Tutorial in Econometrics Part II: Sieve Inferences on Semi-nonparametric Models

Xiaohong Chen (Yale)

NUS, IMS, May 16, 2014

Outline of the Tutorial in Econometrics Part II

1. Introduction; Motivating empirical examples.
2. Sieve extremum (M, MD, GMM...) estimation; Sieve two-step.
3. Asymp. normality of sieve estimates.
4. Sieve Wald statistic; sieve variance estimation.
5. Sieve QLR statistics.
6. Sieve F statistic for weakly dependent data.
7. Concluding remarks.

1. Introduction

Introduction

An econometric (or statistical) model is a family of probability distributions indexed by unknown parameters. A model is called

- **parametric** if all of its parameters are in finite-dimensional parameter spaces;
- **nonparametric** if all of its parameters are in infinite-dimensional parameter spaces;
- **semiparametric** if its parameters of interest are in finite-dimensional spaces but its nuisance parameters are in infinite-dimensional spaces;
- **semi-nonparametric** if it contains both finite-dimensional and infinite-dimensional unknown parameters of interest.

Duration Model with Unobserved Heterogeneity

- $\{T_i, X_i\}_{i=1}^n$ a random sample from

$$p(T|X, \beta_0, h_0) = \int_{\mathcal{U}} g(T|X, u, \beta_0) f_U(u) du,$$

- $g(T|X, u, \beta_0)$: the density of duration T conditional on a scalar unobserved heterogeneity U and observed X. Ex. $g(T|X, u, \beta_0)$ can be Weibull density as in Heckman and Singer (84):

$$g(T|X, u, \beta_0) = \theta_{0,1} T^{\theta_{0,1}-1} \exp\left[\theta_{0,2}' X + u - T^{\theta_{0,1}} \exp\left(\theta_{0,2}' X + u\right)\right].$$

- U is indep. of X. Misspecifying density $f_U(u) \equiv h_0^2(u)$ leads to inconsistent estimation of θ_0.
- Let $\alpha_0 = (\beta_0, h_0) \in B \times \mathcal{H}$, which can estimated by sieve MLE:

$$\widehat{\alpha}_n = \arg \max_{\beta \in B,\ h \in \mathcal{H}_n} \sum_{i=1}^n \log\{\int_{\mathcal{U}} g(T_i|X_i, u, \beta) h^2(u) du\}$$

where $\mathcal{H}_n$ is a sieve space that becomes dense in $\mathcal{H}$ as $n \to \infty$. Semiparametric mixture models are widely used.

Shape-invariant system of Engel curves with endogenous expenditure

- Blundell et al. (03) show that a system of Engel curves satisfying Slutsky's symmetry and allowing for demographic effects on budget shares in a given year must take the form:

$$Y_{1\ell i} = h_{1\ell}(Y_{2i} - h_0(X_{1i})) + h_{2\ell}(X_{1i}) + \varepsilon_{\ell i}, \quad \ell = 1, ..., N,$$

  where $Y_{1\ell i}$ is the $i-th$ household budget share on $\ell - th$ goods, Y_{2i} is the $i-th$ household log-total non-durable expenditure, X_{1i} is a vector of the $i-th$ household demographic variables.
- Blundell-Chen-Kristensen (07) consider a semi-nonparametric mean instrumental variables (IV) regression:

$$E[Y_{1\ell i} - \{h_{1\ell}(Y_{2i} - g(X_{1i}'\beta_1)) + X_{1i}'\beta_{2\ell}\}|X_{1i}, X_{2i}] = 0,$$

- Chen-Pouzo (09, 12) estimate a semi-nonparametric quantile IV:

$$E[1\left(Y_{1\ell i} \leq h_{1\ell}(Y_{2i} - g(X_{1i}'\beta_1)) + X_{1i}'\beta_{2\ell}\right)|X_{1i}, X_{2i}] = \gamma \in (0, 1).$$

- Both are estimated via sieve minimum distance (MD).

Nonlinear habit-based asset pricing models

- Consumption based asset pricing models: $E(M_{t+1}R_{j,t+1} - 1|\mathcal{I}_t) = 0$, $j = 1, ..., N$, $M_{t+1} = \frac{\partial U/\partial C_{t+1}}{\partial U/\partial C_t}$ is IMRS (intertemporal marginal rate of substitution in consumption), and is a pricing kernel or SDF.

Nonlinear habit-based asset pricing models

- Consumption based asset pricing models: $E(M_{t+1}R_{j,t+1} - 1|\mathcal{I}_t) = 0$, $j = 1, ..., N$, $M_{t+1} = \frac{\partial U/\partial C_{t+1}}{\partial U/\partial C_t}$ is IMRS (intertemporal marginal rate of substitution in consumption), and is a pricing kernel or SDF.
- Hansen-Singleton (82): $U = \sum_{t=0}^{\infty} \delta^t \left[(C_t^{1-\gamma} - 1)/(1-\gamma) \right]$, $M_{t+1} = \delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}$. GMM with *unconditional* moment restrictions

$$E\left(\left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{j,t+1} - 1 \right] \mathbf{Z}_t \right) = 0, \quad j = 1, ..., N,$$

$\mathbf{Z}_t$=(1, lagged cons growth, lagged EWR and lagged VWR).

Nonlinear habit-based asset pricing models

- Consumption based asset pricing models: $E(M_{t+1}R_{j,t+1} - 1|\mathcal{I}_t) = 0$, $j = 1, ..., N$, $M_{t+1} = \frac{\partial U/\partial C_{t+1}}{\partial U/\partial C_t}$ is IMRS (intertemporal marginal rate of substitution in consumption), and is a pricing kernel or SDF.
- Hansen-Singleton (82): $U = \sum_{t=0}^{\infty} \delta^t \left[(C_t^{1-\gamma} - 1)/(1-\gamma)\right]$, $M_{t+1} = \delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$. GMM with *unconditional* moment restrictions

$$E\left(\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{j,t+1} - 1\right]\mathbf{Z}_t\right) = 0, \quad j = 1, ..., N,$$

  $\mathbf{Z}_t$=(1, lagged cons growth, lagged EWR and lagged VWR).
- Many finance and macro economists suspect misspecification of time separable utility in consumption.

