Conditional inference: need $(K+L)/n \to 0 +$ "balanced design" $+$ bias negligible (for IJIVE1, $KL/n\hat{r}_n^{1/2} \to 0$)
  - First two conditions similar to those for consistency of EHW s.e. in linear regression with many covariates (Cattaneo, Jansson, and Newey, 2016b)
- Unconditional case: need $(K+L)\log(K+L)/n^{1/2} \to 0$
  - Work in progress on relaxing rate condition

(When) should we run IV with many instruments?

- Our paper provides inference for particular weighted average of LATEs
  - Weighting may not be most policy relevant
- Alternative: use MTE framework of Heckman and Vytlacil (1999, 2005) and do inference on more policy-relevant estimands
  - In general need CI for *each* individual LATE to be informative, or rely on parametric assumptions (e.g. Brinch, Mogstad, and Wiswall, 2015; Cattaneo, Jansson, and Ma, 2016a)
- In "judges" designs, each LATE weakly identified in general: less ambitious goal of inference for *some* TE may be necessary
- IV regressions still ubiquitous: useful to report valid s.e.

Some References

- Nagar '59; Phillips '89; Nelson and Starz '90; Angrist and Krueger '91; Bound, Jaeger, and Baker '95; Staiger and Stock '97; Dufour '97; Andrews, Moreira, and Stock '07; Horowitz '11; Chen and Christensen '15; Maasoumi and Phillips '82.
- Phillips and Hale '77; Angrist, Imbens, and Krueger '99; Blomquist and Dahlberg '99; Davidson and MacKinnon '04,'07; Ackerberg and Devereux '09.
- Bekker '94; Chao, Swanson, Hausman, Newey, and Woutersen '12; Andrews and Stock '05; Newey and Windmeijer '09.
- Andrews '99; Gautier and Tsybakov '11; Belloni, Chernozhukov, Fernandez-val, and Hansen '13.
- Belloni, Chernozhukov, Chetverikov, and Kato '15; Chen and Christensen '15; Tropp '12.
- Imbens and Angrist '94; Angrist, Grady, and Imbens '00; Heckman and Vytlacil '01, '05; Carneiro, Heckman and Vytlacil '11; Kolesár '13; Evdokimov and Lee '13; Kolesár, Chetty, Friedman, Glaeser, and Imbens '14.

General setup

- Interested in effect of treatment X_i on outcome Y_i
- Instruments $Z_i \in \mathbb{R}^K$, valid conditional on controls/exogenous variables $W_i \in \mathbb{R}^L$
- Reduced form:

$$Y_i = \underbrace{Z_i'\pi_Y + W_i'\psi_Y}_{R_{Y,i}} + \zeta_i, \qquad E[\zeta_i \mid W_i, Z_i] = 0,$$

$$X_i = \underbrace{Z_i'\pi + W_i'\psi}_{R_i} + \eta_i, \qquad E[\eta_i \mid W_i, Z_i] = 0.$$

- Under constant TE, $\pi_{Y,k}/\pi_k =$ "the causal effect"

TSLS and concentration parameter

- For matrix A, let $H_A = A(A'A)^{-1}A'$ denote projection (hat) matrix
- Define $\widehat{Y} = (I_n - H_W)Y$, and define $\widehat{X}, \widehat{Z}, \widehat{R}, \widehat{R}_Y$ similarly, e.g. $\widehat{R} = (I_n - H_W)Z\pi + W\psi = \widehat{Z}\pi$.
- Can think $\widehat{Y}_i$ as estimator of $\tilde{Y}_i = Y_i - E[Y_i \mid W_i]$; Define $\tilde{X}, \tilde{Z}, \tilde{R}, \tilde{R}_Y$ similarly ($\tilde{R} = \tilde{Z}\pi$)
- Denote predictor from the first-stage regression by $\widehat{R}_{\mathrm{TSLS}} = H_{\widehat{Z}}\widehat{X} = \widehat{R} + H_{\widehat{Z}}\eta$, so that TSLS is given by

$$\hat{\beta}_{\mathrm{TSLS}} = \frac{\sum_{i=1}^n \widehat{R}_{\mathrm{TSLS},i}\widehat{Y}_i}{\sum_{i=1}^n \widehat{R}_{\mathrm{TSLS},i}\widehat{X}_i}$$

- Define the (sample) concentration parameter $\widehat{r}_n = \sum_{i=1}^n \widehat{R}_i^2$

