Inference for Two-Stage Extremum Estimators

Aristide Houndetoungan[1] and Abdoul Haki Maoude[2]

[1]Cy Cergy Paris Université

[2]Concordia University

EDHEC Business School

November 8, 2023

Sequential Estimation Methods

- When to use two-stage estimation approaches?
  - Endogeneity issues (e.g., instrumental variable models),
  - Missing data (e.g., survey data, Network data),
  - Selection problem,
  - Many DGPs (e.g., multivariate time series modeling),
  - Latent regressors (e.g., expectation about a decision, willingness to pay).
- Procedure:
  1. **First stage:** Estimation of a parameter or a function $\boldsymbol{\beta}_0$.
  2. **Second stage:** The estimator $\hat{\boldsymbol{\beta}}_n$ is plugged into another model to estimate a second parameter $\boldsymbol{\theta}_0$.

     The estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$ is called *plug-in* or *two-stage* estimator.

Sequential Estimation Methods

- Asymptotic properties of $\hat{\boldsymbol{\theta}}_n$
  - Challenging as it depends on the sampling error from the first stage.
  - Is the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ necessarily normal?
  - Even in such a case, the asymptotic variance is difficult to compute.
  - What if the first-stage estimator converges slowly?
  - What if the first stage is a Bayesian estimator, which may not be normally distributed (Zellner and Rossi 1984)?
- Solutions:
  1. Both estimators are asymptotically normally distributed. But computing the variance of $\hat{\boldsymbol{\theta}}_n$ may not be easy (Ackerberg, Chen, and Hahn 2012).
  2. Bootstrap approach (Efron 1992; Gonçalves and White 2005).
     Time-consuming and sometimes infeasible for complex models. Theoretical justification may not be easy (e.g., LASSO models).

Sequential Estimation Methods

- Asymptotic properties of $\hat{\boldsymbol{\theta}}_n$
  - Challenging as it depends on the sampling error from the first stage.
  - Is the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ necessarily normal?
  - Even in such a case, the asymptotic variance is difficult to compute.
  - What if the first-stage estimator converges slowly?
  - What if the first stage is a Bayesian estimator, which may not be normally distributed (Zellner and Rossi 1984)?
- Solutions:
  1. Both estimators are asymptotically normally distributed. But computing the variance of $\hat{\boldsymbol{\theta}}_n$ may not be easy (Ackerberg, Chen, and Hahn 2012).
  2. Bootstrap approach (Efron 1992; Gonçalves and White 2005).
     Time-consuming and sometimes infeasible for complex models. Theoretical justification may not be easy (e.g., LASSO models).

This Paper

- Two- (or multiple-) stage estimation strategy where the second stage leads to an extremum estimator.
- The first-stage estimator, $\hat{\boldsymbol{\beta}}_n$, is general (but consistent): M-estimator, GMM estimator, Minimum distance estimator, nonparametric estimator, Bayesian estimator (e.g., posterior mean).
- Objective function at the second stage:

$$Q_n(\boldsymbol{\theta}, y_{1:n}, \boldsymbol{x}_{1:n}, \hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n} q_{n,i}(\boldsymbol{\theta},\ \hat{\boldsymbol{\beta}}_n),$$

where i can be time for time series models.

- $\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ Q_n(\boldsymbol{\theta}, y_{1:n}, \boldsymbol{x}_{1:n}, \hat{\boldsymbol{\beta}})$. We refer to this class as the *conditional extremum estimator.*

This Paper

- Two- (or multiple-) stage estimation strategy where the second stage leads to an extremum estimator.
- The first-stage estimator, $\hat{\boldsymbol{\beta}}_n$, is general (but consistent): M-estimator, GMM estimator, Minimum distance estimator, nonparametric estimator, Bayesian estimator (e.g., posterior mean).
- Objective function at the second stage:

$$Q_n(\boldsymbol{\theta}, y_{1:n}, \boldsymbol{x}_{1:n}, \hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n} q_{n,i}(\boldsymbol{\theta},\ \hat{\boldsymbol{\beta}}_n),$$

  where i can be time for time series models.
- $\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ Q_n(\boldsymbol{\theta}, y_{1:n}, \boldsymbol{x}_{1:n}, \hat{\boldsymbol{\beta}})$. We refer to this class as the *conditional extremum estimator*.

This Paper

- Novel simulation-based approach to estimate the asymptotic variance and asymptotic CDF of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$.
- Why this method?
  1. Versatility: We do not impose a specific class for $\hat{\boldsymbol{\beta}}_n$, nor a specific convergence rate.
  2. Accounts for the sampling error from the first stage.
  3. Computationally more attractive than the bootstrap method. It eliminates the need for multiple computations of $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\theta}}_n$.
  4. $\hat{\boldsymbol{\beta}}_n$ may not be normally distributed asymptotically (Bayesian estimators in the first stage).
  5. Consequently, $\hat{\boldsymbol{\theta}}_n$ may also not be normally distributed asymptotically.
  6. $\mathbb{E}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))$ may not converge to zero, but $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1)$ (biased estimators).
- Empirical application: Peer effects on adolescent smoking habits when network data are partially observed.

This Paper

- Novel simulation-based approach to estimate the asymptotic variance and asymptotic CDF of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$.
- Why this method?
  1. Versatility: We do not impose a specific class for $\hat{\boldsymbol{\beta}}_n$, nor a specific convergence rate.
  2. Accounts for the sampling error from the first stage.
  3. Computationally more attractive than the bootstrap method. It eliminates the need for multiple computations of $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\theta}}_n$.
  4. $\hat{\boldsymbol{\beta}}_n$ may not be normally distributed asymptotically (Bayesian estimators in the first stage).
  5. Consequently, $\hat{\boldsymbol{\theta}}_n$ may also not be normally distributed asymptotically.
  6. $\mathbb{E}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))$ may not converge to zero, but $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1)$ (biased estimators).
- Empirical application: Peer effects on adolescent smoking habits when network data are partially observed.

