ON ESTIMATION OF SEMI/NONPARAMETRIC CONDITIONAL MOMENT MODELS

Xiaohong Chen (Yale University)

Talk Based on Two Papers

- Chen, X. and D. Pouzo (08): “Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals”.
- Chen, X. and D. Pouzo (07): “Estimation of nonparametric conditional moment models with possibly nonsmooth moments”.

Talk Based on Two Papers

- Chen, X. and D. Pouzo (08): “Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals”.
- Chen, X. and D. Pouzo (07): “Estimation of nonparametric conditional moment models with possibly nonsmooth moments”.

- Closely related work
  - Blundell, Chen and Kristensen (07, Econometrica) on shape-invariant semi/nonparametric Engel curves with endogenous total expenditure.
  - Ai and Chen (03, Econometrica) on efficient estimation with smooth residuals.
  - Chen (07, Handbook of Econometrics, vol. 6B) survey on method of sieves.

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.
- Examples, Brief Literature Review.

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.
- Examples, Brief Literature Review.
- Penalized Sieve Minimum Distance (SMD) Estimation.

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.
- Examples, Brief Literature Review.
- Penalized Sieve Minimum Distance (SMD) Estimation.
- Simulation and Empirical Illustration

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.
- Examples, Brief Literature Review.
- Penalized Sieve Minimum Distance (SMD) Estimation.
- Simulation and Empirical Illustration
- Asymptotic Properties of Penalized SMD Estimators
  - Convergence Rate of Nonparametric Parts.
  - Asymptotic Normality of Smooth Functionals.
  - Semiparametric Efficiency, Confidence Region.

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.
- Examples, Brief Literature Review.
- Penalized Sieve Minimum Distance (SMD) Estimation.
- Simulation and Empirical Illustration
- Asymptotic Properties of Penalized SMD Estimators
  - Convergence Rate of Nonparametric Parts.
  - Asymptotic Normality of Smooth Functionals.
  - Semiparametric Efficiency, Confidence Region.
- Ex 1: Partially Linear Quantile IV Regression

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.
- Examples, Brief Literature Review.
- Penalized Sieve Minimum Distance (SMD) Estimation.
- Simulation and Empirical Illustration
- Asymptotic Properties of Penalized SMD Estimators
  - Convergence Rate of Nonparametric Parts.
  - Asymptotic Normality of Smooth Functionals.
  - Semiparametric Efficiency, Confidence Region.
- Ex 1: Partially Linear Quantile IV Regression
- Ex 2: Weighted average derivative of nonparametric quantile IV Regression

Outline of the Talk

- Semi/Nonparametric Conditional Moment Models.
- Examples, Brief Literature Review.
- Penalized Sieve Minimum Distance (SMD) Estimation.
- Simulation and Empirical Illustration
- Asymptotic Properties of Penalized SMD Estimators
  - Convergence Rate of Nonparametric Parts.
  - Asymptotic Normality of Smooth Functionals.
  - Semiparametric Efficiency, Confidence Region.
- Ex 1: Partially Linear Quantile IV Regression
- Ex 2: Weighted average derivative of nonparametric quantile IV Regression
- Conclusion and Future Work

Semi/nonparametric Conditional Moment Models

- The model: $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0(\cdot))|X] = 0$,

Semi/nonparametric Conditional Moment Models

- The model: $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0(\cdot))|X] = 0$,
- $Z \equiv (Y, X_z)$, Y **endogenous**, $X_z \subset X$, X **conditioning variables** (IV).
- $F_{Y|X}$ is unspecified, (nuisance function).

Semi/nonparametric Conditional Moment Models

- The model: $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0(\cdot))|X] = 0$,
- $Z \equiv (Y, X_z)$, Y **endogenous**, $X_z \subset X$, X **conditioning variables** (IV).
- $F_{Y|X}$ is unspecified, (nuisance function).
- $\alpha_0 \equiv (\theta_0, h_0(.))$ are unknown parameters of interest,
- θ are finite dimensional parameters,
- $h(.) = (h_1(.), ..., h_q(.))$ are functions, $h_j(.)$ could depend on Y, X, θ, other h_{-j} or latent variables.

Semi/nonparametric Conditional Moment Models

- The model: $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0(\cdot))|X] = 0$,
- $Z \equiv (Y, X_z)$, Y **endogenous**, $X_z \subset X$, X **conditioning variables** (IV).
- $F_{Y|X}$ is unspecified, (nuisance function).
- $\alpha_0 \equiv (\theta_0, h_0(.))$ are unknown parameters of interest,
- θ are finite dimensional parameters,
- $h(.) = (h_1(.), ..., h_q(.))$ are functions, $h_j(.)$ could depend on Y, X, θ, other h_{-j} or latent variables.
- $\rho(.)$ is a $d_\rho \times 1$ -vector of generalized residual functions, with known functional form up to unknown $\alpha \equiv (\theta, h(.))$.
- $\rho(.)$ may be **nonlinear**, pointwise **non-smooth** w.r.t. α.

Examples

- Ex 1 (Shape-invariant Engel curve IV regression, BCK):

$$E[Y_{1l} - h_{1l}(Y_2 - X_1'\theta_1) - X_1'\theta_{2,l}|X_1, X_2] = 0, l = 1, ..., d_\rho,$$

- $\rho_l(Z, \alpha) = Y_{1l} - h_{1l}(Y_2 - X_1'\theta_1) - X_1'\theta_{2,l}$.
- $E[\rho(Z, \alpha_0(\cdot))|X] = 0$, $\rho = (\rho_1, ..., \rho_{d_\rho})'$; Para. of interest are $\alpha = (\theta_1, \theta_{2,1}, ..., \theta_{2,d_\rho}, h_1, ..., h_{d_\rho})'$.

Examples

- Ex 1 (Shape-invariant Engel curve IV regression, BCK):

$$E[Y_{1l} - h_{1l}(Y_2 - X_1'\theta_1) - X_1'\theta_{2,l}|X_1, X_2] = 0, l = 1, ..., d_\rho,$$

- $\rho_l(Z,\alpha) = Y_{1l} - h_{1l}(Y_2 - X_1'\theta_1) - X_1'\theta_{2,l}$.
- $E[\rho(Z,\alpha_0(\cdot))|X] = 0$, $\rho = (\rho_1, ..., \rho_{d_\rho})'$; Para. of interest are $\alpha = (\theta_1, \theta_{2,1}, ..., \theta_{2,d_\rho}, h_1, ..., h_{d_\rho})'$.

- Ex 2 (Engel curve quantile IV regression): for $\gamma \in (0,1)$,

$$E[1\{Y_{1l} \leq h_{1l}(Y_2 - X_1'\theta_1) + X_1'\theta_{2,l}\}|X_1, X_2] = \gamma$$

- $\rho_l(Z,\alpha) = 1\{Y_{1l} \leq h_{1l}(Y_2 - X_1'\theta_1) + X_1'\theta_{2,l}\}$.

Examples (cont.)

- Ex 3 (Consumption-based asset pricing models):

$$E(M_{t+1}R_{l,t+1}|w_t) = 1, l = 1, ..., d_\rho,$$

- $M_{t+1} = \delta \frac{mu_{t+1}}{mu_t}$ is the intertemporal marginal rate of substitution or stochastic discount factor.

Examples (cont.)

- Ex 3 (Consumption-based asset pricing models):

$$E(M_{t+1}R_{l,t+1}|w_t) = 1, l = 1, ..., d_\rho,$$

- $M_{t+1} = \delta \frac{mu_{t+1}}{mu_t}$ is the intertemporal marginal rate of substitution or stochastic discount factor.
- Hansen - Singleton (82) assume power utility $E\sum_{t=1}^{\infty} \delta^t \frac{(C_t)^{1-\gamma}-1}{1-\gamma}$; hence $M_{t+1} = \delta(\frac{C_{t+1}}{C_t})^{-\gamma}$.

Examples (cont.)

- Ex 3 (Consumption-based asset pricing models):

$$E(M_{t+1}R_{l,t+1}|w_t) = 1, l = 1, ..., d_\rho,$$

- $M_{t+1} = \delta\frac{mu_{t+1}}{mu_t}$ is the intertemporal marginal rate of substitution or stochastic discount factor.
- Hansen - Singleton (82) assume power utility $E\sum_{t=1}^{\infty}\delta^t\frac{(C_t)^{1-\gamma}-1}{1-\gamma}$; hence $M_{t+1} = \delta(\frac{C_{t+1}}{C_t})^{-\gamma}$.
- Chen - Ludvigson (04) consider a semiparametric utility $E\sum_{t=1}^{\infty}\delta^t\frac{(C_t-H_t)^{1-\gamma}-1}{1-\gamma}$, where $H_t = C_t h(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t})$ is unknown habit level at time t.
- Para. of interest: $\alpha = (\delta, \gamma, h())$ and $E[\frac{\partial^2 h(x_1,...,x_L)}{\partial x_1^2}]$.

Example 3 (cont.)

- rewrite semiparametric asset pricing model as $E[\rho_i(\mathbf{z}_{t+1}, \delta_0, \gamma_0, h_0)|\mathbf{w}_t] = 0$, $i = 1, ..., d_\rho$, where

$$\rho_i(\mathbf{z}_{t+1}, \delta_0, \gamma_0, h_0) \equiv \delta \left(\frac{C_{t+1} - H_{t+1}}{C_t - H_t} \right)^{-\gamma} R_{i,t+1} \widetilde{F}_{i,t+1} - 1,$$

$$\begin{aligned} \widetilde{F}_{i,t+1} \quad &\equiv \quad 1 - \sum_{j=0}^{L} \delta^j \left(\frac{C_{t+1+j} - H_{t+1+j}}{C_{t+1} - H_{t+1}} \right)^{-\gamma} \frac{\partial H_{t+1+j}}{\partial C_{t+1}} \\ &\quad + \sum_{j=0}^{L} \delta^{j-1} \left(\frac{C_{t+j} - H_{t+j}}{C_{t+1} - H_{t+1}} \right)^{-\gamma} \frac{\partial H_{t+j}}{\partial C_t} \frac{1}{R_{i,t+1}}. \end{aligned}$$

More General Class of Models

- $m_j(X_{j,t}, \alpha_0) \equiv E[\rho_j(Z_t, \theta_0, h_0(\cdot))|X_{j,t}] = 0$, $j = 1, ..., d_\rho$,
- $\{(Y_t, X_t) : t = 1, ..., n\}$ either i.i.d. or stationary weakly dependent time series data.
- X_j is "IV" for j-th equation, but may be endogenous to j' equation for $j' \neq j$. Some of the X_j may be constant.

More General Class of Models

- $m_j(X_{j,t}, \alpha_0) \equiv E[\rho_j(Z_t, \theta_0, h_0(\cdot))|X_{j,t}] = 0$, $j = 1, ..., d_\rho$,
- $\{(Y_t, X_t) : t = 1, ..., n\}$ either i.i.d. or stationary weakly dependent time series data.
- X_j is “IV” for j-th equation, but may be endogenous to j' equation for $j' \neq j$. Some of the X_j may be constant.
- **Examples**: structural models of incomplete information, simultaneous equations, control function approach, panel data models, missing data, measurement errors via IV approach, treatment effects. Estimation of smooth functionals defined via expectations: $E[Y_1 - h_0(Y_2)|X] = 0$ and $E[\theta_0 - a(Y_2)\partial h_0(Y_2)] = 0$.

More General Class of Models

- $m_j(X_{j,t}, \alpha_0) \equiv E[\rho_j(Z_t, \theta_0, h_0(\cdot))|X_{j,t}] = 0$, $j = 1, ..., d_\rho$,
- $\{(Y_t, X_t) : t = 1, ..., n\}$ either i.i.d. or stationary weakly dependent time series data.
- X_j is "IV" for j-th equation, but may be endogenous to j' equation for $j' \neq j$. Some of the X_j may be constant.
- **Examples**: structural models of incomplete information, simultaneous equations, control function approach, panel data models, missing data, measurement errors via IV approach, treatment effects. Estimation of smooth functionals defined via expectations: $E[Y_1 - h_0(Y_2)|X] = 0$ and $E[\theta_0 - a(Y_2)\partial h_0(Y_2)] = 0$.
- Ai - Chen (05) on efficiency under correct specification; Ai - Chen (07) on estimation under misspecification.