JIVEs

$$
\begin{aligned}
\hat{\beta}_{\text{TSLS}} &\equiv \frac{\hat{X}'\hat{Z}\left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}'\hat{Y}}{\hat{X}'\hat{Z}\left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}'\hat{X}} = \frac{\sum_{i,j=1}^{n}\hat{X}_i\hat{Z}_i'\left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}_j\hat{Y}_j}{\sum_{i,j=1}^{n}\hat{X}_i\hat{Z}_i'\left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}_j\hat{X}_j} \\
\hat{\beta}_{\text{IJIVE2}} &\equiv \frac{\sum_{i\neq j}\hat{X}_i\hat{Z}_i'\left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}_j\hat{Y}_j}{\sum_{i\neq j}\hat{X}_i\hat{Z}_i'\left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}_j\hat{X}_j} \\
\hat{\beta}_{\text{IJIVE1}} &\equiv \frac{\sum_{i\neq j}\hat{X}_i\hat{Z}_i'\left(\hat{Z}'\hat{Z}-\hat{Z}_j\hat{Z}_j'\right)^{-1}\hat{Z}_j\hat{Y}_j}{\sum_{i\neq j}\hat{X}_i\hat{Z}_i'\left(\hat{Z}'\hat{Z}-\hat{Z}_j\hat{Z}_j'\right)^{-1}\hat{Z}_j\hat{X}_j}
\end{aligned}
$$

... and more. Phillips and Hale, '77; Angrist, Imbens, and Krueger, '99; Blomquist and Dahlberg '99; Ackerberg and Devereux '09, Kolesár '13.

Estimands

- Define β_{U} as plim of estimator under standard asymptotics, and β_{C} as what we would estimate if reduced-form errors were 0.
- For TSLS, JIVE1, IJIVE1:

$$\beta_{\text{U}} = \frac{E[\tilde{X}_i \tilde{Z}_i] E[\tilde{Z}_i \tilde{Z}_i']^{-1} E[\tilde{Z}_i' \tilde{Y}_i]}{E[\tilde{X}_i \tilde{Z}_i] E[\tilde{Z}_i \tilde{Z}_i']^{-1} E[\tilde{Z}_i' \tilde{X}_i]} = \frac{\pi' E[\tilde{Z}_i \tilde{Z}_i'] \pi_Y}{\pi' E[\tilde{Z}_i \tilde{Z}_i'] \pi} = \frac{E\left[\tilde{R}_i \tilde{R}_{Yi}\right]}{E\left[\tilde{R}_i^2\right]}$$

$$\beta_{\text{C}} = \frac{E_n[\widehat{R}_i \widehat{Z}_i] E_n[\widehat{Z}_i \widehat{Z}_i']^{-1} E_n[\widehat{Z}_i' \widehat{R}_{Y,i}]}{E_n[\widehat{X}_i \widehat{Z}_i] E_n[\widehat{Z}_i \widehat{Z}_i']^{-1} E_n[\widehat{Z}_i' \widehat{R}_i]} = \frac{\pi' E_n[\widehat{Z}_i \widehat{Z}_i'] \pi_Y}{\pi' E_n[\widehat{Z}_i \widehat{Z}_i'] \pi} = \frac{E_n\left[\hat{R}_i \hat{R}_{Yi}\right]}{E_n\left[\hat{R}_i^2\right]}$$

  - β_{C} replaces E with E_n (sample average) and tildes with hats
  - For IJIVE2, β_{C} is slightly different
  - $(K + L)\log(K + L) = o(n^{1/2})$
- $\beta_{\text{U}}, \beta_{\text{C}}$ may change with n under many-instrument asymptotics since distribution of data allowed to change with n.

Example: interactions of a single instrument with group dummies

- Suppose $S_i \in \{1, \ldots, K\}$ with $P(S_i = k) = \lambda_k$.
- The K instruments are constructed from a scalar instrument $\mathcal{Z}_i$: $Z_i = (Z_{i1}, \ldots, Z_{iK})'$, $Z_{ik} = \mathcal{Z}_i 1\{S_i = k\}$
- $\tilde{Z}_{ik} = (\mathcal{Z}_i - E[\mathcal{Z}_i | S_i = k]) 1\{S_i = k\}$
- $\hat{Z}_{ik} = (\mathcal{Z}_i - E_n[\mathcal{Z}_i | S_i = k]) 1\{S_i = k\}$
- $\tilde{R}_i = \pi' \tilde{Z}_i$, $\hat{R}_i = \pi' \hat{Z}_i$
- the estimands of TSLS, JIVE1, IJIVE1:

$$
\begin{aligned}
\beta_U &= \frac{\sum_{k=1}^K \beta_k \omega_k}{\sum_{k=1}^K \omega_k}, \ \omega_k = V_k\left[\tilde{R}\right], \ V_k\left[\tilde{R}\right] = V\left[\tilde{R}_i | S_i = k\right] \\
\beta_C &= \frac{\sum_{k=1}^K \beta_k \hat{\omega}_k}{\sum_{k=1}^K \hat{\omega}_k}, \text{ where}
\end{aligned}
$$

$$
\hat{\omega}_k^{\mathrm{TSLS}} = \frac{n_k}{n} \hat{V}_k\left[\tilde{R}\right], \ \hat{\omega}_k^{\mathrm{IJIVE2}} = \frac{n_k}{n} \hat{V}_k\left[\tilde{R}\right] \left(1 - \frac{1}{n_k} \hat{\kappa}_k\left[\tilde{R}\right]\right)
$$

Estimand with many instruments and covariates

Suppose $(K+L)\log(K+L) = o\left(n^{3/4}\right)$ and regularity conditions hold.

Let $\Sigma_{WW} = E\left[W_i W_i'\right]$. Then, for TSLS, JIVE1, IJIVE1:

$$\beta_{\mathrm{U}} \equiv \frac{E\left[\tilde{R}_{Yi}\tilde{R}_i\left(1-\frac{1}{n}W_i'\Sigma_{WW}^{-1}W_i\right)\right]}{E\left[\tilde{R}_i^2\left(1-\frac{1}{n}W_i'\Sigma_{WW}^{-1}W_i\right)\right]}.$$

So, in the previous example

$$\omega_k = \left(\lambda_k - \frac{1}{n}\right) V_k\left[\tilde{R}\right].$$

[[[TBC...]]]

Desirable Property of the Estimand

For any estimand $\overline{\beta}$ we would like to have

$$\overline{\beta} \in \left[\min(\overrightarrow{\beta}), \max(\overrightarrow{\beta})\right] \qquad (*)$$

For the unconditional estimand, Imbens and Angrist '94 show that the of IV-type estimators satisfy $(*)$ under weak conditions, but Kolesár '13 shows that minimum distance estimators in general do not.

Assumption 1 (LATE Model)

(Independence) $\{Y(x), X(z)\}_{x,z} \perp Z \mid W$

(Monotonicity) *For all* z, z', $P(X(z) \geq X(z') \mid W) = 1$ *a.s., or* $P(X(z) \leq X(z') \mid W) = 1$ *a.s.*

(Linearity) $E\left[Z \mid W\right]$ *is linear in* W *and* $E\left[(\zeta, \eta) \mid Z, W\right] = 0$.

Lemma 1

Suppose that the covariates contain an intercept.

- *Then the estimands of TSLS, JIVE1, IJIVE1, JIVE2, IJIVE2 are weighted averages of LATEs.*
- *If the covariates are group dummies, then the weights of TSLS, JIVE1, IJIVE1, and IJIVE2 are non-negative.*

K	$\widehat{r}_n$	β_{sd}	OLS	TSLS	LIML	JIVE1	IJIVE1	IJIVE2
4	16	0	-0.792	-0.031	0.003	0.063	0.014	0.014
4	16	1	-0.792	-0.017	**0.183**	0.077	0.027	0.027
4	64	0	-0.717	-0.007	0.002	0.015	0.004	0.004
4	64	1	-0.716	-0.007	**0.173**	0.015	0.004	0.004
16	16	0	-0.792	**-0.133**	0.008	**0.261**	0.016	0.016
16	16	1	-0.792	**-0.131**	**0.177**	**0.268**	0.022	0.022
16	64	0	-0.717	-0.038	0.003	0.052	0.005	0.005
16	64	1	-0.717	-0.034	**0.175**	0.055	0.008	0.008
64	64	0	-0.717	**-0.145**	0.003	**0.241**	-0.001	-0.001
64	64	1	-0.716	**-0.146**	**0.172**	**0.242**	0.000	0.001

Table: **median bias**; $n = 2000$, $\rho = 0.50$, Normal X, Y

Estimators

$$\hat{\beta}_G = \frac{Y'GX}{X'GX} = \frac{\sum_{i,j} Y_i G_{ij} X_j}{\sum_{i,j} X_i G_{ij} X_j},$$

$$G_{\mathrm{TSLS}} = H_{\widehat{Z}},$$

$$G_{\mathrm{IJIVE1}} = (I - H_W)(I - \operatorname{diag}(H_{\widehat{Z}}))^{-1}(H_{\widehat{Z}} - \operatorname{diag}(H_{\widehat{Z}}))(I - H_W),$$

$$G_{\mathrm{IJIVE2}} = H_{\widehat{Z}} - (I - H_W)\operatorname{diag}(H_{\widehat{Z}})(I - H_W),$$

$$G_{\mathrm{JIVE1}} = (I - H_W)(I - \operatorname{diag}(H_Q))^{-1}(H_Q - \operatorname{diag}(H_Q)).$$