Conditional Extremum Estimator: An Example

- IV approach in nonparametric models:

$$y_i = \rho(d_i,\ \boldsymbol{x}_i) + \varepsilon_i,$$

where ρ is an unknown function and d_i is an endogenous treatment.

- Approximate ρ using a series: $\rho(\boldsymbol{w}_i) = \sum_{j=1}^{J} p_j(\boldsymbol{w}_i)\boldsymbol{\theta}_{0,j}$, where $\boldsymbol{w}_i = (d_i,\ \boldsymbol{x}_i')'$, $p_1,\ \dots,\ p_J$ are polynomial functions, J is an integer, and $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,1}', \dots, \boldsymbol{\theta}_{0,J}')'$ (Johnsson and Moon 2021).

- GMM method in the second stage using the moment function $\boldsymbol{m}_i = \boldsymbol{z}_i'\{y_i - \sum_{j=1}^{J} \mathbb{E}(p_j(\boldsymbol{w}_i)|\boldsymbol{z}_i)\boldsymbol{\theta}_{0,j}\}$, where $\boldsymbol{z}_i$ is a vector of instruments.

- First stage: $\mathbb{E}(p_j(\boldsymbol{w}_i)|\boldsymbol{z}_i)$ is unknown and should be estimated.

Conditional Extremum Estimator: An Example

- IV approach in nonparametric models:

$$y_i = \rho(d_i,\ \boldsymbol{x}_i) + \varepsilon_i,$$

where ρ is an unknown function and d_i is an endogenous treatment.

- Approximate ρ using a series: $\rho(\boldsymbol{w}_i) = \sum_{j=1}^{J} p_j(\boldsymbol{w}_i)\boldsymbol{\theta}_{0,j}$, where $\boldsymbol{w}_i = (d_i,\ \boldsymbol{x}_i')'$, $p_1,\ \ldots,\ p_J$ are polynomial functions, J is an integer, and $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,1}', \ldots, \boldsymbol{\theta}_{0,J}')'$ (Johnsson and Moon 2021).

- GMM method in the second stage using the moment function $\boldsymbol{m}_i = \boldsymbol{z}_i'\{y_i - \sum_{j=1}^{J} \mathbb{E}(p_j(\boldsymbol{w}_i)|\boldsymbol{z}_i)\boldsymbol{\theta}_{0,j}\}$, where $\boldsymbol{z}_i$ is a vector of instruments.

- First stage: $\mathbb{E}(p_j(\boldsymbol{w}_i)|\boldsymbol{z}_i)$ is unknown and should be estimated.

Conditional Extremum Estimator: An Example

- IV approach in nonparametric models:

$$y_i = \rho(d_i,\ \boldsymbol{x}_i) + \varepsilon_i,$$

where ρ is an unknown function and d_i is an endogenous treatment.

- Approximate ρ using a series: $\rho(\boldsymbol{w}_i) = \sum_{j=1}^{J} p_j(\boldsymbol{w}_i)\boldsymbol{\theta}_{0,j}$, where $\boldsymbol{w}_i = (d_i,\ \boldsymbol{x}_i')'$, $p_1,\ \dots,\ p_J$ are polynomial functions, J is an integer, and $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,1}', \dots, \boldsymbol{\theta}_{0,J}')'$ (Johnsson and Moon 2021).

- GMM method in the second stage using the moment function $\boldsymbol{m}_i = \boldsymbol{z}_i'\{y_i - \sum_{j=1}^{J} \mathbb{E}(p_j(\boldsymbol{w}_i)|\boldsymbol{z}_i)\boldsymbol{\theta}_{0,j}\}$, where $\boldsymbol{z}_i$ is a vector of instruments.

- First stage: $\mathbb{E}(p_j(\boldsymbol{w}_i)|\boldsymbol{z}_i)$ is unknown and should be estimated.

Related to the Literature

- Classical inference methods:
  1. $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\theta}}_n$ are extremum estimators that converge at the same rate (Newey 1984; Hotz and Miller 1993).
  2. $\hat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$-consistent and asymptotically normally distributed (Murphy and Topel 2002).
  3. $\hat{\boldsymbol{\theta}}_n$ is asymptotically invariant to infinitesimal variations in the first-stage estimator (Chernozhukov, Chetverikov, et al. 2018; Chernozhukov, Escanciano, et al. 2022).
- Resampling methods:
  1. Bootstrap (Efron 1982; Chen, Linton, and Van Keilegom 2003).
  2. Poor (wo)man's bootstrap: relies only on the estimation of one-dimensional parameters (Honoré and Hu 2017)
  3. Fast bootstrap (Armstrong, Bertanha, and Hong 2014; Gonçalves, Hounyo, et al. 2022).
- High-dimensional modeling (Belloni, Chernozhukov, and Hansen 2014; Cattaneo, Jansson, and Ma 2019).

