Tutorial in Econometrics Part IIb: Sieve Semiparametric Two-Step GMM Estimation and Inference

Xiaohong Chen (Yale)

NUS, IMS, May 16, 2014

Purpose of this Lecture

- **Purpose of this lecture**: Many nonlinear dynamic models in labor, IO and asset pricing can be estimated via semiparametric two-step or multi-step procedures. This lecture will focus on *simple* ways to conduct inference for *general* models estimated via semiparametric two-step or multi-step GMM in which unknown functions are estimated via the method of *sieves*.

Existing General Theory on Semiparametric Two-step GMM

- If nuisance functions $h_o()$ were known, the finite dimensional parameter θ_o is (over-)identified by d_g $(\geq d_\theta)$ moment conditions:
$$E\left[T^{-1}\sum_{t=1}^{T} g\left(Z_t,\theta_o,h_o(\cdot)\right)\right]=0.$$

Existing General Theory on Semiparametric Two-step GMM

- If nuisance functions $h_o()$ were known, the finite dimensional parameter θ_o is (over-)identified by d_g $(\geq d_\theta)$ moment conditions: $E\left[T^{-1}\sum_{t=1}^{T} g\left(Z_t, \theta_o, h_o(\cdot)\right)\right] = 0.$
- $h_o()$ is in fact unknown, but can be consistently estimated by $\widehat{h}_T(\cdot)$. Then θ_o is estimated by a semiparametric two-step GMM

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta}\left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right]' W_T\left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right].$$

Existing General Theory on Semiparametric Two-step GMM

- If nuisance functions $h_o()$ were known, the finite dimensional parameter θ_o is (over-)identified by d_g ($\geq d_\theta$) moment conditions: $E\left[T^{-1}\sum_{t=1}^{T} g\left(Z_t, \theta_o, h_o(\cdot)\right)\right] = 0.$
- $h_o()$ is in fact unknown, but can be consistently estimated by $\widehat{h}_T(\cdot)$. Then θ_o is estimated by a semiparametric two-step GMM

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta}\left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right].$$

- Newey (94), Andrews (94), Pakes and Olley (95), Chen, Linton and van Keilegom (03), Chen (07): $\sqrt{T}$ consistency and asymptotic normality (CAN) of $\widehat{\theta}_T$

Existing General Theory on Semiparametric Two-step GMM

- If nuisance functions $h_o()$ were known, the finite dimensional parameter θ_o is (over-)identified by d_g ($\geq d_\theta$) moment conditions: $E\left[T^{-1}\sum_{t=1}^{T} g\left(Z_t, \theta_o, h_o(\cdot)\right)\right] = 0.$
- $h_o()$ is in fact unknown, but can be consistently estimated by $\widehat{h}_T(\cdot)$. Then θ_o is estimated by a semiparametric two-step GMM

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta}\left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right].$$

- Newey (94), Andrews (94), Pakes and Olley (95), Chen, Linton and van Keilegom (03), Chen (07): $\sqrt{T}$ consistency and asymptotic normality (CAN) of $\widehat{\theta}_T$
- Newey (94): if $\widehat{\theta}_T$ is $\sqrt{T}$ CAN, then $Avar\left(\widehat{\theta}_T\right)$ does not depend on how $h_o()$ is estimated in the first step.

Advantage and Difficulty of Semiparametric Two-step GMM

- **Advantage**: Easy to compute. One could apply all kinds of nonparametric methods to estimate $h_o()$ in the first-step.

Advantage and Difficulty of Semiparametric Two-step GMM

- **Advantage**: Easy to compute. One could apply all kinds of nonparametric methods to estimate $h_o()$ in the first-step.
- **Too many applications** in labor, IO, asset pricing and other structural models to mention them during the talk.

Advantage and Difficulty of Semiparametric Two-step GMM

- **Advantage**: Easy to compute. One could apply all kinds of nonparametric methods to estimate $h_o()$ in the first-step.
- **Too many applications** in labor, IO, asset pricing and other structural models to mention them during the talk.
- **Difficulty**: Generally no closed form expression for $Avar(\widehat{\theta})$.

Advantage and Difficulty of Semiparametric Two-step GMM

- **Advantage**: Easy to compute. One could apply all kinds of nonparametric methods to estimate $h_o()$ in the first-step.
- **Too many applications** in labor, IO, asset pricing and other structural models to mention them during the talk.
- **Difficulty**: Generally no closed form expression for $Avar(\widehat{\theta})$.
- **Questions**: **(1)** assuming $\sqrt{T}$ CAN, when is the procedure semiparametrically efficient? **(2)** how to check $\sqrt{T}$ rate? **(3)** assuming $\sqrt{T}$ CAN, how to estimate $Avar(\widehat{\theta})$? **(4)** how to conduct overidentification test? **(5)** how to conduct inference robust to slower than $\sqrt{T}$ rate?

The Rest of This Lecture

- **Answer to Question (1)**: Semiparametric efficiency: Ai and Chen (12), Chen, Hahn and Liao (13).

The Rest of This Lecture

- **Answer to Question (1)**: Semiparametric efficiency: Ai and Chen (12), Chen, Hahn and Liao (13).
- **Answers to Questions (2), (3) and (4)**: Chen, Hahn and Liao (12).

The Rest of This Lecture

- **Answer to Question (1)**: Semiparametric efficiency: Ai and Chen (12), Chen, Hahn and Liao (13).
- **Answers to Questions (2), (3) and (4)**: Chen, Hahn and Liao (12).
- **Answer to Question (5)**: Chen, Hahn, Liao and Ridder (12): Semi- and non-parametric multistep procedures, and inference robust to slower than root-T rate.

The Rest of This Lecture

- **Answer to Question (1)**: Semiparametric efficiency: Ai and Chen (12), Chen, Hahn and Liao (13).
- **Answers to Questions (2), (3) and (4)**: Chen, Hahn and Liao (12).
- **Answer to Question (5)**: Chen, Hahn, Liao and Ridder (12): Semi- and non-parametric multistep procedures, and inference robust to slower than root-T rate.
- **An Empirical Example: Multivariate semi-nonparametric GARCH + Residual copula model**: Chen (13).

I Will Mention Results from the Following Papers

- Ai and Chen (AC, 12): "Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions containing unknown functions", 2012 *Journal of Econometrics*.
- Chen, Hahn and Liao (CHL, 13): "Asymptotic Efficiency of Semiparametric Two-step GMM".
- Chen, Hahn and Liao (CHL, 12): "Semiparametric Two-step GMM with Weakly Dependent Data".
- Chen, Hahn, Liao and Ridder (CHLR, 12): "nonparametric Two-step sieve estimation and inference".

Review: Root-T CAN of the Second-step GMM

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^T$ is stationary, ergodic.
- Recall semiparametric two-step GMM

$$\widehat{\theta}_T = \arg\min_{\theta \in \Theta} \left[\frac{1}{T} \sum_{t=1}^T g\left(Z_t, \theta, \widehat{h}_T(\cdot) \right) \right]' W_T \left[\frac{1}{T} \sum_{t=1}^T g\left(Z_t, \theta, \widehat{h}_T(\cdot) \right) \right].$$

- Let $G(\theta, h) = E\left[g\left(Z, \theta, h \right) \right]$, and $\Gamma_1(\theta, h)$ be the ordinary derivative of $G(\theta, h)$ wrt θ. Let $\Gamma_1 = \Gamma_1(\theta_o, h_o)$.

Review: Root-T CAN of the Second-step GMM

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^T$ is stationary, ergodic.
- Recall semiparametric two-step GMM

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta} \left[\frac{1}{T}\sum_{t=1}^T g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^T g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right].$$

- Let $G(\theta, h) = E\left[g\left(Z, \theta, h\right)\right]$, and $\Gamma_1(\theta, h)$ be the ordinary derivative of $G(\theta, h)$ wrt θ. Let $\Gamma_1 = \Gamma_1(\theta_o, h_o)$.
- Assume $\Gamma_1' W \Gamma_1$ is non-singular, with $W = p\lim_{T\to\infty} W_T$.