Back to Semi/nonparametric Conditional Moment Models

- $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0(\cdot))|X] = 0$,
- $\{(Y_t, X_t) : t = 1, ..., n\}$ i.i.d. sample from $F_{Y,X}$ (nuisance function).

Back to Semi/nonparametric Conditional Moment Models

- $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0(\cdot))|X] = 0$,
- $\{(Y_t, X_t) : t = 1, ..., n\}$ i.i.d. sample from $F_{Y,X}$ (nuisance function).
- **Issues**: when $h(.)$ may depend on endogenous Y,
  - identification of $\alpha_0 = (\theta_0, h_0(.))$;
  - estimation of h_0 at nonparametric rate;
  - $\sqrt{n}$ normality of estimators of smooth functionals;
  - efficient estimation of θ_0 under correct specification;
  - misspecified models, model comparison, testing.

Back to Semi/nonparametric Conditional Moment Models

- $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0(\cdot))|X] = 0$,
- $\{(Y_t, X_t) : t = 1, ..., n\}$ i.i.d. sample from $F_{Y,X}$ (nuisance function).
- **Issues**: when $h(.)$ may depend on endogenous Y,
  - identification of $\alpha_0 = (\theta_0, h_0(.))$;
  - estimation of h_0 at nonparametric rate;
  - $\sqrt{n}$ normality of estimators of smooth functionals;
  - efficient estimation of θ_0 under correct specification;
  - misspecified models, model comparison, testing.
- **Difficulty**: estimation of h may be ill-posed, and $\rho()$ may not be pointwise smooth wrt α.

Brief Literature Review

- The model **without** unknown h: $E[\rho(Z,\theta_0)|X]=0$.

Brief Literature Review

- The model **without** unknown h: $E[\rho(Z, \theta_0)|X] = 0$.
- Lots of papers about theoretical and practical issues on estimating θ_0 and huge amount of applications !!!
- Sargan (?), Hansen (82, 85, 05), Hansen - Singleton (82), Hansen et al. (95), Hansen et al. (96), Chamberlain (87), Robinson (88), Newey (93), Imbens (97), Imbens et al. (98), Kitamura et al. (04), Antoine et al. (06), Smith (00), Zellner (91), Newey - Smith (04), Newey - McFadden (94), Pakes - Pollard (89), Manski (94), Mantzkin (94), Powell (94), Carrasco - Florens (00), Gallant - Tauchen (00), Stock - Wright (00), Andrews - Stock (05),...
- Estimating equations in statistics: Hyde, Owen, van der Vaart, ...

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$.
- A large special class (**no endogeneity**): $E[\rho(Z, \theta, h) - \rho(Z, \theta_0, h_0)|X] = \rho(Z, \theta, h) - \rho(Z, \theta_0, h_0)$.

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$.
- A large special class (**no endogeneity**): $E[\rho(Z, \theta, h) - \rho(Z, \theta_0, h_0)|X] = \rho(Z, \theta, h) - \rho(Z, \theta_0, h_0)$.
- Semiparametric M-estimation problem, including MLE, Least Squares, nonlinear LS, quantile regression, etc
- Asymptotic theory on consistency, convergence rate, semiparametric efficiency, limiting distribution have been developed.
- Horowitz (98), Pagan - Ullan (99), Robinson (88, 93), Ichimura (93), Powell (94), Hardle - Linton (94), Andrews (94), Manski (94), Newey (94), Chen - Shen (98), Linton - Mammen (05), Ichimura - Lee (06), etc.
- BKRW (93), van der Vaart - Wellner (96), Fan-Gijbels, Fan-Yao, van de Geer, ...

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$,
- $E[\rho(Z, \alpha) - \rho(Z, \alpha_0)|X] \neq \rho(Z, \alpha) - \rho(Z, \alpha_0)$ **(endog.)**
- θ is para. of interest; h is **nuisance**, may depend on Y.
- **Aim**: root-n consistency, normality and efficiency of $\hat{\theta}$.

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$,
- $E[\rho(Z, \alpha) - \rho(Z, \alpha_0)|X] \neq \rho(Z, \alpha) - \rho(Z, \alpha_0)$ **(endog.)**
- θ is para. of interest; h is **nuisance**, may depend on Y.
- **Aim**: root-n consistency, normality and efficiency of $\hat{\theta}$.
- Estimating θ assuming there is a consistent $\hat{h}$: Pakes-Olley (95), Newey (94), Andrews (94), NPV (99), CLvK (03), Chen (07), etc.

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$,
- $E[\rho(Z, \alpha) - \rho(Z, \alpha_0)|X] \neq \rho(Z, \alpha) - \rho(Z, \alpha_0)$ (**endog**.)
- θ is para. of interest; h is **nuisance**, may depend on Y.
- **Aim**: root-n consistency, normality and efficiency of $\hat{\theta}$.
- Estimating θ assuming there is a consistent $\hat{h}$: Pakes-Olley (95), Newey (94), Andrews (94), NPV (99), CLvK (03), Chen (07), etc.
- Estimating both h and θ by imposing the model.
  - Smooth but possibly nonlinear $\rho()$: Ai - Chen (03, 05), Chen - Ludvigson (04), Otsu (07).
  - Linear $\rho()$: FJvB (07), Severini - Tripathi (07).

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$,
- $E[\rho(Z, \alpha) - \rho(Z, \alpha_0)|X] \neq \rho(Z, \alpha) - \rho(Z, \alpha_0)$ **(endog.)**
- h and θ are parameters of interest, $h()$ depends on Y.

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$,
- $E[\rho(Z, \alpha) - \rho(Z, \alpha_0)|X] \neq \rho(Z, \alpha) - \rho(Z, \alpha_0)$ (**endog**.)
- h and θ are parameters of interest, $h()$ depends on Y.
- **Linear** $\rho()$ (e.g., nonparametric IV regression): DFR (06), BCK (07), Blundell - Horowitz (04), Hall - Horowitz (05), CFR (06), Horowitz (06), Gagliardini - Scaillet (07).

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z, \theta_0, h_0)|X] = 0$,
- $E[\rho(Z, \alpha) - \rho(Z, \alpha_0)|X] \neq \rho(Z, \alpha) - \rho(Z, \alpha_0)$ **(endog.)**
- h and θ are parameters of interest, $h()$ depends on Y.
- **Linear** $\rho()$ (e.g., nonparametric IV regression): DFR (06), BCK (07), Blundell - Horowitz (04), Hall - Horowitz (05), CFR (06), Horowitz (06), Gagliardini - Scaillet (07).
- **Smooth** possibly **nonlinear** $\rho()$: Newey - Powell (03).

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z,\theta_0,h_0)|X]=0$,
- $E[\rho(Z,\alpha)-\rho(Z,\alpha_0)|X]\neq\rho(Z,\alpha)-\rho(Z,\alpha_0)$ (**endog**.)
- h and θ are parameters of interest, $h()$ depends on Y.
- **Linear** $\rho()$ (e.g., nonparametric IV regression): DFR (06), BCK (07), Blundell - Horowitz (04), Hall - Horowitz (05), CFR (06), Horowitz (06), Gagliardini - Scaillet (07).
- **Smooth** possibly **nonlinear** $\rho()$: Newey - Powell (03).
- **Non-smooth** $\rho()$: CIN (07) and Horowitz - Lee (07) on nonparametric quantile IV; Chen - Pouzo (07, 08) on general possibly non-smooth $\rho()$.

Literature Review (cont.)

- The model **with** unknown h: $E[\rho(Z,\theta_0,h_0)|X]=0$,
- $E[\rho(Z,\alpha)-\rho(Z,\alpha_0)|X]\neq\rho(Z,\alpha)-\rho(Z,\alpha_0)$ (**endog**.)
- h and θ are parameters of interest, $h()$ depends on Y.
- **Linear** $\rho()$ (e.g., nonparametric IV regression): DFR (06), BCK (07), Blundell - Horowitz (04), Hall - Horowitz (05), CFR (06), Horowitz (06), Gagliardini - Scaillet (07).
- **Smooth** possibly **nonlinear** $\rho()$: Newey - Powell (03).
- **Non-smooth** $\rho()$: CIN (07) and Horowitz - Lee (07) on nonparametric quantile IV; Chen - Pouzo (07, 08) on general possibly non-smooth $\rho()$.
- **Difficulty**: recovering h is nonlinear, may be ill-posed.

New Results in Chen-Pouzo (07, 08)

- For model $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0)|X] = 0$, allow for:

New Results in Chen-Pouzo (07, 08)

- For model $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0)|X] = 0$, allow for:
  - Possibly non-pointwise smooth residual $\rho()$ wrt α.
  - Possibly non-compact functional parameter space $\mathcal{H}$ and sieve spaces $\mathcal{H}_n$.
  - Unknown function h could depend on endog. Y.

New Results in Chen-Pouzo (07, 08)

- For model $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0)|X] = 0$, allow for:
  - Possibly non-pointwise smooth residual $\rho()$ wrt α.
  - Possibly non-compact functional parameter space $\mathcal{H}$ and sieve spaces $\mathcal{H}_n$.
  - Unknown function h could depend on endog. Y.
- Consider penalized Sieve Minimum Distance (SMD).

New Results in Chen-Pouzo (07, 08)

- For model $m(X, \alpha_0) \equiv E[\rho(Z, \theta_0, h_0)|X] = 0$, allow for:
  - Possibly non-pointwise smooth residual $\rho()$ wrt α.
  - Possibly non-compact functional parameter space $\mathcal{H}$ and sieve spaces $\mathcal{H}_n$.
  - Unknown function h could depend on endog. Y.
- Consider penalized Sieve Minimum Distance (SMD).
- Establish consistency, convergence rates of h that may be (nonlinear) ill-posed.
- Obtain asymp. normality of $\hat{\theta}$, and weighted bootstrap.
- Show efficiency of optimally weighted $\tilde{\theta}$, and profile criterion is asymp. Chi-square.
- Ex 1: Partially linear quantile IV regression.
- Ex 2: Average derivative of nonparametric quantile IV.

Review: Sieve Minimum Distance

- $m(X, \alpha) \equiv E[\rho(Z, \theta, h)|X]$, $\Sigma(X)$ is a p.d. matrix.
- Then $m(X, \alpha_0) = 0$ iff $\alpha_0 \in \mathcal{A}$ is the unique solution to

$$\inf_{\alpha \in \mathcal{A}} E\left[m(X, \alpha)'[\Sigma(X)]^{-1} m(X, \alpha)\right].$$

Review: Sieve Minimum Distance

- $m(X,\alpha) \equiv E[\rho(Z,\theta,h)|X]$, $\Sigma(X)$ is a p.d. matrix.
- Then $m(X,\alpha_0) = 0$ iff $\alpha_0 \in \mathcal{A}$ is the unique solution to