Related to the Literature

- Classical inference methods:
  1. $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\theta}}_n$ are extremum estimators that converge at the same rate (Newey 1984; Hotz and Miller 1993).
  2. $\hat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$-consistent and asymptotically normally distributed (Murphy and Topel 2002).
  3. $\hat{\boldsymbol{\theta}}_n$ is asymptotically invariant to infinitesimal variations in the first-stage estimator (Chernozhukov, Chetverikov, et al. 2018; Chernozhukov, Escanciano, et al. 2022).
- Resampling methods:
  1. Bootstrap (Efron 1982; Chen, Linton, and Van Keilegom 2003).
  2. Poor (wo)man's bootstrap: relies only on the estimation of one-dimensional parameters (Honoré and Hu 2017)
  3. Fast bootstrap (Armstrong, Bertanha, and Hong 2014; Gonçalves, Hounyo, et al. 2022).
- High-dimensional modeling (Belloni, Chernozhukov, and Hansen 2014; Cattaneo, Jansson, and Ma 2019).

Related to the Literature

- Classical inference methods:
  1. $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\theta}}_n$ are extremum estimators that converge at the same rate (Newey 1984; Hotz and Miller 1993).
  2. $\hat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$-consistent and asymptotically normally distributed (Murphy and Topel 2002).
  3. $\hat{\boldsymbol{\theta}}_n$ is asymptotically invariant to infinitesimal variations in the first-stage estimator (Chernozhukov, Chetverikov, et al. 2018; Chernozhukov, Escanciano, et al. 2022).
- Resampling methods:
  1. Bootstrap (Efron 1982; Chen, Linton, and Van Keilegom 2003).
  2. Poor (wo)man's bootstrap: relies only on the estimation of one-dimensional parameters (Honoré and Hu 2017)
  3. Fast bootstrap (Armstrong, Bertanha, and Hong 2014; Gonçalves, Hounyo, et al. 2022).
- High-dimensional modeling (Belloni, Chernozhukov, and Hansen 2014; Cattaneo, Jansson, and Ma 2019).

Outline

Asymptotic Theory: Why Having a Sequential Estimator is an Issue?

- First-order condition at the second stage:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\theta}} q_{n,i}\big(\hat{\boldsymbol{\theta}}_n,\ \hat{\boldsymbol{\beta}}_n\big) = 0. \tag{1}$$

- First-order Taylor approximation of (1) around $\boldsymbol{\theta}_0$:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}(1/\sqrt{n})\sum_{i=1}^{n}\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n), \tag{2}$$

where $\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n) = \frac{\partial}{\partial \boldsymbol{\theta}} q_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)$ and $\mathbf{A}_0 = -(1/n)\sum_{i=1}^{n}\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} q_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)$.

Asymptotic Theory: Why Having a Sequential Estimator is an Issue?

- If $\boldsymbol{\beta}_0$ were known (single-step estimation strategy):

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}(1/\sqrt{n}) \textstyle\sum_{i=1}^n \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0), \tag{3}$$

- A central limit theorem (CLT) can be applied to

$$\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) := (1/\sqrt{n}) \textstyle\sum_{i=1}^n \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \overset{\text{a}}{\sim} N(0, \boldsymbol{\Sigma}_0).$$

$$\Longrightarrow \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \overset{\text{a}}{\sim} N(0,\ \mathbf{A}_0^{-1}\boldsymbol{\Sigma}_0\mathbf{A}_0^{-1}).$$

- **Conditions:** Either the variables $\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$'s are independent across i, or the correlation between $\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$ and $\dot{\boldsymbol{q}}_{n,j}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$ vanishes as $|i - j| \to \infty$ (weak dependence).

Asymptotic Theory: Why Having a Sequential Estimator is an Issue?

- If $\boldsymbol{\beta}_0$ were known (single-step estimation strategy):

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}(1/\sqrt{n}) \textstyle\sum_{i=1}^n \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0), \tag{3}$$

- A central limit theorem (CLT) can be applied to

$$\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) := (1/\sqrt{n}) \textstyle\sum_{i=1}^n \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \overset{\text{a}}{\sim} N(0, \boldsymbol{\Sigma}_0).$$

$$\implies \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \overset{\text{a}}{\sim} N(0,\ \mathbf{A}_0^{-1}\boldsymbol{\Sigma}_0\mathbf{A}_0^{-1}).$$

- **Conditions:** Either the variables $\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$'s are independent across i, or the correlation between $\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$ and $\dot{\boldsymbol{q}}_{n,j}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$ vanishes as $|i-j| \to \infty$ (weak dependence).

Asymptotic Theory: Why Having a Sequential Estimator is an Issue?

- Two-stage estimators:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}(1/\sqrt{n})\sum_{i=1}^{n} \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n), \tag{4}$$

- For any i and j, $\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)$ and $\dot{\boldsymbol{q}}_{n,j}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)$ are dependent.
- The weak dependence assumption does not hold, and a CLT cannot be applied without setting new conditions.
- We characterize the asymptotic behavior of $(1/\sqrt{n})\sum_{i=1}^{n} \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)$ in a general context.

Outline

Asymptotic Variance

- Taylor approximation:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1} \underbrace{(1/\sqrt{n}) \sum_{i=1}^{n} \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)}_{\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)}. \tag{5}$$

- Assumptions:
  1. $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator.
  2. $\mathcal{E}_n = \mathbb{E}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)$, $\mathbf{V}_n = \mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)$, $\mathbb{V}(\mathcal{E}_n)$, and $\mathbb{E}(\mathbf{V}_n)$ exist.
  3. $\lim_{n\to\infty} \mathbb{V}(\mathcal{E}_n)$ and $\lim_{n\to\infty} \mathbb{E}(\mathbf{V}_n)$ exist.
- Implications:
  1. $\mathbb{V}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)) = \mathbf{A}_0^{-1}\boldsymbol{\Sigma}_0\mathbf{A}_0^{-1}$, where $\boldsymbol{\Sigma}_0 = \lim \mathbb{V}(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n))$.
  2. $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1)$, root-$n$ consistent.
  3. However, it is possible that $\sqrt{n}(\mathbb{E}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\theta}_0) \nrightarrow 0$. We can accommodate situations in which $\hat{\boldsymbol{\beta}}_n$ is a high-dimensional vector, which implies a biased $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$.