Review: Root-T CAN of the Second-step GMM

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^T$ is stationary, ergodic.
- Recall semiparametric two-step GMM

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta} \left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right].$$

- Let $G(\theta, h) = E\left[g\left(Z, \theta, h\right)\right]$, and $\Gamma_1(\theta, h)$ be the ordinary derivative of $G(\theta, h)$ wrt θ. Let $\Gamma_1 = \Gamma_1(\theta_o, h_o)$.
- Assume $\Gamma_1' W \Gamma_1$ is non-singular, with $W = p\lim_{T\to\infty} W_T$.
- Then: $\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \to_d \mathcal{N}[0, V_\theta]$ with

$$V_\theta = \left(\Gamma_1' W \Gamma_1\right)^{-1} \left(\Gamma_1' W V_1 W \Gamma_1\right) \left(\Gamma_1' W \Gamma_1\right)^{-1},$$

Review: Root-T CAN of the Second-step GMM

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^T$ is stationary, ergodic.
- Recall semiparametric two-step GMM

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta}\left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}_T(\cdot)\right)\right].$$

- Let $G(\theta, h) = E\left[g\left(Z, \theta, h\right)\right]$, and $\Gamma_1(\theta, h)$ be the ordinary derivative of $G(\theta, h)$ wrt θ. Let $\Gamma_1 = \Gamma_1(\theta_o, h_o)$.
- Assume $\Gamma_1' W \Gamma_1$ is non-singular, with $W = p\lim_{T\to\infty} W_T$.
- Then: $\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \to_d \mathcal{N}[0, V_\theta]$ with

$$V_\theta = \left(\Gamma_1' W \Gamma_1\right)^{-1}\left(\Gamma_1' W V_1 W \Gamma_1\right)\left(\Gamma_1' W \Gamma_1\right)^{-1},$$

- iff $T^{-\frac{1}{2}}\sum_{t=1}^{T} g\left(Z_t, \theta_o, \widehat{h}_T\right) \to_d \mathcal{N}[0, V_1]$ for a positive definite V_1.

Semiparametric efficiency?

- If second-step GMM is optimally weighted ($W = V_1^{-1}$), then

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \to_d \mathcal{N}[0, \left(\Gamma_1' V_1^{-1} \Gamma_1\right)^{-1}] \stackrel{d}{=} \mathcal{N}(0, V_\theta^o).$$

Semiparametric efficiency?

- If second-step GMM is optimally weighted ($W = V_1^{-1}$), then

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \to_d \mathcal{N}[0, \left(\Gamma_1' V_1^{-1} \Gamma_1\right)^{-1}] \stackrel{d}{=} \mathcal{N}(0, V_\theta^o).$$

- **Question**: Is this limited information optimality the same as full semiparametric efficiency bound?

Semiparametric efficiency?

- If second-step GMM is optimally weighted ($W = V_1^{-1}$), then

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \to_d \mathcal{N}[0, \left(\Gamma_1' V_1^{-1} \Gamma_1\right)^{-1}] \stackrel{d}{=} \mathcal{N}(0, V_\theta^o).$$

- **Question**: Is this limited information optimality the same as full semiparametric efficiency bound?
- **Answer**: It depends on how true unknown functions $h_o()$ is specified in the semiparametric structural model.

Semiparametric efficiency?

- If second-step GMM is optimally weighted ($W = V_1^{-1}$), then

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \rightarrow_d \mathcal{N}[0, \left(\Gamma_1' V_1^{-1} \Gamma_1\right)^{-1}] \stackrel{d}{=} \mathcal{N}(0, V_\theta^o).$$

- **Question**: Is this limited information optimality the same as full semiparametric efficiency bound?
- **Answer**: It depends on how true unknown functions $h_o()$ is specified in the semiparametric structural model.
- AC (12): semiparametric efficiency bound for sequential moment restriction. If $h_o()$ depends on "endogenous" variables, then this limited information optimality is NOT fully efficient in general.

Semiparametric efficiency?

- If second-step GMM is optimally weighted ($W = V_1^{-1}$), then

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \to_d \mathcal{N}[0, \left(\Gamma_1' V_1^{-1}\Gamma_1\right)^{-1}] \overset{d}{=} \mathcal{N}(0, V_\theta^o).$$

- **Question**: Is this limited information optimality the same as full semiparametric efficiency bound?
- **Answer**: It depends on how true unknown functions $h_o()$ is specified in the semiparametric structural model.
- AC (12): semiparametric efficiency bound for sequential moment restriction. If $h_o()$ depends on "endogenous" variables, then this limited information optimality is NOT fully efficient in general.
- CHL (13): semiparametric efficiency bound for overlapping moment restriction when $h_o()$ is "exactly identified", then this limited information optimality is fully efficient.

AC (12)'s Result on Semiparametric efficiency

- AC characterize the semiparametric efficiency bound for the *sequential* moment restrictions containing unknown functions:

$$E[\rho_t(Z;\theta_o,h_o(\cdot))|X^{(t)}]=0 \quad \text{for } t=1,...,T \quad \text{almost surely,} \quad (1)$$

where $\{1\}\subseteq\{X^{(1)}\}\subset\cdots\subset\{X^{(T)}\}$. When $X^{(1)}$ is constant then $E[\rho_1(Z;\theta,h(\cdot))|X^{(1)}]=E[\rho_1(Z;\theta,h(\cdot))]$. $Z=(Y',X^{(T)\prime})'$. $h_o(\cdot)=(h_{o1}(\cdot),...,h_{oL}(\cdot))$ may depend on endogenous variables Y and other unknown parameters.

AC (12)'s Result on Semiparametric efficiency

- AC characterize the semiparametric efficiency bound for the *sequential* moment restrictions containing unknown functions:

$$E[\rho_t(Z;\theta_o,h_o(\cdot))|X^{(t)}]=0 \quad \text{for } t=1,...,T \quad \text{almost surely,} \qquad (1)$$

where $\{1\} \subseteq \{X^{(1)}\} \subset \cdots \subset \{X^{(T)}\}$. When $X^{(1)}$ is constant then $E[\rho_1(Z;\theta,h(\cdot))|X^{(1)}]=E[\rho_1(Z;\theta,h(\cdot))]$. $Z=(Y',X^{(T)\prime})'$. $h_o(\cdot)=(h_{o1}(\cdot),...,h_{oL}(\cdot))$ may depend on endogenous variables Y and other unknown parameters.

- A special case of the sequential moment model (1) is:

$$E\left[g\left(Z;\theta_o,h_o(\cdot)\right)\right]=0,\; E[\rho_2(Z;h_o(\cdot))|X^{(2)}]=0, \qquad (2)$$

AC (12)'s Result on Semiparametric efficiency

- AC characterize the semiparametric efficiency bound for the *sequential* moment restrictions containing unknown functions:

$$E[\rho_t(Z;\theta_o,h_o(\cdot))|X^{(t)}]=0 \quad \text{for } t=1,...,T \quad \text{almost surely,} \qquad (1)$$

where $\{1\}\subseteq\{X^{(1)}\}\subset\cdots\subset\{X^{(T)}\}$. When $X^{(1)}$ is constant then $E[\rho_1(Z;\theta,h(\cdot))|X^{(1)}]=E[\rho_1(Z;\theta,h(\cdot))]$. $Z=(Y',X^{(T)\prime})'$. $h_o(\cdot)=(h_{o1}(\cdot),...,h_{oL}(\cdot))$ may depend on endogenous variables Y and other unknown parameters.

- A special case of the sequential moment model (1) is:

$$E\left[g\left(Z;\theta_o,h_o(\cdot)\right)\right]=0,\ E[\rho_2(Z;h_o(\cdot))|X^{(2)}]=0, \qquad (2)$$

- AC (12): For model (2), the limited information optimally weighted second-step GMM is NOT fully efficient whenever $E[g\left(Z;\theta_o,h_o(\cdot)\right)\rho_2(Z;h_o(\cdot))|X^{(2)}]\neq 0$.

CHL (13)'s Positive Result on Semiparametric efficiency

- Semiparametric model: θ_o is (over-) identified by (3):

$$E\left[g\left(Z;\theta_o, h_{1,o}(\cdot), ..., h_{L,o}(\cdot)\right)\right] = 0, \tag{3}$$

where the functions $h_o(\cdot) = (h_{1,o}(\cdot), ..., h_{L,o}(\cdot))$ are identified by (4):

$$E\left[\rho_\ell(Z, h_{\ell,o}(X_\ell)) \middle| X_\ell\right] = 0 \text{ almost surely } X_\ell, \quad \ell = 1, ..., L, \tag{4}$$

where the conditioning variables X_ℓ, $\ell = 1, ..., L$, could be *nested, overlapping or non-nested*.