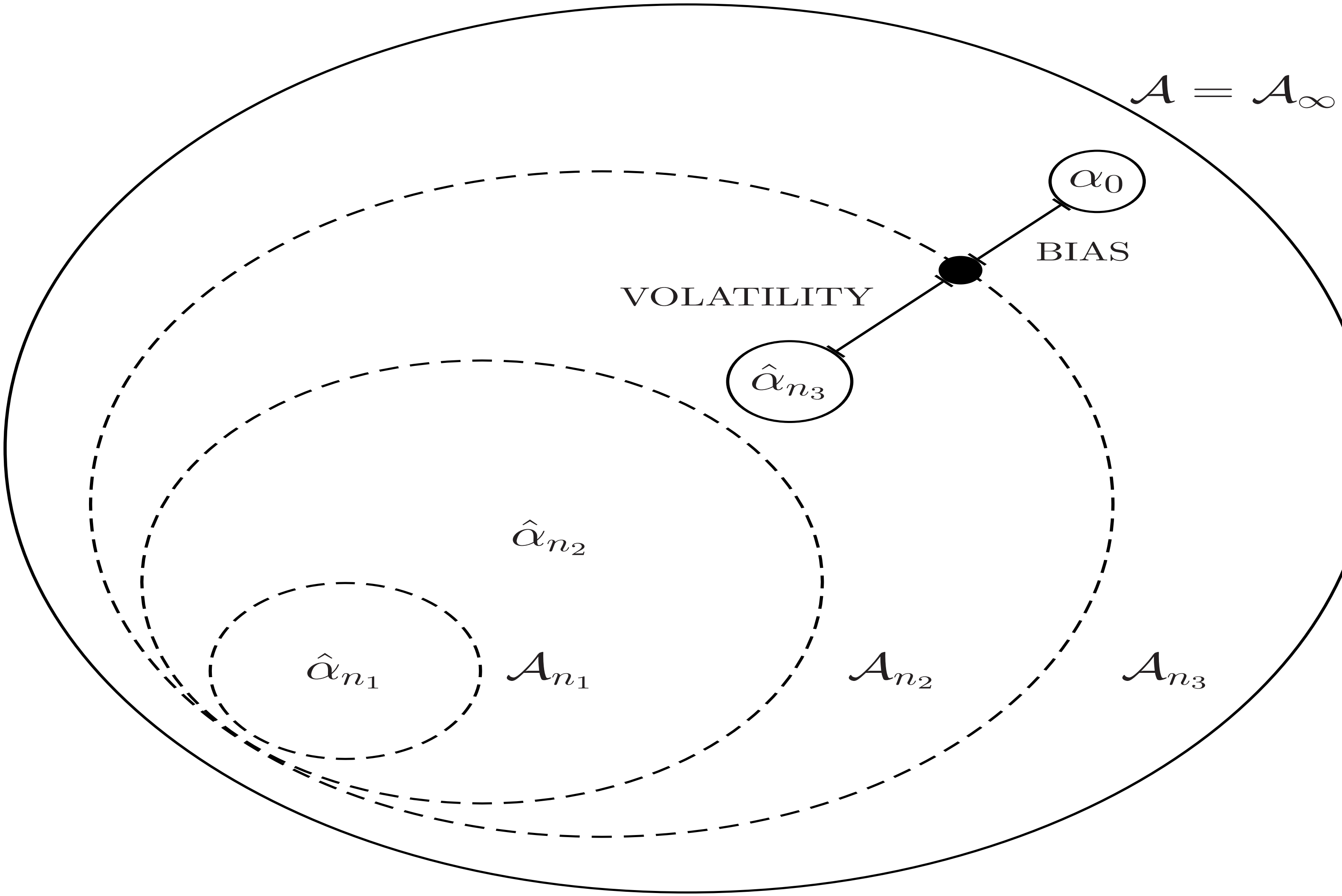
$$\inf_{\alpha \in \mathcal{A}} E\left[m(X,\alpha)'[\Sigma(X)]^{-1} m(X,\alpha)\right].$$

- Newey-Powell (89, 03), Ai-Chen (99, 03) propose SMD estimator $\hat{\alpha}_n$ that solves

$$\min_{\alpha \in \mathcal{A}_n} \frac{1}{n} \sum_{t=1}^{n} \left[\widehat{m}(X_t,\alpha)'[\widehat{\Sigma}(X_t)]^{-1} \widehat{m}(X_t,\alpha)\right]$$

- $\widehat{m}(X,\alpha)$ and $\widehat{\Sigma}(X)$ are any consistent estimators of $m(X,\alpha)$ and $\Sigma(X)$ respectively.
- $\mathcal{A}_n$ is a finite dimensional compact sieve space for $\mathcal{A}$.

SMD Estimation (cont.)



Examples of Sieves

- Finite-dimensional linear sieves $\mathcal{H}_n$ is of the form $\{h(.) = \sum_{k=1}^{k_n} \beta_k p_k(.)\}$, with $p_k(.)$ a known basis, e.g.
  1. Polynomials: $p_k(Y) = Y^k$
  2. Sine (Cosine): $p_k(Y) = Sin(k\pi Y)\ (Cos(k\pi Y))$
  3. B-Splines: $p_k(X) = 2^{k_{1n}/2} B_r(2^{k_{1n}} Y - k)$
  4. Polynomial splines, wavelets, finite elements, Hermite poly., Laguerre poly.

Examples of Sieves

- Finite-dimensional linear sieves $\mathcal{H}_n$ is of the form $\{h(.) = \sum_{k=1}^{k_n} \beta_k p_k(.)\}$, with $p_k(.)$ a known basis, e.g.
  1. Polynomials: $p_k(Y) = Y^k$
  2. Sine (Cosine): $p_k(Y) = Sin(k\pi Y)\ (Cos(k\pi Y))$
  3. B-Splines: $p_k(X) = 2^{k_{1n}/2} B_r(2^{k_{1n}} Y - k)$
  4. Polynomial splines, wavelets, finite elements, Hermite poly., Laguerre poly.
- Finite-dimensional compact sieves $\mathcal{H}_n$ could take the form $\{h(.) = \sum_{k=1}^{k_n} \beta_k p_k(.), ||D^r h||_{L^p} \leq log(k_n)\}$.

Examples of Sieves

- Finite-dimensional linear sieves $\mathcal{H}_n$ is of the form $\{h(.) = \sum_{k=1}^{k_n} \beta_k p_k(.)\}$, with $p_k(.)$ a known basis, e.g.
  1. Polynomials: $p_k(Y) = Y^k$
  2. Sine (Cosine): $p_k(Y) = Sin(k\pi Y)\ (Cos(k\pi Y))$
  3. B-Splines: $p_k(X) = 2^{k_{1n}/2} B_r(2^{k_{1n}} Y - k)$
  4. Polynomial splines, wavelets, finite elements, Hermite poly., Laguerre poly.
- Finite-dimensional compact sieves $\mathcal{H}_n$ could take the form $\{h(.) = \sum_{k=1}^{k_n} \beta_k p_k(.), ||D^r h||_{L^p} \leq log(k_n)\}$.
- Infinite-dimensional compact sieves $\mathcal{H}_n$ could take the form $\{h(.) = \sum_{k=1}^{\infty} \beta_k p_k(.), ||D^r h||_{L^p} \leq log(n)\}$.

Penalized SMD Estimators

- The penalized SMD estimator: $\widehat{\alpha}_n =$

$$\arg\min_{\alpha\in\mathcal{A}_n} n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\left[\widehat{\Sigma}(X_i)\right]^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\hat{P}_n(h).$$

Penalized SMD Estimators

- The penalized SMD estimator: $\widehat{\alpha}_n =$

$$\arg\min_{\alpha\in\mathcal{A}_n} n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\left[\widehat{\Sigma}(X_i)\right]^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\hat{P}_n(h).$$

1. $\mathcal{A}_n \equiv \Theta \times \mathcal{H}_n$, Θ compact subset of $\mathbb{R}^{d_\theta}$, $\mathcal{H}_n$ sieves for a normed function $\mathcal{H}$ (Hölder, Sobolev, Besov). Denote $k(n) = dim(\mathcal{H}_n)$ for finite-dimensional sieves.
2. $\hat{P}_n() \geq 0$: Penalty, either lower semicompact (e.g., Sobolev norm) or convex (e.g., L^2), may be random.
3. $\lambda_n \geq 0$, $\lambda_n \to 0$.

Penalized SMD Estimators

- The penalized SMD estimator: $\widehat{\alpha}_n =$

$$\arg\min_{\alpha\in\mathcal{A}_n} n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\left[\widehat{\Sigma}(X_i)\right]^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\hat{P}_n(h).$$

1. $\mathcal{A}_n \equiv \Theta \times \mathcal{H}_n$, Θ compact subset of $\mathbb{R}^{d_\theta}$, $\mathcal{H}_n$ sieves for a normed function $\mathcal{H}$ (Hölder, Sobolev, Besov). Denote $k(n) = dim(\mathcal{H}_n)$ for finite-dimensional sieves.
2. $\hat{P}_n() \geq 0$: Penalty, either lower semicompact (e.g., Sobolev norm) or convex (e.g., L^2), may be random.
3. $\lambda_n \geq 0$, $\lambda_n \to 0$.

- If $\lambda_n = 0$ and $\mathcal{H}_n$ compact, penalized SMD becomes the SMD of Newey-Powell and Ai-Chen.

Penalized SMD (cont.)

- If $n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha)$ is convex in $h \in \mathcal{H}$, and $\mathcal{H}$ is closed convex (but not compact under $||\cdot||_s$),
- $\mathcal{H}_n^c = \{h \in \mathcal{H} : \widehat{P}_n(h) \leq B_n\}$, $\widehat{P}_n(h)$ convex, $B_n \to \infty$.

Penalized SMD (cont.)

- If $n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha)$ is convex in $h\in\mathcal{H}$, and $\mathcal{H}$ is closed convex (but not compact under $||\cdot||_s$),
- $\mathcal{H}_n^c=\{h\in\mathcal{H}:\widehat{P}_n(h)\le B_n\}$, $\widehat{P}_n(h)$ convex, $B_n\to\infty$.
- Then the SMD using the compact sieve $\mathcal{H}^c$

$$\inf_{\theta\in\Theta}\left\{\inf_{h\in\mathcal{H}_n^c} n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha)\right\},$$

Penalized SMD (cont.)

- If $n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha)$ is convex in $h \in \mathcal{H}$, and $\mathcal{H}$ is closed convex (but not compact under $||\cdot||_s$),
- $\mathcal{H}_n^c = \{h \in \mathcal{H} : \widehat{P}_n(h) \leq B_n\}$, $\widehat{P}_n(h)$ convex, $B_n \to \infty$.
- Then the SMD using the compact sieve $\mathcal{H}^c$

$$\inf_{\theta\in\Theta}\left\{\inf_{h\in\mathcal{H}_n^c} n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha)\right\},$$

- is equivalent to the penalized SMD with a linear sieve $clsp(\mathcal{H}_n^c)$:

$$\inf_{\theta\in\Theta}\left\{\inf_{h\in clsp(\mathcal{H}_n^c)} n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\widehat{P}_n(h)\right\}$$

Penalized SMD (cont.)

- $n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\left[\widehat{\Sigma}(X_i)\right]^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\hat{P}_n(h)$.

Penalized SMD (cont.)

- $n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\left[\widehat{\Sigma}(X_i)\right]^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\hat{P}_n(h)$.
- $\widehat{m}(X,\alpha)$ a series LS estimator for $m(X,\alpha)$:

$$\widehat{m}(X,\alpha)=p^{J_n}(X)'(P'P)^{-1}\sum_{j=1}^{n}p^{J_n}(X_j)\rho(Z_j,\alpha),$$

- $p^{J_n}(X)\equiv(p_{o1}(X),\ldots,p_{oJ_n}(X))'$ a linear sieve basis (e.g. splines). $P=(p^{J_n}(X_1),...,p^{J_n}(X_n))'$.
- J_n : # of terms of the series LS regression for $m()$.

Penalized SMD (cont.)

- $n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\left[\widehat{\Sigma}(X_i)\right]^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\hat{P}_n(h)$.
- $\widehat{m}(X,\alpha)$ a series LS estimator for $m(X,\alpha)$:

$$\widehat{m}(X,\alpha)=p^{J_n}(X)'(P'P)^{-1}\sum_{j=1}^{n}p^{J_n}(X_j)\rho(Z_j,\alpha),$$

- $p^{J_n}(X)\equiv(p_{o1}(X),\ldots,p_{oJ_n}(X))'$ a linear sieve basis (e.g. splines). $P=(p^{J_n}(X_1),...,p^{J_n}(X_n))'$.
- J_n : # of terms of the series LS regression for $m()$.
- $\widehat{\Sigma}(x,\alpha)$: series LS estimator $[\widehat{\sigma}(x,\alpha)]_{jk}=\sum_{s=1}^{n}\rho_j(Z_s,\alpha)\rho_k(Z_s,\alpha)p^{J_n}(X_s)'(P'P)^{-1}p^{J_n}(x)$, or any consistent nonparametric estimator of $Var[\rho(Z,\alpha)|X]$.