Asymptotic Variance

- Taylor approximation:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1} \underbrace{(1/\sqrt{n}) \sum_{i=1}^{n} \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)}_{\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)}. \tag{5}$$

- Assumptions:
  1. $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator.
  2. $\mathcal{E}_n = \mathbb{E}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)$, $\mathbf{V}_n = \mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)$, $\mathbb{V}(\mathcal{E}_n)$, and $\mathbb{E}(\mathbf{V}_n)$ exist.
  3. $\lim_{n\to\infty} \mathbb{V}(\mathcal{E}_n)$ and $\lim_{n\to\infty} \mathbb{E}(\mathbf{V}_n)$ exist.

- Implications:
  1. $\mathbb{V}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)) = \mathbf{A}_0^{-1}\boldsymbol{\Sigma}_0\mathbf{A}_0^{-1}$, where $\boldsymbol{\Sigma}_0 = \lim \mathbb{V}(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n))$.
  2. $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1)$, root-$n$ consistent.
  3. However, it is possible that $\sqrt{n}(\mathbb{E}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\theta}_0) \nrightarrow 0$. We can accommodate situations in which $\hat{\boldsymbol{\beta}}_n$ is a high-dimensional vector, which implies a biased $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$.

Asymptotic Variance

- $\mathbb{V}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathbf{A}_0^{-1}\boldsymbol{\Sigma}_0\mathbf{A}_0^{-1}$, where $\boldsymbol{\Sigma}_0 = \lim \mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)\big)$.
- The usual estimator of $\mathbf{A}_0$ is $\hat{\mathbf{A}}_n = -(1/\sqrt{n})\sum_{i=1}^n \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} q_{n,i}(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)$. How to estimate $\boldsymbol{\Sigma}_0$?
- Law of iterated variances. Let $\boldsymbol{\Sigma}_n = \mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)\big)$.

$$\boldsymbol{\Sigma}_n = \mathbb{E}\Big(\mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)\Big) + \mathbb{V}\Big(\mathbb{E}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)\Big),$$
$$\boldsymbol{\Sigma}_n = \mathbb{E}\big(\mathbf{V}_n\big) + \mathbb{V}\big(\mathcal{E}_n\big). \tag{6}$$

- Sampling errors in both stages are disentangled in (6). This makes it easier to construct a consistent estimator for $\boldsymbol{\Sigma}_0$.

Asymptotic Variance

- $\mathbb{V}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathbf{A}_0^{-1}\boldsymbol{\Sigma}_0\mathbf{A}_0^{-1}$, where $\boldsymbol{\Sigma}_0 = \lim \mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)\big)$.
- The usual estimator of $\mathbf{A}_0$ is $\hat{\mathbf{A}}_n = -(1/\sqrt{n})\sum_{i=1}^n \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} q_{n,i}(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)$. How to estimate $\boldsymbol{\Sigma}_0$?
- Law of iterated variances. Let $\boldsymbol{\Sigma}_n = \mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)\big)$.

$$\boldsymbol{\Sigma}_n = \mathbb{E}\Big(\mathbb{V}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)\Big) + \mathbb{V}\Big(\mathbb{E}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)|\hat{\boldsymbol{\beta}}_n\big)\Big),$$
$$\boldsymbol{\Sigma}_n = \mathbb{E}\big(\mathbf{V}_n\big) + \mathbb{V}\big(\mathcal{E}_n\big). \tag{6}$$

- Sampling errors in both stages are disentangled in (6). This makes it easier to construct a consistent estimator for $\boldsymbol{\Sigma}_0$.

Asymptotic Variance: Estimation

- Assumption: The practitioner possesses $\hat{\mathcal{D}}_n$, a valid estimator of the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$. They can also simulate from $\hat{\mathcal{D}}_n$.
- For some large integer κ, let $\hat{\boldsymbol{\beta}}_n^{(1)}, \ \dots, \ \hat{\boldsymbol{\beta}}_n^{(\kappa)}$ be independent simulations from $\hat{\mathcal{D}}_n$.
- We define $\hat{\mathbf{V}}_{n,s}$ and $\hat{\mathcal{E}}_{n,s}$ as the empirical counterparts of $\mathbf{V}_n$ and $\mathcal{E}_n$ by replacing $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_n$, $\boldsymbol{\beta}_0$ with $\hat{\boldsymbol{\beta}}_n$, and $\hat{\boldsymbol{\beta}}_n$ with $\hat{\boldsymbol{\beta}}_n^{(s)}$, for $s = 1, \ \dots, \ \kappa$.
- We show that a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\theta}}_n$ is:

$$\hat{\mathbb{V}}_{\text{asym}} = \frac{\hat{\mathbf{A}}_n^{-1}\hat{\mathbf{\Sigma}}_n^{\kappa}\hat{\mathbf{A}}_n^{-1}}{n}, \tag{7}$$

where $\hat{\mathbf{\Sigma}}_n^{\kappa} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{\mathbf{V}}_{n,s} + \frac{1}{\kappa-1}\sum_{s=1}^{\kappa}(\hat{\mathcal{E}}_{n,s} - \hat{\mathbf{\Omega}}_n^{\kappa})(\hat{\mathcal{E}}_{n,s} - \hat{\mathbf{\Omega}}_n^{\kappa})'$ and $\hat{\mathbf{\Omega}}_n^{\kappa} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{\mathcal{E}}_{n,s}$.