- One popular theoretical fix is to let period t utility depend on habit level H_t, which is some function of current and lagged consumptions. Constantinides (90), Abel (90), Campbell-Cochrane (99), ...

- One popular theoretical fix is to let period t utility depend on habit level H_t, which is some function of current and lagged consumptions. Constantinides (90), Abel (90), Campbell-Cochrane (99), ...
- Is habit linear or nonlinear, is habit internal or external? Economic theories offered no guidance, but different welfare implications.

- One popular theoretical fix is to let period t utility depend on habit level H_t, which is some function of current and lagged consumptions. Constantinides (90), Abel (90), Campbell-Cochrane (99), ...
- Is habit linear or nonlinear, is habit internal or external? Economic theories offered no guidance, but different welfare implications.
- Chen-Ludvigson (04, 09):
$U = \sum_{t=0}^{\infty} \delta^t \left[\left((C_t - H_t)^{1-\gamma} - 1 \right) / (1-\gamma) \right]$, here $H_t = C_t g\left(c_t^*\right)$ is unknown habit level, $0 \leq g < 1$, g nondecreasing in first argument of $c_t^* = \left(\frac{C_{t-1}}{C_t}, \ldots, \frac{C_{t-L}}{C_t} \right)$. $M_{t+1} = \frac{\partial U / \partial C_{t+1}}{\partial U / \partial C_t}$. For external habit, $\partial U / \partial C_t = C_t^{-\gamma} \left(1 - g\left(c_t^*\right)\right)^{-\gamma}$; for internal habit, $\partial U / \partial C_t =$

$$C_t^{-\gamma} \left[\left(1 - g\left(c_t^*\right)\right)^{-\gamma} - E_t \{ \sum_{j=0}^{L} \delta^j \left(\frac{C_{t+j}}{C_t} \right)^{-\gamma} \left(1 - g(c_{t+j}^*)\right)^{-\gamma} \frac{\partial H_{t+j}}{\partial C_t} \} \right].$$

- Chen-Ludvigson (04, 09): **Sieve minimum distance** (SMD) with *conditional* moment restrictions:

$$E(M_{t+1}R_{j,t+1} - 1|\mathbf{w}_t) = 0, \ j = 1, ..., N, \quad \mathbf{w}_t \subset \mathcal{I}_t,$$

Let $\{\frac{C_t}{C_{t-1}}, R_{j,t}, \mathbf{w}_t\}$ be stationary ergodic. Do not specify parametric LOM. $\mathbf{w}_t = \left[\widehat{cay}_t, RREL_t, SPEX_t, \frac{C_t}{C_{t-1}}\right]'$ in empirical work.

- Chen-Ludvigson (04, 09): **Sieve minimum distance** (SMD) with *conditional* moment restrictions:

$$E(M_{t+1}R_{j,t+1}-1|\mathbf{w}_t)=0,\ j=1,...,N,\quad \mathbf{w}_t\subset\mathcal{I}_t,$$

  Let $\{\frac{C_t}{C_{t-1}},R_{j,t},\mathbf{w}_t\}$ be stationary ergodic. Do not specify parametric LOM. $\mathbf{w}_t=\left[\widehat{cay}_t,RREL_t,SPEX_t,\frac{C_t}{C_{t-1}}\right]'$ in empirical work.

- Using quarterly data, some empirical findings are: (1) estimated *habit is nonlinear*; (2) *internal habit fits data significantly better* than external habit; (3) estimated δ,γ are sensible; (4) estimated habit generated SDF performs well in explaining cross-sectional stock returns; (5) more findings about pricing errors, and model comparison in terms of HJ pricing errors.

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.

Chen-Fan (06b) SCOMDY models:

$$Y_{j,t+1} = E[Y_{j,t+1}|\mathcal{I}_t] + \sqrt{Var(Y_{j,t+1}|\mathcal{I}_t)}\epsilon_{j,t+1},\ j = 1, ..., N,$$

- $\{\epsilon_{t+1} \equiv (\epsilon_{1t+1}, \ldots, \epsilon_{Nt+1})' : t \geq 0\}$ indep. of $\mathcal{I}_t = \sigma(\{\mathbf{Y}^t, \mathbf{X}^t\})$, i.i.d., $E(\epsilon_{jt}) = 0$, $E(\epsilon_{jt}^2) = 1$, each ϵ_{jt} has unknown marginal cdf $F_j^o(\cdot)$,
- ϵ_t has a joint dist. $F^o(\epsilon) = C(F_1^o(\epsilon_1), \ldots, F_N^o(\epsilon_N); \alpha_o)$, where $C(\cdot) : [0,1]^N \rightarrow [0,1]$ is a copula with unknown parameter α_o.

Chen-Fan (06b) SCOMDY models:
$Y_{j,t+1} = E[Y_{j,t+1}|\mathcal{I}_t] + \sqrt{Var(Y_{j,t+1}|\mathcal{I}_t)}\epsilon_{j,t+1}$, $j = 1, ..., N$,

- $\{\epsilon_{t+1} \equiv (\epsilon_{1t+1}, \ldots, \epsilon_{Nt+1})' : t \geq 0\}$ indep. of $\mathcal{I}_t = \sigma(\{\mathbf{Y}^t, \mathbf{X}^t\})$, i.i.d., $E(\epsilon_{jt}) = 0$, $E(\epsilon_{jt}^2) = 1$, each ϵ_{jt} has unknown marginal cdf $F_j^o(\cdot)$,
- ϵ_t has a joint dist. $F^o(\epsilon) = C(F_1^o(\epsilon_1), \ldots, F_N^o(\epsilon_N); \alpha_o)$, where $C(\cdot) : [0,1]^N \to [0,1]$ is a copula with unknown parameter α_o.
- Different specifications of $E[Y_{j,t+1}|\mathcal{I}_t]$, $Var(Y_{j,t+1}|\mathcal{I}_t)$ and $C(\cdot; \alpha_o)$ lead to many different examples of SCOMDY models.

