A NPIV Method:Penalized Seive Extremum Estimation

Presenter Qiaoling Li

School of Statistics and Management
Shanghai University of Finance and Economics, China

February 15, 2023

Outline

Introduction I

Economic models ofen imply a set of semi/nonparametric conditional moment restrictions of the following form:

$$E\left[\rho\left(Y, X; \theta_0, h_0\right) \mid X\right] = 0 \quad \text{a.s.} \; - X \tag{1}$$

Where

- $\rho(\cdot; \theta_0, h_0)$ is a $d_p \times 1$ vector of generalized residual function whose functional form are known up to the true but unknown parameter (θ_0', h_0);
- Y is a vector of endogenous variables;
- X is a vector of instrumental variables.

Introduction II

Let $\alpha \equiv (\theta', h) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$ denote the parameters of interest, with $\theta \in \Theta$ a $d_\theta \times 1$ vector of finite-dimensional parameters and $h \equiv (h_1(), \ldots, h_q()) \in \mathcal{H}$ a $1 \times d_q$ vector valued function.

The model in Equation 1 nests many widely used semi/nonparametric generalized regression models.

eg: NPIV regression:

$$E\left[Y_1 - h_0(Y_2) \mid X\right] = 0 \quad \text{a.s.} \quad -X$$

NPQIV regression:

$$E\left[1\left\{Y_1 - h_0(Y_2)\right\} - \gamma \mid X\right] = 0 \quad \text{a.s.} \quad -X$$

Let $\left\{Z_i \equiv (Y_i', X_i')\right\}_{i=1}^n$ be a random sample from the probability distribution P_0 of $Z \equiv (Y', X')'$ that satisfies the conditional monetbrestrictions in Equation (1).

Let the infinite-dimensional function space $\mathcal{A} \equiv \Theta \times \mathcal{H}$ be endowed with a metric $\|\cdot\|_s = \|\cdot\|_e + \|\cdot\|_H$, where :

Introduction III

- $\|\cdot\|_e$ is the Euclidean norm on Θ,
- $\|\cdot\|_H$ denotes a norm on the infinite-dimensional space $\mathcal{H}$.

Set of parameters identified by the model in Equation(1)

$$\mathcal{A}_I(P_0) \equiv \{\alpha \equiv (\theta', b) \in (\mathcal{A}, \|\cdot\|_s) : E[\rho(Y, X; \alpha) \mid X] = 0 \quad \text{a.s.} \; - X\}$$

When $\mathcal{A}_I(P_0) = \{\alpha_0 \equiv (\theta_0', h)\}$ is a singleton in $(\mathcal{A}\;\|\cdot\|_s)$, the parameter α_0 is identified by the model in Equation (1).
One can consider estimation of the identified set $\mathcal{A}_I(P_0)$ by recasting it as a set of minimizers to a nonrandom criterion function $Q(\cdot) : (\mathcal{A}, \|\cdot\|_s) \to \mathbb{R}$ such that $Q(\alpha) = 0$ whenever $\alpha \in \mathcal{A}_I(P_0)$, and $Q(\alpha) > 0$ for all $\alpha \in \mathcal{A} \backslash \mathcal{A}_I(P_0)$.

Introduction IV

Estimate $\mathcal{A}_I(P_0)$

- Let $\hat{Q}_n$ be a random criterion function that convergence to Q in probability uniformly over totally bounded subsets of $(\mathcal{A}, \|\cdot\|_s)$.
- Estimate $\mathcal{A}_I(P_0)$ by an extremum estimator :$\arg\inf_{\alpha\in\mathcal{A}} \hat{Q}_n(\alpha)$.

Introduction V

Ill-posed problem:
Because

- the parameter space $\mathcal{A}$ is infinite dimensional and possibly noncompact in $\|\cdot\|_s$, $\arg\inf_{\alpha\in\mathcal{A}} \hat{Q}_n(\alpha)$ may be difficult to compute and not well defined,
- or even if it exists, it may be inconsistent for $\mathcal{A}_I(P_0)$ under $\|\cdot\|_s$ when $\mathcal{A}$ is not compact in $\|\cdot\|_s$

The method of sieves and method of penalization are two general approaches to solve possibly ill-posed, infinite-dimensional optimization problems.

