

Estimation for Conditional Moment Restrictions

徐士勛

中央研究院經濟研究所

2008.2.23

Outline

1 Unconditional Moment Restrictions

- Generalized Method of Moment
- Empirical Likelihood Estimation
- Comparisons

2 Conditional Moment Restrictions

- Efficiency Bound
- Nonparametric Approach
- Nuisance Parameter Approach
- Identification Problem
- The Proposed Approach

Unconditional Moment Restrictions

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)] = \mathbf{0}, \quad \text{with probability one (w.p.1).}$$

where \mathbf{h} is a $p \times 1$ vector of functions, \mathbf{Y} is a $r \times 1$ vector of data variables, $\boldsymbol{\theta}_o$, the $q \times 1$ vector of unknown parameters.

Just-identified: $p = q$

Method of Moment

$$\left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \boldsymbol{\theta}) \right] = 0 \Rightarrow \hat{\boldsymbol{\theta}}_{MM}.$$

Over-identified: $p > q$

Generalized Method of Moment (GMM)

$$\hat{\theta}_{GMM} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]' W_N \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]$$

Over-identified: $p > q$

Generalized Method of Moment (GMM)

$$\hat{\theta}_{GMM} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]' W_N \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]$$

Asymptotics

$$\hat{\theta}_{GMM} \xrightarrow{P} \theta_o$$

$$\sqrt{N}(\hat{\theta}_{GMM} - \theta_o) \xrightarrow{D} N(\mathbf{0}, (\Gamma' W \Gamma)^{-1} \Gamma' W \Sigma_o W' \Gamma (\Gamma' W \Gamma)^{-1})$$

where $\Gamma = \mathbb{E}[\frac{\partial}{\partial \theta} \mathbf{h}(\mathbf{y}_i, \theta_o)]$, $\Sigma_o = \mathbb{E}[\mathbf{h}(\mathbf{y}_i, \theta_o) \mathbf{h}(\mathbf{y}_i, \theta_o)']$.

Over-identified: $p > q$

Generalized Method of Moment (GMM)

$$\hat{\theta}_{GMM} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]' W_N \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]$$

2-step GMM :

- ① Choose $W_N = I_N$, the identify matrix, to get $\hat{\theta}^0$
- ② Let $W_N = \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \hat{\theta}^0) \mathbf{h}(\mathbf{y}_i, \hat{\theta}^0)' \right]^{-1}$ to get $\hat{\theta}_{GMM}$

Over-identified: $p > q$

Generalized Method of Moment (GMM)

$$\hat{\theta}_{GMM} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]' W_N \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \right]$$

Continuous Updating Estimator (CUE) :

- Let $W_N = \left[\frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \mathbf{h}(\mathbf{y}_i, \theta)' \right]^{-1}$ to get $\hat{\theta}_{CUE}$

Unconditional Moment Restrictions:

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \theta_o)] = \mathbf{0}, \quad \text{with probability one (w.p.1).}$$

Let $p_i = \mathbb{P}(\mathbf{Y} = \mathbf{y}_i)$

- Moment restriction:

$$\sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta_o) = 0.$$

- Empirical likelihood function:

$$L(p_1, p_2, \dots, p_N) = \prod_{i=1}^N p_i$$

for $0 \leq p_i \leq 1$, and $\sum_{i=1}^N p_i = 1$.

Unconditional Moment Restrictions:

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \theta_o)] = \mathbf{0}, \quad \text{with probability one (w.p.1).}$$

Let $p_i = \mathbb{P}(\mathbf{Y} = \mathbf{y}_i)$

- Moment restriction:

$$\sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta_o) = 0.$$

- Empirical likelihood function:

$$L(p_1, p_2, \dots, p_N) = \prod_{i=1}^N p_i$$

for $0 \leq p_i \leq 1$, and $\sum_{i=1}^N p_i = 1$.

Empirical Likelihood Estimation (EL):

maximizing the constrained log-likelihood:

$$\max_{\theta, p_i} \sum_{i=1}^N \ln p_i \quad s.t. \quad \sum_{i=1}^N p_i = 1, \quad \sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta) = 0.$$

Empirical Likelihood Estimation (EL):

maximizing the constrained log-likelihood:

$$\max_{\theta, p_i} \sum_{i=1}^N \ln p_i \quad s.t. \quad \sum_{i=1}^N p_i = 1, \quad \sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta) = 0.$$

Lagrangian for this optimization problem is

$$\mathcal{L}(\theta, p_1, \dots, p_N, \lambda, \mu)$$

$$= \sum_{i=1}^N \ln p_i - \mu \left(\sum_{i=1}^N p_i - 1 \right) - N\lambda' \left(\sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta) \right).$$

Empirical Likelihood Estimation (EL):

maximizing the constrained log-likelihood:

$$\max_{\theta, p_i} \sum_{i=1}^N \ln p_i \quad s.t. \quad \sum_{i=1}^N p_i = 1, \quad \sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta) = 0.$$

Empirical Probability (Implied Probability):

$$p_i = \frac{1}{N[1 + \lambda' \mathbf{h}(\mathbf{y}_i, \theta)]}.$$

Empirical Likelihood Estimation (EL):

maximizing the constrained log-likelihood:

$$\max_{\theta, p_i} \sum_{i=1}^N \ln p_i \quad s.t. \quad \sum_{i=1}^N p_i = 1, \quad \sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta) = 0.$$

The optimization problem, concentrated in p_i 's, now is

$$\begin{aligned} & \max_{\theta, \lambda} \sum_{i=1}^N \ln \left(\frac{1}{N[1 + \lambda' \mathbf{h}(\mathbf{y}_i, \theta)]} \right) \\ &= \min_{\theta, \lambda} \sum_{i=1}^N \ln (1 + \lambda' \mathbf{h}(\mathbf{y}_i, \theta)) \end{aligned}$$

Empirical Likelihood Estimation (EL):

maximizing the constrained log-likelihood:

$$\max_{\theta, p_i} \sum_{i=1}^N \ln p_i \quad s.t. \quad \sum_{i=1}^N p_i = 1, \quad \sum_{i=1}^N p_i \mathbf{h}(\mathbf{y}_i, \theta) = 0.$$

Saddle point optimization problem

$$\widehat{\boldsymbol{\theta}}_{\text{EL}} = \arg \min_{\boldsymbol{\theta}} \max_{\lambda} \sum_{i=1}^N \ln (1 + \lambda' \mathbf{h}(\mathbf{y}_i, \boldsymbol{\theta})).$$

Let $\rho \in C^2$ a concave function with normalization
 $\rho_1(0) = \rho_2(0) = -1$, where $\rho_j(v) := \partial^j \rho(v)/\partial v^j$.

Generalized Empirical Likelihood Estimation

$$\hat{\boldsymbol{\theta}}_{\text{GEL}} = \arg \min_{\boldsymbol{\theta}} \max_{\lambda} \sum_{i=1}^N \rho(\lambda' \mathbf{h}(\mathbf{y}_i, \boldsymbol{\theta})).$$

Let $\rho \in C^2$ a concave function with normalization
 $\rho_1(0) = \rho_2(0) = -1$, where $\rho_j(v) := \partial^j \rho(v)/\partial v^j$.

Generalized Empirical Likelihood Estimation

$$\hat{\boldsymbol{\theta}}_{\text{GEL}} = \arg \min_{\boldsymbol{\theta}} \max_{\lambda} \sum_{i=1}^N \rho(\lambda' \mathbf{h}(\mathbf{y}_i, \boldsymbol{\theta})).$$

- Empirical Probability:

$$p_i = \frac{\rho_1(\lambda' \mathbf{h}(\mathbf{y}_i, \boldsymbol{\theta}))}{\sum_{j=1}^N \rho_1(\lambda' \mathbf{h}(\mathbf{y}_j, \boldsymbol{\theta}))}$$

Let $\rho \in C^2$ a concave function with normalization
 $\rho_1(0) = \rho_2(0) = -1$, where $\rho_j(v) := \partial^j \rho(v) / \partial v^j$.

Generalized Empirical Likelihood Estimation

$$\hat{\boldsymbol{\theta}}_{\text{GEL}} = \arg \min_{\boldsymbol{\theta}} \max_{\lambda} \sum_{i=1}^N \rho(\lambda' \mathbf{h}(\mathbf{y}_i, \boldsymbol{\theta})).$$

- EL: $\rho(v) = \ln(1 - v)$.
- Exponential tilting estimator: $\rho(v) = -\exp(v)$.
- CUE: $\rho(v) = -0.5v^2 - v$.

Comparisons

GMM vs. EL vs. GEL

- EL estimator and the optimal GMM estimator are first-order asymptotically equivalent.
- Newey and Smith (2004): the bias of optimal GMM estimator grows with the number of over-identified restrictions, but the bias of EL estimator is bounded.
- Compared with the EL estimator, the other GEL estimators have the same first-order asymptotics, but their second-order biases are usually larger.

Comparisons

GMM vs. EL vs. GEL

- EL estimator and the optimal GMM estimator are first-order asymptotically equivalent.
- Newey and Smith (2004): the bias of optimal GMM estimator grows with the number of over-identified restrictions, but the bias of EL estimator is bounded.
- Compared with the EL estimator, the other GEL estimators have the same first-order asymptotics, but their second-order biases are usually larger.

Comparisons

GMM vs. EL vs. GEL

- EL estimator and the optimal GMM estimator are first-order asymptotically equivalent.
- Newey and Smith (2004): the bias of optimal GMM estimator grows with the number of over-identified restrictions, but the bias of EL estimator is bounded.
- Compared with the EL estimator, the other GEL estimators have the same first-order asymptotics, but their second-order biases are usually larger.

Conditional Moment Restrictions

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \theta_o) | \mathbf{X}] = \mathbf{0}, \quad \text{with probability one (w.p.1).}$$

Example

CAPM model:

$$\mathbb{E}[U_t | I_t] = 0, U_t = \delta R_{t+1} x_{t+1}^{-\gamma_o} - 1,$$

where δ is discount factor, $0 < \delta < 1$; R_t is the gross stock return; x_t is consumption growth; γ_o is the coefficient of relative risk aversion and I_t is information set at time t .

Conditional Moment Restrictions

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \theta_o) | \mathbf{X}] = \mathbf{0}, \quad \text{with probability one (w.p.1).}$$

Example

CAPM model:

$$\mathbb{E}[U_t | I_t] = 0, U_t = \delta R_{t+1} x_{t+1}^{-\gamma_o} - 1,$$

where δ is discount factor, $0 < \delta < 1$; R_t is the gross stock return; x_t is consumption growth; γ_o is the coefficient of relative risk aversion and I_t is information set at time t .