Asymptotic Theory

Regularity conditions

- For both conditional and unconditional inference need
  - reduced form errors are independent, and $\max_i E[\eta_i^4 + \zeta_i^4 \mid Z_i, W_i]$ bounded.
  - Other regularity conditions (stated in paper) that hold in running example if π, π_Y are bounded
- For unconditional inference, assume also that sampling is i.i.d., and eigenvalues of $E[\tilde{Z}_i \tilde{Z}_i']$ and $E[W_i W_i']$ are bounded above and below (balanced design condition)
- Some results go through under weaker conditions

Many instrument asymptotics

- Many weak instrument asymptotics of Bekker (1994) and Chao and Swanson (2005)
- Two important changes:
  1. Allow TE to be heterogeneous
  2. Allow $L \to \infty$
- Distributional results generalize those in Newey and Windmeijer (2009) and Chao, Swanson, Hausman, Newey, and Woutersen (2012)
- Regularity conditions in paper, will focus on rate conditions and substantive restrictions

Theorem 2 (Conditional asymptotic distribution)

Suppose regularity conditions hold, $\max_i (H_{\widehat{Z},W})_{ii}/n \to 0$ *and* $L \max_i (H_{\widehat{Z}})_{ii}/\widehat{r}_n^{1/2} \to 0$ *a.s. Then for* IJIVE1 *and* IJIVE2*,* $\mathcal{V}_C^{-1/2}(\hat{\beta} - \beta_C) \xrightarrow{d} \mathcal{N}(0,1)$*, where* $\mathcal{V}_C = (\mathcal{V}_{TEXT} + \mathcal{V}_{LATE} + \mathcal{V}_{MW})/\widehat{r}_n^2$*,*

$$\mathcal{V}_{TEXT} = \sum_{i=1}^{n} \widehat{R}_i^2 \sigma_{\nu,i}^2 \qquad \nu_i = \zeta_i - \beta_C \eta_i$$

$$\mathcal{V}_{LATE} = \sum_{i=1}^{n} (\widehat{R}_{\Delta,i}^2 \sigma_{\eta,i}^2 + 2\widehat{R}_i \widehat{R}_{\Delta,i} \sigma_{\nu\eta,i}) \qquad \widehat{R}_{\Delta,i} = \widehat{R}_{Y,i} - \widehat{R}_i \beta_C$$

$$\mathcal{V}_{MW} = \sum_{i \neq j} (H_{\widehat{Z}})_{ij}^2 \left(\sigma_{\eta,j}^2 \sigma_{\nu,i}^2 + \sigma_{\nu\eta,i} \sigma_{\nu\eta,j} \right)$$

- $\mathcal{V}_{TEXT}$ is usual "textbook" variance of TSLS, $\mathcal{V}_{LATE}$ reflects variability of LATEs, and $\mathcal{V}_{MW}$ is many-instruments component
- Under homogeneous TE, ν_i is "structural error" and $\widehat{R}_{\Delta,i} = 0$

Theorem 3 (Unconditional asymptotic distribution)

Suppose regularity conditions hold, and that $K + L = o(\sqrt{n/\log(n)})$*.* *Then for* IJIVE1 *and* IJIVE2 ${\mathcal{V}_U}^{-1/2}(\hat{\beta} - \beta_U) \xrightarrow{d} \mathcal{N}(0, 1)$*, where*

$$\mathcal{V}_U = \mathcal{V}_C + \mathcal{V}_E / \widehat{r}_n^2$$

$$\mathcal{V}_E = \sum_{i=1}^{n} \widehat{R}_{\Delta,i}^2 \widehat{R}_i^2$$

and $\left(\frac{\mathcal{V}_E}{\widehat{r}_n}\right)^{-1} \frac{E[\tilde{R}_i^2 \tilde{R}_{\Delta,i}^2]}{E[\tilde{R}_i^2]} \to_p 1.$

- Extra term $\mathcal{V}_E$ accounts for variability of conditional estimand around the unconditional estimand.
- Asymptotic normality continues to hold with $(K + L)\log(K + L) = o(n^{3/4})$