Chen-Fan (06b) SCOMDY models:
$Y_{j,t+1} = E[Y_{j,t+1}|\mathcal{I}_t] + \sqrt{Var(Y_{j,t+1}|\mathcal{I}_t)}\epsilon_{j,t+1}$, $j = 1, ..., N$,

- $\{\epsilon_{t+1} \equiv (\epsilon_{1t+1}, \ldots, \epsilon_{Nt+1})' : t \geq 0\}$ indep. of $\mathcal{I}_t = \sigma(\{\mathbf{Y}^t, \mathbf{X}^t\})$, i.i.d., $E(\epsilon_{jt}) = 0$, $E(\epsilon_{jt}^2) = 1$, each ϵ_{jt} has unknown marginal cdf $F_j^o(\cdot)$,
- ϵ_t has a joint dist. $F^o(\epsilon) = C(F_1^o(\epsilon_1), \ldots, F_N^o(\epsilon_N); \alpha_o)$, where $C(\cdot) : [0,1]^N \to [0,1]$ is a copula with unknown parameter α_o.
- Different specifications of $E[Y_{j,t+1}|\mathcal{I}_t]$, $Var(Y_{j,t+1}|\mathcal{I}_t)$ and $C(\cdot;\alpha_o)$ lead to many different examples of SCOMDY models.
- Easy to estimate; useful for conditional VaR, contagion, comovements.

Chen-Fan (06b) SCOMDY models:
$Y_{j,t+1} = E[Y_{j,t+1}|\mathcal{I}_t] + \sqrt{Var(Y_{j,t+1}|\mathcal{I}_t)}\epsilon_{j,t+1}$, $j = 1, ..., N$,

- $\{\epsilon_{t+1} \equiv (\epsilon_{1t+1}, \ldots, \epsilon_{Nt+1})' : t \geq 0\}$ indep. of $\mathcal{I}_t = \sigma(\{\mathbf{Y}^t, \mathbf{X}^t\})$, i.i.d., $E(\epsilon_{jt}) = 0$, $E(\epsilon_{jt}^2) = 1$, each ϵ_{jt} has unknown marginal cdf $F_j^o(\cdot)$,
- ϵ_t has a joint dist. $F^o(\epsilon) = C(F_1^o(\epsilon_1), \ldots, F_N^o(\epsilon_N); \alpha_o)$, where $C(\cdot) : [0,1]^N \to [0,1]$ is a copula with unknown parameter α_o.
- Different specifications of $E[Y_{j,t+1}|\mathcal{I}_t]$, $Var(Y_{j,t+1}|\mathcal{I}_t)$ and $C(\cdot; \alpha_o)$ lead to many different examples of SCOMDY models.
- Easy to estimate; useful for conditional VaR, contagion, comovements.
- Cherubini et al (10) apply SCOMDY to build term structure of multivariate equity derivatives models.

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):

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

$E\left(\varepsilon_{i,t}\right) = 0$ 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., $\mathbf{x} = (x_S, x_M, x_B)$, $x_i = T_v^{-1}(u_i)$.

- 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.

2. Sieve Extremum Estimation

Sieve extremum estimation

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^n$ is stationary, ergodic.

Sieve extremum estimation

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^n$ is stationary, ergodic.
- Parameter space, (Θ, d), is a possibly infinite dimensional (often non-compact) metric space.

Sieve extremum estimation

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^n$ is stationary, ergodic.
- Parameter space, (Θ, d), is a possibly infinite dimensional (often non-compact) metric space.
- A *sieve* $\{\Theta_n\}$ is a sequence of approximating parameter spaces that become dense in Θ as $n \to \infty$, i.e., for any $\theta \in \Theta$ there is an element $\pi_n\theta$ in Θ_n satisfying $d(\theta, \pi_n\theta) \to 0$ as $n \to \infty$.

Sieve extremum estimation

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^{n}$ is stationary, ergodic.
- Parameter space, (Θ, d), is a possibly infinite dimensional (often non-compact) metric space.
- A *sieve* $\{\Theta_n\}$ is a sequence of approximating parameter spaces that become dense in Θ as $n \to \infty$, i.e., for any $\theta \in \Theta$ there is an element $\pi_n\theta$ in Θ_n satisfying $d(\theta, \pi_n\theta) \to 0$ as $n \to \infty$.
- A population criterion $Q : \Theta \to \mathcal{R}$ is maximized at a "pseudo-true" parameter $\theta_o \in \Theta$.

Sieve extremum estimation

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^n$ is stationary, ergodic.
- Parameter space, (Θ, d), is a possibly infinite dimensional (often non-compact) metric space.
- A *sieve* $\{\Theta_n\}$ is a sequence of approximating parameter spaces that become dense in Θ as $n \to \infty$, i.e., for any $\theta \in \Theta$ there is an element $\pi_n\theta$ in Θ_n satisfying $d(\theta, \pi_n\theta) \to 0$ as $n \to \infty$.
- A population criterion $Q : \Theta \to \mathcal{R}$ is maximized at a "pseudo-true" parameter $\theta_o \in \Theta$.
- $\widehat{Q}_n$ is a sample criterion, which is a joint measurable function of θ and data $\{Z_t\}_{t=1}^n$, and converges to Q in some sense as $n \to \infty$.

Sieve extremum estimation

- Data: $\{Z_t = (Y_t', X_t')'\}_{t=1}^n$ is stationary, ergodic.
- Parameter space, (Θ, d), is a possibly infinite dimensional (often non-compact) metric space.
- A *sieve* $\{\Theta_n\}$ is a sequence of approximating parameter spaces that become dense in Θ as $n \to \infty$, i.e., for any $\theta \in \Theta$ there is an element $\pi_n\theta$ in Θ_n satisfying $d(\theta, \pi_n\theta) \to 0$ as $n \to \infty$.
- A population criterion $Q : \Theta \to \mathcal{R}$ is maximized at a "pseudo-true" parameter $\theta_o \in \Theta$.
- $\widehat{Q}_n$ is a sample criterion, which is a joint measurable function of θ and data $\{Z_t\}_{t=1}^n$, and converges to Q in some sense as $n \to \infty$.
- **Sieve extremum estimator**: any estimator $\hat{\theta}_n$ that solves

$$\widehat{Q}_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta_n} \widehat{Q}_n(\theta) - O_P(\eta_n), \quad \text{with } \eta_n \to 0 \text{ as } n \to \infty.$$

- Examples of sieves Θ_n: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).

- Examples of sieves Θ_n: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).
- Examples of criterion function $\widehat{Q}_n()$: ML, MD, GMM, GEL, ... virtually all the existing criterion function for estimating nonlinear parametric models are valid choices; see Chen (07).