- By the law of iterated expectations, we have

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) Z(\mathbf{X})] = 0$$

for any measurable function $Z(\mathbf{X})$.

- The asymptotic variance-covariance matrix of the resulting GMM estimator is

$$\begin{aligned}\Lambda_o &= \left(\mathbb{E}[Z(\mathbf{X}) \nabla_{\boldsymbol{\theta}} \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)]' \mathbb{E}[Z(\mathbf{X}) \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)' Z(\mathbf{X})']^{-1} \right. \\ &\quad \times \left. \mathbb{E}[Z(\mathbf{X}) \nabla_{\boldsymbol{\theta}} \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)] \right)^{-1}.\end{aligned}$$

Remark

- Given $Z(\mathbf{X})$, Λ_o is the best result.
- The identifiability of $Z(\mathbf{X})$?

- By the law of iterated expectations, we have

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) Z(\mathbf{X})] = 0$$

for any measurable function $Z(\mathbf{X})$.

- The asymptotic variance-covariance matrix of the resulting GMM estimator is

$$\begin{aligned}\Lambda_o = & \left(\mathbb{E}[Z(\mathbf{X}) \nabla_{\boldsymbol{\theta}} \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)]' \mathbb{E}[Z(\mathbf{X}) \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)' Z(\mathbf{X})']^{-1} \right. \\ & \left. \times \mathbb{E}[Z(\mathbf{X}) \nabla_{\boldsymbol{\theta}} \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)] \right)^{-1}.\end{aligned}$$

Remark

- Given $Z(\mathbf{X})$, Λ_o is the best result.
- The identifiability of $Z(\mathbf{X})$?

- By the law of iterated expectations, we have

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) Z(\mathbf{X})] = 0$$

for any measurable function $Z(\mathbf{X})$.

- The asymptotic variance-covariance matrix of the resulting GMM estimator is

$$\begin{aligned}\Lambda_o = & \left(\mathbb{E}[Z(\mathbf{X}) \nabla_{\boldsymbol{\theta}} \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)]' \mathbb{E}[Z(\mathbf{X}) \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)' Z(\mathbf{X})']^{-1} \right. \\ & \times \left. \mathbb{E}[Z(\mathbf{X}) \nabla_{\boldsymbol{\theta}} \mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)] \right)^{-1}.\end{aligned}$$

Remark

- Given $Z(\mathbf{X})$, Λ_o is the best result.
- The identifiability of $Z(\mathbf{X})$?

Efficiency bound: Chamberlain (1987)

The semiparametric efficiency bound Λ_o^* :

$$\Lambda_o^* = \mathbf{E} \left[D_o(\mathbf{X})' \Omega_o^{-1}(\mathbf{X}) D_o(\mathbf{X}) \right]^{-1},$$

where

$$D_o(\mathbf{X}) = \mathbf{E}[\nabla_{\theta'} \mathbf{h}(\mathbf{Y}, \theta_o) | \mathbf{X}], \quad \Omega_o(\mathbf{X}) = \mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta_o) \mathbf{h}(\mathbf{Y}, \theta_o)' | \mathbf{X}].$$

- Optimal Instrument: $Z^*(\mathbf{X}) = D_o(\mathbf{X})' \Omega_o(\mathbf{X})^{-1}$.

Efficiency bound: Chamberlain (1987)

The semiparametric efficiency bound Λ_o^* :

$$\Lambda_o^* = \mathbf{E} \left[D_o(\mathbf{X})' \Omega_o^{-1}(\mathbf{X}) D_o(\mathbf{X}) \right]^{-1},$$

where

$$D_o(\mathbf{X}) = \mathbf{E}[\nabla_{\theta'} \mathbf{h}(\mathbf{Y}, \theta_o) | \mathbf{X}], \quad \Omega_o(\mathbf{X}) = \mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta_o) \mathbf{h}(\mathbf{Y}, \theta_o)' | \mathbf{X}].$$

- Optimal Instrument: $Z^*(\mathbf{X}) = D_o(\mathbf{X})' \Omega_o(\mathbf{X})^{-1}$.

Efficiency bound: Chamberlain (1987)

The semiparametric efficiency bound Λ_o^* :

$$\Lambda_o^* = \mathbf{E} \left[D_o(\mathbf{X})' \Omega_o^{-1}(\mathbf{X}) D_o(\mathbf{X}) \right]^{-1},$$

where

$$D_o(\mathbf{X}) = \mathbf{E}[\nabla_{\theta'} \mathbf{h}(\mathbf{Y}, \theta_o) | \mathbf{X}], \quad \Omega_o(\mathbf{X}) = \mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta_o) \mathbf{h}(\mathbf{Y}, \theta_o)' | \mathbf{X}].$$

- Optimal Instrument: $Z^*(\mathbf{X}) = D_o(\mathbf{X})' \Omega_o(\mathbf{X})^{-1}$.
- It is not unique.
- The resulting unconditional moment restriction model is just-identified.

Efficiency bound: Chamberlain (1987)

The semiparametric efficiency bound Λ_o^* :

$$\Lambda_o^* = \mathbf{E} \left[D_o(\mathbf{X})' \Omega_o^{-1}(\mathbf{X}) D_o(\mathbf{X}) \right]^{-1},$$

where

$$D_o(\mathbf{X}) = \mathbf{E}[\nabla_{\theta'} \mathbf{h}(\mathbf{Y}, \theta_o) | \mathbf{X}], \quad \Omega_o(\mathbf{X}) = \mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta_o) \mathbf{h}(\mathbf{Y}, \theta_o)' | \mathbf{X}].$$

- Optimal Instrument: $Z^*(\mathbf{X}) = D_o(\mathbf{X})' \Omega_o(\mathbf{X})^{-1}$.

Without more information or assumptions about underlying data generating process, this optimal instrumental variable, in general, is infeasible.

Nonparametric Approach

Newey(1990, 1993)

- Nonparametrically estimate Optimal IV: nearest neighbor and series approximation.
- Solve the just-identified resulting unconditional moment restrictions directly by method of moment.

Nonparametric Approach

Ai and Chen (2003)

$$\theta_o = \underset{\theta \in \Theta}{\operatorname{argmin}} \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) | \mathbf{X}]' \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) \mathbf{h}(\mathbf{Y}, \theta)' | \mathbf{X}]^{-1} \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) | \mathbf{X}] .$$

- Estimate each component by a sequence of known basis functions, such as power series, spline or Fourier series.
- Plug these consistent linear sieve estimators into the objective function, we will get a sieve minimum distance (SMD) estimator of θ_o .

Nonparametric Approach

Kitamura et al. (2004)

Constrained “smoothed” empirical likelihood:

$$\max_{\theta, p_{ij}} \sum_{j=1}^N \sum_{i=1}^N w_{ij} \ln p_{ij}$$

$$s.t. p_{ij} \geq 0, \sum_{i=1}^N p_{ij} = 1, \sum_{i=1}^N p_{ij} \mathbf{h}(\mathbf{y}_i, \theta) = 0, \text{ for } j, i = 1, \dots, N.$$

- $p_{ij} := \mathbb{P}(\mathbf{y}_i | \mathbf{x}_j)$.
- w_{ij} is a preliminary weight measuring the importance of \mathbf{x}_i to \mathbf{x}_j and carries out the localization at \mathbf{x}_j .

Nonparametric Approach

Disadvantages

- They usually need to select bandwidth parameters in estimation, which may cause poor finite sample performance when samples are small.
- It suffers from the curse of dimensionality because the convergence rate for nonparametric approximation depends on the dimension of \mathbf{X} .

Nuisance Parameter Approach

Conditional Moment Restrictions

$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \theta_o) | \mathbf{X}] = \mathbf{0}$, with probability one (w.p.1).

Possible choice of $Z(\mathbf{X})$:

- Chamberlain (1987) : polynomials of infinite order of \mathbf{X} .
- Bierens (1990): $\exp(\tau \mathbf{X})$ for almost all τ in a subset of \mathbb{R}^m
- Donald et al. (2003) : Fourier series and splines of \mathbf{X} .
- Domínguez and Lobato (2004): $I(\mathbf{X} \leq \tau)$ for almost all $\tau \in \mathbb{R}^m$.

Nuisance Parameter Approach

The implied unconditional moment restrictions

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) Z(\mathbf{X})] = 0.$$

Possible choice of $Z(\mathbf{X})$:

- Chamberlain (1987) : polynomials of infinite order of \mathbf{X} .
- Bierens (1990): $\exp(\tau \mathbf{X})$ for almost all τ in a subset of \mathbb{R}^m
- Donald et al. (2003) : Fourier series and splines of \mathbf{X} .
- Domínguez and Lobato (2004): $I(\mathbf{X} \leq \tau)$ for almost all $\tau \in \mathbb{R}^m$.

Nuisance Parameter Approach

The implied unconditional moment restrictions

$$\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) Z(\mathbf{X})] = 0.$$

Possible choice of $Z(\mathbf{X})$:

- Chamberlain (1987) : polynomials of infinite order of \mathbf{X} .
- Bierens (1990): $\exp(\tau \mathbf{X})$ for almost all τ in a subset of \mathbb{R}^m
- Donald et al. (2003) : Fourier series and splines of \mathbf{X} .
- Domínguez and Lobato (2004): $I(\mathbf{X} \leq \tau)$ for almost all $\tau \in \mathbb{R}^m$.

Typical Estimation Approach

- Select finitely many unconditional restrictions
(with or without prior knowledge)
- Apply some estimation methods to these selected unconditional restrictions

Typical Estimation Approach

- Select **finitely** many unconditional restrictions
(with or without prior knowledge)
- Apply some estimation methods to these **selected** unconditional restrictions

Typical Estimation Approach

- Select **finitely** many unconditional restrictions
(with or without prior knowledge)
- Apply some estimation methods to these **selected** unconditional restrictions

Typical Estimation Approach

- Select **finitely** many unconditional restrictions
(with or without prior knowledge)
- Apply some estimation methods to these **selected** unconditional restrictions

Hansen (1982)

Generalized method of moments (GMM)

Typical Estimation Approach

- Select **finitely** many unconditional restrictions
(with or without prior knowledge)
- Apply some estimation methods to these **selected** unconditional restrictions

Qin and Lawless (1994)

Empirical likelihood (EL) method

Parameter Identifiability

- The parameter of interest may NOT be identified by the selected (finitely) unconditional moments.
- The resulting estimator is INCONSISTENT.