CHL (13)'s Positive Result on Semiparametric efficiency

- Semiparametric model: θ_o is (over-) identified by (3):

$$E\left[g\left(Z;\theta_o,h_{1,o}(\cdot),...,h_{L,o}(\cdot)\right)\right]=0, \tag{3}$$

where the functions $h_o(\cdot)=(h_{1,o}(\cdot),...,h_{L,o}(\cdot))$ are identified by (4):

$$E\left[\rho_\ell(Z,h_{\ell,o}(X_\ell))\middle|X_\ell\right]=0 \text{ almost surely } X_\ell, \quad \ell=1,...,L, \tag{4}$$

where the conditioning variables X_ℓ, $\ell=1,...,L$, could be *nested*, *overlapping or non-nested*.

- CHL (13): For model (3)-(4), if (4) "exactly identifies" $h_{\ell,o}$ for $\ell=1,...,L$, then the limited information optimally weighted the limited information optimally weighted second-step GMM is fully efficient.

CHL (13)'s Positive Result on Semiparametric efficiency

- Semiparametric model: θ_o is (over-) identified by (3):

$$E\left[g\left(Z;\theta_o,h_{1,o}(\cdot),...,h_{L,o}(\cdot)\right)\right]=0, \tag{3}$$

where the functions $h_o(\cdot)=(h_{1,o}(\cdot),...,h_{L,o}(\cdot))$ are identified by (4):

$$E\left[\rho_\ell(Z,h_{\ell,o}(X_\ell))\middle|X_\ell\right]=0 \text{ almost surely } X_\ell,\quad \ell=1,...,L, \tag{4}$$

where the conditioning variables X_ℓ, $\ell=1,...,L$, could be *nested*, *overlapping or non-nested*.

- CHL (13): For model (3)-(4), if (4) "exactly identifies" $h_{\ell,o}$ for $\ell=1,...,L$, then the limited information optimally weighted the limited information optimally weighted second-step GMM is fully efficient.
- This covers many recent papers in labor and IO on semiparametric two-step GMM estimation of structure models.

CHL (12)'s answers to Questions (2), (3) and (4)

- Alternative characterization of $Avar(\widehat{\theta})$ of the second-step GMM $\widehat{\theta}$ with first-step sieve extremum estimation.

CHL (12)'s answers to Questions (2), (3) and (4)

- Alternative characterization of $Avar(\widehat{\theta})$ of the second-step GMM $\widehat{\theta}$ with first-step sieve extremum estimation.
- Consistent kernel estimation of long-run variance: (1) Wald tests; (2) overidentification J tests.

CHL (12)'s answers to Questions (2), (3) and (4)

- Alternative characterization of $Avar(\widehat{\theta})$ of the second-step GMM $\widehat{\theta}$ with first-step sieve extremum estimation.
- Consistent kernel estimation of long-run variance: (1) Wald tests; (2) overidentification J tests.
- Robust orthonormal series estimation of long-run variance: (1) F tests; (2) robust overidentification J tests

CHL (12)'s answers to Questions (2), (3) and (4)

- Alternative characterization of $Avar(\widehat{\theta})$ of the second-step GMM $\widehat{\theta}$ with first-step sieve extremum estimation.
- Consistent kernel estimation of long-run variance: (1) Wald tests; (2) overidentification J tests.
- Robust orthonormal series estimation of long-run variance: (1) F tests; (2) robust overidentification J tests
- Numerical equivalence of asymptotic variance estimates.

Sieve Semiparametric Two–step GMM

- A semiparametric model specifies that

$$E\left[g\left(Z_i, \theta, h_o(\cdot, \theta)\right)\right] = 0 \quad \text{at } \theta = \theta_o \in \Theta, \tag{5}$$

and for any fixed $\theta \in \Theta$, $h_o(\cdot, \theta) \in \mathcal{H}$ solves

$$Q(h_o) = \sup_{h \in \mathcal{H}} Q(h). \tag{6}$$

If $h_o()$ were known, the finite dimensional structural parameter θ_o is (over-)identified by d_g ($\geq d_\theta$) moment conditions (5). But $h_o()$ is unknown, except that it is identified as a maximizer of $Q()$ over $\mathcal{H}$. We suppress the arguments of the function h_o; thus $(\theta, h) \equiv (\theta, h(\cdot, \theta))$, $(\theta, h_o) \equiv (\theta, h_o(\cdot, \theta))$, $(\theta_o, h_o) \equiv (\theta_o, h_o(\cdot, \theta_o))$.

Sieve Semiparametric Two–step GMM

- A semiparametric model specifies that

$$E\left[g\left(Z_i, \theta, h_o(\cdot, \theta)\right)\right] = 0 \quad \text{at } \theta = \theta_o \in \Theta, \tag{5}$$

and for any fixed $\theta \in \Theta$, $h_o(\cdot, \theta) \in \mathcal{H}$ solves

$$Q(h_o) = \sup_{h \in \mathcal{H}} Q(h). \tag{6}$$

If $h_o()$ were known, the finite dimensional structural parameter θ_o is (over-)identified by d_g ($\geq d_\theta$) moment conditions (5). But $h_o()$ is unknown, except that it is identified as a maximizer of $Q()$ over $\mathcal{H}$. We suppress the arguments of the function h_o; thus $(\theta, h) \equiv (\theta, h(\cdot, \theta))$, $(\theta, h_o) \equiv (\theta, h_o(\cdot, \theta))$, $(\theta_o, h_o) \equiv (\theta_o, h_o(\cdot, \theta_o))$.

- Parameter spaces: Θ is a compact subset in $\mathcal{R}^{d_\theta}$. $(\mathcal{H}, d_s(,))$ is a possibly infinite dimensional (often non-compact) metric space.

Sieve Semiparametric Two–step GMM

- In the first-step unknown function $h_o()$ is estimated via a sieve extremum estimator $\widehat{h}$ that solves

$$\widehat{Q}_T(\widehat{h}) \geq \sup_{h \in \mathcal{H}_T} \widehat{Q}_T(h) - o_P(T^{-1})$$

where $\widehat{Q}_T$ is a random criterion that converges to Q over the sieve $\mathcal{H}_T = \mathcal{H}_{K(T)}$ as $T \to \infty$. A *sieve* $\{\mathcal{H}_{K(T)}\}$ is a sequence of approximating parameter spaces that become dense in $(\mathcal{H}, d_s(,))$ as $K(T) \to \infty$.

Sieve Semiparametric Two–step GMM

- In the first-step unknown function $h_o()$ is estimated via a sieve extremum estimator $\widehat{h}$ that solves

$$\widehat{Q}_T(\widehat{h}) \geq \sup_{h \in \mathcal{H}_T} \widehat{Q}_T(h) - o_P(T^{-1})$$

where $\widehat{Q}_T$ is a random criterion that converges to Q over the sieve $\mathcal{H}_T = \mathcal{H}_{K(T)}$ as $T \to \infty$. A *sieve* $\{\mathcal{H}_{K(T)}\}$ is a sequence of approximating parameter spaces that become dense in $(\mathcal{H}, d_s(,))$ as $K(T) \to \infty$.

- In the second-step θ_o is estimated by GMM with plugged-in $\widehat{h}()$:

$$\widehat{\theta}_T = \arg\min_{\theta \in \Theta} \left[\frac{1}{T} \sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}(\cdot) \right) \right]' W_T \left[\frac{1}{T} \sum_{t=1}^{T} g\left(Z_t, \theta, \widehat{h}(\cdot) \right) \right].$$

- The first-step sieve extremum estimation is very general. Different choices of criterion functions $Q()$, $\widehat{Q}_T$ and different choices of the sieve spaces allow for all kinds of nonparametric first step.

- The first-step sieve extremum estimation is very general. Different choices of criterion functions $Q()$, $\widehat{Q}_T$ and different choices of the sieve spaces allow for all kinds of nonparametric first step.
- Examples of sieves $\mathcal{H}_T$: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).

- The first-step sieve extremum estimation is very general. Different choices of criterion functions $Q()$, $\widehat{Q}_T$ and different choices of the sieve spaces allow for all kinds of nonparametric first step.
- Examples of sieves $\mathcal{H}_T$: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).
- Examples of criterion functions $Q()$, $\widehat{Q}_T$: ML, QML, MD, GMM, GEL, ... virtually all the existing criterion function for estimating nonlinear parametric models are valid choices; see Chen (07).