Asymptotic Variance: Estimation

- Assumption: The practitioner possesses $\hat{\mathcal{D}}_n$, a valid estimator of the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$. They can also simulate from $\hat{\mathcal{D}}_n$.
- For some large integer κ, let $\hat{\boldsymbol{\beta}}_n^{(1)}, \ \dots, \ \hat{\boldsymbol{\beta}}_n^{(\kappa)}$ be independent simulations from $\hat{\mathcal{D}}_n$.
- We define $\hat{\mathbf{V}}_{n,s}$ and $\hat{\mathcal{E}}_{n,s}$ as the empirical counterparts of $\mathbf{V}_n$ and $\mathcal{E}_n$ by replacing $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_n$, $\boldsymbol{\beta}_0$ with $\hat{\boldsymbol{\beta}}_n$, and $\hat{\boldsymbol{\beta}}_n$ with $\hat{\boldsymbol{\beta}}_n^{(s)}$, for $s = 1, \ \dots, \ \kappa$.
- We show that a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\theta}}_n$ is:

$$\hat{\mathbb{V}}_{\text{asym}} = \frac{\hat{\mathbf{A}}_n^{-1}\hat{\boldsymbol{\Sigma}}_n^{\kappa}\hat{\mathbf{A}}_n^{-1}}{n}, \tag{7}$$

where $\hat{\boldsymbol{\Sigma}}_n^{\kappa} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{\mathbf{V}}_{n,s} + \frac{1}{\kappa-1}\sum_{s=1}^{\kappa}(\hat{\mathcal{E}}_{n,s} - \hat{\boldsymbol{\Omega}}_n^{\kappa})(\hat{\mathcal{E}}_{n,s} - \hat{\boldsymbol{\Omega}}_n^{\kappa})'$ and $\hat{\boldsymbol{\Omega}}_n^{\kappa} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{\mathcal{E}}_{n,s}$.

Asymptotic Variance: Example with the IV Approach

- Instrumental variable model: $y_i = \theta_0 d_i + \varepsilon_i$, where d_i is an endogenous treatment variable, for which we have an instrument $\boldsymbol{z}_i$.
- **First stage:** $\mathbb{E}(d_i|\boldsymbol{z}_i) = \boldsymbol{z}_i'\boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ is estimated by OLS.
- **Second stage:** We estimate an OLS regression, where the objective function is $Q_n(\theta, \mathbf{y}_n, \hat{\boldsymbol{\beta}}_n)) = \frac{1}{n}\sum_{i=1}^n (y_i - \theta(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))^2$.

$$\dot{\boldsymbol{q}}_{n,i}(y_i, \hat{\boldsymbol{\beta}}_n) = -2(y_i - \theta_0(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n) \quad \text{and} \quad \hat{\mathbf{A}}_n = \frac{2}{n}\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2,$$

$$\Longrightarrow \mathbf{V}_n = \frac{4}{n}\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2\sigma_0^2 \quad \text{and} \quad \mathcal{E}_n = \frac{2\theta_0}{\sqrt{n}}\sum_{i=1}^n (\beta_0(\boldsymbol{z}_i) - (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n).$$

- We estimate $\mathbb{E}(\mathbf{V}_n)$ by $\hat{\mathbb{E}}(\mathbf{V}_n) = \frac{1}{\kappa}\sum_{s=1}^{\kappa} \hat{v}_s$ where $\hat{v}_s = \frac{4}{n}\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})^2\hat{\sigma}_n^2$.
- We estimate $\mathbb{V}(\mathcal{E}_n)$ by $\hat{\mathbb{V}}(\mathcal{E}_n) = \frac{1}{\kappa-1}\sum_{s=1}^{\kappa}(\hat{e}_s - \bar{\hat{e}})^2$, where $\hat{e}_s = \frac{2\hat{\theta}_n}{\sqrt{n}}\sum_{i=1}^n (\hat{\beta}_n(\boldsymbol{z}_i) - \boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)}$, $\bar{\hat{e}} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{e}_s$.
- A consistent estimator of the asymptotic variance of $\hat{\theta}_n$ is $\frac{\hat{\mathbb{E}}(\hat{v}) + \hat{\mathbb{V}}(\hat{e})}{n\hat{\mathbf{A}}_n^2}$.

Asymptotic Variance: Example with the IV Approach

- Instrumental variable model: $y_i = \theta_0 d_i + \varepsilon_i$, where d_i is an endogenous treatment variable, for which we have an instrument $\boldsymbol{z}_i$.
- **First stage:** $\mathbb{E}(d_i|\boldsymbol{z}_i) = \boldsymbol{z}_i'\boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ is estimated by OLS.
- **Second stage:** We estimate an OLS regression, where the objective function is $Q_n\big(\theta, \mathbf{y}_n, \hat{\boldsymbol{\beta}}_n)\big) = \frac{1}{n}\sum_{i=1}^n (y_i - \theta(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))^2$.

$$\dot{\boldsymbol{q}}_{n,i}(y_i, \hat{\boldsymbol{\beta}}_n) = -2(y_i - \theta_0(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n) \quad \text{and} \quad \hat{\mathbf{A}}_n = \frac{2}{n}\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2,$$

$$\Longrightarrow \mathbf{V}_n = \tfrac{4}{n}\textstyle\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2\sigma_0^2 \quad \text{and} \quad \mathcal{E}_n = \tfrac{2\theta_0}{\sqrt{n}}\textstyle\sum_{i=1}^n (\beta_0(\boldsymbol{z}_i) - (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n).$$