Parameter Identifiability

- The parameter of interest may **NOT** be identified by the selected (finitely) unconditional moments.
- The resulting estimator is **INCONSISTENT**.

Parameter Identifiability

- The parameter of interest may **NOT** be identified by the selected (finitely) unconditional moments.
- The resulting estimator is **INCONSISTENT**.

Parameter Identifiability

- The parameter of interest may **NOT** be identified by the selected (finitely) unconditional moments.
- The resulting estimator is **INCONSISTENT**.

Why?

All implied unconditional moment restrictions should be considered.

Parameter Identifiability

- The parameter of interest may **NOT** be identified by the selected (finitely) unconditional moments.
- The resulting estimator is **INCONSISTENT**.

Typical solutions:

- Ignore this problem.
- Assume that θ_0 can be (globally) identified by these selected moment restrictions

Parameter Identifiability

- The parameter of interest may **NOT** be identified by the selected (finitely) unconditional moments.
- The resulting estimator is **INCONSISTENT**.

Example

Newey(1990): Assumption 3.3 (a) $\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta})Z^*(\mathbf{X})] = 0$ has a **unique** solution on $\boldsymbol{\Theta}$ at $\boldsymbol{\theta}_o$, where $Z^*(\mathbf{X})$ is an optimal instrument variable.

Two examples (DL, 2004)

Example

DGP: $E[Y|X] = X^{\theta_o}$, $\theta_o = 4$, X is symmetric s.t. $E[X^4] = E[X^6]$.

- Instruments: $Z(X) = (1, X)'$.
- Model: $E[(Y - X^\theta)Z(X)] = 0$.
- Solutions: $\theta = 4$ and $\theta = 6$.
- θ_o can **Not** be globally identified by the selected moments.

Two examples (DL, 2004)

Example

DGP: $E[Y|X] = X^{\theta_o}$, $\theta_o = 4$, X is symmetric s.t. $E[X^4] = E[X^6]$.

- Instruments: $Z(X) = (1, X)'$.
- Model: $E[(Y - X^\theta)Z(X)] = 0$.
- Solutions: $\theta = 4$ and $\theta = 6$.
- θ_o can Not be globally identified by the selected moments.

Two examples (DL, 2004)

Example

DGP: $E[Y|X] = \theta_o^2 X + \theta_o X^2$, $\theta_o = 5/4$, $X \sim N(-1, 1)$, $V(Y|X)$ is constant.

- Instruments: $Z(X) = 2\theta X + X^2$. (This is an optimal instrument.)
- Model: $E[(Y - \theta^2 X - \theta X^2)Z(X)] = 0$.
- Solutions: $\theta = 5/4$, $\theta = -5/4$ and $\theta = -3$.
- θ_o can Not be globally identified by the selected moments.

Two examples (DL, 2004)

Example

DGP: $E[Y|X] = \theta_o^2 X + \theta_o X^2$, $\theta_o = 5/4$, $X \sim N(-1, 1)$, $V(Y|X)$ is constant.

- Instruments: $Z(X) = 2\theta X + X^2$. (This is an optimal instrument.)
- Model: $E[(Y - \theta^2 X - \theta X^2)Z(X)] = 0$.
- Solutions: $\theta = 5/4$, $\theta = -5/4$ and $\theta = -3$.
- θ_o can Not be globally identified by the selected moments.

CAPM model:

$$\mathbb{E}[U_t | I_t] = 0, U_t = \delta R_{t+1} x_{t+1}^{-\gamma_o} - 1.$$

CAPM model:

$$\mathbb{E}[U_t | I_t] = 0, U_t = \delta R_{t+1} x_{t+1}^{-\gamma_o} - 1.$$

Assume $R_t = R$ is time-invariant, and generate data by

$$\ln(x_t) = 0.021 - 0.161 \ln(x_{t-1}) + v_t, \quad v_t \sim N(0, 0.0012),$$

$$U_t = x_{t+1}^{-\gamma_o} - 1 + \epsilon_t, \quad \epsilon_t \sim N(0, 0.001),$$

where $\gamma_o = 13.7$.

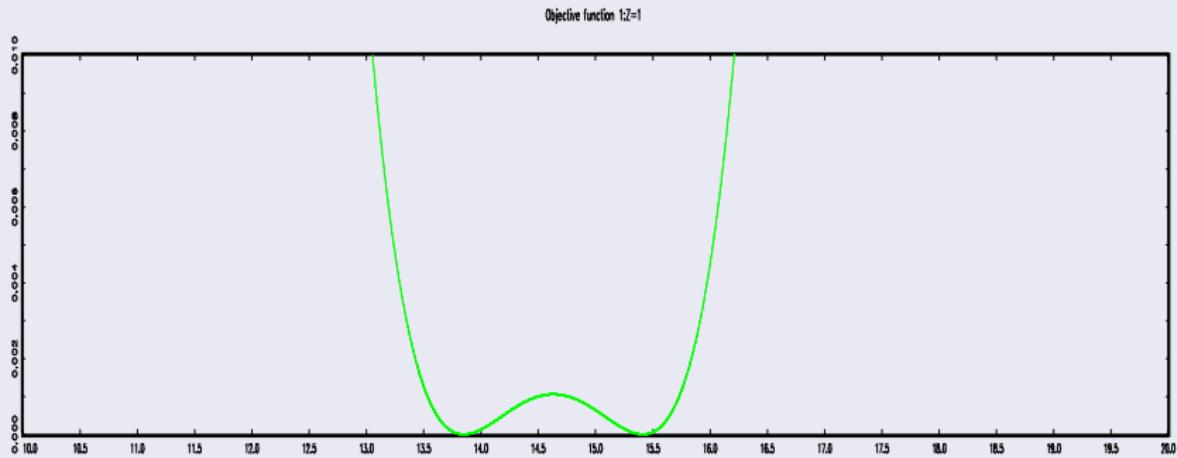
CAPM model:

$$\mathbb{E}[U_t | I_t] = 0, U_t = \delta R_{t+1} x_{t+1}^{-\gamma_o} - 1.$$

- $Z_t = 1: (\mathbb{E}[U_t - x_{t+1}^{-\gamma_o} + 1])^2.$
- $Z_t = [1 \ x_t]':$
 $(\mathbb{E}[(U_t - x_{t+1}^{-\gamma_o} + 1)])^2 + (\mathbb{E}[(U_t - x_{t+1}^{-\gamma_o} + 1)x_t])^2.$

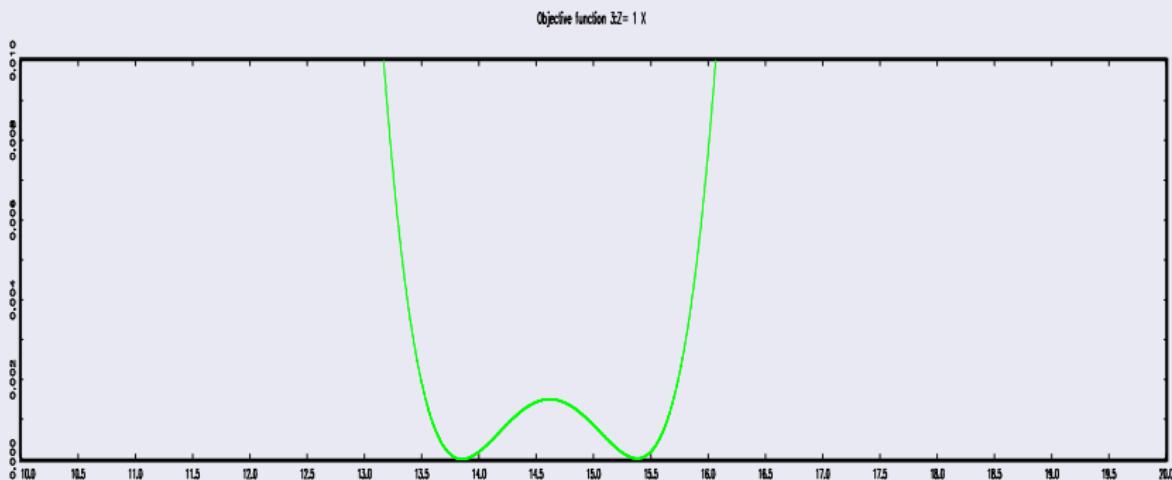
CAPM model:

$$\mathbb{E}[U_t | I_t] = 0, U_t = \delta R_{t+1} x_{t+1}^{-\gamma_o} - 1.$$



CAPM model:

$$\mathbb{E}[U_t | I_t] = 0, U_t = \delta R_{t+1} x_{t+1}^{-\gamma_o} - 1.$$



Remarks:

- In the nonlinear conditional moment restrictions, the parameters of interest may not be identified exactly by the selected moments (even the implied unconditional restrictions are over-identified).
- If without identification, the estimator is **inconsistent**.
- How to select more (all) implied restrictions?

Remarks:

- In the nonlinear conditional moment restrictions, the parameters of interest may not be identified exactly by the selected moments (even the implied unconditional restrictions are over-identified).
- If without identification, the estimator is **inconsistent**.
- How to select more (all) implied restrictions?

Remarks:

- In the nonlinear conditional moment restrictions, the parameters of interest may not be identified exactly by the selected moments (even the implied unconditional restrictions are over-identified).
- If without identification, the estimator is **inconsistent**.
- How to select more (all) implied restrictions?

The Proposed Approach

- We propose an **efficient** and **systematical** way to consider all implied unconditional moment restrictions.
- No additional assumptions for parameter identifiability are needed.
- The proposed estimator is consistent.
- Asymptotics are provided.
- Good finite sample properties are revealed by some Monte Carlo simulations.

The Proposed Approach

- We propose an **efficient** and **systematical** way to consider all implied unconditional moment restrictions.
- No additional assumptions for parameter identifiability are needed.
- The proposed estimator is consistent.
- Asymptotics are provided.
- Good finite sample properties are revealed by some Monte Carlo simulations.

The Proposed Approach

- We propose an **efficient** and **systematical** way to consider all implied unconditional moment restrictions.
- No additional assumptions for parameter identifiability are needed.
- The proposed estimator is consistent.
- Asymptotics are provided.
- Good finite sample properties are revealed by some Monte Carlo simulations.