- The first-step sieve extremum estimation is very general. Different choices of criterion functions $Q()$, $\widehat{Q}_T$ and different choices of the sieve spaces allow for all kinds of nonparametric first step.
- Examples of sieves $\mathcal{H}_T$: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).
- Examples of criterion functions $Q()$, $\widehat{Q}_T$: ML, QML, MD, GMM, GEL, ... virtually all the existing criterion function for estimating nonlinear parametric models are valid choices; see Chen (07).
- **Sieve M**: $\widehat{Q}_T(h) = \frac{1}{T}\sum_{t=1}^{T} \varphi(Z_t, h)$. ML, QML, LS, NLS, QR.

- The first-step sieve extremum estimation is very general. Different choices of criterion functions $Q()$, $\widehat{Q}_T$ and different choices of the sieve spaces allow for all kinds of nonparametric first step.
- Examples of sieves $\mathcal{H}_T$: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).
- Examples of criterion functions $Q()$, $\widehat{Q}_T$: ML, QML, MD, GMM, GEL, ... virtually all the existing criterion function for estimating nonlinear parametric models are valid choices; see Chen (07).
- **Sieve M**: $\widehat{Q}_T(h) = \frac{1}{T}\sum_{t=1}^{T} \varphi(Z_t, h)$. ML, QML, LS, NLS, QR.
- **Sieve MD**: $\widehat{Q}_T(h) = \frac{-1}{T}\sum_{t=1}^{T} \widehat{m}(X_t, h)'\widehat{m}(X_t, h)$ for conditional moment restriction $E[\rho(Z, h_o)|X] = 0$, where $\widehat{m}(X, h)$ is a consistent nonparametric estimator of $E[\rho(Z, h)|X]$, Newey-Powell (03), Ai-Chen (03).

Asy. Normality of Sieve Semiparametric Two-step GMM

- Under mild conditions we have: $d_s(\widehat{h}, h_o) = o_p(1)$.

Asy. Normality of Sieve Semiparametric Two-step GMM

- Under mild conditions we have: $d_s(\widehat{h}, h_o) = o_p(1)$.
- Define a metric within a small $d_s(\cdot, \cdot)$ neighborhood $\mathcal{B}(h_o)$ of h_o:

$$\|h - h_o\| = \left\{ - \left[\frac{\partial^2}{\partial \tau^2} Q(h_o + \tau(h - h_o)) \right] |_{\tau=0} \right\}^{1/2} \text{ for any } h \in \mathcal{B}(h_o),$$

Asy. Normality of Sieve Semiparametric Two-step GMM

- Under mild conditions we have: $d_s(\widehat{h}, h_o) = o_p(1)$.
- Define a metric within a small $d_s(\cdot,\cdot)$ neighborhood $\mathcal{B}(h_o)$ of h_o:

$$\|h - h_o\| = \left\{ -\left[\frac{\partial^2}{\partial \tau^2} Q(h_o + \tau(h - h_o)) \right] |_{\tau=0} \right\}^{1/2} \text{ for any } h \in \mathcal{B}(h_o),$$

- Let $\mathcal{B}_T(h_o) = \mathcal{B}(h_o) \cap \mathcal{H}_T$ and $h_{o,T} \in \arg\min_{h \in \mathcal{B}_T(h_o)} \|h - h_o\|$.

Asy. Normality of Sieve Semiparametric Two-step GMM

- Under mild conditions we have: $d_s(\widehat{h}, h_o) = o_p(1)$.
- Define a metric within a small $d_s(\cdot,\cdot)$ neighborhood $\mathcal{B}(h_o)$ of h_o:

$$\|h - h_o\| = \left\{ - \left[\frac{\partial^2}{\partial \tau^2} Q(h_o + \tau(h - h_o)) \right] |_{\tau=0} \right\}^{1/2} \text{ for any } h \in \mathcal{B}(h_o),$$

- Let $\mathcal{B}_T(h_o) = \mathcal{B}(h_o) \cap \mathcal{H}_T$ and $h_{o,T} \in \arg\min_{h \in \mathcal{B}_T(h_o)} \|h - h_o\|$.
- Let $\mathcal{V}$ (and $\mathcal{V}_T$) be the closed linear span of $\mathcal{B}(h_o) - \{h_o\}$ (and $\mathcal{B}_T(h_o) - \{h_{o,T}\}$) under $\|\cdot\|$. Let $\langle \cdot, \cdot \rangle$ be the inner-product induced by $\|\cdot\|$.

Asy. Normality of Sieve Semiparametric Two-step GMM

- Under mild conditions we have: $d_s(\widehat{h}, h_o) = o_p(1)$.
- Define a metric within a small $d_s(\cdot, \cdot)$ neighborhood $\mathcal{B}(h_o)$ of h_o:

$$\|h - h_o\| = \left\{ - \left[\frac{\partial^2}{\partial \tau^2} Q(h_o + \tau(h - h_o)) \right] |_{\tau=0} \right\}^{1/2} \text{ for any } h \in \mathcal{B}(h_o),$$

- Let $\mathcal{B}_T(h_o) = \mathcal{B}(h_o) \cap \mathcal{H}_T$ and $h_{o,T} \in \arg\min_{h \in \mathcal{B}_T(h_o)} \|h - h_o\|$.
- Let $\mathcal{V}$ (and $\mathcal{V}_T$) be the closed linear span of $\mathcal{B}(h_o) - \{h_o\}$ (and $\mathcal{B}_T(h_o) - \{h_{o,T}\}$) under $\|\cdot\|$. Let $\langle \cdot, \cdot \rangle$ be the inner-product induced by $\|\cdot\|$.
- Under mild conditions we have: $\left\|\widehat{h} - h_o\right\| = o_p(T^{-1/4})$. See, e.g., Chen-Shen (98) for sieve M estimation for weakly dependent data; Chen-Pouzo (12) for sieve MD estimation.

Asymptotic normality of the second-step GMM

- Let $\Gamma_2(\theta, h)[v] \equiv \left.\frac{\partial G[\theta, h(\cdot)+\tau v(\cdot)]}{\partial \tau}\right|_{\tau=0}$ be the pathwise derivative of $G(\theta, h) = E\left[g\left(Z, \theta, h\right)\right]$ at $h \in \mathcal{H}$ in the direction $v \in \mathcal{V}$.

Asymptotic normality of the second-step GMM

- Let $\Gamma_2(\theta, h)[v] \equiv \left.\frac{\partial G[\theta, h(\cdot)+\tau v(\cdot)]}{\partial \tau}\right|_{\tau=0}$ be the pathwise derivative of $G(\theta, h) = E\left[g\left(Z, \theta, h\right)\right]$ at $h \in \mathcal{H}$ in the direction $v \in \mathcal{V}$.
- A necessary condition for $T^{-\frac{1}{2}} \sum_{t=1}^{T} g\left(Z_t, \theta_o, \widehat{h}\right) \rightarrow_d \mathcal{N}[0, V_1]$ is that $\Gamma_{2,j}(\theta_o, h_o)[] : \mathcal{V} \rightarrow \mathcal{R}$ is a bounded functional for all $j = 1, ..., d_g$

Asymptotic normality of the second-step GMM

- Let $\Gamma_2(\theta, h)[v] \equiv \frac{\partial G[\theta, h(\cdot)+\tau v(\cdot)]}{\partial \tau}\Big|_{\tau=0}$ be the pathwise derivative of $G(\theta, h) = E\left[g\left(Z, \theta, h\right)\right]$ at $h \in \mathcal{H}$ in the direction $v \in \mathcal{V}$.
- A necessary condition for $T^{-\frac{1}{2}} \sum_{t=1}^{T} g\left(Z_t, \theta_o, \widehat{h}\right) \to_d \mathcal{N}[0, V_1]$ is that $\Gamma_{2,j}(\theta_o, h_o)[] : \mathcal{V} \to \mathcal{R}$ is a bounded functional for all $j = 1, ..., d_g$
- Iff $\Gamma_{2,j}(\theta_o, h_o)[] : \mathcal{V} \to \mathcal{R}$ is a bounded functional, there is a Riesz representer $v_j^* \in \mathcal{V}$ such that

$$\Gamma_{2,j}(\theta_o, h_o)[v] = \left\langle v, v_j^* \right\rangle \text{ for all } v \in \mathcal{V},$$

$$\left\| v_j^* \right\|^2 = \sup_{v \in \mathcal{V}, v \neq 0} \frac{\left|\Gamma_{2,j}(\theta_o, h_o)[v]\right|^2}{\left\| v \right\|^2} < \infty.$$