- We estimate $\mathbb{E}(\mathbf{V}_n)$ by $\hat{\mathbb{E}}(\mathbf{V}_n) = \frac{1}{\kappa}\sum_{s=1}^{\kappa} \hat{v}_s$ where $\hat{v}_s = \frac{4}{n}\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})^2\hat{\sigma}_n^2$.
- We estimate $\mathbb{V}(\mathcal{E}_n)$ by $\hat{\mathbb{V}}(\mathcal{E}_n) = \frac{1}{\kappa-1}\sum_{s=1}^{\kappa}(\hat{e}_s - \bar{\hat{e}})^2$, where $\hat{e}_s = \frac{2\hat{\theta}_n}{\sqrt{n}}\sum_{i=1}^n (\hat{\beta}_n(\boldsymbol{z}_i) - \boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)}$, $\bar{\hat{e}} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{e}_s$.
- A consistent estimator of the asymptotic variance of $\hat{\theta}_n$ is $\frac{\hat{\mathbb{E}}(\hat{v}) + \hat{\mathbb{V}}(\hat{e})}{n\hat{\mathbf{A}}_n^2}$.

Asymptotic Variance: Example with the IV Approach

- Instrumental variable model: $y_i = \theta_0 d_i + \varepsilon_i$, where d_i is an endogenous treatment variable, for which we have an instrument $\boldsymbol{z}_i$.
- **First stage:** $\mathbb{E}(d_i|\boldsymbol{z}_i) = \boldsymbol{z}_i'\boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ is estimated by OLS.
- **Second stage:** We estimate an OLS regression, where the objective function is $Q_n\big(\theta, \mathbf{y}_n, \hat{\boldsymbol{\beta}}_n)\big) = \frac{1}{n}\sum_{i=1}^n (y_i - \theta(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))^2$.

$$\dot{\boldsymbol{q}}_{n,i}(y_i, \hat{\boldsymbol{\beta}}_n) = -2(y_i - \theta_0(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n) \quad \text{and} \quad \hat{\mathbf{A}}_n = \frac{2}{n}\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2,$$

$$\Longrightarrow \mathbf{V}_n = \tfrac{4}{n}\textstyle\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2\sigma_0^2 \quad \text{and} \quad \mathcal{E}_n = \tfrac{2\theta_0}{\sqrt{n}}\textstyle\sum_{i=1}^n (\beta_0(\boldsymbol{z}_i) - (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n).$$

- We estimate $\mathbb{E}(\mathbf{V}_n)$ by $\hat{\mathbb{E}}(\mathbf{V}_n) = \frac{1}{\kappa}\sum_{s=1}^\kappa \hat{v}_s$ where $\hat{v}_s = \frac{4}{n}\sum_{i=1}^n (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})^2\hat{\sigma}_n^2$.
- We estimate $\mathbb{V}(\mathcal{E}_n)$ by $\hat{\mathbb{V}}(\mathcal{E}_n) = \frac{1}{\kappa-1}\sum_{s=1}^\kappa (\hat{e}_s - \bar{\hat{e}})^2$, where $\hat{e}_s = \frac{2\hat{\theta}_n}{\sqrt{n}}\sum_{i=1}^n (\hat{\beta}_n(\boldsymbol{z}_i) - \boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)}$, $\bar{\hat{e}} = \frac{1}{\kappa}\sum_{s=1}^\kappa \hat{e}_s$.
- A consistent estimator of the asymptotic variance of $\hat{\theta}_n$ is $\frac{\hat{\mathbb{E}}(\hat{v}) + \hat{\mathbb{V}}(\hat{e})}{n\hat{\mathbf{A}}_n^2}$.

Asymptotic Distribution

- Taylor approximation:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1} \underbrace{(1/\sqrt{n}) \sum_{i=1}^{n} \dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)}_{\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)}. \tag{8}$$

- **Conditional CLT**: We apply a CLT to $\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)$ conditional on $\hat{\boldsymbol{\beta}}_n$.
- Conditional on $\hat{\boldsymbol{\beta}}_n$, $\mathbf{V}_n^{-1/2}\big(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n) - \mathcal{E}_n\big)$ converges in distribution to $N(0, \boldsymbol{I})$.
- If $\hat{\boldsymbol{\beta}}_n$ is no longer a random variable, the dependence across i in $\dot{\boldsymbol{q}}_{n,i}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n)$ only depends on the outcome $\boldsymbol{y}_{1:n}$. The weak dependence condition holds if it does in the case of a single-step estimation approach.

Asymptotic Distribution

- Taylor approximation:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n). \tag{9}$$

Theorem

Let $\boldsymbol{\psi}_n = \mathbf{A}_0^{-1}\mathbf{V}_n^{1/2}\boldsymbol{\zeta} + \mathbf{A}_0^{-1}\mathcal{E}_n$, where $\boldsymbol{\zeta} \sim N(0, \boldsymbol{I})$. Let F be asymptotic CDF of $\boldsymbol{\psi}_n$. We have $\lim \mathbb{P}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \preceq \boldsymbol{t}) = F(\boldsymbol{t})$.