Semiparametric Efficient Estimation

- Method I: 3-step optimally weighted Penalized SMD:

Semiparametric Efficient Estimation

- Method I: 3-step optimally weighted Penalized SMD:
  1. solve $\hat{\alpha}$ by $\min_{\alpha\in\mathcal{A}_n}\left\{n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'\widehat{m}(X_i,\alpha)+\lambda_n\widehat{P}_n(h)\right\}$.
  2. Start at $\hat{\alpha}^{(1)}=\hat{\alpha}$, and solve $\hat{\alpha}^{(2)}$ by

$$\min_{\alpha\in\mathcal{N}_{0n}}\left\{n^{-1}\sum_{i=1}^{n}\widehat{m}(X_i,\alpha)'[\widehat{\Sigma}(X_i,\hat{\alpha}^{(1)})]^{-1}\widehat{m}(X_i,\alpha)+\lambda_n\widehat{P}_n(h)\right.$$

  3. Set $\hat{\alpha}^{(1)}=\hat{\alpha}^{(2)}$ and go to Step 2,

Semiparametric Efficient Estimation

- Method I: 3-step optimally weighted Penalized SMD:

1. solve $\hat{\alpha}$ by
$\min_{\alpha \in \mathcal{A}_n} \left\{ n^{-1} \sum_{i=1}^{n} \widehat{m}(X_i, \alpha)' \widehat{m}(X_i, \alpha) + \lambda_n \widehat{P}_n(h) \right\}$.

2. Start at $\hat{\alpha}^{(1)} = \hat{\alpha}$, and solve $\hat{\alpha}^{(2)}$ by

$$\min_{\alpha \in \mathcal{N}_{0n}} \left\{ n^{-1} \sum_{i=1}^{n} \widehat{m}(X_i, \alpha)' [\widehat{\Sigma}(X_i, \hat{\alpha}^{(1)})]^{-1} \widehat{m}(X_i, \alpha) + \lambda_n \widehat{P}_n(h)\right.$$

3. Set $\hat{\alpha}^{(1)} = \hat{\alpha}^{(2)}$ and go to Step 2,

- $\widehat{\Sigma}(x, \alpha)$ is any consistent estimator of $Var[\rho(Z, \alpha)|X]$.
Optimal weighting: $\Sigma_0(X) = Var[\rho(Z, \alpha_0)|X]$

Semiparametric Efficient Estimation

- Method I: 3-step optimally weighted Penalized SMD:
  1. solve $\hat{\alpha}$ by $\min_{\alpha \in \mathcal{A}_n} \left\{ n^{-1} \sum_{i=1}^{n} \widehat{m}(X_i, \alpha)' \widehat{m}(X_i, \alpha) + \lambda_n \widehat{P}_n(h) \right\}$.
  2. Start at $\hat{\alpha}^{(1)} = \hat{\alpha}$, and solve $\hat{\alpha}^{(2)}$ by

$$\min_{\alpha \in \mathcal{N}_{0n}} \left\{ n^{-1} \sum_{i=1}^{n} \widehat{m}(X_i, \alpha)' [\widehat{\Sigma}(X_i, \hat{\alpha}^{(1)})]^{-1} \widehat{m}(X_i, \alpha) + \lambda_n \widehat{P}_n(h)\right.$$

  3. Set $\hat{\alpha}^{(1)} = \hat{\alpha}^{(2)}$ and go to Step 2,
- $\widehat{\Sigma}(x, \alpha)$ is any consistent estimator of $Var[\rho(Z, \alpha)|X]$. Optimal weighting: $\Sigma_0(X) = Var[\rho(Z, \alpha_0)|X]$
- Method II: locally continuous updated Penalized SMD on $\mathcal{N}_{0n}$.

Simulation: Partially linear quantile IV

- The model: $E[1\{Y_1 \leq X_1\theta + h(Y_2))\}|X_1, X_2] = \gamma$.

Simulation: Partially linear quantile IV

- The model: $E[1\{Y_1 \leq X_1\theta + h(Y_2))\}|X_1, X_2] = \gamma$.
- MC design: $X_1 \sim U[0,1]$, $(Y_2, X_2) \sim f$ with:(1) f is Gaussian Density (G-DEN); (2) f is Gaussian kernel density estimate using BCK "no kids" subsample (G-KER).
- $$Y_1 = X_1 + \overbrace{\Phi\left(\frac{Y_2 - \mu_{y_2}}{\sigma_{y_2}}\right)}^{h_0} + U.$$
  $U = \sqrt{0.075}(-\Phi^{-1}((E[h_0(Y_2)|X] - h_0(Y_2))/10 + \gamma) + v)$, $v \sim N(0,1)$.

Simulation: Partially linear quantile IV

- The model: $E[1\{Y_1 \leq X_1\theta + h(Y_2))\}|X_1, X_2] = \gamma$.
- MC design: $X_1 \sim U[0,1]$, $(Y_2, X_2) \sim f$ with:(1) f is Gaussian Density (G-DEN); (2) f is Gaussian kernel density estimate using BCK "no kids" subsample (G-KER).
- $Y_1 = X_1 + \overbrace{\Phi\left(\frac{Y_2 - \mu_{y_2}}{\sigma_{y_2}}\right)}^{h_0} + U$.
  $U = \sqrt{0.075}(-\Phi^{-1}((E[h_0(Y_2)|X] - h_0(Y_2))/10 + \gamma) + v)$, $v \sim N(0,1)$.
- $\hat{m}$: P-SPL(3,3) $\times$ P-COL(9). $\lambda_n \in \{0.001, 0.01, 0.1\}$.
- # of MC iter: 500, # of Obs: 1000.

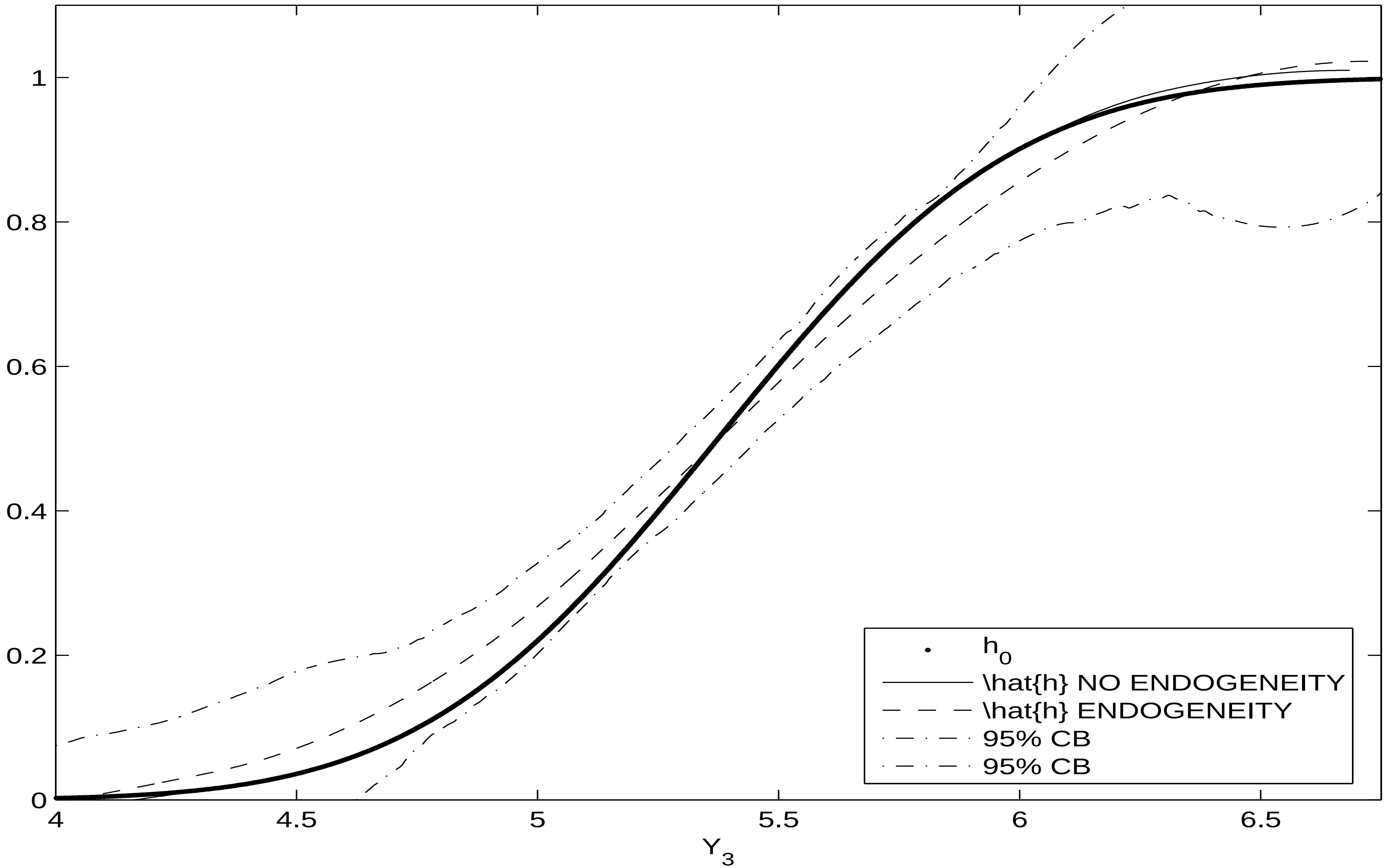
Monte Carlo: Robustness Analysis f = G-DEN and $\gamma = 0.5$

Endogeneity:	No	Yes	Yes	Yes
Basis h:	PSpl(2,6)	PSpl(2,6)	PSpl(2,6)	BSpl(8)
Penalization:	$\|\|D^1 h\|\|_2$	$\|\|D^1 h\|\|_2$	$\|\|D^2 h\|\|_2$	$\|\|D^1 h\|\|_2$
$E[\theta]$	0.9999	1.0009	1.0015	1.0081
$V[\theta]$	0.0002	0.0011	0.0013	0.0067
2.5% CI	0.96	0.93	0.93	0.90
97.5% CI	1.02	1.07	1.07	1.19
$BIAS^2[\theta] \times 10^3$	0.0000	0.0008	0.0023	0.0060
$MISE[h]$	0.0017	0.0087	0.0144	0.0960
$IBIAS^2[h]$	0.0000	0.0030	0.0067	0.0139
$IVAR[h]$	0.0016	0.0056	0.0137	0.0821

Monte Carlo: Robustness Analysis f = G-DEN and $\gamma = 0.5$

Endogeneity:	No	Yes	Yes	Yes
Basis h:	PSpl(2,6)	PSpl(2,6)	PSpl(2,6)	BSpl(8)
Penalization:	$\|\|D^1 h\|\|_2$	$\|\|D^1 h\|\|_2$	$\|\|D^2 h\|\|_2$	$\|\|D^1 h\|\|_2$
$E[\theta]$	0.9999	1.0009	0.9981	1.0081
$V[\theta]$	0.0002	0.0011	0.0020	0.0067
2.5% CI	0.96	0.93	0.90	0.90
97.5% CI	1.02	1.07	1.08	1.19
$BIAS^2[\theta] \times 10^3$	0.0000	0.0008	0.0040	0.0060
$MISE[h]$	0.0017	0.0087	0.0100	0.0960
$IBIAS^2[h]$	0.0000	0.0030	0.0031	0.0139
$IVAR[h]$	0.0016	0.0056	0.0072	0.0821