- Examples of sieves Θ_n: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).
- Examples of criterion function $\widehat{Q}_n()$: ML, MD, GMM, GEL, ... virtually all the existing criterion function for estimating nonlinear parametric models are valid choices; see Chen (07).
- **Sieve M-estimation**: a special case of sieve extremum estimation when $\widehat{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l(\theta, Z_t)$. E.g., sieve maximum likelihood (ML), sieve least squares (LS), sieve nonlinear least squares (NLS), sieve quantile regression (QR).

- Examples of sieves Θ_n: Polynomial series, Hermite polynomial series (SNP), Fourier series, Splines, Wavelets, Neural Networks, etc. See e.g. Judd (98), Chen (07).
- Examples of criterion function $\widehat{Q}_n()$: ML, MD, GMM, GEL, ... virtually all the existing criterion function for estimating nonlinear parametric models are valid choices; see Chen (07).
- **Sieve M-estimation**: a special case of sieve extremum estimation when $\widehat{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l(\theta, Z_t)$. E.g., sieve maximum likelihood (ML), sieve least squares (LS), sieve nonlinear least squares (NLS), sieve quantile regression (QR).
- Ex: $Y_t = \theta_o(X_t) + \varepsilon_t$, $E[\varepsilon_t|X_t] = 0$. Let $\{p_j(X), j = 1, 2, ...\}$ be a sequence of known basis functions that can approximate any $\theta \in \Theta$ well. $p^{k_n}(X) = (p_1(X), ..., p_{k_n}(X))'$. Then $\Theta_n = \{h : h(x) = p^{k_n}(x)'A : A \in \mathcal{R}^{k_n}\}$, with $k_n \to \infty$ slowly as $n \to \infty$, is a finite-dimensional linear sieve for Θ. And $\widehat{\theta}$ is a sieve (or series) LS estimator of θ_o:

$$\widehat{\theta} = \arg\max_{\theta\in\Theta_n} \frac{-1}{n}\sum_{t=1}^{n}[Y_t - \theta(X_t)]^2 = p^{k_n}(\cdot)'(P'P)^{-}\sum_{t=1}^{n} p^{k_n}(X_t)Y_t.$$

- **Sieve MD estimation**: a special case of sieve extremum estimation when $-\widehat{Q}_n(\theta)$ can be expressed as some distance from zero.
- For conditional moment restriction $E[\rho(Z,\theta_o)|X]=0$, one typical quadratic distance is

$$-\widehat{Q}_n(\theta)=\frac{1}{n}\sum_{t=1}^{n}\widehat{m}(X_t,\theta)'\{\widehat{\Sigma}(X_t)\}^{-1}\widehat{m}(X_t,\theta),$$

where $\widehat{m}(X_t,\theta)$ is a nonparametrically estimated moment condition of fixed, finite dimension and $\widehat{\Sigma}(X_t)\to\Sigma(X_t)$ in prob., where $\Sigma(X_t)$ is a psd weighting matrix of the same fixed, finite dimension as that of $\widehat{m}(X_t,\theta)$. For example, $\widehat{m}(X_t,\theta)$ could be any series estimate of the conditional mean function $m(X_t,\theta)=E[\rho(Z,\theta)|X=X_t]$; see Newey-Powell (03) and Ai-Chen (03).

- For conditional moment restriction $E[\rho(Z, \theta_o)|X] = 0$, another typical quadratic distance is

$$-\widehat{Q}_n(\theta) = \widehat{g}_n(\theta)'\widehat{W}\widehat{g}_n(\theta),$$

with $\widehat{g}_n(\theta_o) \to 0$ in prob. Here $\widehat{g}_n(\theta)$ is a sample average of some unconditional moment conditions of increasing dimension and $\widehat{W} \to W$ in prob., where W is a psd weighting matrix of the same increasing dimension as that of $\widehat{g}_n(\theta)$. This is "**sieve GMM**".

- $E[\rho(Z, \theta_o)|X] = 0$ iff the increasing number of unconditional moment restrictions hold:

$$E[\rho(Z_t, \theta_o)p_{0j}(X_t)] = 0, j = 1, 2, ..., k_{m,n},$$

where $\{p_{0j}(X), j = 1, 2, ..., k_{m,n}\}$ is a sequence of known basis functions that can approximate any real-valued square integrable functions of X well as $k_{m,n} \to \infty$. Let $p^{k_{m,n}}(X) = (p_{01}(X), ..., p_{0k_{m,n}}(X))'$. Then $E[\rho(Z, \theta_o)|X] = 0$ can be estimated via the above sieve GMM using $\widehat{g}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \rho(Z_t, \theta) \otimes p^{k_{m,n}}(X_t)$.

Why method of sieves?

1. Easy to compute. Once when the unknown functions are approximated by finite dimensional sieves, the implementation is the same as any parametric nonlinear extremum estimation.
2. Easier to impose shape (monotonicity, concavity), additivity, non-negativity and other restrictions on unknown functions.
3. Can *simultaneously* obtain optimal convergence rates for unknown functions and root-n normality for *regular* functionals (such as finite dimensional parameter); see Chen-Shen (98, sieve M estimation for time series); Chen-Pouzo (09, sieve MD for iid)

Estimation methods for unknown functionals

- Joint estimation procedure: Simultaneous estimation of all the unknown parameters of interests.

Profile sieve extremum estimator: For a semi-nonparametric model, $\Theta = B \times \mathcal{H}$, with B a finite-dimensional compact space, $\mathcal{H}$ an infinite-dimensional function space. Then $\Theta_n = B \times \mathcal{H}_n$. The *profile sieve extremum estimator* consists of two steps:

- Step 1, for fixed β, compute $\widehat{Q}_n(\beta, \widetilde{h}(\beta)) \geq \sup_{h \in \mathcal{H}_n} \widehat{Q}_n(\beta, h) - o_P(1)$;
- Step 2, estimate β_o by $\widehat{\beta}_n = \arg\max_{\beta \in B} \widehat{Q}_n(\beta, \widetilde{h}(\beta))$, and estimate h_o by $\widehat{h}_n = \widetilde{h}(\widehat{\beta}_n)$.