Asymptotic normality of the second-step GMM

- Let $\Gamma_2(\theta,h)[v] \equiv \left.\frac{\partial G[\theta,h(\cdot)+\tau v(\cdot)]}{\partial \tau}\right|_{\tau=0}$ be the pathwise derivative of $G(\theta,h) = E\left[g\left(Z,\theta,h\right)\right]$ at $h \in \mathcal{H}$ in the direction $v \in \mathcal{V}$.
- A necessary condition for $T^{-\frac{1}{2}} \sum_{t=1}^{T} g\left(Z_t, \theta_o, \widehat{h}\right) \to_d \mathcal{N}[0, V_1]$ is that $\Gamma_{2,j}(\theta_o, h_o)[] : \mathcal{V} \to \mathcal{R}$ is a bounded functional for all $j = 1, ..., d_g$
- Iff $\Gamma_{2,j}(\theta_o, h_o)[] : \mathcal{V} \to \mathcal{R}$ is a bounded functional, there is a Riesz representer $v_j^* \in \mathcal{V}$ such that

$$\begin{aligned} \Gamma_{2,j}(\theta_o,h_o)[v] &= \left\langle v, v_j^* \right\rangle \text{ for all } v \in \mathcal{V}, \\ \left\| v_j^* \right\|^2 &= \sup_{v\in\mathcal{V}, v\neq 0} \frac{\left|\Gamma_{2,j}(\theta_o,h_o)[v]\right|^2}{\left\| v \right\|^2} < \infty. \end{aligned}$$

- Then $T^{-\frac{1}{2}} \sum_{t=1}^{T} g\left(Z_t, \theta_o, \widehat{h}\right) =$
$T^{-\frac{1}{2}} \sum_{t=1}^{T} g\left(Z_t, \theta_o, h_o\right) + \sqrt{T}\left\langle \widehat{h} - h_o, \mathbf{v}^* \right\rangle + o_p(1).$

Asy. Normality of Sieve Semiparametric Two-step GMM

- Difficult to solve for $v_j^* \in \mathcal{V}$ in closed form. But can compute a sieve Riesz representer $v_{j,T}^* \in \mathcal{V}_T$ in closed form such that

  $\Gamma_{2,j}(\theta_o, h_o)[v] = \left\langle v, v_{j,T}^* \right\rangle$ for all $v \in \mathcal{V}_T$, and

$$\left\| v_{j,T}^* \right\|^2 = \sup_{v \in \mathcal{V}_T, v \neq 0} \frac{\left| \Gamma_{2,j}(\theta_o, h_o)[v] \right|^2}{\left\| v \right\|^2} \to \left\| v_j^* \right\|^2 < \infty.$$

Asy. Normality of Sieve Semiparametric Two-step GMM

- Difficult to solve for $v_j^* \in \mathcal{V}$ in closed form. But can compute a sieve Riesz representer $v_{j,T}^* \in \mathcal{V}_T$ in closed form such that

  $\Gamma_{2,j}(\theta_o, h_o)[v] = \left\langle v, v_{j,T}^* \right\rangle$ for all $v \in \mathcal{V}_T$, and

$$\left\| v_{j,T}^* \right\|^2 = \sup_{v \in \mathcal{V}_T, v \neq 0} \frac{\left| \Gamma_{2,j}(\theta_o, h_o)[v] \right|^2}{\left\| v \right\|^2} \to \left\| v_j^* \right\|^2 < \infty.$$

- **Proposition 1.** Sieve semiparametric two-step GMM satisfies $\sqrt{T}\left(\widehat{\theta}_T - \theta_o\right) \to_d \mathcal{N}[0, V_\theta]$ with

$$Avar(\widehat{\theta}_T) = V_\theta = \left(\Gamma_1' W \Gamma_1\right)^{-1} \left(\Gamma_1' W V_1 W \Gamma_1\right) \left(\Gamma_1' W \Gamma_1\right)^{-1},$$

$$V_1 = \lim_{n \to \infty} E\left[T^{-1} \sum_{i=1}^{T} \sum_{j=1}^{T} \{S_i(\alpha_o)[\mathbf{v}_T^*]\}\{S_j(\alpha_o)[\mathbf{v}_T^*]\}' \right],$$

$S_i(\alpha_o)[\mathbf{v}_T^*] =$

$g(Z_i, \theta_o, h_o) + \left(\Delta(Z_i, h_o)[v_{1,T}^*], ..., \Delta(Z_i, h_o)[v_{d_g,T}^*]\right)'$ is a sieve score.

Consistent kernel estimation of the LRV

- Newey-West type kernel estimate of V_1 is defined as

$$\widehat{V}_{1,T} = \sum_{t=-T+1}^{T-1} \mathcal{K}\left(\frac{t}{M_T}\right) \Upsilon_{T,t}(\widehat{\alpha}_T)\left[\widehat{\mathbf{v}}_T^*, \widehat{\mathbf{v}}_T^*\right],$$

where $M_T \to \infty$ as $T \to \infty$, and $\Upsilon_{T,t}(\widehat{\alpha}_T)\left[\widehat{\mathbf{v}}_T^*, \widehat{\mathbf{v}}_T^*\right]$ is defined as

$$\begin{cases} \frac{1}{T}\sum_{l=t+1}^{T} S_l\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\{S_{l-t}\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\}' & \text{for } t \geq 0 \\ \frac{1}{T}\sum_{l=-t+1}^{T} S_l\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\{S_{l+t}\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\}' & \text{for } t < 0 \end{cases}.$$

Consistent kernel estimation of the LRV

- Newey-West type kernel estimate of V_1 is defined as

$$\widehat{V}_{1,T} = \sum_{t=-T+1}^{T-1} \mathcal{K}\left(\frac{t}{M_T}\right) \Upsilon_{T,t}(\widehat{\alpha}_T)\left[\widehat{\mathbf{v}}_T^*, \widehat{\mathbf{v}}_T^*\right],$$

where $M_T \to \infty$ as $T \to \infty$, and $\Upsilon_{T,t}(\widehat{\alpha}_T)\left[\widehat{\mathbf{v}}_T^*, \widehat{\mathbf{v}}_T^*\right]$ is defined as

$$\begin{cases} \frac{1}{T}\sum_{l=t+1}^{T} S_l\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\left\{S_{l-t}\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\right\}' & \text{for } t \geq 0 \\ \frac{1}{T}\sum_{l=-t+1}^{T} S_l\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\left\{S_{l+t}\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]\right\}' & \text{for } t < 0 \end{cases}.$$

- $\mathcal{K}(\cdot)$ is the kernel function and M_T is the bandwidth. $\widehat{\mathbf{v}}_T^*$ is the estimate of $\mathbf{v}_T^*$.

Kernel auto-correlation robust inference

- **Theorem 1.** Under some regularity conditions, the kernel LRV estimate $\widehat{V}_1$ is consistent, i.e. $\widehat{V}_{1,T} \to_p V_1$. Also,

$$\widehat{\Gamma}_1 = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(Z_t, \widehat{\theta}_T, \widehat{h}_T)}{\partial \theta} \to_p \Gamma_1.$$

Kernel auto-correlation robust inference

- **Theorem 1.** Under some regularity conditions, the kernel LRV estimate $\widehat{V}_1$ is consistent, i.e. $\widehat{V}_{1,T} \to_p V_1$. Also,

$$\widehat{\Gamma}_1 = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(Z_t, \widehat{\theta}_T, \widehat{h}_T)}{\partial \theta} \to_p \Gamma_1.$$

- $\widehat{V}_\theta = \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \left(\widehat{\Gamma}_1' W_T \widehat{V}_{1,T} W_T \widehat{\Gamma}_1\right) \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \to_p V_\theta$

Kernel auto-correlation robust inference

- **Theorem 1.** Under some regularity conditions, the kernel LRV estimate $\widehat{V}_1$ is consistent, i.e. $\widehat{V}_{1,T} \to_p V_1$. Also,

$$\widehat{\Gamma}_1 = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(Z_t, \widehat{\theta}_T, \widehat{h}_T)}{\partial \theta} \to_p \Gamma_1.$$