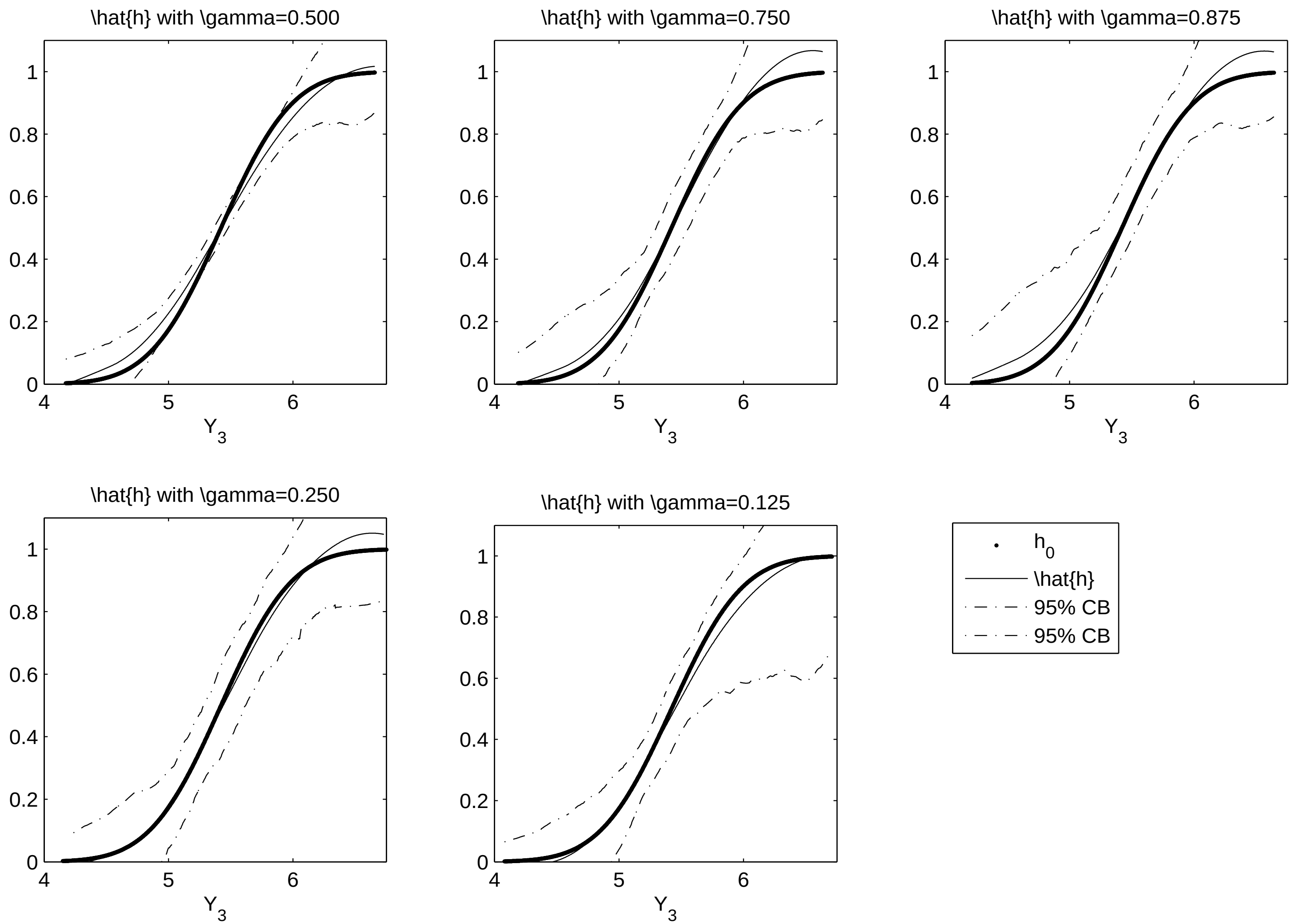
$\widehat{h}$: f = G-DEN and $\gamma = 0.5$



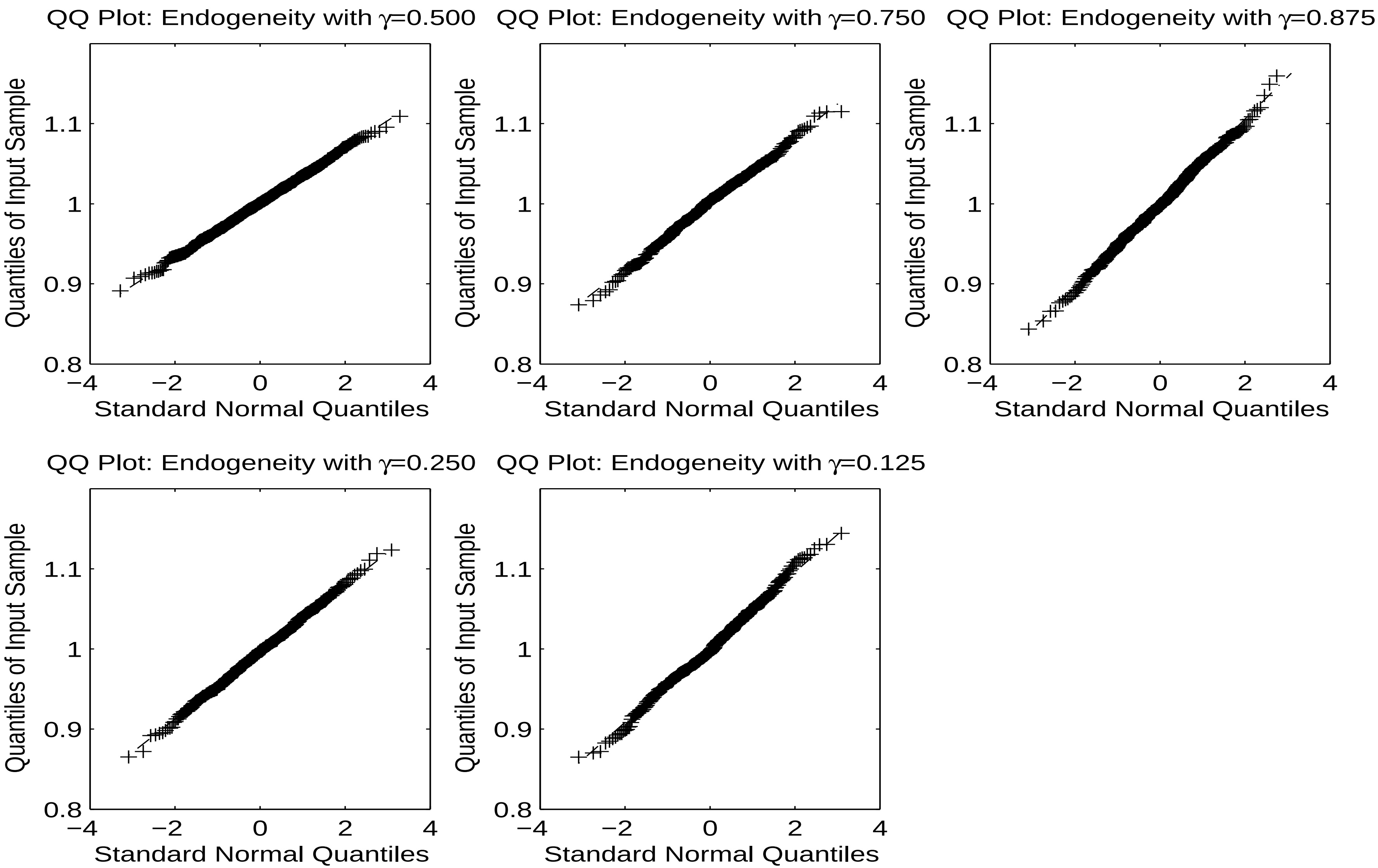
Monte Carlo: f = G-DEN and $\gamma \in \{0.125, 0.25, 0.50, 0.75, 0.875\}$

γ	0.125	0.250	0.500	0.750	0.875
$E[\theta]$	1.0009	0.9981	1.0009	1.0008	0.9992
$V[\theta]$	0.0023	0.0018	0.0011	0.0017	0.0028
$BIAS^2[\theta] \times 10^3$	0.0008	0.0034	0.0008	0.0006	0.0007
CI 2.5%	0.90	0.91	0.93	0.91	0.89
CI 97.5%	1.10	1.07	1.07	1.08	1.09
$IBIAS^2_{MC}[h]$	0.0022	0.0015	0.0030	0.0030	0.0044
$IVar_{MC}[h]$	0.0221	0.0287	0.0056	0.0147	0.0173
$IMSE^2_{MC}[h]$	0.0244	0.0302	0.0087	0.0177	0.0217

$\widehat{h}$: f = G-DEN and $\gamma \in \{0.125, 0.25, 0.5, 0.75, 0.825\}$



QQ-Plot: f = G-DEN and $\gamma \in \{0.125, 0.25, 0.5, 0.75, 0.825\}$



Monte Carlo: f = G-DEN, f = G-KER and $\gamma \in \{0.25, 0.50, 0.75\}$

γ f:	0.500 G-DEN	0.500 G-KER	0.750 G-KER	0.250 G-KER
$E[\theta]$	1.0009	1.0016	1.0013	0.9974
$V[\theta]$	0.0011	0.0010	0.0019	0.0016
2.5% CI	0.93	0.94	0.91	0.91
97.5% CI	1.07	1.06	1.08	1.09
$BIAS^2[\theta] \times 10^3$	0.0008	0.0021	0.0015	0.0064
$MISE[h]$	0.0087	0.0152	0.0499	0.0400
$IBIAS^2[h]$	0.0030	0.0048	0.0050	0.0058
$IVAR[h]$	0.0056	0.0094	0.0449	0.0341

Monte Carlo: f = G-DEN, $n = 125, 250, 500, 1000$ for $\gamma = 0.75$

n	125	250	500	1000
$E[\theta]$	1.0364	0.9926	1.0028	1.0008
$V[\theta]$	0.0278	0.0099	0.0039	0.0017

Monte Carlo: Estimators* for C.I. of $\hat{\theta}_n$ for f = G-DEN

Quantile:	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 0.25$
$E_{MC}[\theta]$	1.0009	1.0008	0.9981
$V_{MC}[\theta]$	0.0011	0.0017	0.0018
2.5% CI	0.93	0.91	0.91
97.5% CI	1.07	1.08	1.07
2.5% CI - BOOT	0.92	0.90	0.91
97.5% CI - BOOT	1.08	1.09	1.08
2.5% CI - χ^2	0.93	0.91	0.91
97.5% CI - χ^2	1.05	1.07	1.06

Application: Quantile IV Engel Curves

- Data is from BCK, “No Kids” sample (n=628) and the “Pooled” sample (n=1655).

Application: Quantile IV Engel Curves

- Data is from BCK, “No Kids” sample (n=628) and the “Pooled” sample (n=1655).
- The Model: $Y_{1il} = h_l(Y_{2i} - \theta_1 X_{1i}) + \theta_{2l} X_{1i} + \varepsilon_{il}$, $F_{\varepsilon_l | X}(0) = \gamma \in \{0.25, 0.50, 0.75\}$, $l = 1, ..., L$. $X = (X_1, X_2)$.

Application: Quantile IV Engel Curves

- Data is from BCK, “No Kids” sample (n=628) and the “Pooled” sample (n=1655).
- The Model: $Y_{1il} = h_l(Y_{2i} - \theta_1 X_{1i}) + \theta_{2l} X_{1i} + \varepsilon_{il}$, $F_{\varepsilon_l|X}(0) = \gamma \in \{0.25, 0.50, 0.75\}$, $l = 1, ..., L$. $X = (X_1, X_2)$.
- X_1: No Kids ($X_1 = 0$), Kids ($X_1 = 1$); Y_2: Total Expenditures. X_2 : Gross Earnings. $L = 7$ (food-in, food-out, alcohol, fares, fuel, leisure, travel).