Estimation methods for unknown functionals

- **Semiparametric two-step** procedure:
- Step 1: for fixed β, estimate unknown $h()$ using whatever nonparametric methods, say, using a sieve estimator $\widetilde{h}(\beta) = \arg\max_{h \in \mathcal{H}_n} \widehat{Q}_{1,n}(\beta, h)$
- Step 2, estimate unknown β_o using one of existing nonlinear extremum procedure with plugged in estimated $h()$, say, $\widehat{\beta}_n = \arg\max_{\beta \in B} \widehat{Q}_{2,n}(\beta, \widetilde{h}(\beta))$.

- **Advantages** of 2-step: easier to compute; easier to establish root-n asymptotic normality of regular functionals (β).
- **Disadvantages** of 2-step: generally inefficient; difficult to obtain a consistent estimator of Avar$(\widehat{\beta}_n)$.
- **Advantages** of 2-step with **sieve as 1st step**: easy to compute a consistent estimator of Avar$(\widehat{\beta}_n)$ even when there is no closed form expression of Avar$(\widehat{\beta}_n)$. Ai-Chen (07); Ackerberg-Chen-Hahn (11), Chen-Hahn-Liao (12, time series)

Example: Semi-nonparametric copula-based Markov model

- True cond. density, $p^0(\cdot|Y^{t-1})$ of Y_t given $Y^{t-1} \equiv (Y_{t-1}, ..., Y_1)$ is:

$$p^0(\cdot|Y^{t-1}) = g_0(\cdot) c(F_0(Y_{t-1}), F_0(\cdot); \alpha_0),$$

  the $q-$th, $q \in (0,1)$, conditional quantile of Y_t given Y^{t-1} is:

$$Q_q^Y(y) = F_0^{-1}\left(C_{2|1}^{-1}\left[q|F_0(y); \alpha_0\right]\right)$$

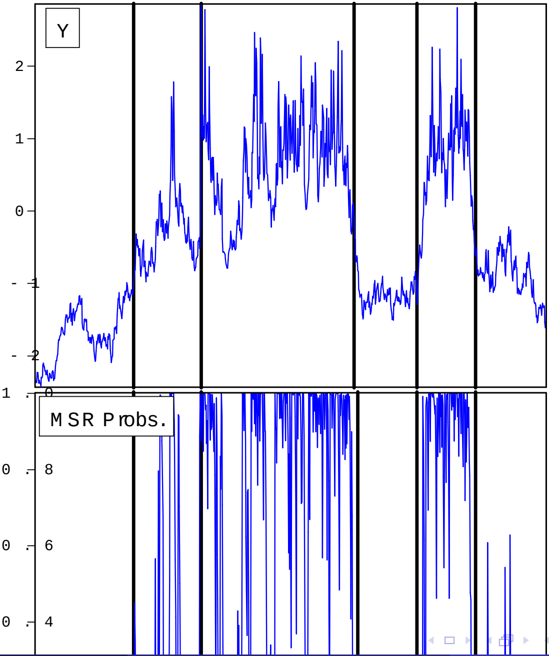
  where $C_{2|1}[\cdot|u; \alpha_0] \equiv \frac{\partial}{\partial u} C(u, \cdot; \alpha_0) \equiv C_1(u, \cdot; \alpha_0)$ is the cond. dist of $U_t \equiv F_0(Y_t)$ given $U_{t-1} = u$; and $C_{2|1}^{-1}\left[q|u; \alpha_0\right]$ is the $q-$th conditional quantile of U_t given $U_{t-1} = u$.
- Chen-Fan (06a) two-step estimation: step1: $\widehat{F}(y) = \frac{1}{n+1} \sum_{t=1}^n 1\{Y_t \leq y\}$: rescaled empirical cdf; step 2: pseudo-MLE $\widehat{\alpha}$:

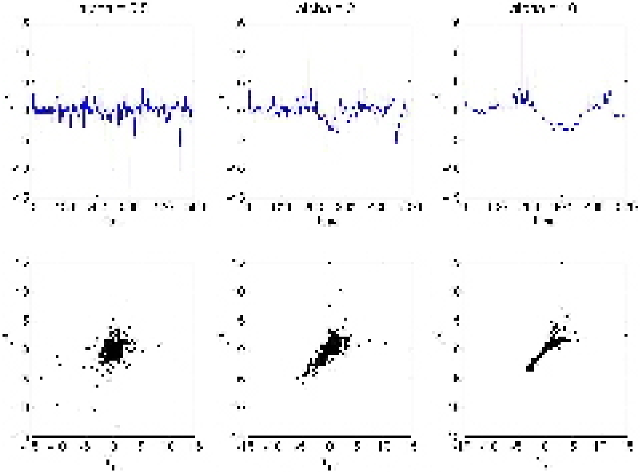
$$\max_{\alpha} \frac{1}{n} \sum_{t=2}^{n} \log c(\widehat{F}(Y_{t-1}), \widehat{F}(Y_t); \alpha)$$

- Ex:, any strictly stationary 1st order Markov series $\{Y_t\}_{t=1}^n$ can be equivalently expressed as: $f(Y_t|Y_{t-1}) = c(G(Y_{t-1}), G(Y_t))f(Y_t)$, i.e., can be generated using a copula $C(u_1, u_2; \alpha)$ with a marginal cdf F: (i) generate n independent $U(0,1)$ r.v. $\{X_t\}_{t=1}^n$; (ii) $U_1 = X_1$, $U_t = C_{2|1}^{-1}(X_t|U_{t-1}; \alpha)$, and $Y_t \triangleq G^{-1}(U_t)$.

- Ex:, any strictly stationary 1st order Markov series $\{Y_t\}_{t=1}^n$ can be equivalently expressed as: $f(Y_t|Y_{t-1}) = c(G(Y_{t-1}), G(Y_t))f(Y_t)$, i.e., can be generated using a copula $C(u_1, u_2; \alpha)$ with a marginal cdf F: (i) generate n independent $U(0,1)$ r.v. $\{X_t\}_{t=1}^n$; (ii) $U_1 = X_1$, $U_t = C_{2|1}^{-1}(X_t|U_{t-1}; \alpha)$, and $Y_t \triangleq G^{-1}(U_t)$.
- In the next graph, $\{Y_t\}_{t=1}^n$ is generated using "Clayton(15) + t(3)": $C_{2|1}^{-1}(X_t|U_{t-1}; \alpha) = [(X_t^{-\alpha/(1+\alpha)} - 1)U_{t-1}^{-\alpha} + 1]^{-1/\alpha}$, with $\alpha = 15$, G = cdf of t(3). However, structural break test of Davis et al. (05) detects several breaks; Markov switching model also fits well.

