Generalized Method of Tail Trimmed Moments

Jonathan B. Hill
Dept. of Economics
University of North Carolina – Chapel Hill

Eric Renault
Dept. of Economics
University of North Carolina – Chapel Hill
CIRANO, CIREQ

OBJECTIVES

- Asymptotically normal (robust) GMM for stationary heavy tailed data.

- Characterize rate of convergence : super-$n^{1/2}$ -convergence, heterogeneous

  Extensions

  - EL and Efficient Tests (Hill and Renault '10, Hill and Prokhorov '10)
  - Robust Moment Condition Tests (Hill '10, Hill and Aguilar '10)
  - weak limit theory : dependent trimmed arrays (Hill '10)

  Result - asymptotic normality and super-$n^{1/2}$ -convergence

  - convergence rate achieves maximum of all M-estimators for stationary data.
  - robust moment condition tests with arbitrary plug-in

OBJECTIVES

- Estimating equations:

$$m_t(\theta):\Theta \to \Re^q \quad \text{compact}\,\Theta \subset \Re^k,\ q \ge k \ge 1$$

$$\text{Identification : } E\left[m_t(\theta)\right]=0 \text{ if and only if } \theta=\theta^0$$

Asymptotic theory requires $E\left[m_{i,t}^2(\theta^0)\right]<\infty$.

- **EXAMPLE : Strong-ARCH(1) with least squares equations**

$$y_t=\left(\alpha^0+\beta^0 y_{t-1}^2\right)^{1/2}\varepsilon_t \text{ and } m_t(\theta)=\begin{bmatrix} y_t^2-\left(\alpha+\beta y_{t-1}^2\right) \\ \left\{y_t^2-\left(\alpha+\beta y_{t-1}^2\right)\right\}\times y_{t-1}^2 \end{bmatrix}$$

$$E\left[m_{2,t}^2(\theta^0)\right]<\infty \text{ if only if } E\left[\varepsilon_t^4\right]<\infty \text{ and } E\left[y_t^8\right]<\infty$$

OBJECTIVES

- Estimating equations:

$$m_t(\theta):\Theta \to \Re^q \quad \text{compact}\,\Theta \subset \Re^k,\ q \geq k \geq 1$$

$$\text{Identification}: E[m_t(\theta)] = 0 \text{ if and only if } \theta = \theta^0$$

Asymptotic theory requires $E\left[m_{i,t}^2(\theta^0)\right] < \infty$.

- **EXAMPLE : Strong-GARCH with QML equations**

$$y_t = h_t(\theta^0)\varepsilon_t \text{ and } m_t(\theta) = \left\{\varepsilon_t^2 - 1\right\}\frac{1}{h_t^2(\theta)}\frac{\partial}{\partial\theta}h_t^2(\theta)$$

$$E\left[m_{2,t}^2(\theta^0)\right] < \infty \text{ if only if } E\left[\varepsilon_t^4\right] < \infty$$

Hall/Yao ('03), Berkes/Horvarth ('03), Meddahi/Renault ('04), Linton et al ('10)

OBJECTIVES

- Estimating equations:

$$m_t(\theta) : \Theta \to \Re^q \quad \text{compact}\, \Theta \subset \Re^k,\ q \geq k \geq 1$$

Identification : $E[m_t(\theta)] = 0$ if and only if $\theta = \theta^0$

Asymptotic theory requires $E\left[m_{i,t}^2(\theta^0)\right] < \infty$.

- **EXAMPLE : AR(1)**

$$y_t = \theta^0 y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid, \quad E[\varepsilon_t] = 0, \quad E[\varepsilon_t^2] = \infty, \text{ and } \quad m_t(\theta) = \left[\left(y_t - \theta y_{t-1}\right) y_{t-1}\right]$$

$$E\left[m_t(\theta^0)\right] = 0, \text{ yet } m_t(\theta) \text{ is non-integrable} : \left|E[m_t(\theta)]\right| = \left|\theta - \theta^0\right| E[y_t^2] = \infty$$

$$E\left[m_t^2(\theta^0)\right] < \infty \text{ if only if } E\left[\varepsilon_t^2\right] < \infty$$

LITERATURE

- **Three camps:**

1. M- estimation with heavy tails

   non-standard limits, super-$n^{1/2}$-convergent

   (Hannan and Kanter, Cline, Hall and Yao)

2. Robust M- and GMM-estimation: *robust to outliers* (*errors*)

   non-standard limits if heavy tails, super-$n^{1/2}$-convergent (M-estimation)

   non-identification if asymmetric (GMM: Cizek '05, '08, '10)

3. Heavy-tail robust M-estimation - $n^{1/2}$-convergent, symmetric DGP

   LAWD, QMWL (Ling '05, '07), R-estimation (Andrews '08)

LITERATURE

- **Outlier Robust M-estimators** : Diminish high breakdown points

  Least Trimmed Squares (Rousseeuw 83, 84, 85; Čižek 07, 08)

  Maximum Trimmed Likelihood (Neykov and Neytchev 90; Čižek 07, 08)

  Least Absolute Trimmed Deviations (Bassett 91; Čižek 07, 08)

  Fixed quantile trimming, cross-sectional data, thin-tails.

  Linear models.