The Proposed Approach

- We propose an **efficient** and **systematical** way to consider all implied unconditional moment restrictions.
- No additional assumptions for parameter identifiability are needed.
- The proposed estimator is consistent.
- Asymptotics are provided.
- Good finite sample properties are revealed by some Monte Carlo simulations.

The Proposed Approach

- We propose an **efficient** and **systematical** way to consider all implied unconditional moment restrictions.
- No additional assumptions for parameter identifiability are needed.
- The proposed estimator is consistent.
- Asymptotics are provided.
- Good finite sample properties are revealed by some Monte Carlo simulations.

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta_o) G(A(\mathbf{X}, \tau))] = 0, \text{ for almost all } \tau \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

- A is the affine transformation such that
$$A(\mathbf{X}, \tau) = \tau_0 + \sum_{j=1}^m X_j \tau_j.$$
- G is generically comprehensive revealing functions.
- They are real analytic (but not a polynomial) of \mathbf{X} : $\exp(\cdot)$, the logistic, the hyperbolic tangent, the sine and cosine, etc.

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta_o) G(A(\mathbf{X}, \tau))] = 0, \text{ for almost all } \tau \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

Advantages:

- Any nonempty subset of \mathbb{R}^{m+1} could be selected as \mathcal{T} .
- If the conditional moment restriction is not true, then for all $\theta \in \Theta$,

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] \neq 0, \text{ essential all } \tau \in \mathcal{T}.$$

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \tau))] = 0, \text{ for almost all } \tau \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

Advantages:

- Any nonempty subset of \mathbb{R}^{m+1} could be selected as \mathcal{T} .
- If the conditional moment restriction is not true, then for all $\boldsymbol{\theta} \in \Theta$,

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) G(A(\mathbf{X}, \tau))] \neq 0, \text{ essential all } \tau \in \mathcal{T}.$$

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta_o) G(A(\mathbf{X}, \tau))] = 0, \text{ for almost all } \tau \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

Disadvantages:

- The dimension of τ is larger than that of conditioning variable by one.
- These unconditional moment restrictions indexed by τ are infinitely many and **uncountable**.

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))] = 0, \text{ for almost all } \boldsymbol{\tau} \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

L_2 -norm

$$\int_{\mathcal{T}} |\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))]|^2 d\boldsymbol{\tau} = 0.$$

- For all $\boldsymbol{\theta} \in \Theta$, only $\boldsymbol{\theta}_o$ leads zero L_2 -norm.
- $\boldsymbol{\theta}_o$ can be identified by this equality.
- The numerical method of integration is always needed, if solving this with sample counterparts directly.

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))] = 0, \text{ for almost all } \boldsymbol{\tau} \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

L_2 -norm

$$\int_{\mathcal{T}} |\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))]|^2 d\boldsymbol{\tau} = 0.$$

- For all $\boldsymbol{\theta} \in \Theta$, only $\boldsymbol{\theta}_o$ leads zero L_2 -norm.
- $\boldsymbol{\theta}_o$ can be identified by this equality.
- The numerical method of integration is always needed, if solving this with sample counterparts directly.

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))] = 0, \text{ for almost all } \boldsymbol{\tau} \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

L_2 -norm

$$\int_{\mathcal{T}} |\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))]|^2 d\boldsymbol{\tau} = 0.$$

- For all $\boldsymbol{\theta} \in \Theta$, only $\boldsymbol{\theta}_o$ leads zero L_2 -norm.
- $\boldsymbol{\theta}_o$ can be identified by this equality.**
- The numerical method of integration is always needed, if solving this with sample counterparts directly.

Generically Comprehensive Revealing

(Stinchcombe and White 1998)

$$\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))] = 0, \text{ for almost all } \boldsymbol{\tau} \in \mathcal{T} \subset \mathbb{R}^{m+1}.$$

L_2 -norm

$$\int_{\mathcal{T}} |\mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o) G(A(\mathbf{X}, \boldsymbol{\tau}))]|^2 d\boldsymbol{\tau} = 0.$$

- For all $\boldsymbol{\theta} \in \Theta$, only $\boldsymbol{\theta}_o$ leads zero L_2 -norm.
- $\boldsymbol{\theta}_o$ can be identified by this equality.
- The numerical method of integration is always needed, if solving this with sample counterparts directly.

We project the induced unconditional moments along the exponential Fourier series and obtain

$$\mathbf{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) G(A(\mathbf{X}, \tau))] = \frac{1}{(2\pi)^{m+1}} \sum_{\mathbf{k} \in \mathcal{S}} C_{G,\mathbf{k}}(\boldsymbol{\theta}) \exp(i\mathbf{k}'\tau),$$

where $\mathcal{S} := \{\mathbf{k} = [k_0, k_1, \dots, k_m]' \in \mathbb{Z}^{m+1}\}$ with $k_i = 0, \pm 1, \pm 2, \dots, \pm \infty$.

$C_{G,\mathbf{k}}(\boldsymbol{\theta})$ is a Fourier coefficient:

$$\begin{aligned} C_{G,\mathbf{k}}(\boldsymbol{\theta}) &= \int_T \mathbf{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) G(A(\mathbf{X}, \tau))] \exp(-i\mathbf{k}'\tau) d\tau \\ &= \mathbf{E} \left[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) \int_T G(A(\mathbf{X}, \tau)) \exp(-i\mathbf{k}'\tau) d\tau \right], \quad \mathbf{k} \in \mathcal{S}. \end{aligned}$$

We project the induced unconditional moments along the exponential Fourier series and obtain

$$\mathbf{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) G(A(\mathbf{X}, \tau))] = \frac{1}{(2\pi)^{m+1}} \sum_{\mathbf{k} \in \mathcal{S}} C_{G,\mathbf{k}}(\boldsymbol{\theta}) \exp(i\mathbf{k}'\tau),$$

where $\mathcal{S} := \{\mathbf{k} = [k_0, k_1, \dots, k_m]' \in \mathbb{Z}^{m+1}\}$ with $k_i = 0, \pm 1, \pm 2, \dots, \pm \infty$.

$C_{G,\mathbf{k}}(\boldsymbol{\theta})$ is a Fourier coefficient:

$$\begin{aligned} C_{G,\mathbf{k}}(\boldsymbol{\theta}) &= \int_{\mathcal{T}} \mathbf{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) G(A(\mathbf{X}, \tau))] \exp(-i\mathbf{k}'\tau) d\tau \\ &= \mathbf{E} \left[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) \int_{\mathcal{T}} G(A(\mathbf{X}, \tau)) \exp(-i\mathbf{k}'\tau) d\tau \right], \quad \mathbf{k} \in \mathcal{S}. \end{aligned}$$

- Each $C_{G,\mathbf{k}}(\boldsymbol{\theta})$ incorporates the continuum of the original instruments $G(A(\mathbf{X}, \tau))$ into a new instrument:

$$\varphi_{G,\mathbf{k}}(\mathbf{X}) = \int_{\mathcal{T}} G(A(\mathbf{X}, \tau)) \exp(-i\mathbf{k}'\tau) d\tau.$$

We project the induced unconditional moments along the exponential Fourier series and obtain

$$\mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] = \frac{1}{(2\pi)^{m+1}} \sum_{\mathbf{k} \in \mathcal{S}} C_{G,\mathbf{k}}(\theta) \exp(i\mathbf{k}'\tau),$$

where $\mathcal{S} := \{\mathbf{k} = [k_0, k_1, \dots, k_m]' \in \mathbb{Z}^{m+1}\}$ with $k_i = 0, \pm 1, \pm 2, \dots, \pm \infty$.

$C_{G,\mathbf{k}}(\theta)$ is a Fourier coefficient:

$$\begin{aligned} C_{G,\mathbf{k}}(\theta) &= \int_T \mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] \exp(-i\mathbf{k}'\tau) d\tau \\ &= \mathbf{E} \left[\mathbf{h}(\mathbf{Y}, \theta) \int_T G(A(\mathbf{X}, \tau)) \exp(-i\mathbf{k}'\tau) d\tau \right], \quad \mathbf{k} \in \mathcal{S}. \end{aligned}$$

Parseval's Theorem:

$$\int_T |\mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))]|^2 d\tau = \sum_{\mathbf{k} \in \mathcal{S}} |\mathbf{E}[\mathbf{h}(\mathbf{Y}, \theta) \varphi_{G,\mathbf{k}}(\mathbf{X})]|^2.$$

Identification:

$$\begin{aligned}\theta_o &= \operatorname{argmin}_{\theta \in \Theta} \int_{\mathcal{T}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] \right|^2 d\tau \\ &= \operatorname{argmin}_{\theta \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) \varphi_{G, \mathbf{k}}(\mathbf{X})] \right|^2.\end{aligned}$$

Identification:

$$\begin{aligned}\boldsymbol{\theta}_o &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \int_{\mathcal{T}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) G(A(\mathbf{X}, \boldsymbol{\tau}))] \right|^2 d\boldsymbol{\tau} \\ &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) \varphi_{G,\mathbf{k}}(\mathbf{X})] \right|^2.\end{aligned}$$

Remarks:

- Project a continuum moment restrictions along exponential Fourier series.
- A new sequence of unconditional moment restrictions induced by $\varphi_{\mathbf{k}}(\mathbf{X})$. There are **countable**.

Identification:

$$\begin{aligned}\theta_o &= \operatorname{argmin}_{\theta \in \Theta} \int_{\mathcal{T}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] \right|^2 d\tau \\ &= \operatorname{argmin}_{\theta \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) \varphi_{G,\mathbf{k}}(\mathbf{X})] \right|^2.\end{aligned}$$

Advantages:

- The resulting extreme estimator is thus easy to obtained.
- Each $\varphi_{G,\mathbf{k}}(\mathbf{X})$ incorporates all information revealed by $G(A(\mathbf{X}, \tau))$ for all τ .
- Bessel's inequality: $\varphi_{G,\mathbf{k}}(\mathbf{X}) \rightarrow 0$ as $|\mathbf{k}| \rightarrow \infty$. The corresponding moment restriction is less informative in identifying θ_o .