- Ex:, any strictly stationary 1st order Markov series $\{Y_t\}_{t=1}^n$ can be equivalently expressed as: $f(Y_t|Y_{t-1}) = c(G(Y_{t-1}), G(Y_t))f(Y_t)$, i.e., can be generated using a copula $C(u_1, u_2; \alpha)$ with a marginal cdf F: (i) generate n independent $U(0,1)$ r.v. $\{X_t\}_{t=1}^n$; (ii) $U_1 = X_1$, $U_t = C_{2|1}^{-1}(X_t|U_{t-1}; \alpha)$, and $Y_t \triangleq G^{-1}(U_t)$.
- In the next graph, $\{Y_t\}_{t=1}^n$ is generated using "Clayton(15) + t(3)": $C_{2|1}^{-1}(X_t|U_{t-1}; \alpha) = [(X_t^{-\alpha/(1+\alpha)} - 1)U_{t-1}^{-\alpha} + 1]^{-1/\alpha}$, with $\alpha = 15$, G = cdf of t(3). However, structural break test of Davis et al. (05) detects several breaks; Markov switching model also fits well.
- If one cares about conditional VaR or tail dependence, then copula-based Markov model is better; see Chen-Fan (06a), Chen-Koenker-Xiao (09), Bouye-Salmon (09).


Y
2
1
0
-1
-2
1.0
MSR Probs.
0.8
0.6
0.4



- Chen-Wu-Yi (09) sieve MLE: Let $Z_t = (Y_{t-1}, Y_t)$, and $\ell(\alpha, g, Z_t) \equiv \log p(Y_t | Y^{t-1}) =$

$$\log f(Y_t) + \log c\left(F(Y_{t-1}), F(Y_t); \alpha\right)$$
$$= \log f(Y_t) + \log c\left(\int 1(y \leq Y_{t-1}) f(y) dy, \int 1(y \leq Y_t) f(y) dy; \alpha\right)$$

Then the joint log-likelihood function of the data $\{Y_t\}_{t=1}^n$ is

$$L_n(\alpha, f) \equiv \frac{1}{n} \sum_{t=2}^{n} \ell(\alpha, f, Z_t) + \frac{1}{n} \log f(Y_1).$$

The sieve MLE $\widehat{\theta}_n \equiv (\widehat{\alpha}_n, \widehat{g}_n)$ is defined as

$$L_n(\widehat{\alpha}_n, \widehat{f}_n) \geq \max_{\alpha \in \mathcal{A}, f \in \mathcal{F}_n} L_n(\alpha, f) - O_p(1),$$

$$\mathcal{F}_n = \left\{ f_{K_n} \in \mathcal{F} : f_{K_n}(y) = [\sum_{k=1}^{K_n} a_k A_k(y)]^2, \ \int f_{K_n}(y) dy = 1 \right\},$$

or

$$\mathcal{F}_n = \left\{ f_{K_n} \in \mathcal{F} : f_{K_n}(y) = \exp\{\sum_{k=1}^{K_n} a_k A_k(y)\}, \ \int f_{K_n}(y) dy = 1 \right\},$$

Table: Clayton, true $F = t_3$: estimation of α

		Sieve	Ideal	2step	Para
$\alpha = 2$	Mean	1.969	2.002	1.912	1.989
τ	Bias	-0.031	0.002	-0.088	-0.011
(0.500)	Var	0.019	0.007	0.101	0.012
λ	MSE	0.020	0.007	0.109	0.012
(0.707)	$\alpha^{MC}_{(2.5,97.5)}$	(1.70, 2.25)	(1.83, 2.17)	(1.36, 2.60)	(1.76,2.19)
$\alpha = 10$	Mean	9.687	10.00	7.115	9.967
τ	Bias	-0.313	0.004	-2.886	-0.033
(0.833)	Var	0.351	0.085	4.852	0.129
λ	MSE	0.449	0.085	13.18	0.130
(0.933)	$\alpha^{MC}_{(2.5,97.5)}$	(8.68, 10.87)	(9.43, 10.6)	(3.87, 12.5)	(9.26,10.6)
$\alpha = 12$	Mean	11.62	12.01	7.896	11.98
τ	Bias	-0.382	0.012	-4.104	-0.016
(0.857)	Var	0.541	0.119	5.656	0.222
λ	MSE	0.687	0.120	22.50	0.222
(0.944)	$\alpha^{MC}_{(2.5,97.5)}$	(10.5, 13.3)	(11.3,12.7)	(4.35, 13.6)	(11.0, 12.9)

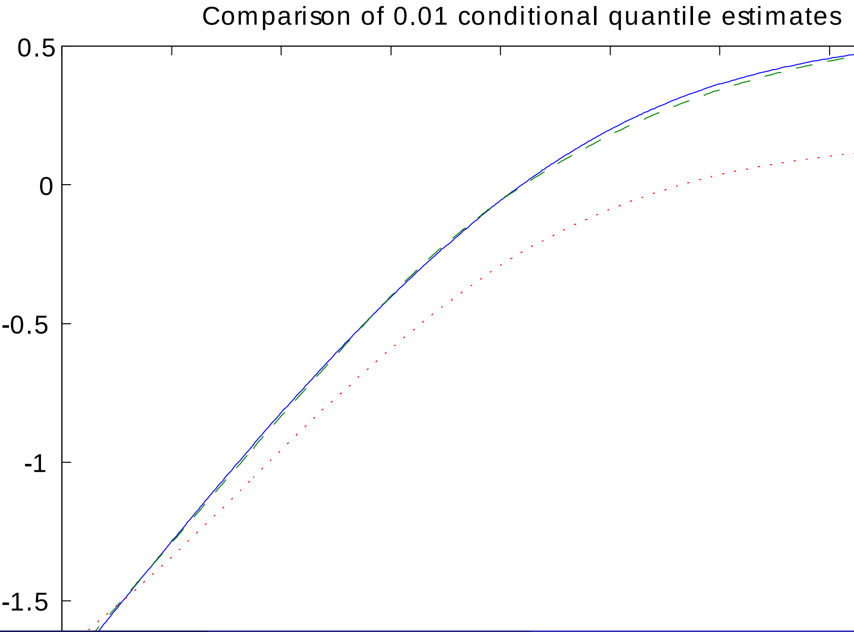
Table: Clayton, true $F = t_3$: estimation of F. Reported $Bias^2$, Var and MSE are the true ones multiplied by 1000.