  No feedback (e.g. ADL, GARCH, SV, Markov Chains)

- **Heavy Tail Robust M-estimators**

  Weighted LAD, QML - AR, GARCH (Ling '05, '07)

  R-estimation for ARMA (Andrews '08)

  $n^{1/2}$-convergence - (rank weights, fixed quantile weight, symmetric DGP)

❑ Example: Least Trimmed Squares

$$y_t = \theta' x_t + \varepsilon_t$$

Criterion equations $s_t(\theta) := (y_t - \theta' x_t)^2$.

$$\text{Criterion : } \hat{Q}_n(\theta) := \frac{1}{n}\sum_{t=1}^{n} s_t(\theta) \times I\left(s_t(\theta) \le s_{([\lambda n])}(\theta)\right) = \frac{1}{n}\sum_{t=1}^{n} s_{n,t}(\theta)$$

$s_{(1)}(\theta) \geq s_{(2)}(\theta) \geq \ldots$ and $[\mathrm{z}]$ denotes the integer part

$\lambda \in (0,1)$ is the "fixed quantile" basis of trimming.

Ex. $\lambda = .05 \quad \Rightarrow \quad s_{([.05n])}(\theta)$ is the 5% two-tailed quantile.

❑ Example: Least Trimmed Squares

If the marginal distributions of $\{\varepsilon_t, x_t\}$ are sufficiently smooth (Čižek '05, '08)

$$\frac{\partial}{\partial\theta}\hat{Q}_n(\theta)\,|_{\theta=\hat{\theta}} = -2\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t(\hat{\theta})x_t \times I\left(\left|\varepsilon_t(\hat{\theta})\right| \leq \varepsilon^{(a)}_{([\lambda n])}(\hat{\theta})\right) = 0$$

$$= -2\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t x_t \times I\left(\left|\varepsilon_t\right| \leq \varepsilon^{(a)}_{([\lambda n])}\right) + o(n^{-1/2})$$

where $\varepsilon^{(a)}_t := |\,\varepsilon_t\,|$

Asymptotics governed by $\varepsilon_t x_t \times I\left(\left|\varepsilon_t\right| \leq \varepsilon^{(a)}_{([\lambda n])}\right)$.

Asymptotic normality requires finite variance errors *and* regressors.

Ignores vast array of time series models.

- **Robust GMM Estimators**

  Diminish high breakdown points (Roncheetti and Trojani 01, Čižek 10)

  Trim or truncate $m_t(\theta)$ by fixed λ-quantile

  Fixed quantile forces bias if asymmetric DGP

    Assume identification (Čižek) or SMM for known distribution (Ronchetti)

- **Example: GMTM on Linear Regression (**Čižek 10**)**

$$E\left[\varepsilon_t x_{i,t} \times I\left(\left|\varepsilon_t x_{i,t}\right| \leq m^{(a)}_{i,([\lambda n])}(\theta^0)\right)\right] = 0 \text{ assumed to hold}$$

  Fails to hold for asymmetric distributions.

❑ Estimating Equations and Identification

$$m_t(\theta) : \Theta \rightarrow \Re^q \quad \text{compact } \Theta \subset \Re^k,\ q \geq k \geq 1$$

stationary	: simplify thresholds
continuous, differentiable	: allow non-differentiable
absolutely continuous marginal dist.	: aids asymptotic expansion
geometrically β-mixing	: allow hyperbolic memory

Identification $E[m_t(\theta)] = 0$ *if and only if* $\theta = \theta_0$, unique interior Θ.

❑ Stochastically Trimmed Equations

Define *tail-specific* observations and *order statistics*:

$$m_{i,t}^{(-)} := m_{i,t} \times I\left(m_{i,t} < 0\right) \text{ and } m_{i,t}^{(+)} := m_{i,t} \times I\left(m_{i,t} \geq 0\right)$$

$$m_{i,(1)}^{(-)} \leq m_{i,(2)}^{(-)} \leq \ldots \leq m_{i,(n)}^{(-)} \text{ and } m_{i,(1)}^{(+)} \geq m_{i,(2)}^{(+)} \geq \ldots \geq m_{i,(n)}^{(+)}$$

Trim $k_{1,i,n}$ left-tail and $k_{2,i,n}$ right-tail equations:

$$\hat{m}_{n,t}^{*}(\theta) := \left[m_{i,t}(\theta) \times I\left(m_{i,(k_{1,i,n})}^{(-)}(\theta) \leq m_{i,t}(\theta) \leq m_{i,(k_{2,i,n})}^{(+)}(\theta)\right)\right]_{i=1}^{q}$$

Intermediate order trimming (Leadbetter et al '83):

$k_{j,i,n} \rightarrow \infty$ and $k_{j,i,n}/n \rightarrow 0$: trim *many* for normality, *few* for ident'ion.

❑ GMTTM Criterion and Estimator

$$\hat{Q}_n(\theta) := \left(\frac{1}{n}\sum_{t=1}^{n}\hat{m}^{*}_{n,t}(\theta)'\right) \times \hat{\Omega}_n \times \left(\frac{1}{n}\sum_{t=1}^{n}\hat{m}^{*}_{n,t}(\theta)\right) \text{ where } \hat{\Omega}_n \xrightarrow{p} \Omega \text{ p.d.}$$

$$\text{The GMTTM estimator}: \hat{\theta}_n = \underset{\theta\in\Theta}{\arg\inf}\, \hat{Q}_n(\theta)$$

- Symmetric trimming for symmetric DGP's: **AR with symmetric errors**

$$\hat{m}^{*}_{n,i,t}(\theta) := m_{i,t}(\theta) \times I\left(\left|m_{i,t}(\theta)\right| \leq m^{(a)}_{i,(k_{i,n})}(\theta)\right) \text{ where } m^{(a)}_{i,t} := \left|m_{i,t}\right|$$

- Do not trim $m_{i,t}(\theta^0)$ if known to have *finite variance*

- "Irrelevant" if *unnecessary* trimming (e.g. unknown $m_{i,t}$ moments)