- $\widehat{V}_\theta = \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \left(\widehat{\Gamma}_1' W_T \widehat{V}_{1,T} W_T \widehat{\Gamma}_1\right) \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \to_p V_\theta$
- Thus $\sqrt{T} \widehat{V}_\theta^{-1/2} (\widehat{\theta}_T - \theta_o) \to_d \mathcal{N}(0, I_{d_\theta})$

Kernel auto-correlation robust inference

- **Theorem 1.** Under some regularity conditions, the kernel LRV estimate $\widehat{V}_1$ is consistent, i.e. $\widehat{V}_{1,T} \to_p V_1$. Also,

$$\widehat{\Gamma}_1 = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(Z_t, \widehat{\theta}_T, \widehat{h}_T)}{\partial \theta} \to_p \Gamma_1.$$

- $\widehat{V}_\theta = \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \left(\widehat{\Gamma}_1' W_T \widehat{V}_{1,T} W_T \widehat{\Gamma}_1\right) \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \to_p V_\theta$
- Thus $\sqrt{T}\widehat{V}_\theta^{-1/2}(\widehat{\theta}_T - \theta_o) \to_d \mathcal{N}(0, I_{d_\theta})$
- $T(\widehat{\theta}_T - \theta_o)' \widehat{V}_\theta^{-1} (\widehat{\theta}_T - \theta_o) \to_d \chi^2_{d_\theta}$.

Kernel auto-correlation robust inference

- **Theorem 1.** Under some regularity conditions, the kernel LRV estimate $\widehat{V}_1$ is consistent, i.e. $\widehat{V}_{1,T} \to_p V_1$. Also,

$$\widehat{\Gamma}_1 = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(Z_t, \widehat{\theta}_T, \widehat{h}_T)}{\partial \theta} \to_p \Gamma_1.$$

- $\widehat{V}_\theta = \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \left(\widehat{\Gamma}_1' W_T \widehat{V}_{1,T} W_T \widehat{\Gamma}_1\right) \left(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1\right)^{-1} \to_p V_\theta$
- Thus $\sqrt{T}\widehat{V}_\theta^{-1/2}(\widehat{\theta}_T - \theta_o) \to_d \mathcal{N}(0, I_{d_\theta})$
- $T(\widehat{\theta}_T - \theta_o)' \widehat{V}_\theta^{-1}(\widehat{\theta}_T - \theta_o) \to_d \chi^2_{d_\theta}$.
- Auto-covariance robust inference about θ_o can be conducted.

Kernel overidentification test of moment conditions

- Suppose the structural parameter θ_o is over-identified, i.e. $d_g > d_\theta$.

Kernel overidentification test of moment conditions

- Suppose the structural parameter θ_o is over-identified, i.e. $d_g > d_\theta$.
- Given the first-step sieve extremum estimate $\widehat{h}_T$ and the second-step GMM estimate $\widehat{\theta}_T$, we would like to test the validities of the moment conditions

$$E\left[g\left(Z_t, \theta_o, h_o\right)\right] \stackrel{?}{=} 0.$$

Kernel overidentification test of moment conditions

- Suppose the structural parameter θ_o is over-identified, i.e. $d_g > d_\theta$.
- Given the first-step sieve extremum estimate $\widehat{h}_T$ and the second-step GMM estimate $\widehat{\theta}_T$, we would like to test the validities of the moment conditions

$$E\left[g\left(Z_t, \theta_o, h_o\right)\right] \stackrel{?}{=} 0.$$

- Define $\widehat{W}_T^{-1} = \sum_{t=-T+1}^{T-1} \mathcal{K}\left(\frac{t}{M_T}\right) \overline{\Upsilon}_{T,t}(\widehat{\alpha}_T)\left[\widehat{\mathbf{v}}_T^*, \widehat{\mathbf{v}}_T^*\right]$, where $\overline{\Upsilon}_{T,t}(\widehat{\alpha}_T)\left[\widehat{\mathbf{v}}_T^*, \widehat{\mathbf{v}}_T^*\right]$ is defined as

$$\begin{cases} \frac{1}{T}\sum\limits_{l=t+1}^{T}\left(\widehat{S}_{l,T}^* - \overline{S}_T\right)\left(\widehat{S}_{l-t,T}^* - \overline{S}_T\right)' & \text{for } t \geq 0 \\ \frac{1}{T}\sum\limits_{l=-t+1}^{T}\left(\widehat{S}_{l,T}^* - \overline{S}_T\right)\left(\widehat{S}_{l+t,T}^* - \overline{S}_T\right)' & \text{for } t < 0 \end{cases}$$

with $\widehat{S}_{l,T}^* = S_l\left(\widehat{\alpha}_T\right)\left[\widehat{\mathbf{v}}_T^*\right]$ and $\overline{S}_T = \frac{1}{T}\sum\limits_{l=1}^{T}\widehat{S}_{l,T}^*$.

Kernel overidentification test of moment conditions

- Our over-identification test statistic is

$$J_T = \left[T^{-\frac{1}{2}} \sum_{t=1}^{T} g(Z_t, \widehat{\theta}_T, \widehat{h}_T) \right] \widehat{W}_T \left[T^{-\frac{1}{2}} \sum_{t=1}^{T} g(Z_t, \widehat{\theta}_T, \widehat{h}_T) \right].$$

Kernel overidentification test of moment conditions

- Our over-identification test statistic is

$$J_T = \left[T^{-\frac{1}{2}} \sum_{t=1}^{T} g(Z_t, \widehat{\theta}_T, \widehat{h}_T) \right] \widehat{W}_T \left[T^{-\frac{1}{2}} \sum_{t=1}^{T} g(Z_t, \widehat{\theta}_T, \widehat{h}_T) \right].$$

- **Proposition 2.** Under the null hypothesis $E\left[g\left(Z_t, \theta_o, h_o\right)\right] = 0$ with $d_g > d_\theta$, we have:

$$J_T \to_d \chi^2_{d_g - d_\theta}.$$

Numerical Equivalence of LRV Estimates

- Suppose a researcher believes $h_o(\cdot) = P_K'(\cdot)\beta_{o,K}$ with fixed K.

Numerical Equivalence of LRV Estimates

- Suppose a researcher believes $h_o(\cdot) = P_K'(\cdot)\beta_{o,K}$ with fixed K.
- He/she estimates $\beta_{o,K}$ and h_o by the parametric extremum estimation

$$\widehat{\beta}_{T,K} = argmax_{\beta_K \in \mathcal{B}_K} \widehat{Q}_n \left(P_K'(\cdot)\beta \right)$$

where $\mathcal{B}_K$ is a nonempty compact subset in $\mathcal{R}^K$.

Numerical Equivalence of LRV Estimates

- Suppose a researcher believes $h_o(\cdot) = P'_K(\cdot)\beta_{o,K}$ with fixed K.
- He/she estimates $\beta_{o,K}$ and h_o by the parametric extremum estimation

$$\widehat{\beta}_{T,K} = argmax_{\beta_K \in \mathcal{B}_K} \widehat{Q}_n \left(P'_K(\cdot)\beta \right)$$

where $\mathcal{B}_K$ is a nonempty compact subset in $\mathcal{R}^K$.

- θ_o is estimated in GMM

$$\widehat{\theta}_{T,K} = argmin_{\theta \in \Theta} \left[\sum_{t=1}^{T} g_K(Z_t, \theta, \widehat{\beta}_{T,K}) \right] \frac{W_T}{T} \left[\sum_{t=1}^{T} g_K(Z_t, \theta, \widehat{\beta}_{T,K})' \right],$$

where $g_K\left(Z, \theta, \beta_{o,K}\right) \equiv g\left(Z, \theta, P'_K\left(\cdot\right)\beta_{o,K}\right)$.