		Sieve		2step		Para		Mis-N	
		$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	$Q_{2/3}$	$Q_{1/3}$	Q
$\alpha = 2$	Mean	0.325	0.673	0.333	0.666	0.333	0.667	0.347	0.5
	$Bias^2_{10^3}$	0.026	0.007	0.011	0.013	0.009	0.009	0.282	12
$\tau(0.500)$	Var_{10^3}	0.054	0.049	1.430	0.801	0.002	0.002	1.921	5.6
$\lambda(0.707)$	MSE_{10^3}	0.080	0.056	1.441	0.814	0.011	0.011	2.203	18
$\alpha = 10$	Mean	0.319	0.664	0.331	0.666	0.333	0.667	0.364	0.5
	$Bias^2_{10^3}$	0.128	0.042	0.001	0.013	0.009	0.009	1.132	7.4
$\tau(0.833)$	Var_{10^3}	0.109	0.137	22.28	9.800	0.003	0.003	0.711	3.4
$\lambda(0.933)$	MSE_{10^3}	0.236	0.178	22.29	9.813	0.012	0.012	1.843	10
$\alpha = 12$	Mean	0.318	0.661	0.331	0.665	0.333	0.667	0.374	0.5
	$Bias^2_{10^3}$	0.154	0.079	0.001	0.023	0.010	0.010	1.903	5.2
$\tau(0.857)$	Var_{10^3}	0.127	0.141	28.83	12.08	0.003	0.003	0.950	2.6
$\lambda(0.944)$	MSE_{10^3}	0.281	0.220	28.83	12.10	0.013	0.013	2.853	7.9

Table: Clayton, true $F = t_3$: estimation of 0.01 conditional quantile

		Sieve	Ideal	2step	Para	Mis-N	Mis-EV
$\alpha = 5$	Int$Bias^2_{10^3}$	36.26	0.000	150.0	0.172	900.7	704.8
$\tau(0.714)$	IntVar_{10^3}	32.15	5.450	985.3	10.18	463.7	313.4
$\lambda(0.871)$	IntMSE_{10^3}	68.41	5.450	1135	10.35	1364	1018
$\alpha = 10$	Int$Bias^2_{10^3}$	7.712	0.000	527.3	0.040	815.3	427.4
$\tau(0.833)$	IntVar_{10^3}	19.36	2.475	855.3	3.716	361.7	202.7
$\lambda(0.933)$	IntMSE_{10^3}	27.07	2.475	1383	3.756	1177	630.1
$\alpha = 12$	Int$Bias^2_{10^3}$	2.851	0.000	367.7	0.004	181.1	175.9
$\tau(0.857)$	IntVar_{10^3}	6.236	1.068	590.9	1.578	59.44	46.12
$\lambda(0.944)$	IntMSE_{10^3}	9.086	1.069	958.7	1.582	240.5	222.0

For each α, evaluation is based on the common support of 1000 MC simulated data. Reported integrated $Bias^2$, integrated Var and the integrated MSE are the true ones multiplied by 1000.


Comparison of 0.01 conditional quantile estimates
0.5
0
-0.5
-1
-1.5

3. Limiting distributions of sieve estimates.

Limiting dist. of plug-in sieve estimates of regular functionals

- $\sqrt{n}$ normality of semiparametric 2-step GMM estimators: Newey (94), Chen-Linton-Keilegom (03), Chen (07, beta-mixing, non-smooth criterion).
- $\sqrt{n}$ normality of sieve simultaneous M-estimator; efficiency of sieve MLE: Chen-Shen (98, beta-mixing)
- $\sqrt{n}$ normality and inference of sieve MD estimator of semi-nonparametric conditional moment restrictions: Ai-Chen (03, iid), Ai-Chen (07, could be misspecified, iid), Chen-Pouzo (09, iid).

Limiting dist. of plug-in sieve estimates of irregular functionals

- Irregular functionals are also called nonsmooth functionals or unbounded functionals, which have singular semiparamertic information bound, and hence can not be estimated at a root-n rate.
- Asym normality of sieve M-estimators of possibly irregular functionals: Chen-Liao-Sun (2013, time series).
- Asym normality of sieve MD estimators of possibly irregular functionals: Chen-Pouzo (2010, iid); allowing for nonparametric endogeneity.

4. Sieve Wald statistics; consistent sieve variance estimation.

- Consistent variance estimation of semiparametric 2-step GMM estimators with sieve as 1st step: Ai-Chen (07), Ackerberg-Chen-Hahn (12), Chen, Hahn and Liao (12, time series).
- Robust sieve long-run variance estimation of sieve M estimator of possibly irregular functionals: Chen-Liao-Sun (13, time series).
- Consistent sieve variance estimation of sieve MD estimators of possibly irregular functionals: Chen-Pouzo (10, iid).

Sieve MD estimates of irregular functionals: Chen-Pouzo

- Nonpara conditional moment model: $E[\rho(Z; h_0(Y))|X] = 0$.
- Functionals of interest: $\phi(h)$, e.g., $\phi(h) = h(\overline{y})$ (for $\overline{y} \in \text{supp}(Y)$), $\int w(y)\nabla h(y)dy$ or $\int w(y)\left|\nabla h(y)\right|^2 dy$.
- Asymp normality: $\frac{\sqrt{n}\left\{\phi(\widehat{h}_n)-\phi(h_0)\right\}}{||v_n^*||_{sd}} \Rightarrow N(0,1)$,
- $||v_n^*||_{sd}^2 = \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_n^{-1} \mho_n D_n^{-1} \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$,
- $D_n = E\left[\left(\frac{dm(X,h_0)}{dh}[q^{k(n)}(\cdot)']\right)' \Sigma(X)^{-1}\left(\frac{dm(X,h_0)}{dh}[q^{k(n)}(\cdot)']\right)\right]$,
- $\mho_n = E\left[\left(\frac{dm(X,h_0)}{dh}[q^{k(n)}(\cdot)']\right)' W \left(\frac{dm(X,h_0)}{dh}[q^{k(n)}(\cdot)']\right)\right]$
- $W = \Sigma(X)^{-1}\rho(Z,h_0)\rho(Z,h_0)'\Sigma(X)^{-1}$.