❑ GMTTM Criterion and Estimator

- Symmetric Trimming for symmetric *equations*

$$\text{AR}: \; y_t = \theta^0 y_{t-1} + \varepsilon_t \quad \Rightarrow \quad m_{i,t}(\theta^0) = \varepsilon_t y_{t-i} \; \sim \; \text{symmetric...}$$

- Asymmetric Trimming for asymmetric *equations*

$$\text{ARCH}: \; y_t = \left(\omega^0 + \alpha^0 y_{t-1}^2\right)^2 \varepsilon_t$$

$$\Rightarrow \quad m_t(\theta^0) = \left(\varepsilon_t^2 - 1\right) \frac{1}{\omega^0 +^0 \alpha y_{t-1}^2} \left[1, y_{t-1}^2\right]' \; \sim \; \text{asymmetric...}$$

❑ Deterministically Trimmed Equations

Intermediate order thresholds $\{l_{i,n}(\theta), u_{i,n}(\theta)\}$

$$P\left(m_{i,t}(\theta) < -l_{i,n}(\theta)\right) = \frac{k_{1,i,n}}{n} \to 0 \text{ and } P\left(m_{i,t}(\theta) > u_{i,n}(\theta)\right) = \frac{k_{2,i,n}}{n} \to 0$$

(**exist for smooth distributions**)

Deterministically trimmed equations

$$m^*_{n,t}(\theta) := \left[m_{i,t}(\theta) \times I\left(-l_{i,n}(\theta) \leq m_{i,t}(\theta) \leq u_{i,n}(\theta)\right)\right]_{i=1}^{q}$$

Asymptotics based on $m^*_{n,t}(\theta)$.

GMTTM THEORY

❑ Assumptions

A. Identification by Tail-Trimmed Equations

$$E\left[m^*_{n,t}(\theta^0)\right] \to 0 \text{ sufficiently fast (Lebesgue's dom. conv. ensures} \to 0)$$

If symmetric equations are symmetrically trimmed: *trivial*.

$$\text{Else } \left\| n^{1/2} S_n^{-1/2} E\left[m^*_{n,t}(\theta^0)\right]\right\| \to 0 \text{ where } S_n = E\left(\frac{1}{n}\sum_{s,t=1}^{n} m^*_{n,s}(\theta^0) m^*_{n,t}(\theta^0)'\right)$$

Intuitively, show $\hat{m}^*_{n,t}(\theta) \approx m^*_{n,t}(\theta)$ and expand $m^*_{n,t}(\theta)$ around θ^0

$$S_n^{-1/2}\sum_{t=1}^{n} m^*_{n,t}(\theta^0) = S_n^{-1/2}\sum_{t=1}^{n}\left\{m^*_{n,t}(\theta^0) - E\left[m^*_{n,t}(\theta^0)\right]\right\} + \underline{n^{1/2} S_n^{-1/2} E\left[m^*_{n,t}(\theta^0)\right]}$$

Trivial if symmetric: $E[m^*_{n,t}(\theta^0)] = 0.$

GMTTM THEORY

- **Assumptions**

A. Identification

Sufficient smoothness:

$$\inf_{n \geq N} \inf_{\|\theta-\theta^0\|>\delta} \frac{1}{\|\mathbf{m}_n\|} \left\| E\left[m^*_{n,t}(\theta)\right] \right\| > 0 \ \ \forall \delta > 0, \ \ N \geq 1 \text{ sufficiently large}$$

where

$$\mathbf{m}_n = \sup_{\theta \in \Theta} \left\| E\left[m^*_{n,t}(\theta)\right] \right\| \text{ the moment envelope.}$$

$E[m_{n,t}(\theta)]$ are *non-zero* near θ^0 for large n,

$||E[m_{n,t}(\theta)]||$ is *possibly divergent* : if $m_t(\theta)$ is non-integrable (e.g. AR).

Holds for linear in parameters : deviation from θ^0 does not alter rate $E[m^*_{n,t}(\theta)]$

❑ Assumptions

B. Equation Properties

i. $\{m_t(\theta)\}$ is L_p-bounded, $p > 0$, geometrically β-mixing.

ii. $\{m_t(\theta)\}$ is continuous, differentiable

iii. $m_{i,t}(\theta)$ have absolutely continuous distributions

iv. $m_{i,t}(\theta)$ have power-law tails if they have infinite variance.

- **Geometric** mixing simplifies asymptotic arguments.
- **β-mixing** for uniform laws for empirical processes (Doukhan et al '95)
- **Nonlinear AR-nonlinear GARCH** (e.g. Meitz and Saikkonen 08)
- **Differentiability** easily relaxed (e.g. Pakes/Pollard 89, Newey/McFaddn 94)
- **Power-laws** ease solving rate of convergence, bounding matrices...

Feller class '67: moment/tail-probability link.

(Pruit '83, Hahn et al '90, Whalen '93)

❑ Assumptions

C. Jacobia and Covariance

$$J_n(\theta) = \frac{\partial}{\partial\theta} E\left[m^*_{n,t}(\theta)\right] \qquad\qquad S_n = \frac{1}{n}\sum_{s,t=1}^{n} E\left[m^*_{n,s}(\theta^0)m^*_{n,t}(\theta^0)'\right]$$

$$J^*_n(\theta) = \left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta} m_{i,t}(\theta) \times I\left(-l_{i,n}(\theta) \le m_{i,t}(\theta) \le u_{i,n}(\theta)\right)\right]_{i=1}^{q}$$

$$\hat{J}_n(\theta) = \left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta} m_{i,t}(\theta) \times I\left(m^{(-)}_{i,(k_{l,i,n})}(\theta) \le m_{i,t}(\theta) \le m^{(+)}_{i,(k_{u,i,n})}(\theta)\right)\right]_{i=1}^{q}$$

J's have full column rank for all $n \ge N$; $\qquad S_n$ is p.d. for all $n \ge N$.