Identification:

$$\begin{aligned}\boldsymbol{\theta}_o &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \int_{\mathcal{T}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) G(A(\mathbf{X}, \boldsymbol{\tau}))] \right|^2 d\boldsymbol{\tau} \\ &= \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) \varphi_{G,\mathbf{k}}(\mathbf{X})] \right|^2.\end{aligned}$$

Advantages:

- The resulting extreme estimator is thus easy to obtained.
- Each $\varphi_{G,\mathbf{k}}(\mathbf{X})$ incorporates all information revealed by $G(A(\mathbf{X}, \boldsymbol{\tau}))$ for all $\boldsymbol{\tau}$.
- Bessel's inequality: $\varphi_{G,\mathbf{k}}(\mathbf{X}) \rightarrow 0$ as $|\mathbf{k}| \rightarrow \infty$. The corresponding moment restriction is less informative in identifying $\boldsymbol{\theta}_o$.

Identification:

$$\begin{aligned}\theta_o &= \operatorname{argmin}_{\theta \in \Theta} \int_{\mathcal{T}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] \right|^2 d\tau \\ &= \operatorname{argmin}_{\theta \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) \varphi_{G,\mathbf{k}}(\mathbf{X})] \right|^2.\end{aligned}$$

Advantages:

- The resulting extreme estimator is thus easy to obtained.
- Each $\varphi_{G,\mathbf{k}}(\mathbf{X})$ incorporates all information revealed by $G(A(\mathbf{X}, \tau))$ for all τ .
- Bessel's inequality: $\varphi_{G,\mathbf{k}}(\mathbf{X}) \rightarrow 0$ as $|\mathbf{k}| \rightarrow \infty$. The corresponding moment restriction is less informative in identifying θ_o .

Identification:

$$\begin{aligned}\theta_o &= \operatorname{argmin}_{\theta \in \Theta} \int_{\mathcal{T}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] \right|^2 d\tau \\ &= \operatorname{argmin}_{\theta \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) \varphi_{G, \mathbf{k}}(\mathbf{X})] \right|^2.\end{aligned}$$

The Proposed Estimator:

$$\hat{\theta}(G, \mathcal{K}_N) = \operatorname{argmin}_{\theta \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}(\mathcal{K}_N)} \left| \frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \varphi_{G, \mathbf{k}}(\mathbf{x}_i) \right|^2$$

- $\mathcal{S}(\mathcal{K}_N)$ is a subset of \mathcal{S} with $k_i = 0, \pm 1, \dots, \pm \mathcal{K}_N$.
- The proposed estimator can be viewed as a GMM estimator based on $(2\mathcal{K}_N + 1)^{m+1}$ unconditional moments with the identity weighting matrix.

Identification:

$$\begin{aligned}\theta_o &= \operatorname{argmin}_{\theta \in \Theta} \int_{\mathcal{T}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) G(A(\mathbf{X}, \tau))] \right|^2 d\tau \\ &= \operatorname{argmin}_{\theta \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}} \left| \mathbb{E} [\mathbf{h}(\mathbf{Y}, \theta) \varphi_{G, \mathbf{k}}(\mathbf{X})] \right|^2.\end{aligned}$$

The Proposed Estimator:

$$\hat{\theta}(G, \mathcal{K}_N) = \operatorname{argmin}_{\theta \in \Theta} \sum_{\mathbf{k} \in \mathcal{S}(\mathcal{K}_N)} \left| \frac{1}{N} \sum_{i=1}^N \mathbf{h}(\mathbf{y}_i, \theta) \varphi_{G, \mathbf{k}}(\mathbf{x}_i) \right|^2$$

- $\mathcal{S}(\mathcal{K}_N)$ is a subset of \mathcal{S} with $k_i = 0, \pm 1, \dots, \pm \mathcal{K}_N$.
- The proposed estimator can be viewed as a GMM estimator based on $(2\mathcal{K}_N + 1)^{m+1}$ unconditional moments with the identity weighting matrix.

A Specific Estimator

Set $G(A(\mathbf{X}, \tau)) = \exp(A(\mathbf{X}, \tau))$.

- In the class of generically comprehensively revealing functions, $\exp(\tau \mathbf{X})$ and $\exp(A(\mathbf{X}, \tau))$ play the same role.
- The estimator is called the specific estimator.
- The exponential function results in an analytic form for $\varphi_k(\mathbf{X})$:

$$\begin{aligned}\varphi_{\exp, k}(\mathbf{X}) &= \int_{-\pi}^{\pi} \exp(\mathbf{X}' \tau) \exp(-ik' \tau) d\tau \\ &= \varphi_{\exp, k_1}(X_1) \cdots \varphi_{\exp, k_m}(X_m), \quad k \in \mathcal{S},\end{aligned}$$

where

$$\begin{aligned}\varphi_{\exp, k_j}(X_j) &= \int_{-\pi}^{\pi} \exp(X_j \tau_j) \exp(-ik_j \tau_j) d\tau_j \\ &= \frac{(-1)^{k_j} \cdot 2 \sinh(\pi X_j)}{(X_j - ik_j)}, \quad j = 1, \dots, m.\end{aligned}$$

A Specific Estimator

Set $G(A(\mathbf{X}, \tau)) = \exp(A(\mathbf{X}, \tau))$.

- In the class of generically comprehensively revealing functions, $\exp(\tau \mathbf{X})$ and $\exp(A(\mathbf{X}, \tau))$ play the same role.
We reduce the dimension of τ by one!
- The exponential function results in an analytic form for $\varphi_k(\mathbf{X})$:

$$\begin{aligned}\varphi_{\exp, k}(\mathbf{X}) &= \int_{\mathcal{T}} \exp(\mathbf{X}' \tau) \exp(-ik' \tau) d\tau \\ &= \varphi_{\exp, k_1}(X_1) \cdots \varphi_{\exp, k_m}(X_m), \quad k \in \mathcal{S},\end{aligned}$$

where

$$\begin{aligned}\varphi_{\exp, k_j}(X_j) &= \int_{-\pi}^{\pi} \exp(X_j \tau_j) \exp(-ik_j \tau_j) d\tau_j \\ &= \frac{(-1)^{k_j} \cdot 2 \sinh(\pi X_j)}{(X_j - ik_j)}, \quad j = 1, \dots, m.\end{aligned}$$

A Specific Estimator

Set $G(A(\mathbf{X}, \tau)) = \exp(A(\mathbf{X}, \tau))$.

- In the class of generically comprehensively revealing functions, $\exp(\tau \mathbf{X})$ and $\exp(A(\mathbf{X}, \tau))$ play the same role.
We reduce the dimension of τ by one!
- The exponential function results in an analytic form for $\varphi_{\mathbf{k}}(\mathbf{X})$:

$$\begin{aligned}\varphi_{\exp, \mathbf{k}}(\mathbf{X}) &= \int_{\mathcal{T}} \exp(\mathbf{X}' \tau) \exp(-i \mathbf{k}' \tau) d\tau \\ &= \varphi_{\exp, k_1}(X_1) \cdots \varphi_{\exp, k_m}(X_m), \quad \mathbf{k} \in \mathcal{S},\end{aligned}$$

where

$$\begin{aligned}\varphi_{\exp, k_j}(X_j) &= \int_{-\pi}^{\pi} \exp(X_j \tau_j) \exp(-ik_j \tau_j) d\tau_j \\ &= \frac{(-1)^{k_j} \cdot 2 \sinh(\pi X_j)}{(X_j - ik_j)}, \quad j = 1, \dots, m.\end{aligned}$$

A Specific Estimator

Set $G(A(\mathbf{X}, \tau)) = \exp(A(\mathbf{X}, \tau))$.

- In the class of generically comprehensively revealing functions, $\exp(\tau \mathbf{X})$ and $\exp(A(\mathbf{X}, \tau))$ play the same role.
We reduce the dimension of τ by one!
- The exponential function results in an analytic form for $\varphi_{\mathbf{k}}(\mathbf{X})$:

$$\begin{aligned}\varphi_{\exp, \mathbf{k}}(\mathbf{X}) &= \int_{\mathcal{T}} \exp(\mathbf{X}' \tau) \exp(-i \mathbf{k}' \tau) d\tau \\ &= \varphi_{\exp, k_1}(X_1) \cdots \varphi_{\exp, k_m}(X_m), \quad \mathbf{k} \in \mathcal{S},\end{aligned}$$

where

$$\begin{aligned}\varphi_{\exp, k_j}(X_j) &= \int_{-\pi}^{\pi} \exp(X_j \tau_j) \exp(-ik_j \tau_j) d\tau_j \\ &= \frac{(-1)^{k_j} \cdot 2 \sinh(\pi X_j)}{(X_j - ik_j)}, \quad j = 1, \dots, m.\end{aligned}$$

A Specific Estimator

Set $G(A(\mathbf{X}, \tau)) = \exp(A(\mathbf{X}, \tau))$.

- In the class of generically comprehensively revealing functions, $\exp(\tau \mathbf{X})$ and $\exp(A(\mathbf{X}, \tau))$ play the same role.
We reduce the dimension of τ by one!
- The exponential function results in an analytic form for $\varphi_{\mathbf{k}}(\mathbf{X})$:

$$\begin{aligned}\varphi_{\exp, \mathbf{k}}(\mathbf{X}) &= \int_{\mathcal{T}} \exp(\mathbf{X}' \tau) \exp(-i \mathbf{k}' \tau) d\tau \\ &= \varphi_{\exp, k_1}(X_1) \cdots \varphi_{\exp, k_m}(X_m), \quad \mathbf{k} \in \mathcal{S},\end{aligned}$$

where

$$\begin{aligned}\varphi_{\exp, k_j}(X_j) &= \int_{-\pi}^{\pi} \exp(X_j \tau_j) \exp(-ik_j \tau_j) d\tau_j \\ &= \frac{(-1)^{k_j} \cdot 2 \sinh(\pi X_j)}{(X_j - ik_j)}, \quad j = 1, \dots, m.\end{aligned}$$

Comparisons (I)

Carrasco and Florens (2000):

$$\hat{\theta}_{\text{CF}}(N) = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{k=1}^{\textcolor{red}{N}} \left| \frac{1}{T} \sum_{i=1}^N h(y_i, \theta) \psi_k(x_i) \right|^2,$$

$$\psi_k(x_i) = w_N(k) \int_{-\pi}^{\pi} \exp(\tau x_i) \phi_N^{(k)}(\tau) d\tau, \quad w_N(k) = \frac{\sqrt{\lambda_N^{(k)}}}{\sqrt{(\lambda_N^{(k)})^2 + \alpha_N}}$$

Comparisons (I)

Carrasco and Florens (2000):

$$\hat{\theta}_{\text{CF}}(N) = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{k=1}^N \left| \frac{1}{T} \sum_{i=1}^N h(y_i, \theta) \psi_k(x_i) \right|^2,$$

$$\psi_k(x_i) = w_N(k) \int_{-\pi}^{\pi} \exp(\tau x_i) \phi_N^{(k)}(\tau) d\tau, \quad w_N(k) = \frac{\sqrt{\lambda_N^{(k)}}}{\sqrt{(\lambda_N^{(k)})^2 + \alpha_N}}$$

Comparisons (I)

Carrasco and Florens (2000):

$$\hat{\theta}_{\text{CF}}(N) = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{k=1}^N \left| \frac{1}{T} \sum_{i=1}^N h(y_i, \theta) \psi_k(x_i) \right|^2,$$

$$\psi_k(x_i) = w_N(k) \int_{-\pi}^{\pi} \exp(\tau x_i) \phi_N^{(k)}(\tau) d\tau, \quad w_N(k) = \frac{\sqrt{\lambda_N^{(k)}}}{\sqrt{(\lambda_N^{(k)})^2 + \alpha_N}}$$

- A preliminary estimation of a covariance operator and its eigenvalues ($\lambda_N^{(k)}$) and eigen-functions ($\phi_N^{(k)}(\tau)$) is needed.
- A regularization parameter (α_N) is needed.