Inference

- Naive plug-in estimators of $\mathcal{V}_C, \mathcal{V}_U$ are *upward* biased under many instrument asymptotics
- Recall tsls estimator $\widehat{R}_{\mathrm{TSLS},i} = \widehat{Z}_i(\widehat{Z}'\widehat{Z})^{-1}\widehat{Z}'\widehat{X} = \sum_j (H_{\widehat{Z}})_{ij}\widehat{X}_j$ of $\widehat{R}_i$
- Define jackknife estimator of $\widehat{R}_i\widehat{R}_{\Delta,i}$:

$$\hat{J}_i(\widehat{X}, \widehat{Y} - \widehat{X}\hat{\beta}) = \sum_{j:\, j\neq i} (H_{\widehat{Z}})_{ij}\widehat{X}_j \sum_{k:\, k\notin\{i,j\}} (\widehat{Y}_i - \widehat{X}_i\hat{\beta})(H_{\widehat{Z}})_{ik}$$

  $\hat{J}_i(\widehat{X}, \widehat{X})$ and $\hat{J}_i(\widehat{Y} - \widehat{X}\hat{\beta}, \widehat{Y} - \widehat{X}\hat{\beta})$ estimate $\widehat{R}_i^2$ and $\widehat{R}_{\Delta,i}^2$
- Estimate ν_i and η_i by $\hat{\zeta}_i - \eta_i\hat{\beta}$ and $\hat{\eta}$, where $\hat{\nu}_i$ and $\hat{\zeta}_i$ denote OLS residuals from reduced form regressions.
- Plug these estimators into asymptotic variance formulas $\mathcal{V}_C$ and $\mathcal{V}_U$

Leads to estimators:

$$\hat{\mathcal{V}}_{TEXT} = \sum_{i=1}^{n} \hat{J}_i(\widehat{X}, \widehat{X})\hat{\upsilon}_i^2$$

$$\hat{\mathcal{V}}_{LATE} = \sum_{i=1}^{n} (\hat{J}_i(\widehat{Y} - \widehat{X}\hat{\beta}, \widehat{Y} - \widehat{X}\hat{\beta})\hat{\eta}_i^2 + 2\hat{J}_i(\widehat{Y} - \widehat{X}\hat{\beta}, \widehat{X})\hat{\upsilon}_i\hat{\eta}_i)$$

$$\hat{\mathcal{V}}_{MW} = \sum_{i \neq j} (H_{\widehat{Z}})_{ij}^2 \left(\hat{\upsilon}_j^2\hat{\eta}_i^2 + \hat{\upsilon}_j\hat{\eta}_j\hat{\upsilon}_i\hat{\eta}_i\right)$$

$$\hat{\mathcal{V}}_E = \sum_i \widehat{R}_{\mathrm{IJIVE1},i}^2 \hat{J}_i(\widehat{Y} - \widehat{X}\hat{\beta}, \widehat{Y} - \widehat{X}\hat{\beta})$$

$$\hat{\mathcal{V}}_C = \frac{\hat{\mathcal{V}}_{TEXT} + \hat{\mathcal{V}}_{LATE} + \hat{\mathcal{V}}_{MW}}{\sum_{i=1}^{n} \widehat{X}_i \widehat{R}_{\mathrm{IJIVE1},i}}$$

$$\hat{\mathcal{V}}_U = \hat{\mathcal{V}}_C + \frac{\hat{\mathcal{V}}_E}{\sum_{i=1}^{n} \widehat{X}_i \widehat{R}_{\mathrm{IJIVE1},i}}$$

For estimators other than IJIVE1, replace $\widehat{R}_{\mathrm{IJIVE1}}$ by corresponding first-stage predictor

Theorem 4 (Standard errors)

Under same conditions leading to asymptotic normality of $\mathcal{V}_C^{-1/2}(\hat{\beta} - \beta_C)$ *and* $\mathcal{V}_U^{-1/2}(\hat{\beta} - \beta_U)$*,*

$$\hat{\mathcal{V}}_C^{-1/2}(\hat{\beta} - \beta_C) \xrightarrow{d} \mathcal{N}(0,1), \quad \textit{and}$$
$$\hat{\mathcal{V}}_U^{-1/2}(\hat{\beta} - \beta_U) \xrightarrow{d} \mathcal{N}(0,1), \quad \textit{respectively.}$$

Takeaways

- When TE are heterogeneous, usual standard errors *not* valid for inference
  - Conditional variance: additional term reflecting variability of LATEs
  - Unconditional variance: conditional variance + term reflecting variability of the conditional estimand
- Need to take a stance on object of interest (conditional/unconditional estimand)
- Precision trade-offs (e.g. Donald and Newey, 2001) different. Changing set of instruments
  - Changes estimand
  - Changes s.e. not only through effect $\widehat{r}_n$ vs K tradeoff, but also through effect on variability of LATEs