❑ Assumptions

C. Jacobian Smootheness

$$\frac{\sup\limits_{\|\theta-\theta^0\|\le\delta_n} \left\| J_n(\theta) - J_n(\theta^0) \right\|}{\left\| J_n(\theta^0) \right\|} \to 0 \text{ for any } \delta_n \to 0$$

Near θ^0, $J_n(\theta)$ does not grow "much faster" than $J_n(\theta^0)$.

D. Metric Entropy

$\{I_{i,n}(\theta) : \theta \in \Theta\}$ satisfies metric entropy with L_2-bracketing

- **Consistency and Asymptotic Normality under A-D**

$$\hat{\theta}_n \xrightarrow{p} \theta^0 \text{ and } \hat{\theta}_n = \theta^0 + O\left(\underline{\left\|V_n\right\|^{-1/2}}\right)$$

and

$$V_n^{1/2}\left(\hat{\theta}_n - \theta^0\right) = A_n \sum_{t=1}^{n} m_{n,t}^{*}(\theta^0) \times \left(1 + o_p(1)\right) + o_p(1) \xrightarrow{d} N(0, I_k)$$

where

$$\underline{V_n = n \times H_n (J_n{}'\Omega_n S_n \Omega_n J_n)^{-1} H_n} \quad \text{- usual quadratic form (Hansen 82,...)}$$

and $H_n = -J_n{}'\Omega_n J_n$ and $J_n := \dfrac{\partial}{\partial\theta} E\left[m_{n,t}^{*}(\theta)\right]|_{\theta=\theta^0}$

- **Consistency : Modified Pakes and Polard ('89) due to ULLN**

Define $Q_n(\theta) = E[m^*_{n,t}(\theta)]'\Omega_n E[m^*_{n,t}(\theta)]$ and $\mathbf{m}_n := \sup_{\theta\in\Theta}\left\| E[m^*_{n,t}(\theta)] \right\|$

Identification *smoothness* and *weight boundedness* imply

$$\varepsilon(\delta) := \inf_{n\geq N} \inf_{\|\theta-\theta^0\|\leq\delta} \frac{Q_n(\theta)}{\mathbf{m}_n^2} > 0 \text{ for large } N \text{ and any } \delta > 0$$

Since $P\left(\left\|\hat{\theta}_n - \theta^0\right\| > \delta\right) \leq P\left(\frac{Q_n(\hat{\theta}_n)}{\mathbf{m}_n^2} > \varepsilon(\delta)\right)$

if we show $Q_n(\hat{\theta}_n) = o_p\left(\mathrm{M}_n^2\right)$ then it follows $\| \hat{\theta}_n - \theta^0 \| \xrightarrow{p} 0.$

GMTTM THEORY

- **Consistency : Modified Pakes and Polard ('89) due to ULLN**

We know $\|\hat{\theta}_n - \theta^0\| \xrightarrow{p} 0$ if $Q_n(\hat{\theta}_n) = o_p\left(\mathrm{M}_n^2\right)$.

By construction and uniform criterion bound (*requires* **)

$$Q_n(\hat{\theta}_n) \leq \hat{Q}_n(\hat{\theta}_n) + \underline{\left|\hat{Q}_n(\hat{\theta}_n) - Q_n(\hat{\theta}_n)\right|} \leq \hat{Q}_n(\theta^0) + \underline{\left(\mathrm{M}_n^2 + Q_n(\hat{\theta}_n)\right) \times o_p(1)}$$

hence $Q_n(\hat{\theta}_n) \times \left(1 + o_p(1)\right) \leq \hat{Q}_n(\hat{\theta}_n) + o_p(\mathbf{m}_n^2)$

Further $\hat{Q}_n(\hat{\theta}_n) \leq \hat{Q}_n(\theta^0) \leq K\left\|\hat{m}_n^*(\theta^0)\right\|^2 \leq K\left\|m_n^*(\theta^0)\right\|^2 + o_p\left(\left\|S_n\right\|^{1/2} / n^{1/2}\right) \xrightarrow{p} 0$

** ULLN for $m_n^*(\theta)$ is delicate if $\left\|E\left[m_n^*(\theta)\right]\right\| \to \infty$

$$\sup_{\theta \in \Theta}\left\|m_n^*(\theta) - E\left[m_n^*(\theta)\right]\right\| = o_p\left(\sup_{\theta \in \Theta}\left\|E[m_{n,t}^*(\theta)]\right\|\right) \text{ for } \beta\text{-mixing}\{m_t(\theta)\}$$

❑ Asymptotic Efficiency - Intermediate vs. Central Order Trimming

GMTTM scale has classic symmetric/quadratic form:

$$V_n = n \times H_n (J_n ' \Omega_n S_n \Omega_n J_n)^{-1} H_n,$$

Due to *negligibility*:

$$J_{i,j,n} = \frac{\partial}{\partial \theta_j} E\left[m_{i,t}(\theta) I\left(\left| m_{i,t}(\theta) \right| \leq c_{i,n}(\theta) \right) \right] |_{\theta^0}$$

$$= E\left[\frac{\partial}{\partial \theta_j} m_{i,t}(\theta) |_{\theta^0} \; I\left(\left| m_{i,t}(\theta) \right| \leq c_{i,n}(\theta) \right) \right] + \frac{\partial}{\partial \theta_j} E\left[m_{i,t}(\theta^0) I\left(\left| m_{i,t}(\theta) \right| \leq c_{i,n}(\theta) \right) \right] |_{\theta^0}$$

$$= E\left[\frac{\partial}{\partial \theta_j} m_{i,t}(\theta) |_{\theta^0} \; I\left(\left| m_{i,t}(\theta) \right| \leq c_{i,n}(\theta) \right) \right] \times \left(1 + o(1) \right) \text{ since } \underline{c_{i,n}(\theta) \to \infty}$$