Comparisons (I)

Carrasco and Florens (2000):

$$\hat{\theta}_{\text{CF}}(N) = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{k=1}^N \left| \frac{1}{T} \sum_{i=1}^N h(y_i, \theta) \psi_k(x_i) \right|^2,$$

$$\psi_k(x_i) = w_N(k) \int_{-\pi}^{\pi} \exp(\tau x_i) \phi_N^{(k)}(\tau) d\tau, \quad w_N(k) = \frac{\sqrt{\lambda_N^{(k)}}}{\sqrt{(\lambda_N^{(k)})^2 + \alpha_N}}$$

- A preliminary estimation of a covariance operator and its eigenvalues ($\lambda_N^{(k)}$) and eigen-functions ($\phi_N^{(k)}(\tau)$) is needed.
- A regularization parameter (α_N) is needed.

Comparisons (I)

Carrasco and Florens (2000):

$$\hat{\theta}_{\text{CF}}(N) = \operatorname{argmin}_{\theta \in \Theta} \sum_{k=1}^N \left| \frac{1}{T} \sum_{i=1}^N h(y_i, \theta) \psi_k(x_i) \right|^2,$$

$$\psi_k(x_i) = w_N(k) \int_{-\pi}^{\pi} \exp(\tau x_i) \phi_N^{(k)}(\tau) d\tau, \quad w_N(k) = \frac{\sqrt{\lambda_N^{(k)}}}{\sqrt{(\lambda_N^{(k)})^2 + \alpha_N}}$$

Remark:

Our approach is much simpler, if consistency is our concern.

Comparisons (II)

Domínguez and Lobato (2004):

$$\hat{\theta}_{DL}(N) = \operatorname{argmin}_{\theta \in \Theta} \sum_{k=1}^N \left| \frac{1}{N} \sum_{i=1}^N h(y_i, \theta) I(x_i \leq \tau_k) \right|^2,$$

$$\tau_1 = x_1, \dots, \tau_k = x_k.$$

- The distribution of τ must be assumed identical to X .
- The indicator function, which takes only the values one and zero, may not well present the information in the conditioning variables.
- The indicator function is “comprehensively revealing”, not “generically comprehensively revealing”.

Comparisons (II)

Domínguez and Lobato (2004):

$$\hat{\theta}_{DL}(N) = \operatorname{argmin}_{\theta \in \Theta} \sum_{k=1}^N \left| \frac{1}{N} \sum_{i=1}^N h(y_i, \theta) I(x_i \leq \tau_k) \right|^2,$$

$$\tau_1 = x_1, \dots, \tau_k = x_k.$$

- The distribution of τ must be assumed identical to X .
- The indicator function, which takes only the values one and zero, may not well present the information in the conditioning variables.
- The indicator function is “comprehensively revealing”, not “generically comprehensively revealing”.

Comparisons (II)

Domínguez and Lobato (2004):

$$\hat{\theta}_{DL}(N) = \operatorname{argmin}_{\theta \in \Theta} \sum_{k=1}^N \left| \frac{1}{N} \sum_{i=1}^N h(y_i, \theta) I(x_i \leq \tau_k) \right|^2,$$

$$\tau_1 = x_1, \dots, \tau_k = x_k.$$

- The distribution of τ must be assumed identical to X .
- The indicator function, which takes only the values one and zero, may not well present the information in the conditioning variables.
- The indicator function is “comprehensively revealing”, not “generically comprehensively revealing”.

Comparisons (II)

Domínguez and Lobato (2004):

$$\hat{\theta}_{DL}(N) = \operatorname{argmin}_{\theta \in \Theta} \sum_{k=1}^N \left| \frac{1}{N} \sum_{i=1}^N h(y_i, \theta) I(x_i \leq \tau_k) \right|^2,$$

$$\tau_1 = x_1, \dots, \tau_k = x_k.$$

- The distribution of τ must be assumed identical to X .
- The indicator function, which takes only the values one and zero, may not well present the information in the conditioning variables.
- The indicator function is “comprehensively revealing”, **not** “generically comprehensively revealing”.

Asymptotic Theory

Weak Consistency:

If $\mathcal{K}_N = o(N^{1/(m+1)})$, then $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_N) \xrightarrow{P} \boldsymbol{\theta}_o$ as $N \rightarrow \infty$.

Asymptotic Normality:

If $\mathcal{K}_N = o(N^{1/(2m+2)})$, then

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_N) - \boldsymbol{\theta}_o) \xrightarrow{D} \mathcal{N}(0, \mathcal{V}),$$

where $\mathcal{V} = \mathcal{M}_q^{-1} \boldsymbol{\Omega}_q \mathcal{M}_q^{-1}$.

Simulations

- We compare the NLS estimator $\hat{\theta}_{\text{NLS}}$, the estimator of Domínguez and Lobato (2004): $\hat{\theta}_{\text{DL}}$, and the proposed estimator: $\hat{\theta}(\mathcal{K}_n)$ with $\mathcal{K}_n = 5$.
- Performance criteria: Bias, SE (standard error) and MSE (mean squared error).

Model in Domínguez and Lobato (2004)

- A nonlinear model:

$$Y = \theta_o^2 X + \theta_o X^2 + \epsilon, \quad \epsilon \sim N(0, 1),$$

where $\theta_o = 5/4$. The conditional moment restriction is
 $E[\epsilon|X] = 0$.

- Two cases: $X \sim N(0, 1)$ and $N(1, 1)$. Using the feasible optimal instrument to identify θ_o , the first case has a unique real solution $\theta_o = 5/4$, but the second case has 3 solutions: $5/4, -5/4$ and -3 .

Table: Model in Domínguez and Lobato (2004).

Sample <i>N</i>	Estimator	$X \sim N(0, 1)$			$X \sim N(1, 1)$		
		Bias	SE	MSE	Bias	SE	MSE
50	$\hat{\theta}_{NLS}$	-0.0006	0.0501	0.0025	-0.0083	0.1881	0.0354
	$\hat{\theta}_{DL}$	-0.0390	0.2282	0.0536	-0.0336	0.3667	0.1355
	$\hat{\theta}(K_N)$	-0.0061	0.1600	0.0256	-0.0249	0.3308	0.1100
	$\hat{\theta}_{OPIV}$	-0.2222	0.6288	0.4447	-1.6922	1.2783	4.4972
100	$\hat{\theta}_{NLS}$	-0.0004	0.0342	0.0012	-0.0071	0.1713	0.0294
	$\hat{\theta}_{DL}$	-0.0152	0.1541	0.0240	-0.0316	0.3595	0.1302
	$\hat{\theta}(K_N)$	-0.0059	0.1511	0.0228	-0.0217	0.3094	0.0962
	$\hat{\theta}_{OPIV}$	-0.1480	0.5096	0.2815	-1.7217	1.2619	4.5564
200	$\hat{\theta}_{NLS}$	-0.0004	0.0239	0.0006	-0.0025	0.1035	0.0107
	$\hat{\theta}_{DL}$	-0.0017	0.0864	0.0075	-0.0191	0.2796	0.0785
	$\hat{\theta}(K_N)$	-0.0045	0.1390	0.0193	-0.0116	0.2278	0.0520
	$\hat{\theta}_{OPIV}$	-0.0931	0.3994	0.1681	-1.6649	1.2859	4.4250

Table: Model in Domínguez and Lobato (2004).

Sample <i>N</i>	Estimator	$X \sim N(0, 1)$			$X \sim N(1, 1)$		
		Bias	SE	MSE	Bias	SE	MSE
50	$\hat{\theta}_{NLS}$	-0.0006	0.0501	0.0025	-0.0083	0.1881	0.0354
	$\hat{\theta}_{DL}$	-0.0390	0.2282	0.0536	-0.0336	0.3667	0.1355
	$\hat{\theta}(K_N)$	-0.0061	0.1600	0.0256	-0.0249	0.3308	0.1100
	$\hat{\theta}_{OPIV}$	-0.2222	0.6288	0.4447	-1.6922	1.2783	4.4972
100	$\hat{\theta}_{NLS}$	-0.0004	0.0342	0.0012	-0.0071	0.1713	0.0294
	$\hat{\theta}_{DL}$	-0.0152	0.1541	0.0240	-0.0316	0.3595	0.1302
	$\hat{\theta}(K_N)$	-0.0059	0.1511	0.0228	-0.0217	0.3094	0.0962
	$\hat{\theta}_{OPIV}$	-0.1480	0.5096	0.2815	-1.7217	1.2619	4.5564
200	$\hat{\theta}_{NLS}$	-0.0004	0.0239	0.0006	-0.0025	0.1035	0.0107
	$\hat{\theta}_{DL}$	-0.0017	0.0864	0.0075	-0.0191	0.2796	0.0785
	$\hat{\theta}(K_N)$	-0.0045	0.1390	0.0193	-0.0116	0.2278	0.0520
	$\hat{\theta}_{OPIV}$	-0.0931	0.3994	0.1681	-1.6649	1.2859	4.4250

Table: Model in Domínguez and Lobato (2004).