- NPIV example: $Y_1 = h_0(Y_2) + U$, $E(U|X) = 0$.
- NPQIV example: $Y_1 = h_0(Y_2) + U$, $\Pr(U \le 0|X) = \gamma$.
- Asymp normality: $\frac{\sqrt{n}\left\{\phi(\widehat{h}_n) - \phi(h_0)\right\}}{||v_n^*||_{sd}} \Rightarrow N(0,1)$,
- $||v_n^*||_{sd}^2 = \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_n^{-1} \mho_n D_n^{-1} \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$,
- NPIV: $D_n = E\left(E[q^{k(n)}(Y_2)|X] E[q^{k(n)}(Y_2)|X]'\right)$, $\mho_n = E\left(E[q^{k(n)}(Y_2)|X] U^2 E[q^{k(n)}(Y_2)|X]'\right)$.
- NPQIV: $D_n = \frac{1}{\gamma(1-\gamma)} E\left(E[f_{U|Y_2,X}(0) q^{k(n)}(Y_2)|X] E[f_{U|Y_2,X}(0) q^{k(n)}(Y_2)|X]'\right)$, $\mho_n = D_n$.
- Operator $Th = E\left[h(Y_2)|X\right]$ (for NPIV) and $Th = E\left[f_{U|Y_2,X}(0) h(Y_2)|X\right]$ (for NPQIV) mapping from $h \in \mathcal{H} \subset L^2(f_{Y_2})$ to $L^2(f_X)$ are compact, with $\left\{\psi_j(\cdot) : j \ge 1\right\}$ the eigenfunctions, and $\mu_1 \ge ... \ge \mu_j \ge \mu_{j+1} \searrow 0$ the singular values.
- $||v_n^*||_{sd}^2 \asymp \sum_{j=1}^{k(n)} \mu_j^{-2} \left(\frac{d\phi(h_0)}{dh}[\psi_j(\cdot)]\right)^2$, which could go to infinity.

- Sieve t statistic: $\frac{\sqrt{n}\left\{\phi(\widehat{h}_n)-\phi(h_0)\right\}}{||\widehat{v}_n^*||_{n,sd}} \Rightarrow N(0,1)$,

- $||\widehat{v}_n^*||_{n,sd}^2 = \frac{d\phi(\widehat{h})}{dh}[q^{k(n)}(\cdot)]'\widehat{D}_n^{-1}\widehat{\mho}_n\widehat{D}_n^{-1}\frac{d\phi(\widehat{h})}{dh}[q^{k(n)}(\cdot)]$,
- $\widehat{D}_n = \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{h})}{dh}[q^{k(n)}(\cdot)']\right)'\widehat{\Sigma}(X_i)^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{h})}{dh}[q^{k(n)}(\cdot)']\right)$,
- $\widehat{\mho}_n = \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{h})}{dh}[q^{k(n)}(\cdot)']\right)'\widehat{W}_i\left(\frac{d\widehat{m}(X_i,\widehat{h})}{dh}[q^{k(n)}(\cdot)']\right)$
- $\widehat{W}_i = \widehat{\Sigma}(X_i)^{-1}\rho(Z_i,\widehat{h})\rho(Z_i,\widehat{h})'\widehat{\Sigma}(X_i)^{-1}$.

- Applying it to NPIV example, $||\widehat{v}_n^*||_{n,sd}^2$ becomes the robust variance estimator of parametric 2SLS.

5. Sieve QLR statistics.

6. Sieve F statistic for weakly dependent data.

7. Conclusion and future research

Concluding remarks

- Semi-nonparametric methods are useful for complicated nonlinear non-Gaussian, nonseparable structural models.

Concluding remarks

- Semi-nonparametric methods are useful for complicated nonlinear non-Gaussian, nonseparable structural models.
- Multi-step estimation is typical for complicated dynamic models.

Concluding remarks

- Semi-nonparametric methods are useful for complicated nonlinear non-Gaussian, nonseparable structural models.
- Multi-step estimation is typical for complicated dynamic models.
- Large sample properties (consistency, rate, limiting distribution) of sieve M-estimation for weakly dependent time series models are relatively complete; sieve Wald, QLR tests are also available and easy to implement

Concluding remarks

- Semi-nonparametric methods are useful for complicated nonlinear non-Gaussian, nonseparable structural models.
- Multi-step estimation is typical for complicated dynamic models.
- Large sample properties (consistency, rate, limiting distribution) of sieve M-estimation for weakly dependent time series models are relatively complete; sieve Wald, QLR tests are also available and easy to implement
- Large sample properties (consistency, rate, limiting distribution) of sieve MD-estimation (or sieve GMM) for cross-section and small-T panel data structural models are relatively complete.

Concluding remarks

- Semi-nonparametric methods are useful for complicated nonlinear non-Gaussian, nonseparable structural models.
- Multi-step estimation is typical for complicated dynamic models.
- Large sample properties (consistency, rate, limiting distribution) of sieve M-estimation for weakly dependent time series models are relatively complete; sieve Wald, QLR tests are also available and easy to implement
- Large sample properties (consistency, rate, limiting distribution) of sieve MD-estimation (or sieve GMM) for cross-section and small-T panel data structural models are relatively complete.
- Sieve Wald, score and QLR tests, and their bootstrap versions based on sieve MD for possible irregular functionals are developed. (Chen-Pouzo, 2013). This allows for inference on general class of semi-nonparametric conditional moment models involving nonparametric endogeneity.

Concluding remarks, future research

- There is few result on higher order refinements.

Concluding remarks, future research

- There is few result on higher order refinements.
- There is few result on simulation based methods for semi-nonparametric time series models with nonlinear non-Gaussian latent structures.

Concluding remarks, future research

- There is few result on higher order refinements.
- There is few result on simulation based methods for semi-nonparametric time series models with nonlinear non-Gaussian latent structures.
- Choice of smoothing parameters and lag length.