Jacobian of trimmed moment is like mean of Jacobian

❑ Asymptotic Efficiency - Intermediate vs. Central Order Trimming

GMTTM scale has classic symmetric/quadratic form:

$$V_n = n^2 \times H_n (J_n{}' \Omega_n S_n \Omega_n J_n)^{-1} H_n$$

Due to *negligibility*: **fixed quantile trimming does not** (Čižek 10):

Asymptotically efficient weight Ω_n is *trimmed equation covariance* (Hansen)

$$\Omega_n = S_n^{-1} \quad \Rightarrow \quad V_n = n \times J_n{}' S_n^{-1} J_n$$

(**Does not exist for GMTM** - Čižek 10)

❑ Scale Estimation for GMTTM Inference

$$\hat{J}_n(\hat{\theta}) = \left[\frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta) \right]_{i=1}^{q}$$

$$\hat{S}_n(\theta) = \sum_{s,t=1}^{n} k((s-t)/\gamma_n)\hat{m}^{*}_{n,s}(\theta)\hat{m}^{*}_{n,t}(\theta)' \ : \ \text{bandwidth } \gamma_n = o(n)$$

$$\hat{H}_n(\theta) = -\hat{J}_n(\theta)'\hat{\Omega}_n \hat{J}_n(\theta)$$

$$\hat{V}_n(\theta) = n \times \hat{H}_n(\theta) \left\{ \hat{J}_n(\theta)'\hat{\Omega}_n \hat{\Sigma}_n^{-1}(\theta)\hat{\Omega}_n \hat{J}_n(\theta) \right\}^{-1} \hat{H}_n(\theta)$$

HAC - kernel self normalized CLT for tail-trimmed NED arrays (Hill '10)

- theory grounded on Davidson and de Jong ('00)

❑ Scale Estimation for GMTTM Inference

Under A-D, and kernel properties in Davidson and de Jong ('00)

$$\hat{J}_n(\tilde{\theta}_n) = J_n(\theta^0)\times(1+o_p(1)) \quad \text{for any } \tilde{\theta}_n \xrightarrow{p} \theta^0 \text{ under A - D.}$$

$$\hat{S}_n(\hat{\theta}_n) = S_n(\theta^0)\times(1+o_p(1) \quad \text{for any } \tilde{\theta}_n = \theta^0 = O_p\left(\left\|V_n\right\|^{-1/2}\right)$$

$$\hat{V}_n(\hat{\theta}_n) = V_n \times (1+o_p(1)) \quad : \quad V_n^{-1}\hat{V}_n(\hat{\theta}_n) \xrightarrow{p} I_k$$

Barlett, Parzen, Tukey-Hanning, Quadratic Spectral...

CONTRIBUTIONS TO LIMIT THEORY

- Pointwise and uniform LLN, CLT for dependent tail-trimmed arrays

- Sharp mixingale inequality without memory/heterog. restrictions (Hill '10):

$$S_n = E\left[\left(\sum_{t=1}^{n}\left\{m_{n,t}^{*}(\theta) - E\left[m_{n,t}^{*}(\theta)\right]\right\}\right)^2\right] \leq K\sum_{t=1}^{n} E\left[m_{n,t}^{*2}(\theta)\right]$$

ULLN and Uniform Approximations (Hill and Renault '10, Hill '10):

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{m_{n,t}^{*}(\theta) - E\left[m_{n,t}^{*}(\theta)\right]\right\}\right\| = o_p\left(\sup_{\theta\in\Theta}\left\|E\left[m_{n,t}^{*}(\theta)\right]\right\|\right)$$

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\hat{m}_{n,t}^{*}(\theta) - m_{n,t}^{*}(\theta)\right\}\right\| = o_p(1) : \left\{I_{i,n,t}(\theta) : \theta \in \Theta\right\} \text{ and } m_{i,(k_{j,i,n})}(\theta)$$

$$\text{CLT (Hill '09, Hill and Renault '10)} : \Sigma_n^{-1/2}\sum_{t=1}^{n} m_{n,t}^{*}(\theta^0) \xrightarrow{d} N\left(0, I_q\right)$$

ASSUMPTION VERIFICATION

- Linear models (ARX, GARCH, ARX-GARCH) satisfy major assumptions.

- **EXAMPLE: Stationary AR(*p*)**

$$y_t = \sum_{i=1}^{p} \theta_i^0 y_{t-i} + \varepsilon_t \ \text{ where } \varepsilon_t \sim P\left(\left|\varepsilon_t\right| > \varepsilon\right) = d\varepsilon^{-\kappa}(1 + o(1)),\ \kappa \in (1,2]$$

ε_t has symmetric, absolutely continuous distribution.

All GMTTM assumptions are *trivial* (identification, smoothness, equation diff., metric entropy) or *regulatory* (thresholds, positive definiteness, rank).

- **EXAMPLE: Stationary ARCH(*p*) with OLS or QML-equations**

RATE OF CONVERGENCE

- **Rate of Convergence : Stationary AR(1),** ***symmetric trimming***

$y_t = \theta^0 y_{t-1} + \varepsilon_t, \left|\theta^0\right| < 1$ ε_t is iid, symmetric, absolutely continuous dist.