Penalized Sieve Extremum Estimation I

- Method of sieves replaces $\inf_{\alpha\in\mathcal{A}} \hat{Q}_n(\alpha)$ by $\inf_{\alpha\in\mathcal{A}_{k(n)}} \hat{Q}_n(\alpha)$, where the sieves $\mathcal{A}_{k(n)}$ are a sequence of approximating parameter spaces that are less complex but dense in $(\mathcal{A}, \|\cdot\|_s)$.
- Method of penalization replaces $\inf_{\alpha\in\mathcal{A}} \hat{Q}_n(\alpha)$ by $\inf_{\alpha\in\mathcal{A}} \left\{\hat{Q}_n(\alpha) + \lambda_n \hat{P}_n(h)\right\}$.
- Chen Pouzo(2012) and Chen (2013) introduce a class of penalized sieve extremun(PSE) estimators, $\widehat{\alpha}_n = \left(\widehat{\theta}_n, \widehat{b}_n\right) \in \mathcal{A}_{k(n)} = \Theta \times \mathcal{H}_{k(n)}$, defined by

$$\left\{\widehat{Q}_n\left(\widehat{\alpha}_n\right) + \lambda_n \widehat{P}_n\left(\widehat{b}_n\right)\right\} \leq \inf_{\alpha\in\Theta\times\mathcal{H}_{k(n)}} \left\{\widehat{Q}_n(\alpha) + \lambda_n \widehat{P}_n(b)\right\} \tag{2}$$

Penalized Sieve Extremum Estimation II

In equation(2) :

- $\mathcal{H}_{k(n)}$ is a sieve parameter space whose compexity, denoted by $k(n) \equiv dim(\mathcal{H}_{k(n)})$, grows with sample size n and becomes dense in the original function space $\mathcal{H}$ under the metric $\|\cdot\|_H$.
- $\lambda_n \geq 0$ is a penalization parameter such that $\lambda_n \to 0$ as $n \to \infty$.
- penalty $\hat{p}_n(\cdot) \leq 0$, which is an empirical analog of a nonrandom penalty function $Pen : \mathcal{H} \to [0, +\infty)$, is jointly measurable in h and the data $\{Z_t\}_{t=1}^n$.

Penalized Sieve Extremum Estimation III

Sieve Space :$\mathcal{H}_{k(n)}$

Commonly used finite-dimensional linear sieves take the following form:

$$\mathcal{H}_{k(n)} = \left\{ h \in \mathcal{H} : h(\cdot) = \sum_{k=1}^{k(n)} \pi_k q_k(\cdot) \right\}, \tag{3}$$

$k(n) \leq \infty, k(n) \to \infty$ slowly as $n \to \infty$,
where $\{q_k\}_{k=1}^{\infty}$ is a sequence of known basis functions of a Banach space $(\mathcal{H}, \|\cdot\|_H)$ such as polynomial splines, B-splines, wavelets, Fourier series, Hermite polynomial series, power series, and Chebyshev series.

Penalized Sieve Extremum Estimation IV

Constrained version

$$\mathcal{H}_{k(n)} = \left\{ h \in \mathcal{H} : h(\cdot) = \sum_{k=1}^{k(n)} \pi_k q_k(\cdot), R_n(h) \leq B_n \right\},$$

$k(n) \leq \infty, B(n) \to \infty$ slowly as $n \to \infty$,
where the constraint $R_n(h) \leq B_n$ reflects prior information about $h_0 \in \mathcal{H}$ such as smoothness properties.

Penalized Sieve Extremum Estimation V

Penalty function $Pen(\cdot)$

The penalty function $Pen(\cdot)$ is typically convex or lower semicompact [i.e., the set $\{h \in \mathcal{H} : Pen(h) \leq M\}$ is compact in $(\mathbf{H}, \|\cdot\|_H)$ for all $M \in (0, \infty)$] and reflects prior information about $h_0 \in \mathcal{H}$.

Criteria function I

Criteria based on nonparametrically estimated conditional moments.

Let $m(X,\alpha) \equiv E[\rho(Y,X;\alpha)|X]$ be the $d_\rho \times 1$ conditional mean function of the residual function $\rho(Y,X;\alpha)$, and $\Sigma(X)$ be any $d_\rho \times d_\rho$ positive-definite weighting matrix. Then the conditional moment restrictions model in Equation 1 is equivalent to $\| [\Sigma(\cdot)]^{-1/2} m(\cdot,\alpha) \|_{L^p(X)} = 0$, when $\alpha = \alpha_0$

- SMD estimation:
  $\min_{\alpha\in\Theta\times\mathcal{H}_{k(n)}} Q_n(\alpha), Q_n(\alpha) = \frac{1}{n}\sum_{t=1}^{n} \hat{m}(X_t,\alpha)' \left\{\Sigma(X)\right\}^{-1} \hat{m}(X_t,\alpha)$,
  where $\hat{m}(X_t,\alpha)$ and $\Sigma(X)$ are any consistent estimators of $\hat{m}(X_t,\alpha)$ and $\hat{\Sigma}(X)$, respectively.