Numerical Equivalence of LRV Estimates

Proposition 3. Suppose that the parametric specification is true, then under some regularity conditions,

$$\sqrt{T}(\widehat{\theta}_{T,K} - \theta_o) \to_d \mathcal{N}\left(0, V_{\theta,K}\right)$$

where

$$V_{\theta,K} = \left(\Gamma'_{1,K} W \Gamma_{1,K}\right)^{-1} \Gamma'_{1,K} W V_{1,K} W \Gamma_{1,K} \left(\Gamma'_{1,K} W \Gamma_{1,K}\right)^{-1},$$

$$V_{1,K} = \lim_{T\to\infty} \text{Var}\left\{ T^{-\frac{1}{2}} \sum_{t=1}^{T} \left[g_K(Z_t, \theta_o, \beta_{o,K}) + \Gamma'_{2,K} R_{o,K}^{-1} \frac{\partial \varphi_K\left(Z_t, \beta_{o,K}\right)}{\partial \beta_K} \right] \right\},$$

$$\Gamma_{1,K} = E\left[\frac{\partial g_K\left(Z, \theta_o, \beta_{o,K}\right)}{\partial \theta'}\right] \text{ and } \Gamma_{2,K} = E\left[\frac{\partial g_K\left(Z, \theta_o, \beta_{o,K}\right)}{\partial \beta'_K}\right].$$

- **Note**: the asymptotic variance of $\widehat{\theta}_{T,K}$ is different from the asymptotic variance of $\widehat{\theta}_T$.

Numerical Equivalence of LRV Estimates

An typical estimator of $V_{1,K}$ is $\widehat{V}_{1,K} = \sum_{t=-T+1}^{T-1} \mathcal{K}\left(\frac{t}{M_T}\right) \widehat{\Upsilon}_K(t)$, where:

$$\widehat{\Upsilon}_K(t) = \begin{cases} \sum_{l=t+1}^{T} \frac{\widehat{S}_{l,K}\widehat{S}'_{l-t,K}}{T} & \text{for } t \geq 0 \\ \sum_{l=-t+1}^{T} \frac{\widehat{S}_{l,K}\widehat{S}'_{l+t,K}}{T} & \text{for } t < 0 \end{cases},$$

$$\widehat{S}_{t,K} = g_K\left(Z_t, \widehat{\theta}_{T,K}, \widehat{\beta}_{T,K}\right) + \widehat{\Gamma}'_{2,K}\widehat{R}_K^{-1} \frac{\partial \varphi_K\left(Z_t, \widehat{\beta}_{T,K}\right)}{\partial \beta'_K},$$

$$\widehat{\Gamma}_{2,K} \equiv \frac{1}{T}\sum_{t=1}^{T} \frac{\partial g_K\left(Z_t, \widehat{\theta}_{T,K}, \widehat{\beta}_{T,K}\right)}{\partial \beta'_K},$$

$$\widehat{R}_K \equiv -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \varphi_P\left(Z_t, \widehat{\beta}_{T,K}\right)}{\partial \beta \partial \beta'}.$$

Numerical Equivalence of LRV Estimates

The consistent estimate of the asymptotic variance of $\widehat{\theta}_{T,K}$ can be defined as

$$\widehat{V}_{\theta,P} = \left(\widehat{\Gamma}'_{1,K} W_T \widehat{\Gamma}_{1,K}\right)^{-1} \widehat{\Gamma}'_{1,K} W_T \widehat{V}_{1,K} W_T \widehat{\Gamma}_{1,K} \left(\widehat{\Gamma}'_{1,K} W_T \widehat{\Gamma}_{1,K}\right)^{-1}$$

where $\widehat{V}_{1,K}$ is defined in the previous slide and

$$\widehat{\Gamma}'_{1,K} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_K(Z_t, \widehat{\theta}_{T,K}, \widehat{\beta}_{T,K})}{\partial \theta}.$$

Proposition 4. If $K = K(T)$, then $\widehat{V}_{\theta,P} = \widehat{V}_{\theta}$ for all T

- For the sieve method, in finite samples, not only the estimation can be viewed as a parametric problem after the number of the basis functions is determined, but also the inference can be conducted as if the model is parametrically specified.

Orthonormal series estimation of LRV

- $\{\phi_m\}_{m=1}^{\infty}$ is a sequence of orthonormal basis in $L_2[0,1]$ with $\int_0^1 \phi_m(s)ds = 0$.

Orthonormal series estimation of LRV

- $\{\phi_m\}_{m=1}^{\infty}$ is a sequence of orthonormal basis in $L_2[0,1]$ with $\int_0^1 \phi_m(s)ds = 0$.
- We define the following series projection

$$\hat{\Lambda}_m = \frac{(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1)^{-1} \widehat{\Gamma}_1' W_T}{\sqrt{T}} \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right) S_t(\widehat{\theta}_T, \widehat{h}_T)\left[\mathbf{v}_T^*\right]$$

for $m = 1, ..., M$, where

$$S_t(\widehat{\theta}_T, \widehat{h}_T)\left[\mathbf{v}_T^*\right] = g(Z_t, \widehat{\theta}_T, \widehat{h}_T) + \Delta(Z_t, \widehat{h}_T)\left[\widehat{\mathbf{v}}_T^*\right].$$

Orthonormal series estimation of LRV

- $\{\phi_m\}_{m=1}^{\infty}$ is a sequence of orthonormal basis in $L_2[0,1]$ with $\int_0^1 \phi_m(s)ds = 0$.
- We define the following series projection

$$\hat{\Lambda}_m = \frac{(\widehat{\Gamma}_1' W_T \widehat{\Gamma}_1)^{-1} \widehat{\Gamma}_1' W_T}{\sqrt{T}} \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right) S_t(\widehat{\theta}_T, \widehat{h}_T)\left[\mathbf{v}_T^*\right]$$

for $m = 1, ..., M$, where

$$S_t(\widehat{\theta}_T, \widehat{h}_T)\left[\mathbf{v}_T^*\right] = g(Z_t, \widehat{\theta}_T, \widehat{h}_T) + \Delta(Z_t, \widehat{h}_T)\left[\widehat{\mathbf{v}}_T^*\right].$$

- Orthonormal series estimator of V_θ is:

$$\widehat{V}_{\theta,M} \equiv \frac{1}{M} \sum_{m=1}^{M} \hat{\Lambda}_m \hat{\Lambda}_m'.$$

A natural extension of Phillips (2005) and Sun (2013) to semiparametric two-step GMM setting.

OS robust inference for scalar parameter

Theorem 2. Under some regularity conditions, we have

$$t_{M,T} \equiv \sqrt{T}\frac{\widehat{\theta}_T - \theta_o}{\sqrt{\widehat{V}_{\theta,M}}} \to_d \ t(M),$$

where $t(M)$ is a student-t random variable with degree of freedom M.

- When the number of the basis functions $M \to \infty$, $t_{M,T}$ will converge in distribution to the standard normal distribution.

OS robust inference for vector parameter

Theorem 3. Under some regularity conditions, we have

$$F_{M,T} \equiv \frac{T}{d_\theta}\left(\widehat{\theta}_T - \theta_o\right)' \widehat{V}_{\theta,M}^{-1}\left(\widehat{\theta}_T - \theta_o\right).$$

$$\frac{M - d_\theta + 1}{M} F_{M,T} \rightarrow_d \mathcal{F}(d_\theta, M - d_\theta + 1),$$

where $\mathcal{F}(a,b)$ denotes the F-distribution with (a,b) degree freedom.

- When the number of the basis functions $M \to \infty$, $F_{M,T}$ will converge in distribution to the chi-square random variable with degree of freedom d_θ.

OS robust overidentification test of the moment conditions

- Let $\widetilde{W}_{r,T}^{-1} = \frac{1}{M}\sum_{m=1}^{M} \widetilde{\Lambda}_m \widetilde{\Lambda}_m'$, where

$$\widetilde{\Lambda}_m = T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right)\left\{g(Z_t, \widetilde{\theta}_T, \widehat{h}_T) + \Delta(Z_t, \widehat{h}_T)\left[\widehat{\mathbf{v}}_T^*\right]\right\}$$

and $\widetilde{\theta}_T$ is some preliminary GMM estimate.

OS robust overidentification test of the moment conditions

- Let $\widetilde{W}_{r,T}^{-1} = \frac{1}{M}\sum_{m=1}^{M} \widetilde{\Lambda}_m \widetilde{\Lambda}_m'$, where

$$\widetilde{\Lambda}_m = T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right)\left\{g(Z_t, \widetilde{\theta}_T, \widehat{h}_T) + \Delta(Z_t, \widehat{h}_T)\left[\widehat{\mathbf{v}}_T^*\right]\right\}$$

  and $\widetilde{\theta}_T$ is some preliminary GMM estimate.