$P\left(\left|\varepsilon_t\right| > \varepsilon\right) = d\varepsilon^{-\kappa}\left(1 + o(1)\right)$ where $d > 0$ and $\kappa \in (1,2)$

and $m_t(\theta) = \left(y_t - \theta y_{t-1}\right) y_{t-1}$

[AR(p) and over-identification ($q > p$) are similar.]

Symmetric trimming : one fractile k_n

RATE OF CONVERGENCE

❑ Rate of Convergence : Stationary AR(1), *symmetric trimming*

$$\text{Efficient weight } \Omega_n = S_n^{-1} \quad \Rightarrow \quad V_n = n \times \frac{J_n^2}{S_n}$$

1. Scale S_n

1.1 $m_t(\theta^0) = \varepsilon_t y_{t\text{-}1}$ is a product convolution with iid ε_t

$$E\left[m_{n,t}^*(\theta^0) \mid \Im_{t-1}\right] = 0 \quad \Rightarrow \quad S_n = E\left[m_{n,t}^{*2}(\theta^0)\right]$$

1.2 Equations have power-law tail (Brockwell/Cline 85, Cline 86)

$$m_t(\theta^0) = \varepsilon_t y_{t-1} \text{ has same tail index } \kappa \in (1,2)$$

$$\text{Karamata's Theorem and } c_n = K(n/k_n)^{1/\kappa} :$$

$$S_n = E\left[m_{n,t}^{*2}(\theta^0)\right] \sim K c_n^2 P\left(\left|m_t(\theta^0)\right| > c_n\right) \sim K(n/k_n)^{2/\kappa - 1}$$

- **Rate of Convergence : Stationary AR(1), *symmetric trimming***

$$\text{Efficient weight } \Omega_n = S_n^{-1} \quad \Rightarrow \quad V_n = n \times \frac{J_n^2}{S_n}$$

2. Jacobian J_n

$$J_n = \frac{\partial}{\partial \theta} E\left[m_t(\theta) I\left(\left|m_t(\theta)\right| \le c_n(\theta)\right)\right]|_{\theta^0}$$

$$= E\left[\frac{\partial}{\partial \theta}\left(y_t - \theta y_{t-1}\right) y_{t-1} \,|_{\theta^0}\; I\left(\left|\varepsilon_t y_{t-1}\right| \le c_n\right)\right] \times \left(1 + o(1)\right)$$

$$= -E\left[y_{t-1}^2 I\left(\left|\varepsilon_t y_{t-1}\right| \le c_n\right)\right] \times \left(1 + o(1)\right) \longrightarrow$$

(works like tail-trimmed variance due to iid ε_t)

$$\sim -K c_n^2 P\left(\left|\varepsilon_t y_{t-1}\right| > c_n\right) \sim -K (n / k_n)^{2/\kappa - 1}$$

RATE OF CONVERGENCE

- **Rate of Convergence : Stationary AR(1), *symmetric trimming***

$$\text{Efficient weight } \Omega_n = S_n^{-1} \quad \Rightarrow \quad V_n = n \times \frac{J_n^2}{S_n}$$

3. Scale V_n has super n-rate:

Regressor as *leverage* point : large values augment rate.

$$V_n^{1/2} = n^{1/2} \frac{|J_n|}{S_n^{1/2}} \sim n^{1/2} \frac{(n/k_n)^{2/\kappa-1}}{\left[n(n/k_n)^{2/\kappa-1}\right]^{1/2}} = n^{1/2}(n/k_n)^{1/\kappa-1/2}$$

Error as *outlier* : diminishes rate when variance is infinite.

Argument fails if there is error/regressor feedback: $S_n \to \infty$ very fast.

GARCH, AR-GARCH, etc.

RATE OF CONVERGENCE

- **Rate of Convergence : Stationary AR(1), *symmetric trimming***

$$\text{Efficient weight } \Omega_n = S_n^{-1} \quad \Rightarrow \quad V_n = n \times \frac{J_n^2}{S_n}$$

3. Scale V_n has super n-rate:

$$\frac{V_n^{1/2}}{n^{1/2}} = (n / k_n)^{1/\kappa - 1/2} \to \infty$$

due to tail-trimming.

GMTTM: best of all worlds

asymptotic normality

super-$n^{1/2}$-convergence.

RATE OF CONVERGENCE

- **Rate of Convergence : Stationary AR(1), *symmetric trimming***

$$\text{Efficient weight } \Omega_n = S_n^{-1} \quad \Rightarrow \quad V_n = n \times \frac{J_n^2}{S_n}$$

3. Scale V_n has super n-rate:

Light trimming augments *leverage* effect, optimizes rate.

$k_n = ln(n)$ or any slowly varying function $L(n)$

$$V_n^{1/2} = n^{1/2}\left(n/\ln(n)\right)^{1/\kappa - 1/2} = n^{1/\kappa} / L(n) \text{ for some s.v. } L(n)$$

The *highest rate* amongst M-estimators for stationary data

(Hannan and Kanter 77, Cline 89, Davis et al 92)

RATE OF CONVERGENCE

❑ Rate of Convergence : ARCH(q) with QML Equations

$$y_t = h_t(\theta_0)\varepsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} (0,1), \quad \text{and} \quad h_t^2(\theta) = \alpha + \sum_{i=1}^{q} \beta_i y_{t-i}^2$$

$$m_{1,t}(\theta) = \left(y_t^2 - h_t^2(\theta)\right)\frac{1}{h_t^4(\theta)} \quad \text{and} \quad m_{i,t}(\theta) = \left(y_t^2 - h_t^2(\theta)\right)\frac{y_{t-i}^2}{h_t^4(\theta)}$$

$$\text{If } E[\varepsilon_t^4] < \infty \text{ then } n_\alpha, n_{\beta_i} = K n^{1/2}.$$

$$\text{If } E[\varepsilon_t^4] = \infty \text{ then } n_\alpha, n_{\beta_i} = o\left(n^{1/2}\right).$$

Allow arbitrarily heavy tails for GARCH $\{y_t\}$.