Sample N	Estimator	$X \sim N(0, 1)$			$X \sim N(1, 1)$		
		Bias	SE	MSE	Bias	SE	MSE
50	$\hat{\theta}_{NLS}$	-0.0006	0.0501	0.0025	-0.0083	0.1881	0.0354
	$\hat{\theta}_{DL}$	-0.0390	0.2282	0.0536	-0.0336	0.3667	0.1355
	$\hat{\theta}(K_N)$	-0.0061	0.1600	0.0256	-0.0249	0.3308	0.1100
	$\hat{\theta}_{OPIV}$	-0.2222	0.6288	0.4447	-1.6922	1.2783	4.4972
100	$\hat{\theta}_{NLS}$	-0.0004	0.0342	0.0012	-0.0071	0.1713	0.0294
	$\hat{\theta}_{DL}$	-0.0152	0.1541	0.0240	-0.0316	0.3595	0.1302
	$\hat{\theta}(K_N)$	-0.0059	0.1511	0.0228	-0.0217	0.3094	0.0962
	$\hat{\theta}_{OPIV}$	-0.1480	0.5096	0.2815	-1.7217	1.2619	4.5564
200	$\hat{\theta}_{NLS}$	-0.0004	0.0239	0.0006	-0.0025	0.1035	0.0107
	$\hat{\theta}_{DL}$	-0.0017	0.0864	0.0075	-0.0191	0.2796	0.0785
	$\hat{\theta}(K_N)$	-0.0045	0.1390	0.0193	-0.0116	0.2278	0.0520
	$\hat{\theta}_{OPIV}$	-0.0931	0.3994	0.1681	-1.6649	1.2859	4.4250

Other Experiments

- Model with an endogenous explanatory variable:

$$\begin{cases} Y = \theta_o^2 Z + \theta_o Z^2 + \epsilon \\ Z = X + \nu \end{cases}, \quad \begin{bmatrix} \epsilon \\ \nu \end{bmatrix} \sim N \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

where $\theta_o = 5/4$, and $X \sim N(0, 1)$ and is independent of ϵ and ν .

- Noisy disturbances:

$$Y = \theta_o^2 X + \theta_o X^2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2),$$

where $\theta_o = 5/4$, X is a uniform random variable on $(-1, 1)$, and $\sigma^2 = 0.1, 1, 4, 9$, and 16 .

Other Experiments

- Model with an endogenous explanatory variable:

$$\begin{cases} Y = \theta_o^2 Z + \theta_o Z^2 + \epsilon \\ Z = X + \nu \end{cases}, \quad \begin{bmatrix} \epsilon \\ \nu \end{bmatrix} \sim N \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

where $\theta_o = 5/4$, and $X \sim N(0, 1)$ and is independent of ϵ and ν .

- Noisy disturbances:

$$Y = \theta_o^2 X + \theta_o X^2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2),$$

where $\theta_o = 5/4$, X is a uniform random variable on $(-1, 1)$, and $\sigma^2 = 0.1, 1, 4, 9$, and 16 .

Table: Models with an endogenous explanatory variable.

ρ	Est.	$N = 50$			$N = 100$			$N = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.01	$\hat{\theta}_{NLS}$	0.0009	0.0317	0.0010	0.0005	0.0212	0.0004	0.0011	0.0146	0.0002
	$\hat{\theta}_{DL}$	-0.0103	0.1165	0.0137	-0.0062	0.0809	0.0066	-0.0027	0.0561	0.0032
	$\hat{\theta}(\mathcal{K}_N)$	-0.0003	0.0561	0.0031	-0.0009	0.0365	0.0013	0.0001	0.0245	0.0006
0.1	$\hat{\theta}_{NLS}$	0.0097	0.0313	0.0011	0.0102	0.0210	0.0005	0.0103	0.0146	0.0003
	$\hat{\theta}_{DL}$	-0.0116	0.1153	0.0134	-0.0069	0.0816	0.0067	-0.0036	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_N)$	-0.0021	0.0550	0.0030	-0.0010	0.0358	0.0013	-0.0006	0.0242	0.0006
0.3	$\hat{\theta}_{NLS}$	0.0315	0.0310	0.0020	0.0311	0.0209	0.0014	0.0315	0.0144	0.0012
	$\hat{\theta}_{DL}$	-0.0125	0.1214	0.0149	-0.0061	0.0819	0.0067	-0.0032	0.0585	0.0034
	$\hat{\theta}(\mathcal{K}_N)$	-0.0039	0.0565	0.0032	-0.0016	0.0358	0.0013	-0.0002	0.0244	0.0006
0.5	$\hat{\theta}_{NLS}$	0.0539	0.0311	0.0039	0.0527	0.0207	0.0032	0.0520	0.0143	0.0029
	$\hat{\theta}_{DL}$	-0.0125	0.1231	0.0153	-0.0045	0.0817	0.0067	-0.0017	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_N)$	-0.0056	0.0596	0.0036	-0.0021	0.0366	0.0013	-0.0010	0.0247	0.0006
0.7	$\hat{\theta}_{NLS}$	0.0746	0.0298	0.0064	0.0739	0.0196	0.0058	0.0731	0.0140	0.0055
	$\hat{\theta}_{DL}$	-0.0153	0.1242	0.0156	-0.0083	0.0840	0.0071	-0.0053	0.0588	0.0035
	$\hat{\theta}(\mathcal{K}_N)$	-0.0097	0.0574	0.0034	-0.0038	0.0366	0.0014	-0.0020	0.0247	0.0006
0.9	$\hat{\theta}_{NLS}$	0.0972	0.0285	0.0103	0.0953	0.0190	0.0094	0.0942	0.0134	0.0091
	$\hat{\theta}_{DL}$	-0.0166	0.1288	0.0169	-0.0086	0.0845	0.0072	-0.0042	0.0598	0.0036
	$\hat{\theta}(\mathcal{K}_N)$	-0.0117	0.0947	0.0091	-0.0053	0.0370	0.0014	-0.0019	0.0250	0.0006

Table: Models with an endogenous explanatory variable.

ρ	Est.	$N = 50$			$N = 100$			$N = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.01	$\hat{\theta}_{NLS}$	0.0009	0.0317	0.0010	0.0005	0.0212	0.0004	0.0011	0.0146	0.0002
	$\hat{\theta}_{DL}$	-0.0103	0.1165	0.0137	-0.0062	0.0809	0.0066	-0.0027	0.0561	0.0032
	$\hat{\theta}(\mathcal{K}_N)$	-0.0003	0.0561	0.0031	-0.0009	0.0365	0.0013	0.0001	0.0245	0.0006
0.1	$\hat{\theta}_{NLS}$	0.0097	0.0313	0.0011	0.0102	0.0210	0.0005	0.0103	0.0146	0.0003
	$\hat{\theta}_{DL}$	-0.0116	0.1153	0.0134	-0.0069	0.0816	0.0067	-0.0036	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_N)$	-0.0021	0.0550	0.0030	-0.0010	0.0358	0.0013	-0.0006	0.0242	0.0006
0.3	$\hat{\theta}_{NLS}$	0.0315	0.0310	0.0020	0.0311	0.0209	0.0014	0.0315	0.0144	0.0012
	$\hat{\theta}_{DL}$	-0.0125	0.1214	0.0149	-0.0061	0.0819	0.0067	-0.0032	0.0585	0.0034
	$\hat{\theta}(\mathcal{K}_N)$	-0.0039	0.0565	0.0032	-0.0016	0.0358	0.0013	-0.0002	0.0244	0.0006
0.5	$\hat{\theta}_{NLS}$	0.0539	0.0311	0.0039	0.0527	0.0207	0.0032	0.0520	0.0143	0.0029
	$\hat{\theta}_{DL}$	-0.0125	0.1231	0.0153	-0.0045	0.0817	0.0067	-0.0017	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_N)$	-0.0056	0.0596	0.0036	-0.0021	0.0366	0.0013	-0.0010	0.0247	0.0006
0.7	$\hat{\theta}_{NLS}$	0.0746	0.0298	0.0064	0.0739	0.0196	0.0058	0.0731	0.0140	0.0055
	$\hat{\theta}_{DL}$	-0.0153	0.1242	0.0156	-0.0083	0.0840	0.0071	-0.0053	0.0588	0.0035
	$\hat{\theta}(\mathcal{K}_N)$	-0.0097	0.0574	0.0034	-0.0038	0.0366	0.0014	-0.0020	0.0247	0.0006
0.9	$\hat{\theta}_{NLS}$	0.0972	0.0285	0.0103	0.0953	0.0190	0.0094	0.0942	0.0134	0.0091
	$\hat{\theta}_{DL}$	-0.0166	0.1288	0.0169	-0.0086	0.0845	0.0072	-0.0042	0.0598	0.0036
	$\hat{\theta}(\mathcal{K}_N)$	-0.0117	0.0947	0.0091	-0.0053	0.0370	0.0014	-0.0019	0.0250	0.0006

Table: Models with an endogenous explanatory variable.

ρ	Est.	$N = 50$			$N = 100$			$N = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.01	$\hat{\theta}_{NLS}$	0.0009	0.0317	0.0010	0.0005	0.0212	0.0004	0.0011	0.0146	0.0002
	$\hat{\theta}_{DL}$	-0.0103	0.1165	0.0137	-0.0062	0.0809	0.0066	-0.0027	0.0561	0.0032
	$\hat{\theta}(\mathcal{K}_N)$	-0.0003	0.0561	0.0031	-0.0009	0.0365	0.0013	0.0001	0.0245	0.0006
0.1	$\hat{\theta}_{NLS}$	0.0097	0.0313	0.0011	0.0102	0.0210	0.0005	0.0103	0.0146	0.0003
	$\hat{\theta}_{DL}$	-0.0116	0.1153	0.0134	-0.0069	0.0816	0.0067	-0.0036	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_N)$	-0.0021	0.0550	0.0030	-0.0010	0.0358	0.0013	-0.0006	0.0242	0.0006
0.3	$\hat{\theta}_{NLS}$	0.0315	0.0310	0.0020	0.0311	0.0209	0.0014	0.0315	0.0144	0.0012
	$\hat{\theta}_{DL}$	-0.0125	0.1214	0.0149	-0.0061	0.0819	0.0067	-0.0032	0.0585	0.0034
	$\hat{\theta}(\mathcal{K}_N)$	-0.0039	0.0565	0.0032	-0.0016	0.0358	0.0013	-0.0002	0.0244	0.0006
0.5	$\hat{\theta}_{NLS}$	0.0539	0.0311	0.0039	0.0527	0.0207	0.0032	0.0520	0.0143	0.0029
	$\hat{\theta}_{DL}$	-0.0125	0.1231	0.0153	-0.0045	0.0817	0.0067	-0.0017	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_N)$	-0.0056	0.0596	0.0036	-0.0021	0.0366	0.0013	-0.0010	0.0247	0.0006
0.7	$\hat{\theta}_{NLS}$	0.0746	0.0298	0.0064	0.0739	0.0196	0.0058	0.0731	0.0140	0.0055
	$\hat{\theta}_{DL}$	-0.0153	0.1242	0.0156	-0.0083	0.0840	0.0071	-0.0053	0.0588	0.0035
	$\hat{\theta}(\mathcal{K}_N)$	-0.0097	0.0574	0.0034	-0.0038	0.0366	0.0014	-0.0020	0.0247	0.0006
0.9	$\hat{\theta}_{NLS}$	0.0972	0.0285	0.0103	0.0953	0.0190	0.0094	0.0942	0.0134	0.0091
	$\hat{\theta}_{DL}$	-0.0166	0.1288	0.0169	-0.0086	0.0845	0.0072	-0.0042	0.0598	0.0036
	$\hat{\theta}(\mathcal{K}_N)$	-0.0117	0.0947	0.0091	-0.0053	0.0370	0.0014	-0.0019	0.0250	0.0006

Table: Models with an endogenous explanatory variable.