- Intuition: Equation (9) $\implies$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}\mathbf{V}_n^{1/2}\underbrace{\left(\mathbf{V}_n^{-1/2}\left(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n) - \mathcal{E}_n\right)\right)}_{\text{Asy. Normal by the CLT}} + \mathbf{A}_0^{-1}\mathcal{E}_n.$$

- The first term in ψ_n is the sampling error from the second stage, whereas the second term captures the sampling error from the first stage.
- $\sqrt{n}\left(\mathbb{E}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\theta}_0\right) \to \mathbf{A}_0^{-1}\lim_{n\to\infty}\mathbb{E}(\mathcal{E}_n)$ and may not be zero.
- The $\alpha/2$ and $(1-\alpha/2)$ quantiles of $\hat{\boldsymbol{\theta}}_n - \hat{\psi}_n/\sqrt{n}$ are the bounds of the $(1-\alpha)$ confidence interval (CI) of $\boldsymbol{\theta}_0$.

Asymptotic Distribution

- Taylor approximation:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n). \tag{9}$$

Theorem

Let $\boldsymbol{\psi}_n = \mathbf{A}_0^{-1}\mathbf{V}_n^{1/2}\boldsymbol{\zeta} + \mathbf{A}_0^{-1}\mathcal{E}_n$, where $\boldsymbol{\zeta} \sim N(0, \boldsymbol{I})$. Let F be asymptotic CDF of $\boldsymbol{\psi}_n$. We have $\lim \mathbb{P}(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \preceq \boldsymbol{t}) = F(\boldsymbol{t})$.

- Intuition: Equation (9) $\implies$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \approx \mathbf{A}_0^{-1}\mathbf{V}_n^{1/2}\underbrace{\left(\mathbf{V}_n^{-1/2}\left(\dot{\boldsymbol{q}}_n(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}_n) - \mathcal{E}_n\right)\right)}_{\text{Asy. Normal by the CLT}} + \mathbf{A}_0^{-1}\mathcal{E}_n.$$

- The first term in $\boldsymbol{\psi}_n$ is the sampling error from the second stage, whereas the second term captures the sampling error from the first stage.
- $\sqrt{n}\left(\mathbb{E}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\theta}_0\right) \to \mathbf{A}_0^{-1}\lim_{n\to\infty}\mathbb{E}(\mathcal{E}_n)$ and may not be zero.
- The $\alpha/2$ and $(1-\alpha/2)$ quantiles of $\hat{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\psi}}_n/\sqrt{n}$ are the bounds of the $(1-\alpha)$ confidence interval (CI) of $\boldsymbol{\theta}_0$.

Asymptotic Distribution: Example with the IV Approach (Continued)

- We have

$$\dot{\boldsymbol{q}}_{n,i}(y_i, \hat{\boldsymbol{\beta}}_n) = -2(y_i - \theta_0(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n) \quad \text{and} \quad \hat{\mathbf{A}}_n = \frac{2}{n}\sum_{i=1}^{n}(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2,$$

$$\Longrightarrow \mathbf{V}_n = \tfrac{4}{n}\textstyle\sum_{i=1}^{n}(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n)^2\sigma_0^2 \quad \text{and} \quad \mathcal{E}_n = \tfrac{2\theta_0}{\sqrt{n}}\textstyle\sum_{i=1}^{n}(\beta_0(\boldsymbol{z}_i) - (\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n))(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n).$$

- Let $\zeta_1, \ \dots, \ \zeta_\kappa$ be κ independent draws from $N(0,1)$. We can obtain a simulation of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n$ by

$$\hat{\boldsymbol{\psi}}_{n,s} = \frac{\sqrt{\hat{v}_s}\zeta_s + \hat{e}_s}{\hat{\mathbf{A}}_n},$$

where $\hat{v}_s = \frac{4}{n}\sum_{i=1}^{n}(\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})^2\hat{\sigma}_n^2$ and $\hat{e}_s = \frac{2\hat{\theta}_n}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\beta}_n(\boldsymbol{z}_i) - \boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)})\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_n^{(s)}$, $\bar{\hat{e}} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{e}_s$.

Outline

Monte Carlo Simulations

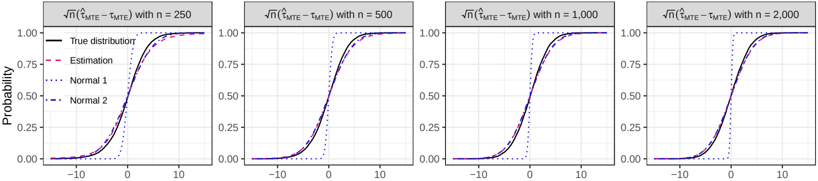
- DGP A: treatment effect model (Roy model):

$$y_i = d_i y_i(1) + (1 - d_i) y_i(0), \quad y_i(0) = u_{0i}, \quad y_i(1) = 0.5 + u_{1i}, \quad d_i = \mathbf{1}\{p_i \geq v_i\},$$

$$p_i = 0.1 + 0.7 z_i, \quad z_i, v_i \sim \text{Uniform}[0,\ 1], \quad u_{0i} \sim \text{Uniform}[-1, 1],$$

$$u_{1i} | v_i \sim \text{Uniform}[-0.5,\ 1.5 - 2v_i].$$

- $\mathbb{E}(y_i | p_i) = \theta_{0,1} + \theta_{0,2} p_i + \theta_{0,3} p_i^2$, where true value of $\boldsymbol{\theta}_0 = (\theta_{0,1},\ \theta_{0,2},\ \theta_{0,3})'$ is $(0,\ 1,\ -1/2)'$. We are interested in the marginal treatment effect at $pi = 0.5$: $\tau_{MTE} = \theta_{0,2} + 2\theta_{0,3} p_i$.