Application: Quantile IV Engel Curves

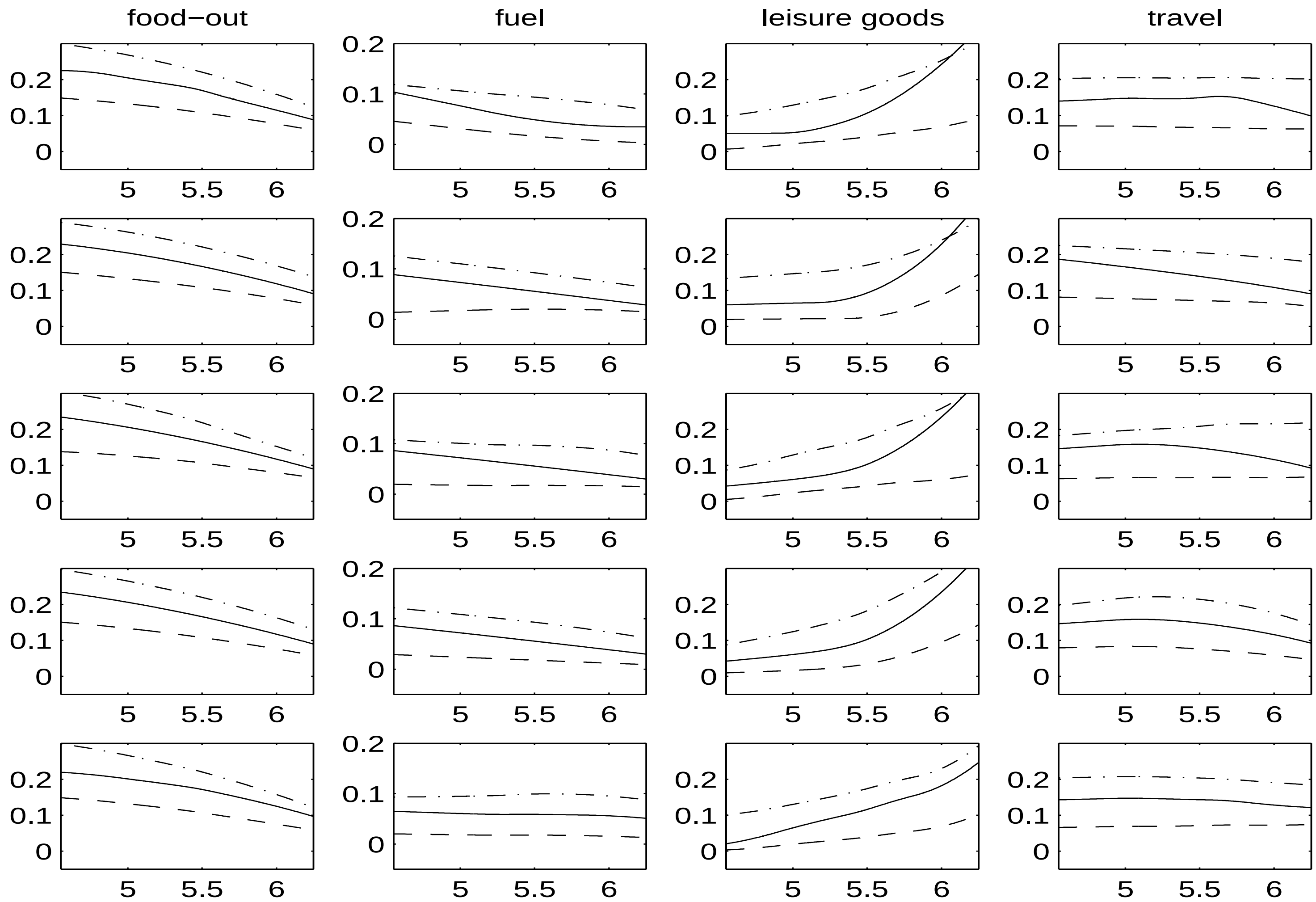
- Data is from BCK, “No Kids” sample (n=628) and the “Pooled” sample (n=1655).
- The Model: $Y_{1il} = h_l(Y_{2i} - \theta_1 X_{1i}) + \theta_{2l} X_{1i} + \varepsilon_{il}$, $F_{\varepsilon_l|X}(0) = \gamma \in \{0.25, 0.50, 0.75\}$, $l = 1, ..., L$. $X = (X_1, X_2)$.
- X_1: No Kids ($X_1 = 0$), Kids ($X_1 = 1$); Y_2: Total Expenditures. X_2 : Gross Earnings. $L = 7$ (food-in, food-out, alcohol, fares, fuel, leisure, travel).
- $\widehat{m}(X, \alpha)$: P-Spline(5,10). $\mathcal{H}_n$: P-Spline(2,5).
- $P(h)$: $||\nabla^k h||^j_{L^j(d\hat{\mu})} \equiv n^{-1} \sum_{i=1}^n |\nabla^k h(Y_{2i})|^j$ for $k = 1, 2$ and $j = 1, 2$.

θ_l **for** $l = 1, \ldots, 7$ **for different penalty and** $\gamma = 0.50$

$\hat{P}_n(h)$	$\lvert\lvert\nabla^2 h\rvert\rvert^2_{L^2(d\hat{\mu})}$	$\lvert\lvert\nabla^2 h\rvert\rvert_{L^1(d\hat{\mu})}$	$\lvert\lvert\nabla h\rvert\rvert^2_{L^2(d\hat{\mu})}$	$\lvert\lvert\nabla^2 h\rvert\rvert^2_{L^2(d\hat{\mu})}$	$\lvert\lvert\nabla$
λ_n	0.001	0.001	0.001	0.0003	0
$\hat{\theta}_1$	0.4133	0.3895	0.5479	0.43136	0.
food-i	0.0200	0.0267	-0.0056	0.00989	0.
food-o	0.0010	0.0006	0.0019	0.00033	0.
alc'ol	-0.0195	-0.0123	-0.0171	-0.02002	-0.
fares	0.0106	-0.0031	-0.0001	-0.00009	-0.
fuel	-0.0027	0.0027	0.0004	-0.00198	-0.
lei're	0.0208	0.0214	0.0380	0.02582	0.
travel	-0.0207	-0.0218	-0.0084	-0.00622	-0.

Quantile IV Engel curves $\gamma = 0.25$ (dash), 0.50 (solid), 0.75 (dot-dash)

(1) $||\nabla^2 h||^2_{L^2(d\hat{\mu})}$, $\lambda_n = 0.001$; (2) $||\nabla^2 h||_{L^1(d\hat{\mu})}$, $\lambda_n = 0.001$; (3) $||\nabla h||^2_{L^2(d\hat{\mu})}$, $\lambda_n = 0.001$, (4) $\lambda_n = 0.003$; (5) $||\nabla h||^2_{L^2(leb)}$, $\lambda_n = 0.005$.



Quantile IV Engel Curves ("Pooled" sample) (cont.)

γ	0.250	0.500 (BCK)	0.750
θ_1	0.669714	0.415483 (0.4088)	0.381019
θ_{21} - Food In	0.002548	0.013011 (0.0191)	0.037018
θ_{22} - Food Out	0.000504	0.000508 (-0.0002)	-0.000270
θ_{23} - Alcohol	-0.001969	-0.005315 (-0.0285)	0.046248
θ_{24} - Fares	-0.026957	-0.001056 (-0.0011)	0.001449
θ_{25} - Fuel	-0.010338	-0.006796 (-0.0038)	0.013448
θ_{26} - Leisure	0.003206	0.035873 (0.0496)	0.052509
θ_{27} - Travel	-0.034212	-0.036183 (-0.0399)	-0.045201

Table 1: θ_1 *and* $\theta_{2l},\ l = 1, ..., 7$

Consistency in Strong Norm

- Strong Norm, $||\cdot||_s$ on $\mathcal{H}$ is the "standard" norm associated to the Banach space $\mathcal{H}$, e.g., L^p norms.

Consistency in Strong Norm

- Strong Norm, $||\cdot||_s$ on $\mathcal{H}$ is the "standard" norm associated to the Banach space $\mathcal{H}$, e.g., L^p norms.
- Chen-Pouzo (07) establish general consistency results of the penalized SMD estimator $\widehat{\alpha}_n$ without imposing identification of α_0, permitting flexible penalization function $\widehat{P}_n(h)$, and allowing for any consistent estimator $\widehat{m}(X,\alpha)$ of $m(X,\alpha)$

Weak Pseudo-Norm

- Under ill-posedness, the convergence rate under $||\cdot||_s$ is typically slower than $n^{-1/4}$.
- Ai - Chen (03) introduce a "weaker" pseudo-metric $||\cdot||$ (i.e., $||\alpha|| \leq ||\alpha||_s$): $||\alpha - \alpha'||^2 \equiv$

$$E\left[\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha']'[\Sigma(X)]^{-1}\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha']\right]$$

where $\frac{dm(X,\alpha_0)}{d\alpha}[u] \equiv \lim_{t\to 0} \frac{E[\rho(Z,(1-t)\alpha_0+t(\alpha_0+u))|X]}{t}$.

Weak Pseudo-Norm

- Under ill-posedness, the convergence rate under $||\cdot||_s$ is typically slower than $n^{-1/4}$.
- Ai - Chen (03) introduce a "weaker" pseudo-metric $||\cdot||$ (i.e., $||\alpha|| \le ||\alpha||_s$): $||\alpha - \alpha'||^2 \equiv$

$$E\left[\frac{dm(X,\alpha_0)}{d\alpha}[\alpha-\alpha']'[\Sigma(X)]^{-1}\frac{dm(X,\alpha_0)}{d\alpha}[\alpha-\alpha']\right]$$

  where $\frac{dm(X,\alpha_0)}{d\alpha}[u] \equiv \lim_{t\to 0}\frac{E[\rho(Z,(1-t)\alpha_0+t(\alpha_0+u))|X]}{t}$.
- Ex: NPIV $E[Y_1 - h_0(Y_2)|X] = 0$. $||\alpha||_s^2 = E[h(Y_2)^2]$, $||\alpha||^2 = E[(E[h(Y_2)|X])^2]$.

Weak Pseudo-Norm

- Under ill-posedness, the convergence rate under $||\cdot||_s$ is typically slower than $n^{-1/4}$.
- Ai - Chen (03) introduce a "weaker" pseudo-metric $||\cdot||$ (i.e., $||\alpha|| \leq ||\alpha||_s$): $||\alpha - \alpha'||^2 \equiv$

$$E\left[\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha']'[\Sigma(X)]^{-1}\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha']\right]$$

  where $\frac{dm(X,\alpha_0)}{d\alpha}[u] \equiv \lim_{t\to 0}\frac{E[\rho(Z,(1-t)\alpha_0 + t(\alpha_0+u))|X]}{t}$.
- Ex: NPIV $E[Y_1 - h_0(Y_2)|X] = 0$. $||\alpha||_s^2 = E[h(Y_2)^2]$, $||\alpha||^2 = E[(E[h(Y_2)|X])^2]$.
- Ai - Chen (03) obtain $||\hat{\alpha}_n - \alpha_0|| = o_p(n^{-1/4})$.

Convergence Rate in Weaker Metric

- **Thm 3.1** $\hat{\alpha}_n$ is penalized SMD with $||\hat{\alpha}_n - \alpha_0||_s = o_P(1)$. Then: For lower semicompact penalty,

$$||\hat{\alpha}_n - \Pi_n\alpha_0|| = O_P\left(\max\left\{\sqrt{\frac{J_n}{n}} + b_{J_n}, ||\Pi_n\alpha_0 - \alpha_0||, \sqrt{\lambda_n}\right\}\right)$$

For convex but non-lower semicompact penalty,

$$||\hat{\alpha}_n - \Pi_n\alpha_0|| = O_P\left(\max\left\{\sqrt{\frac{J_n}{n}} + b_{J_n}, ||\Pi_n\alpha_0 - \alpha_0||, o(\sqrt{\lambda_n})\right\}\right)$$

Convergence Rate in Weaker Metric

- **Thm 3.1** $\hat{\alpha}_n$ is penalized SMD with $||\hat{\alpha}_n - \alpha_0||_s = o_P(1)$. Then: For lower semicompact penalty,

$$||\hat{\alpha}_n - \Pi_n \alpha_0|| = O_P\left(\max\left\{\sqrt{\frac{J_n}{n}} + b_{J_n}, ||\Pi_n \alpha_0 - \alpha_0||, \sqrt{\lambda_n}\right\}\right)$$

For convex but non-lower semicompact penalty,

$$||\hat{\alpha}_n - \Pi_n \alpha_0|| = O_P\left(\max\left\{\sqrt{\frac{J_n}{n}} + b_{J_n}, ||\Pi_n \alpha_0 - \alpha_0||, o(\sqrt{\lambda_n})\right\}\right)$$

- Without nonparametric endogeneity, the weaker and strong metrics are equivalent. Thm 3.1 leads to optimal convergence rates for penalized sieve M-estimators when $\rho()$ could be non-smooth.

Sieve Measure of Ill-posedness

- Define a *sieve measure of ill-posedness* as $\tau_n \equiv$

$$\sup_{\alpha \in \mathcal{A}_{osn}:\alpha \neq \Pi_n \alpha_0} \frac{||\alpha - \Pi_n \alpha_0||_s}{||\alpha - \Pi_n \alpha_0||} \asymp \frac{||\alpha - \Pi_n \alpha_0||_s}{\sqrt{E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \Pi_n \alpha_0]\right)^2\right]}},$$

where $\mathcal{A}_{osn} = \{\alpha \in \mathcal{A}_n : ||\alpha - \Pi_n \alpha_0||_s = o(1)\}$.

Sieve Measure of Ill-posedness

- Define a *sieve measure of ill-posedness* as $\tau_n \equiv$

$$\sup_{\alpha\in\mathcal{A}_{osn}:\alpha\neq\Pi_n\alpha_0} \frac{||\alpha-\Pi_n\alpha_0||_s}{||\alpha-\Pi_n\alpha_0||} \asymp \frac{||\alpha-\Pi_n\alpha_0||_s}{\sqrt{E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha-\Pi_n\alpha_0]\right)^2\right]}},$$

  where $\mathcal{A}_{osn} = \{\alpha\in\mathcal{A}_n : ||\alpha-\Pi_n\alpha_0||_s = o(1)\}$.