ρ	Est.	$N = 50$			$N = 100$			$N = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.01	$\hat{\theta}_{NLS}$	0.0009	0.0317	0.0010	0.0005	0.0212	0.0004	0.0011	0.0146	0.0002
	$\hat{\theta}_{DL}$	-0.0103	0.1165	0.0137	-0.0062	0.0809	0.0066	-0.0027	0.0561	0.0032
	$\hat{\theta}(K_N)$	-0.0003	0.0561	0.0031	-0.0009	0.0365	0.0013	0.0001	0.0245	0.0006
0.1	$\hat{\theta}_{NLS}$	0.0097	0.0313	0.0011	0.0102	0.0210	0.0005	0.0103	0.0146	0.0003
	$\hat{\theta}_{DL}$	-0.0116	0.1153	0.0134	-0.0069	0.0816	0.0067	-0.0036	0.0570	0.0033
	$\hat{\theta}(K_N)$	-0.0021	0.0550	0.0030	-0.0010	0.0358	0.0013	-0.0006	0.0242	0.0006
0.3	$\hat{\theta}_{NLS}$	0.0315	0.0310	0.0020	0.0311	0.0209	0.0014	0.0315	0.0144	0.0012
	$\hat{\theta}_{DL}$	-0.0125	0.1214	0.0149	-0.0061	0.0819	0.0067	-0.0032	0.0585	0.0034
	$\hat{\theta}(K_N)$	-0.0039	0.0565	0.0032	-0.0016	0.0358	0.0013	-0.0002	0.0244	0.0006
0.5	$\hat{\theta}_{NLS}$	0.0539	0.0311	0.0039	0.0527	0.0207	0.0032	0.0520	0.0143	0.0029
	$\hat{\theta}_{DL}$	-0.0125	0.1231	0.0153	-0.0045	0.0817	0.0067	-0.0017	0.0570	0.0033
	$\hat{\theta}(K_N)$	-0.0056	0.0596	0.0036	-0.0021	0.0366	0.0013	-0.0010	0.0247	0.0006
0.7	$\hat{\theta}_{NLS}$	0.0746	0.0298	0.0064	0.0739	0.0196	0.0058	0.0731	0.0140	0.0055
	$\hat{\theta}_{DL}$	-0.0153	0.1242	0.0156	-0.0083	0.0840	0.0071	-0.0053	0.0588	0.0035
	$\hat{\theta}(K_N)$	-0.0097	0.0574	0.0034	-0.0038	0.0366	0.0014	-0.0020	0.0247	0.0006
0.9	$\hat{\theta}_{NLS}$	0.0972	0.0285	0.0103	0.0953	0.0190	0.0094	0.0942	0.0134	0.0091
	$\hat{\theta}_{DL}$	-0.0166	0.1288	0.0169	-0.0086	0.0845	0.0072	-0.0042	0.0598	0.0036
	$\hat{\theta}(K_N)$	-0.0117	0.0947	0.0091	-0.0053	0.0370	0.0014	-0.0019	0.0250	0.0006

Table: Models with different disturbance variances.

σ^2	Est.	$N = 50$			$N = 100$			$N = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\sigma^2 = 0.1$	$\hat{\theta}_{NLS}$	-0.2827	0.7836	0.6939	-0.2511	0.7392	0.6094	-0.2523	0.7433	0.6160
	$\hat{\theta}_{DL}$	-0.4645	0.8678	0.9687	-0.3938	0.8108	0.8124	-0.3701	0.7900	0.7610
	$\hat{\theta}(K_N)$	-0.1129	0.5687	0.3361	-0.0491	0.4016	0.1637	-0.0317	0.3467	0.1212
$\sigma^2 = 1$	$\hat{\theta}_{NLS}$	-0.5845	1.0572	1.4591	-0.4451	0.9481	1.0968	-0.3089	0.8158	0.7608
	$\hat{\theta}_{DL}$	-0.9491	1.0820	2.0711	-0.7899	1.0236	1.6715	-0.6692	0.9776	1.4033
	$\hat{\theta}(K_N)$	-0.3478	1.1121	1.3574	-0.1724	0.8191	0.7005	-0.0880	0.5790	0.3429
$\sigma^2 = 4$	$\hat{\theta}_{NLS}$	-0.8508	1.2109	2.1899	-0.7320	1.1470	1.8513	-0.5897	1.0582	1.4673
	$\hat{\theta}_{DL}$	-1.2441	1.1875	2.9576	-1.1126	1.1187	2.4891	-0.9874	1.0752	2.1307
	$\hat{\theta}(K_N)$	-0.5081	1.5507	2.6623	-0.3767	1.3357	1.9257	-0.2599	1.0442	1.1577
$\sigma^2 = 9$	$\hat{\theta}_{NLS}$	-0.9452	1.2880	2.5519	-0.8877	1.2281	2.2960	-0.7253	1.1491	1.8463
	$\hat{\theta}_{DL}$	-1.3698	1.2738	3.4985	-1.2821	1.1920	3.0644	-1.1359	1.1210	2.5467
	$\hat{\theta}(K_N)$	-0.5013	1.8507	3.6759	-0.4600	1.6048	2.7864	-0.3355	1.3225	1.8612
$\sigma^2 = 16$	$\hat{\theta}_{NLS}$	-1.0299	1.3672	2.9295	-0.9329	1.2814	2.5121	-0.8481	1.2134	2.1915
	$\hat{\theta}_{DL}$	-1.4794	1.3772	4.0848	-1.3361	1.2439	3.3321	-1.2542	1.1681	2.9371
	$\hat{\theta}(K_N)$	-0.5085	2.1182	4.7443	-0.3882	1.7852	3.3372	-0.3912	1.5377	2.5172

Table: Models with different disturbance variances.

σ^2	Est.	$N = 50$			$N = 100$			$N = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\sigma^2 = 0.1$	$\hat{\theta}_{NLS}$	-0.2827	0.7836	0.6939	-0.2511	0.7392	0.6094	-0.2523	0.7433	0.6160
	$\hat{\theta}_{DL}$	-0.4645	0.8678	0.9687	-0.3938	0.8108	0.8124	-0.3701	0.7900	0.7610
	$\hat{\theta}(K_N)$	-0.1129	0.5687	0.3361	-0.0491	0.4016	0.1637	-0.0317	0.3467	0.1212
$\sigma^2 = 1$	$\hat{\theta}_{NLS}$	-0.5845	1.0572	1.4591	-0.4451	0.9481	1.0968	-0.3089	0.8158	0.7608
	$\hat{\theta}_{DL}$	-0.9491	1.0820	2.0711	-0.7899	1.0236	1.6715	-0.6692	0.9776	1.4033
	$\hat{\theta}(K_N)$	-0.3478	1.1121	1.3574	-0.1724	0.8191	0.7005	-0.0880	0.5790	0.3429
$\sigma^2 = 4$	$\hat{\theta}_{NLS}$	-0.8508	1.2109	2.1899	-0.7320	1.1470	1.8513	-0.5897	1.0582	1.4673
	$\hat{\theta}_{DL}$	-1.2441	1.1875	2.9576	-1.1126	1.1187	2.4891	-0.9874	1.0752	2.1307
	$\hat{\theta}(K_N)$	-0.5081	1.5507	2.6623	-0.3767	1.3357	1.9257	-0.2599	1.0442	1.1577
$\sigma^2 = 9$	$\hat{\theta}_{NLS}$	-0.9452	1.2880	2.5519	-0.8877	1.2281	2.2960	-0.7253	1.1491	1.8463
	$\hat{\theta}_{DL}$	-1.3698	1.2738	3.4985	-1.2821	1.1920	3.0644	-1.1359	1.1210	2.5467
	$\hat{\theta}(K_N)$	-0.5013	1.8507	3.6759	-0.4600	1.6048	2.7864	-0.3355	1.3225	1.8612
$\sigma^2 = 16$	$\hat{\theta}_{NLS}$	-1.0299	1.3672	2.9295	-0.9329	1.2814	2.5121	-0.8481	1.2134	2.1915
	$\hat{\theta}_{DL}$	-1.4794	1.3772	4.0848	-1.3361	1.2439	3.3321	-1.2542	1.1681	2.9371
	$\hat{\theta}(K_N)$	-0.5085	2.1182	4.7443	-0.3882	1.7852	3.3372	-0.3912	1.5377	2.5172

The performances of various \mathcal{K}_n

- When \mathcal{K}_n increases, SE decreases in all cases while the bias also does in most cases.
- The percentages of the change are, however, quite slight (less than 0.1%) in all cases when \mathcal{K}_n is greater than 5.

Remark:

In practice, the information in identifying θ_0 is indeed efficiently incorporated in the first few φ_k of the proposed estimator, even though \mathcal{K}_n equal to infinity is needed in asymptotics.

The performances of various \mathcal{K}_N

- When \mathcal{K}_N increases, SE decreases in all cases while the bias also does in most cases.
- The percentages of the change are, however, quite slight (less than 0.1%) in all cases when \mathcal{K}_N is greater than 5.

Remark:

In practice, the information in identifying θ_0 is indeed efficiently incorporated in the first few φ_k of the proposed estimator, even though \mathcal{K}_N equal to infinity is needed in asymptotics.