- The two-step GMM estimate is defined as

$$\widehat{\theta}_{r,T} = argmin_{\theta\in\Theta}\left[\sum_{t=1}^{T} g(Z_t, \theta, \widehat{h}_T)\right] \frac{\widetilde{W}_{r,T}}{T}\left[\sum_{t=1}^{T} g(Z_t, \theta, \widehat{h}_T)\right].$$

OS robust overidentification test of the moment conditions

- Let $\widetilde{W}_{r,T}^{-1} = \frac{1}{M}\sum_{m=1}^{M} \widetilde{\Lambda}_m \widetilde{\Lambda}_m'$, where

$$\widetilde{\Lambda}_m = T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right)\left\{g(Z_t, \widetilde{\theta}_T, \widehat{h}_T) + \Delta(Z_t, \widehat{h}_T)\left[\widehat{\mathbf{v}}_T^*\right]\right\}$$

and $\widetilde{\theta}_T$ is some preliminary GMM estimate.

- The two-step GMM estimate is defined as

$$\widehat{\theta}_{r,T} = argmin_{\theta\in\Theta}\left[\sum_{t=1}^{T} g(Z_t, \theta, \widehat{h}_T)\right]\frac{\widetilde{W}_{r,T}}{T}\left[\sum_{t=1}^{T} g(Z_t, \theta, \widehat{h}_T)\right].$$

- Define $\widehat{W}_{r,n}^{-1} = \frac{1}{M}\sum_{m=1}^{M} \widehat{\Lambda}_{r,m} \widehat{\Lambda}_{r,m}'$, where

$$\widehat{\Lambda}_{r,m} = T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right)\left\{g(Z_t, \widehat{\theta}_{r,T}, \widehat{h}_T) + \Delta(Z_t, \widehat{h}_T)\left[\widehat{\mathbf{v}}_T^*\right]\right\}.$$

OS robust overidentification test of the moment conditions

- Define the J-test statistic

$$J_{M,T} = \left[\sum_{t=1}^{T} g(Z_t, \widehat{\theta}_{r,T}, \widehat{h}_T)\right] \frac{\widehat{W}_{r,T}}{T} \left[\sum_{t=1}^{T} g(Z_t, \widehat{\theta}_{r,T}, \widehat{h}_T)\right].$$

Theorem 4. Under the null hypothesis $E\left[g\left(Z_t, \theta_o, h_o\right)\right] = 0$ with $d_g > d_\theta$, we have for fixed finite $M > (d_g - d_\theta)$,

$$J^*_{M,T} \equiv \frac{M - (d_g - d_\theta) + 1}{M(d_g - d_\theta)} J_{M,T} \to_d \mathcal{F}(d_g - d_\theta, M - (d_g - d_\theta) + 1),$$

where $\mathcal{F}(a, b)$ denotes a F-distributed random variable with degree of freedom (a, b).

Summary of CHL (12)

- Given the first-step sieve extremum estimation, we provide an explicit characterization of the asymptotic variance of the semiparametric two-step GMM estimate.
- We provide consistent kernel LRV estimates; kernel based t test, Wald test, and overidentification test of the moment conditions.
- We provide robust orthonormal series LRV estimates; OS based t test, Wald test and overidentification test of the moment conditions.
- We show that our kernel LRV estimates of semiparametric asymptotic variance are *numerical equivalent* to the kernel LRV estimates of the corresponding two-step parametric models. Such results hold similarly for the series LRV estimates.

Answers to Question (5): CHLR (12))

- Multi-step procedures in which all steps could involve nonparametric sieve M estimation for weakly dependent data.
- nonparametric generated regressors or nonparametric filtered data are examples that fit into the framework.
- Model framework: Assume that $h_o \in \mathcal{H}$ is the unique solution to $\sup_{h \in \mathcal{H}} E[\varphi(Z_1, h)]$ and $g_o \in \mathcal{G}$ is the unique solution to $\sup_{g \in \mathcal{G}} E[\psi(Z_2, g, h_o)]$
- Two-step sieve M estimation: In the first step, we estimate $h_o \in \mathcal{H}$ by $\widehat{h}_n \in \mathcal{H}_n$ defined as

$$\frac{1}{n}\sum_{i=1}^{n} \varphi\left(Z_{1,i}, \widehat{h}_n\right) \geq \sup_{h \in \mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n} \varphi\left(Z_{1,i}, h\right) - O_p(\varepsilon_{1,n}^2); \tag{7}$$

in the second step, we estimate $g_o \in \mathcal{G}$ by $\widehat{g}_n \in \mathcal{G}_n$ defined as

$$\frac{1}{n}\sum_{i=1}^{n} \psi\left(Z_{2,i}, \widehat{g}_n, \widehat{h}_n\right) \geq \sup_{g \in \mathcal{G}_n} \frac{1}{n}\sum_{i=1}^{n} \psi\left(Z_{2,i}, g, \widehat{h}_n\right) - O_p(\varepsilon_{2,n}^2), \tag{8}$$

Example: semi-nonparametric GARCH + residual copula models

- Many explanations of the recent financial crisis have emphasized the role of financial frictions and collateral, “leverage cycle” in Geanakoplos (10) assumes that bad news is accompanied by increased uncertainty (volatility). “News impact curve”.
- Engle (10): “risk assessment” is also important in understanding the financial crisis.
- Our model: semi-nonparametric GARCH + residual copula, slightly modified SCOMDY model of Chen-Fan (06).
- We use daily data from the last 4 years to address both “news impact curve” and risk assessment” based on 3 series: mortgage-backed security (MBS), stock, and bond market returns.

SCOMDY model: Excess returns on Barclays MBS index (S_t^e), excess market (daily Fama-French factor) returns (M_t^e), and excess returns on the Barclays bond index (B_t^e):

$$\begin{aligned} MBS\ Market &: \quad S_t^e = c_S + \rho_S S_{t-1}^e + \beta_S M_{t-1}^e + \sigma_{S,t}\varepsilon_{S,t} \\ Stock\ Market &: \quad M_t^e = c_M + \rho_M M_{t-1}^e + \sigma_{M,t}\varepsilon_{M,t} \\ Bonds\ Market &: \quad B_t^e = c_B + \rho_M B_{t-1}^e + \beta_B M_{t-1}^e + \sigma_{B,t}\varepsilon_{B,t} \\ Volatility &: \quad \sigma_{i,t}^2 = \omega_i + \theta_i \sigma_{i,t-1}^2 + h_i\left(\sigma_{i,t-1}\varepsilon_{i,t-1}\right),\ i \in \{S, M, B\}, \end{aligned}$$

$E\left(\varepsilon_{i,t}\right) = 0$ and $E\left(\varepsilon_{i,t}^2\right) = 1$ for $i \in \{S, M, B\}$. $\left(\varepsilon_{S,t}, \varepsilon_{M,t}, \varepsilon_{B,t}\right)'$ are indep. across time but jointly distributed according to unknown marginals $F_i(\cdot)$, $i \in \{S, M, B\}$, and Student's t-copula, which has copula density $c(\mathbf{u}; \Sigma, v) =$

$$\frac{\Gamma\left(\frac{v+2}{2}\right)\left(\Gamma\left(\frac{v}{2}\right)\right)^2}{\sqrt{\det(\Sigma)}\left(\Gamma\left(\frac{v+1}{2}\right)\right)^3}\left(1 + \frac{\mathbf{x}\Sigma^{-1}\mathbf{x}'}{v}\right)^{-\frac{v+3}{2}} \prod_{i \in \{S,M,B\}}\left(1 + \frac{x_i^2}{v}\right)^{\frac{v+2}{2}},$$

with Σ the correlation matrix, T_v the scalar Student' t dist., $\mathbf{x} = \left(x_S, x_M, x_B\right),\ x_i = T_v^{-1}\left(u_i\right)$.

- All 3 estimated "news impact curves" exhibit the same asymmetry: bad news increases volatility more than does good news. For mortgage-backed securities and stocks, some goods news actually decreases volatility, as in Fostel and Geanakoplos (10). As in Linton and Mammen (05), most good news in the stock market does not have much effect on volatility.
- We find (i) shocks to bonds and shocks to mortgage-backed securities are highly correlated, (ii) shocks to mortgage-backed securities and shocks to stocks are moderately negatively correlated, and (iii) shocks to bonds and shocks to stocks are also moderately negatively correlated.
- With estimated semi-nonparametric GARCH and residual copula dependence parameters, we can easily calculate VaR for a portfolio comprised of mortgage-backed securities, stocks, and bonds.