Sub-$n^{1/2}$-convergence due to "error"-regressor feedback.

Advantage : allow asymmetry (Ling 07: QMWL $n^{1/2}$).

RATE OF CONVERGENCE

❑ Rate of Convergence : ARCH(q) with QML Equations

Feedback: large "leverage" is large "outlier" : heavy trimming

$k_n = n/ln(n)$ or any slowly varying function $L(n)$

$$V_{i,i,n}^{1/2} \sim Kn^{1/2} / L(n) = o(n^{1/2}) \text{ if } E\left[\varepsilon_t^4\right] = \infty$$

$$V_{i,i,n}^{1/2} \sim Kn^{1/2} \text{ if } E\left[\varepsilon_t^4\right] < \infty$$

QML (Hall and Yao 03)

$$V_{i,i,n}^{1/2} \sim Kn^{1-2/\kappa_\varepsilon} \text{ where } \kappa_\varepsilon \text{ is tail index of } \varepsilon_{\mathrm{t}}$$

$$= o(n^{1/2} / L(n)) \text{ if } \kappa_\varepsilon \in (2,4)$$

QMWL : $n^{1/2}$-rate for stationary linear GARCH, iid (0,1) - errors (Ling 07)

FRACTILE SELECTION

- **Rate of Convergence Criterion for Equation Type $\boldsymbol{m_t = \varepsilon_t x_t}$**

Select policy $\{k_{1,i,n}, k_{2,i,n}\}$ to optimize $||V_n|| \to \infty$

No feedback rule of thumb : $k_{j,i,n} = L(n) \to \infty$ slowly vary rate.

Feedback rule of thumb : $k_{j,i,n} = n/L(n) \to \infty$.

❑ Identification Criterion for Asymmetric Equations

Assume one equation $m_t(\theta^0)$: q =1.

Recall we require

$$\left\| n^{1/2} S_n^{-1/2} E\left[m_{n,t}^*(\theta^0)\right]\right\| \to 0 \text{ where } S_n = E\left(\frac{1}{n}\sum_{s,t=1}^{n} m_{n,s}^*(\theta^0) m_{n,t}^*(\theta^0)\right)$$

Assume $m_i(\theta^0)$ has exactly Pareto tails: for all $m > N > 0$

$$P\left(m_t(\theta) < -m\right) = d_1 m^{-\kappa_1} \text{ and } P\left(\left|m_t(\theta)\right| > m\right) = d_2 m^{-\kappa_2},\ d_i > 0, \kappa_i > 1$$

Then

$$E\left[m_t(\theta^0) I\left(-l_n \le m_t(\theta^0) \le u_n\right)\right] = \left(\frac{d_1\kappa_1}{\kappa_1 - 1}\right)\frac{1}{l_n^{\kappa_1 - 1}} - \left(\frac{d_2\kappa_2}{\kappa_2 - 1}\right)\frac{1}{u_n^{\kappa_2 - 1}}$$

FRACTILE SELECTION

- **Identification Criterion for Asymmetric Equations**

$$E\left[m_t(\theta^0)I\left(-l_n \le m_t(\theta^0) \le u_n\right)\right] = \left(\frac{d_1\kappa_1}{\kappa_1 - 1}\right)\frac{1}{l_n^{\kappa_1 - 1}} - \left(\frac{d_2\kappa_2}{\kappa_2 - 1}\right)\frac{1}{u_n^{\kappa_2 - 1}}$$

Force *as close to zero as we choose* for well chosen $\{l_n, u_n\}$.

Fractiles and thresholds linked:

$$l_n = d_1\left(\frac{n}{k_{1,n}}\right)^{1/\kappa_1} \text{ and } u_n = d_2\left(\frac{n}{k_{2,n}}\right)^{1/\kappa_2}$$

Choose $\dfrac{k_{2,n}^{1-1/\kappa_2}}{k_{1,n}^{1-1/\kappa_1}} = n^{1/\kappa_1 - 1/\kappa_2}\dfrac{d_1^{1/\kappa_1}}{d_2^{1/\kappa_2}}\dfrac{(1-1/\kappa_2)}{(1-1/\kappa_1)}$: trim **less** from **heavier** tail.