- This definition is a generalization of that in BCK (03) for nonparametric IV regression $E[Y_1 - h(Y_2)|X] = 0$:

$$\tau_n = \sup_{h_n\in\mathcal{H}_n:h_n\neq 0} \frac{\sqrt{E\{h_n(Y_2)\}^2}}{\sqrt{E\{E[h_n(Y_2)|X]\}^2}},$$

- In BCK, $\tau_n = 1$ iff Y_2 is measurable w.r.t. X.

Modulus of Continuity

- *Modulus of Continuity*:

$$\omega(\delta, \mathcal{A}_{os}) = \sup_{\{\alpha \in \mathcal{A}_{os} : ||\alpha - \alpha_0|| \leq \delta\}} ||\alpha - \alpha_0||_s$$

where $\mathcal{A}_{os} = \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s = o(1)\}$.

Modulus of Continuity

- *Modulus of Continuity*:

$$\omega(\delta, \mathcal{A}_{os}) = \sup_{\{\alpha \in \mathcal{A}_{os}: ||\alpha - \alpha_0|| \leq \delta\}} ||\alpha - \alpha_0||_s$$

  where $\mathcal{A}_{os} = \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s = o(1)\}$.

- The estimation problem is *Ill-posed* if $\frac{\omega(\delta, \mathcal{A}_{os})}{\delta}$ goes to infinity as δ goes to zero. This does not depend on estimation methods.

Modulus of Continuity

- *Modulus of Continuity*:

$$\omega(\delta, \mathcal{A}_{os}) = \sup_{\{\alpha \in \mathcal{A}_{os}: ||\alpha - \alpha_0|| \leq \delta\}} ||\alpha - \alpha_0||_s$$

where $\mathcal{A}_{os} = \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s = o(1)\}$.

- The estimation problem is *Ill-posed* if $\frac{\omega(\delta, \mathcal{A}_{os})}{\delta}$ goes to infinity as δ goes to zero. This does not depend on estimation methods.
- *Sieve Modulus of Continuity*:

$$\omega_n(\delta, \mathcal{A}_{osn}) = \sup_{\{\alpha \in \mathcal{A}_{osn}: ||\alpha - \Pi_n \alpha_0|| \leq \delta\}} ||\alpha - \Pi_n \alpha_0||_s$$

Modulus of Continuity

- *Modulus of Continuity*:

$$\omega(\delta, \mathcal{A}_{os}) = \sup_{\{\alpha \in \mathcal{A}_{os} : ||\alpha - \alpha_0|| \leq \delta\}} ||\alpha - \alpha_0||_s$$

  where $\mathcal{A}_{os} = \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s = o(1)\}$.

- The estimation problem is *Ill-posed* if $\frac{\omega(\delta, \mathcal{A}_{os})}{\delta}$ goes to infinity as δ goes to zero. This does not depend on estimation methods.
- *Sieve Modulus of Continuity*:

$$\omega_n(\delta, \mathcal{A}_{osn}) = \sup_{\{\alpha \in \mathcal{A}_{osn} : ||\alpha - \Pi_n \alpha_0|| \leq \delta\}} ||\alpha - \Pi_n \alpha_0||_s$$

- τ_n and $\omega_n(\delta, \mathcal{A}_{osn})$ measures do depend on choice of sieve space; only useful for finite-dimensional sieves.

Convergence Rate in Strong Metric

- Case I: Finite dimensional sieve dominating case:

Convergence Rate in Strong Metric

- Case I: Finite dimensional sieve dominating case:
- **A3.2**: (i) $\tau_n||\Pi_n\alpha_0 - \alpha_0|| \leq ||\Pi_n\alpha_0 - \alpha_0||_s$; or (ii) $\omega_n(||\Pi_n\alpha_0 - \alpha_0||, \mathcal{A}_{osn}) \leq ||\Pi_n\alpha_0 - \alpha_0||_s$.

Convergence Rate in Strong Metric

- Case I: Finite dimensional sieve dominating case:
- **A3.2**: (i) $\tau_n||\Pi_n\alpha_0 - \alpha_0|| \leq ||\Pi_n\alpha_0 - \alpha_0||_s$; or (ii) $\omega_n(||\Pi_n\alpha_0 - \alpha_0||, \mathcal{A}_{osn}) \leq ||\Pi_n\alpha_0 - \alpha_0||_s$.
- **Thm 3.2** Under conditions of Thm 3.1 and A3.2, if $\max\left\{\frac{J_n}{n} + b^2_{m,J_n}, \lambda_n\right\} = \frac{J_n}{n}$, then:

  $$||\hat{\alpha}_n - \alpha_0||_s = O_P\left(||\alpha_0 - \Pi_n\alpha_0||_s + \tau_n \times \sqrt{\frac{J_n}{n}}\right)$$

  $$= O_P\left(||\alpha_0 - \Pi_n\alpha_0||_s + \omega_n(\sqrt{\frac{J_n}{n}}, \mathcal{A}_{osn})\right).$$

Convergence Rate in Strong Metric

- Case I: Finite dimensional sieve dominating case:
- **A3.2**: (i) $\tau_n||\Pi_n\alpha_0 - \alpha_0|| \leq ||\Pi_n\alpha_0 - \alpha_0||_s$; or (ii) $\omega_n(||\Pi_n\alpha_0 - \alpha_0||, \mathcal{A}_{osn}) \leq ||\Pi_n\alpha_0 - \alpha_0||_s$.
- **Thm 3.2** Under conditions of Thm 3.1 and A3.2, if $\max\left\{\frac{J_n}{n} + b^2_{m,J_n}, \lambda_n\right\} = \frac{J_n}{n}$, then:
  $$||\hat{\alpha}_n - \alpha_0||_s = O_P\left(||\alpha_0 - \Pi_n\alpha_0||_s + \tau_n \times \sqrt{\frac{J_n}{n}}\right)$$
  $$= O_P\left(||\alpha_0 - \Pi_n\alpha_0||_s + \omega_n(\sqrt{\frac{J_n}{n}}, \mathcal{A}_{osn})\right).$$
- Thm 3.2 directly extends BCK (07) on nonparametric IV regression to nonlinear or nonsmooth ill-posed problems.

Convergence Rate in Strong Metric

- Case II: Penalization dominating case:

Convergence Rate in Strong Metric

- Case II: Penalization dominating case:
- Either for lower semicompact penalty with
$$\max\left\{\sqrt{\frac{J_n}{n}} + b_{m,J_n}, \sqrt{\lambda_n}\right\} = \sqrt{\frac{J_n}{n}} + b_{m,J_n} = O(\sqrt{\lambda_n}),$$
- or for convex but non-lower semicompact penalty with
$$\max\{\sqrt{\frac{J_n}{n}} + b_{m,J_n}, \sqrt{\lambda_n ||\hat{\alpha} - \Pi_n \alpha_0||_s}\} = \sqrt{\frac{J_n}{n}} + b_{m,J_n},$$

Convergence Rate in Strong Metric

- Case II: Penalization dominating case:
- Either for lower semicompact penalty with $\max\left\{\sqrt{\frac{J_n}{n}} + b_{m,J_n}, \sqrt{\lambda_n}\right\} = \sqrt{\frac{J_n}{n}} + b_{m,J_n} = O(\sqrt{\lambda_n})$,
- or for convex but non-lower semicompact penalty with $\max\{\sqrt{\frac{J_n}{n}} + b_{m,J_n}, \sqrt{\lambda_n ||\hat{\alpha} - \Pi_n \alpha_0||_s}\} = \sqrt{\frac{J_n}{n}} + b_{m,J_n}$,
- Under conditions of Thm 3.1 and A3.2, we have:

$$||\hat{\alpha}_n - \Pi_n \alpha_0||_s = O_P\left(||h_0 - \Pi_n h_0||_s + \omega_n(\{\sqrt{\frac{J_n}{n}} + b_{m,J_n}\}, \mathcal{A}_{osi}$$

Sufficient Conditions for Convergence Rates

- A3.5: $\{q_j\}_{j=1}^{\infty}$ is a Riesz basis for a separable Hilbert space $(\mathcal{H}, ||\cdot||_s)$, and $\mathcal{H}_{os}$ is a subset of $\mathcal{H}$.

Sufficient Conditions for Convergence Rates

- A3.5: $\{q_j\}_{j=1}^{\infty}$ is a Riesz basis for a separable Hilbert space $(\mathcal{H}, ||\cdot||_s)$, and $\mathcal{H}_{os}$ is a subset of $\mathcal{H}$.
- A3.6: Let $\mathcal{H}_n = clsp\{q_1, ..., q_{k(n)}\}$. There is a non-increasing positive sequence $\{b_j\}_{j=1}^{\infty}$ such that: (i) $||h||^2 \geq c\sum_{j=1}^{\infty} b_j |\langle h, q_j\rangle_s|^2$ for all $h \in \mathcal{H}_{osn}$; (ii) $C\sum_j b_j |\langle h_0 - \Pi_n h_0, q_j\rangle_s|^2 \geq ||h_0 - \Pi_n h_0||^2$.

Sufficient Conditions for Convergence Rates

- A3.5: $\{q_j\}_{j=1}^{\infty}$ is a Riesz basis for a separable Hilbert space $(\mathcal{H}, ||\cdot||_s)$, and $\mathcal{H}_{os}$ is a subset of $\mathcal{H}$.
- A3.6: Let $\mathcal{H}_n = clsp\{q_1, ..., q_{k(n)}\}$. There is a non-increasing positive sequence $\{b_j\}_{j=1}^{\infty}$ such that: (i) $||h||^2 \geq c\sum_{j=1}^{\infty} b_j|\langle h, q_j\rangle_s|^2$ for all $h \in \mathcal{H}_{osn}$; (ii) $C\sum_j b_j|\langle h_0 - \Pi_n h_0, q_j\rangle_s|^2 \geq ||h_0 - \Pi_n h_0||^2$.
- Lemma: Let $\mathcal{H}_n = clsp\{q_1, ..., q_{k(n)}\}$, A3.5 and A3.6 hold. Then: A3.2 is satisfied, and

$$\tau_n \leq const./\sqrt{b_{k(n)}} \quad \text{and} \quad \omega_n(\delta, \mathcal{H}_{osn}) \leq const.\times\delta/\sqrt{b_{k(n)}}.$$

Sufficient Conditions for Convergence Rates

- let $a > 0$ be a finite constant. (i) If $b_j \asymp j^{-2a}$ then $\tau_n \leq const.(k(n))^a$. (ii) If $b_j \asymp \exp\{-j^a\}$ then $\tau_n \leq const. \exp\{\frac{1}{2}(k(n))^a\}$.

Sufficient Conditions for Convergence Rates

- let $a > 0$ be a finite constant. (i) If $b_j \asymp j^{-2a}$ then $\tau_n \leq const.(k(n))^a$. (ii) If $b_j \asymp \exp\{-j^a\}$ then $\tau_n \leq const. \exp\{\frac{1}{2}(k(n))^a\}$.
- assume $||\alpha_0 - \Pi_n \alpha_0||_s = O(k(n)^{-\mu_h})$, $J_n = ck(n)$ for $c \geq 1$,
- if $\tau_n \leq const.(k(n))^a$, then $||\widehat{\alpha}_n - \alpha_0||_s = O_p(n^{-\frac{\mu_h}{2(a+\mu_h)+1}})$;
- if $\tau_n \leq const. \exp\{\frac{1}{2}(k(n))^a\}$, and $\mu_m = \infty$, then $||\widehat{\alpha}_n - \alpha_0||_s = O_p([\log(n)]^{-\mu_h/a})$.