Criteria function II

- PSMD Estimation:

$$\widehat{Q}_n(\widehat{\alpha}) \leq \inf_{\alpha \in \Theta \times \mathcal{H}_{k(n)}} \widehat{Q}_n(\alpha) + \widehat{\eta}_n \tag{4}$$

with $\widehat{\eta}_n \geq 0, \widehat{\eta}_n = O_p(\eta_n)$ and

$$\widehat{Q}_n(\alpha) \equiv \frac{1}{n} \sum_{t=1}^{n} \widehat{m}(X_t, \alpha)' \widehat{W}(X_t) \widehat{m}(X_t, \alpha) + \lambda_n \widehat{P}_n(\alpha),$$

where $\{\eta_n\}_{n=1}^{\infty}$ is a sequence of positive real values such that $\eta_n = o(1)$, and $W(X)$ is a consistent estimator of $W(X)$ that is introduced to address potential heteroskedasticity.

Criteria function III

To solve the ill-posed inverse problem, the PSMD procedure effec-tively combines two types of regularization methods: the regularization bysieves and the regularization by penalization. The family of PSMD procedures consists of two broad subclasses:

- PSMD using slowly growing finite-dimensional sieves ($k(n)/n \to 0$), with small flexible penalty ($\lambda_n P(\cdot) \searrow 0$ fast) or zero penalty ($\lambda_n P(\cdot) = 0$); and
- PSMD using large dimensional sieves($k(n)/n ß const. > 0$), with positive penalty ($\lambda_n P(\cdot) > 0$) that is convex and/orlower semicompact.

The first subclass of PSMD procedures mainly follows the regularization by sieves approach, while the second subclass adopts the regularization by penalizing criterion function approach.

Criteria function IV

Which Class of PSMD Estimators to Use?

In most economics applications, the unknown structural function h_0 is Hölder continuous or has continuous derivatives or satisfies some shape restrictions (such as monotonicity or concavity). To estimate such smooth functions for the model (1), we recommend two classes of PSMD estimators.

- The class of PSMD estimators using slowly growing finite-dimensional sieves with/without small flexible penalty ($k(n) \to \infty$ slowly, $k(n)/n \to 0$; $\lambda_n \searrow 0$ fast or $\lambda_n = 0$).
- The class of PSMD estimators using faster growing finite-dimensional sieves with large lower semicompact penalty ($k(n) \to \infty$ faster, $k(n)/n \to 0$; $\lambda_n = O(k(n)/n)$).

Between these two, the subclass of PSMD estimators using slowly growing finite-dimensional linear sieves (i.e., series (3)) with small flexible penalty is our favorite since it is easier to compute and performs very well in finite samples.

Nonparametric estimation of $m(\cdot, \alpha)$ I

To compute the PSMD estimator $\widehat{\alpha}$ defined in (4), nonparametric estimators of the conditional mean function $m(\cdot\alpha) \equiv E[\rho(Z, \alpha)|X = \cdot]$ is needed.
We use a series LS estimator:

$$\widehat{m}(X, \alpha) = p^{J_n}(X)'(P'P)^{-} \sum_{i=1}^{n} p^{J_n}(X_i)\rho(Z_i, \alpha), \tag{5}$$

where

- $\{p_j(\cdot)\}_{j=1}^{\infty}$ is a sequence of known basis functions that can approximate any square integrable function of X well;
- J_n is the number of approximating terms such that $J_n \to \infty$ slowly as $n \to \infty$;
- $p^{J_n}(X) = (p_1(X), \cdots, p_{J_n}(X))'$, $P = (p^{J_n}(X_1), \cdots, p^{J_n}(X_n))'$
- $(P'P)$ is the generalized inverse of the matrix $P'P$

Nonparametric Additive Quantile IV Regression

$$Y_3 = h_{01}(Y_1) + h_{02}(Y_2) + U, Pr(U \leq 0|X) = \gamma,$$

- $\gamma \in (0, 1)$ is known fixed;
- Map it into general model:
  - let $Z = (Y', X')'$,
  - $h = (h_!, h_2)$,
  - $\rho(Z, h) = 1\{Y_{31}(Y_1) + h_2(Y_2)\} - \gamma$,
  - $m(X, h) = E[F_{Y_3|Y_1,Y_2,X}(h_1(Y_1) + h_2(Y_2))|X] - \gamma$

Application to Nonparametric Additive Quantile IV Regression II

We estimate $h_0 = (h_{01}, h_{02}) \in \mathcal{H} = \mathcal{H}^1 \times \mathcal{H}^2$ using the PSMD estimator $\widehat{h}_n$ given in (4),with

- $\widehat{W} = W = I$,
- $\mathcal{H}_n = \mathcal{H}^1 \times \mathcal{H}^2$ being either a finite-dimensional $(dim(\mathcal{H}_n) \equiv k(n) = k_1(n) + k_2(n)) < \infty$ or an infinite-dimensional $(k(n) = \infty)$ linear sieve,
- $\widehat{P}_n(h) = P(h) \geq 0$,
- the conditional mean function $m(X, h)$ is estimated by the series LS estimator $\widehat{m}(X, h)$ defined in(5).

ACKNOWLEDGEMENT

Thank you all for your attention!