Intuition : symmetric trim when heavier right tail $E\left[m_t(\theta^0)I\left(\left|m_t(\theta^0)\right| \le c_n\right)\right] < 0$

- **Identification Criterion for Asymmetric Equations**

Obtain consistent estimator $\tilde{\theta}_n$

GMTTME functional $\hat{\theta}_n(\delta_{j,i})$: e.g. $k_{j,i,n} = \delta_{j,i} n / \ln(n)$ or $k_{j,i,n} = n^{\delta_{j,i}}$

$$\min_{\delta_{j,i}} \left\| \hat{\theta}_n(\delta_{j,i}) - \tilde{\theta}_n \right\|$$

or indirect inference (e.g. Aguilar, Hill and Renault '10)

SIMULATION

- **AR, GARCH(1,1), IGARCH(1,1), TARCH(1), QARCH(1)**

AR : iid Pareto errors ε_t with index $\kappa \in \{1.5, 2.5\}$

All GARCH : iid $N(0,1)$ errors,

iid Pareto errors with index $\kappa = 2.5$

SIMULATION

- **AR, GARCH(1,1), IGARCH(1,1), TARCH(1), QARCH(1)**

AR : iid Pareto errors ε_t with index $\kappa \in \{1.5, 2.5\}$

All GARCH : iid $N(0,1)$ errors,

iid Pareto errors with index $\kappa = 2.5$

GMTTM : Symmetric trimming for AR

$$k_n = [n^{\lambda}] \text{ over } \lambda \in \{.01, .02, ..., .99\}$$

GMTTM : Asymmetric trimming for all GARCH with QML-equations

$$k_{l,n} = [n^{\lambda_l}] \text{ and } k_{u,n} = [n^{\lambda_u}] \text{ over grid } \lambda_l, \lambda_u \in \{.01, .02, ..., .99\}$$

Minimize k^{th}-GMTTME's Kolmogorov-Smirnov statistic : λ^*.

SIMULATION

- **AR, GARCH(1,1), IGARCH(1,1), TARCH(1), QARCH(1)**

AR : iid Pareto errors ε_t with index $\kappa \in \{1.5, 2.5\}$

All GARCH : iid $N(0,1)$ errors,

iid Pareto errors with index $\kappa = 2.5$

LAWD for <u>AR</u> with Huber ('77) weight, 5% quantile (Ling 05)

$$\sum_{t=1}^{n} w_t \left| y_t - \theta y_{t-1} \right| \text{ where } w_t = \left(\max\left\{ 1, \frac{1}{y^{(a)}_{[.05n]}} \left| y_{t-1} \right| I\left(\left| y_{t-1} \right| > y^{(a)}_{[.05n]} \right) \right\} \right)^{-3}$$

Weight = 1 for errors with "small" level y_{t-1}.

Weight smoothly declines for "large" y_{t-1}.

SIMULATION

- **AR, GARCH(1,1), IGARCH(1,1), TARCH(1), QARCH(1)**

AR : iid Pareto errors ε_t with index $\kappa \in \{1.5, 2.5\}$

All GARCH : iid $N(0,1)$ errors,

iid Pareto errors with index $\kappa = 2.5$

QMWL for GARCH with Huber ('77) weight, 5% quantile (Ling 07)

$$\sum_{t=1}^{n} w_t \ln\phi_t \quad \text{where } w_t = \left(\max\left\{1, \frac{1}{y_{[.05n]}^{(a)}} |y_{t-1}| I\left(|y_{t-1}| > y_{[.05n]}^{(a)}\right)\right\}\right)^{-4}$$

Weight = 1 for errors with "small" level y_{t-1}.

Weight smoothly declines for "large" y_{t-1}.

SIMULATION

- **AR, GARCH(1,1), IGARCH(1,1), TARCH(1), QARCH(1)**

$$\text{AR}(1):\ y_t = .9y_{t-1} + \varepsilon_t \quad \boxed{\theta_k = .9} \quad m_t(\theta) = \left[(y_t - \theta y_{t-1})y_{t-i}\right]_{i=1}^{2}$$

IID pareto error ε_t with index $\kappa = 1.5$

$n = 1000$

	Mean	**StDev**	**KS**	λ^*	k
GMTTM	.900	.007	.066*	.40	16
GMM	.900	.012	.162		
OLS	.889	.013	.253		
LAWD	.900	.007	.061		

KS {10%, 5%, 1%} critical values {.136, .122, .107}

SIMULATION

- **AR, GARCH(1,1), IGARCH(1,1), TARCH(1), QARCH(1)**

$$\text{IGARCH}(1,1): \; y_t = h_t \varepsilon_t \quad h_t^2 = .3 + .4 y_{t-1}^2 + .6 h_{t-1}^2 \quad \boxed{\theta_k = .6} \quad \varepsilon_t \overset{iid}{\sim} N(0,1)$$

$$m_t(\theta) = \left(y_t^2 - \omega - \alpha y_{t-1}^2 - \beta h_{t-1}^2(\theta)\right) \frac{1}{h_{t-1}^4(\theta)} \frac{\partial}{\partial \theta} h_{t-1}^2(\theta)$$

$n = 1000$

	Mean	**StDev**	**KS**	λ^*	k
GMTTM	.608	.179	.064*	.27, .26	26, 6
GMM	.523	.195	.246		
QML	.586	.196	.262*		
QMWL	.602	.094	.095		

KS {10%, 5%, 1%} critical values {.136, .122, .107}

SIMULATION

- **AR, GARCH(1,1), IGARCH(1,1), TARCH(1), QARCH(1)**

$$\text{TARCH}(1,1):\; y_t = h_t\varepsilon_t \quad h_t^2 = .3 + .6 y_{t-1}^2 I(y_{t--1} < 0) \quad \boxed{\theta_k = .6} \quad \varepsilon_t \overset{iid}{\sim} P_{2.5}$$

$$m_t(\theta) = \left(y_t^2 - \omega - \alpha y_{t-1}^2\right)\frac{1}{h_{t-1}^4(\theta)}\frac{\partial}{\partial\theta}h_{t-1}^2(\theta)$$

$n = 1000$

	Mean	**StDev**	**KS**	λ^*	k
GMTTM	.602	.223	.083	.26, .11	6, 2
GMM	.526	.236	.224		
QML	.516	.384	.323		
QMWL	.676	.274	.239		

KS {10%, 5%, 1%} critical values {.136, .122, .107}

The End