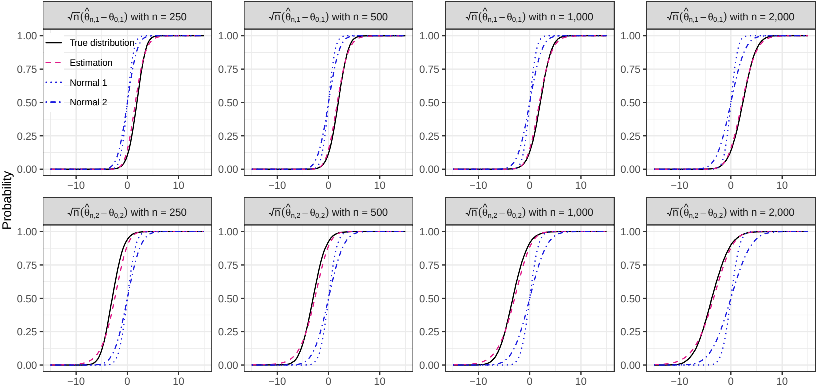


Monte Carlo Simulations

- DGP B is a Poisson model with a latent covariate:

$$y_i \sim \mathcal{P}(\exp(\theta_{0,1} + \theta_{0,2} p_i)), \quad p_i = \sin^2(\pi z_i), \quad z_i \sim \mathcal{U}[0, 10], \quad d_i \sim \mathcal{B}(p_i).$$

- The practitioner observes (y_i, z_i) for all i but only observes d_i for a representative subsample of size $n^* = \lfloor n^{\alpha_n} \rceil$, where $\lfloor . \rceil$ is the rounding to the nearest integer and $\alpha_n \in \{1,\ 0.95,\ 0.90,\ 0.85\}$.

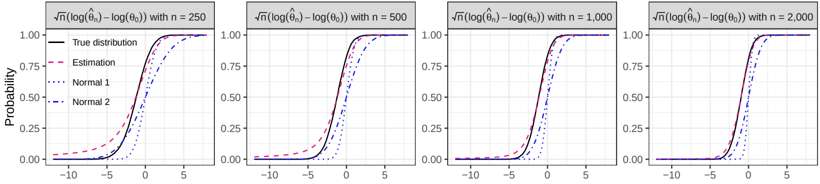


Monte Carlo Simulations

- DGP C is a multivariate time series model of k_n returns $y_{1,i}, \ \ldots, \ y_{k_n,i}$, where i is time. Each $y_{p,i}$, for $p = 2, \ \ldots, \ k_n$, follows an AR(1)-GARCH(1, 1) model:

$$y_{p,i} = \phi_{p,0} + \phi_{p,1} y_{p,i-1} + \sigma_{p,i}\varepsilon_{p,i}, \quad \sigma^2_{p,i} = \beta_{p,0} + \beta_{p,1}\sigma^2_{p,i-1}\varepsilon^2_{p,i-1} + \beta_{p,2}\sigma^2_{p,i-1},$$

  where $\phi_{p,0} = 0$, $\phi_{p,i-1} = 0.4$, $\beta_{p,0} = 0.05$, $\beta_{p,1} = 0.05$, $\beta_{p,2} = 0.9$, and $\varepsilon_{p,i}$ follows a standardized Student distribution of degree of freedom $\nu_p = 6$.
- k_n takes values in $\{2, \ 3, \ 5, \ 8\}$.
- We account for the correlation between the returns using the Clayton copula of parameter $\theta_0 = 4$.



Outline

Peer Effects on Adolescent Smoking Habits

- We study peer effects on adolescent smoking habits using the Add Health survey.
- 26% of the declared best friends are untraceable to schoolmates due to "error codes." The number of friends that a student can declare should not exceed five boys and five girls. 41% of the students declare 5 male friends or 5 female friends.
- Linear probability peer effect model:

$$y_{r,i} = \alpha_{0,r} + \theta_{0,1} \sum_{j=1}^{n_r} \frac{g_{r,ij}}{n_{r,i}} y_{r,j} + \boldsymbol{x}'_{r,i} \boldsymbol{\theta}_{0,2} + \sum_{j=1}^{n_r} \frac{g_{r,ij}}{n_{r,i}} \boldsymbol{x}'_{r,j} \boldsymbol{\theta}_{0,3} + \varepsilon_{r,i},$$

where $g_{r,ij} = 1$ if j is an i's friend and $g_{r,ij} = 0$ otherwise and $y_{r,i}$ is an indicator variable that takes one if student i in school r smokes and zero otherwise.

Peer effects on adolescent smoking habits

- Two-stage simulated GMM, where the network distribution is estimated in the first stage (Boucher and Houndetoungan 2023).
- The simulated GMM proposes a consistent moment function using network simulations from the estimated distribution.

	Standard error		95% confidence interval		
Coef	SdErr 1	SdErr 2	Normal 1	Normal 2	Simulations
Using the network as given					
0.739	0.059		[0.623, 0.855]		
Controlling for missing network data					
0.383	0.123	0.135	[0.141, 0.624]	[0.117, 0.648]	[0.151, 0.674]

Conclusion

- We propose a new simulation-based method to estimate the asymptotic variance and asymptotic CDF of sequential estimators.
- We consider a large class of first-stage estimators.
- The assumption of asymptotic normality at the second stage is unnecessary.
- Conditional on the first-stage estimator, the inference problem is similar to that of single-step extremum estimators, yielding asymptotic normality.
- We account for the sampling error from the first stage using simulations from an estimator of the asymptotic distribution of the first stage.
- The approach is easily implementable and does not require multiple computations.