Root-n Normality and Efficiency

- **Asymptotic Normality of $\hat{\theta}_n$:**

$$\begin{aligned} \sqrt{n} \quad & (\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, V^{-1}) \\ V^{-1} \quad = \quad & E[D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)]^{-1} \times \\ & E[D_{w^*}(X)'\Sigma(X)^{-1}\Sigma_0(X)\Sigma(X)^{-1}D_{w^*}(X)] \times \\ & E[D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)]^{-1}. \end{aligned}$$

with w^* as the minimizer of: $E[D_w(X)'\Sigma(X)^{-1}D_w(X)] =$

$$E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[(\mathbf{1}, -w)]\right)' \Sigma^{-1}(X) \left(\frac{dm(X,\alpha_0)}{d\alpha}[(\mathbf{1}, -w)]\right)\right]$$

Root-n Normality and Efficiency

- **Asymptotic Normality of $\hat{\theta}_n$:**

$$
\begin{aligned}
\sqrt{n} &\quad (\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, V^{-1}) \\
V^{-1} &= E[D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)]^{-1} \times \\
&\quad E[D_{w^*}(X)'\Sigma(X)^{-1}\Sigma_0(X)\Sigma(X)^{-1}D_{w^*}(X)] \times \\
&\quad E[D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)]^{-1}.
\end{aligned}
$$

with w^* as the minimizer of: $E[D_w(X)'\Sigma(X)^{-1}D_w(X)] =$

$$
E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[(\mathbf{1}, -w)]\right)' \Sigma^{-1}(X) \left(\frac{dm(X,\alpha_0)}{d\alpha}[(\mathbf{1}, -w)]\right)\right]
$$

- **Efficiency:** $V_0 = \inf_w E[D_w(X)'\Sigma_0(X)^{-1}D_w(X)]$.

Weighted Bootstrap

- **Thm**: Let $\{W_i > 0\}_{i=1}^n$ be i.i.d. with $E[W_i] = 1$, $Var(W_i) = w_0$, and is indep. of the data $\{(Y_i', X_i')\}_{i=1}^n$.

$$\widehat{\alpha}_n^* \equiv \arg \inf_{\alpha \in \mathcal{N}_{0n}} \left\{ \frac{1}{n} \sum_{i=1}^n W_i \left\{ \widehat{m}(X_i, \alpha)' [\widehat{\Sigma}(X_i)]^{-1} \widehat{m}(X_i, \alpha) \right\} + \lambda_n P(h) \right.$$

Then: Conditional on the data, $\sqrt{\frac{n}{w_0}} \left(\widehat{\theta}_n^* - \widehat{\theta}_n \right)$ has the same limiting dist. as that of $\sqrt{n} \left(\widehat{\theta}_n - \theta_0 \right)$.

Weighted Bootstrap

- **Thm**: Let $\{W_i > 0\}_{i=1}^n$ be i.i.d. with $E[W_i] = 1$, $Var(W_i) = w_0$, and is indep. of the data $\{(Y_i', X_i')\}_{i=1}^n$.

$$\widehat{\alpha}_n^* \equiv \arg \inf_{\alpha \in \mathcal{N}_{0n}} \left\{ \frac{1}{n} \sum_{i=1}^n W_i \left\{ \widehat{m}(X_i, \alpha)' [\widehat{\Sigma}(X_i)]^{-1} \widehat{m}(X_i, \alpha) \right\} + \lambda_n P(h)$$

Then: Conditional on the data, $\sqrt{\frac{n}{w_0}} \left(\widehat{\theta}_n^* - \widehat{\theta}_n \right)$ has the same limiting dist. as that of $\sqrt{n} \left(\widehat{\theta}_n - \theta_0 \right)$.

- **W.B. Algorithm**: (1) Draw an i.i.d. sample $\{W_i > 0\}_{i=1}^n$ with $E(W_i) = 1$, $Var(W_i) = 1$, and compute $\widehat{\alpha}_n^*$; (2) Repeat step (1) many times (say N numbers of times) and compute the empirical quantiles of $(\widehat{\theta}_{n,q}^*)_{q=1}^N$.

Partially Linear Quantile IV

- $Y_{1i} = X_{1i}\theta_0 + h_0(Y_{2i}) + U_i$ with $F_{U|X}(0|X) = \gamma$.
  - $\mathcal{A} = [\underline{\theta}, \overline{\theta}] \times \mathcal{H}$.
  - $\mathcal{A}_n = [\underline{\theta}, \overline{\theta}] \times \{h \ : \ h(y_2) = q^{k_n}(y_2)'\beta\} \cap \mathcal{H}$.
  - $Y = (Y_1, Y_2)$, $X = (X_1, X_2)$, $dim(X) = 2$.

Partially Linear Quantile IV

- $Y_{1i} = X_{1i}\theta_0 + h_0(Y_{2i}) + U_i$ with $F_{U|X}(0|X) = \gamma$.
  - $\mathcal{A} = [\underline{\theta}, \overline{\theta}] \times \mathcal{H}$.
  - $\mathcal{A}_n = [\underline{\theta}, \overline{\theta}] \times \{h \;:\; h(y_2) = q^{k_n}(y_2)'\beta\} \cap \mathcal{H}$.
  - $Y = (Y_1, Y_2)$, $X = (X_1, X_2)$, $dim(X) = 2$.
- $\rho(Z, \alpha) = \gamma - \mathcal{I}\{Y_1 - (X_1\theta + h(Y_2)) \leq 0\}$.
- $m(X, \alpha) = \gamma - \int F_{Y_1|Y_2,X}(x_1\theta + h(y_2)) f_{Y_2|X}(y_2, X) dy_2$.

Partially Linear Quantile IV

- $Y_{1i} = X_{1i}\theta_0 + h_0(Y_{2i}) + U_i$ with $F_{U|X}(0|X) = \gamma$.
  - $\mathcal{A} = [\underline{\theta}, \overline{\theta}] \times \mathcal{H}$.
  - $\mathcal{A}_n = [\underline{\theta}, \overline{\theta}] \times \{h \ : \ h(y_2) = q^{k_n}(y_2)'\beta\} \cap \mathcal{H}$.
  - $Y = (Y_1, Y_2)$, $X = (X_1, X_2)$, $dim(X) = 2$.
- $\rho(Z, \alpha) = \gamma - \mathcal{I}\{Y_1 - (X_1\theta + h(Y_2)) \leq 0\}$.
- $m(X, \alpha) = \gamma - \int F_{Y_1|Y_2,X}(x_1\theta + h(y_2)) f_{Y_2|X}(y_2, X) dy_2$.
- Case I: $(\Lambda_c^{r_h}(\mathcal{R}), ||h||_c = ||h \times w||_{L^\infty})$, $w(y) = (1 + y^2)^{-c}$, and $\lambda_n = 0$ ($\approx$ AC (03)).
- Case II: $(L^2(\mathcal{R}) \cap ||h||_c \leq M, ||h|_c = ||h||_{L^2})$ and $\lambda_n > 0$, $P(h) = ||D^s h||^2_{L^2}$ ($\approx$ HL (07)).

Partially Linear Quantile IV (cont.)

A: Low level standard assumptions:

- Smoothness and boundedness of $F_{Y_1|Y_2,X}$.
- Smoothing parameters, k_n and J_n.
- Identification conditions (CIN (07)).

Partially Linear Quantile IV (cont.)

A: Low level standard assumptions:

- Smoothness and boundedness of $F_{Y_1|Y_2,X}$.
- Smoothing parameters, k_n and J_n.
- Identification conditions (CIN (07)).

- Case I: A + $E[w^{-2}|X] \leq M < \infty$ then $|\widehat{\theta}_n - \theta_0| + \sup_y |(\widehat{h}_n - h_0)w(y)| = o_p(1)$.
- Case II: A + $\frac{n^{-2r_m/(2r_m+1)}}{\lambda_n} = o_p(1)$ then $|\widehat{\theta}_n - \theta_0| + ||\widehat{h}_n - h_0||_{L^2} = o_p(1)$.

Partially Linear Quantile IV (cont.)

- Case I: A + A3.5 - A3.6 + $\int f_{Y|X} w^{-2} \leq M < \infty$ then:
  - (i) If $b_k \leq \mu_k \asymp k^{-2a}$:
    $||\widehat{h}_n - h_0||_{L^2} = O_p(n^{-r_h/(2(a+r_h)+1)})$,
    with $k_n = O(n^{1/(2(a+r_h)+1)})$.
  - (ii) If $b_k \leq \mu_k \asymp \exp\{k^{-a}\}$ and $r_m = \infty$:
    $||\widehat{h}_n - h_0||_{L^2} = O_p([\ln(n)]^{-r_h/a})$,
    with $k_n = O_p([\ln(n)]^{1/a})$.
  - (iii) same rate as that in BCK (07) on nonparametric IV regression.

Partially Linear Quantile IV (cont.)

- Case I: A + A3.5 - A3.6 + $\int f_{Y|X} w^{-2} \le M < \infty$ then:
  - (i) If $b_k \le \mu_k \asymp k^{-2a}$:
    $||\widehat{h}_n - h_0||_{L^2} = O_p(n^{-r_h/(2(a+r_h)+1)})$,
    with $k_n = O(n^{1/(2(a+r_h)+1)})$.
  - (ii) If $b_k \le \mu_k \asymp \exp\{k^{-a}\}$ and $r_m = \infty$:
    $||\widehat{h}_n - h_0||_{L^2} = O_p([\ln(n)]^{-r_h/a})$,
    with $k_n = O_p([\ln(n)]^{1/a})$.
  - (iii) same rate as that in BCK (07) on nonparametric IV regression.
- Case II: polynomial rate coincides with HL (07).

Partially Linear Quantile IV (cont.)

- For Asymptotic normality we need:

Partially Linear Quantile IV (cont.)

- For Asymptotic normality we need:

D: Low level standard assumptions.

Partially Linear Quantile IV (cont.)

- For Asymptotic normality we need:

D: Low level standard assumptions.

- Key Assumptions:

E.1: Specific rate for $\widehat{\alpha}_n$ under "strong" norm, i.e., $\sqrt{J_n/n}||\widehat{\alpha}_n - \alpha_0||_{L^2} = o(n^{-1/2})$.

E.2: $\widehat{h}_\theta$ satisfies stochastic equi-continuity type of restrictions $\forall\theta \; : \; |\theta - \theta_0| = O_p(n^{-1/2})$.

- $\widehat{h}_\theta$ is the profiled estimator of h, i.e., fixing θ, we solve the SMD problem for h.
- E.2 is easy to check in linear problems (e.g. linear IV semiparametric regression) but hard to verify for non-linear problems.

Partially Linear Quantile IV (cont.)

- Both Cases: if A - E hold and the problem is **mildly** ill-posed, then $\widehat{\theta}_n$ is Asymp. Normal with variance

$$\frac{1}{\gamma(1-\gamma)} E\left[\left(\int f_{Y|X}(X_1\theta_0 + h_0; y_2, X)(X_1 - w^*)dy_2\right)^2\right].$$

Partially Linear Quantile IV (cont.)

- Both Cases: if A - E hold and the problem is **mildly** ill-posed, then $\widehat{\theta}_n$ is Asymp. Normal with variance

$$\frac{1}{\gamma(1-\gamma)} E\left[\left(\int f_{Y|X}(X_1\theta_0 + h_0; y_2, X)(X_1 - w^*)dy_2\right)^2\right].$$

- If $Y_2 = X_2$ and $f_{U|X} = f_U$ then $V = f_U^2(0)\frac{E[Var(X_1|X_2)]}{\gamma(1-\gamma)}$ which is optimal, Lee (03).
- Under **severely** ill-posedness some regularity restrictions on 2nd order approx. term are difficult to check.

Conclusion

- We propose penalized SMD estimators for semi/nonparametric conditional moment models, allowing for:
  - Possibly non-smooth generalized residual functions.
  - Possibly non-compact infinite-dimensional parameter space and sieve spaces, with flexible penalties.

Conclusion

- We propose penalized SMD estimators for semi/nonparametric conditional moment models, allowing for:
  - Possibly non-smooth generalized residual functions.
  - Possibly non-compact infinite-dimensional parameter space and sieve spaces, with flexible penalties.
- Obtain efficiency and normality of parametric part.
- Establish convergence rate for nonparametric functions that may depend on endogenous variables.

Conclusion

- We propose penalized SMD estimators for semi/nonparametric conditional moment models, allowing for:
  - Possibly non-smooth generalized residual functions.
  - Possibly non-compact infinite-dimensional parameter space and sieve spaces, with flexible penalties.
- Obtain efficiency and normality of parametric part.
- Establish convergence rate for nonparametric functions that may depend on endogenous variables.
- Future work:
  - Data-driven choice of smoothing parameters.
  - Time series extension.
  - Partially identified semi/nonparametric conditional moment models.