Generalized Method of Moments with Tail Trimming

Jonathan B. Hill* and Eric Renault†
Dept. of Economics
University of North Carolina - Chapel Hill

Feb. 2010

...comments are welcome...

Abstract

We develop a GMM estimator for stationary heavy tailed data by trimming an asymptotically vanishing sample portion of the estimating equations. Trimming ensures the estimator is asymptotically normal, and self-normalization implies we do not need to know the rate of convergence. *Tail*-trimming, however, ensures asymmetric models are covered under rudimentary assumptions about the thresholds; it implies super-$\sqrt{n}$-convergence is achievable depending on regressor and error tail thickness and feedback, with a rate arbitrarily close to the largest possible rate amongst untrimmed estimators in some cases; and it implies possibly heterogeneous convergence rates below, at or above $\sqrt{n}$. Models covered include linear or nonlinear autoregressions with linear or nonlinear GARCH innovations. Simulation evidence shows the new estimator dominates GMM and QML when these estimators are not or have not been shown to be asymptotically normal.

1. INTRODUCTION We develop a Generalized Method of Tail-Trimmed Moments estimator for possibly very heavy tailed time series. Heavy tails could be the result of the underlying shocks (e.g. ARX) and/or the parametric structure (e.g. GARCH), depending on the model. There now exists an abundance of stylized evidence in favor of asymmetry and heavy tails in financial, macroeconomic and actuarial data like exchange rate and asset price fluctuations and insurance claims (Mandelbrot 1963, Campbell and Hentschel 1992, Engle and Ng 1993, Embrechts et al 1997, Finkenstadt and Rootzén 2003); microeconomic data like auction bids and birth weight (Chernozhukov 2005, Hill and Shneyerov 2009); and network traffic (Resnick 1997). Coupled with the necessity for over-identifying restrictions in economic models, a robust GMM methodology will be useful to the analyst unwilling to impose ad hoc error and parameter restrictions. See

*Corresponding author. Dept. of Economics, University of North Carolina -Chapel Hill; www.unc.edu/∼jbhill; jbhill@email.unc.edu.

†Dept. of Economics, University of North Carolina -Chapel Hill; renault@email.unc.edu.

Key words and phrases: GMM, tail trimming, robust estimation, heavy tails, super consistency.

JEL subject classifications : C13, C22.

We thank Memet Caner, Pavel Čižek, Feico Drost, Eric Ghysels, Christian Gouriéroux, Alberto Holly, Shakeeb Khan, Denis Pelletier, Liang Peng, Bas Werker, and seminar participants at the Triangle Econometrics Conference 2008, the Joint Statistical Meeting 2008, the First French Econometrics Conference 2009, the University of North Carolina-Chapel Hill Dept. of Economics, the University of California-Davis Dept. of Statistics, the Georgia Institute of Technology Dept. of Mathematics, the University of Tilburg Dept. of Econometrics, for helpful comments.

Hansen (1982), Renault (1997) and Hall (2005).

1.1 TAIL-TRIMMED ESTIMATING EQUATIONS

Let $m_t(\theta)$ denote estimating equations, a stochastic mapping

$$m_t : \Theta \rightarrow \mathbb{R}^q, \text{ compact } \Theta \subset \mathbb{R}^r, \; q \geq r,$$

induced from some moment condition. The strong global identification condition is

$$E\left[m_t(\theta)\right] = 0 \textit{ if and only if } \theta = \theta^0 \text{ for unique } \theta^0 \in \Theta.$$

As an example consider a strong-ARCH(1) process $\{y_t\}$,

$$y_t = h_t \epsilon_t, \;\; h_t^2 = \alpha^0 + \beta^0 y_{t-1}^2, \, \theta^0 = [\alpha^0, \beta^0]', \;\; \epsilon_t \overset{iid}{\sim} (0,1) \text{ and } \Im_t := \sigma(y_\tau : \tau \leq t) \quad (1)$$

with equations

$$m_t(\theta) = \left\{y_t^2 - \alpha - \beta y_{t-1}^2\right\} z_{t-1}, \;\; z_{t-1} = \left[1, y_{t-1}^2, ...\right] \in \mathbb{R}^q, \; q \geq 2.$$

The GMM estimator solves

$$\hat{\theta}_g = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\{ \left(\frac{1}{n} \sum_{t=1}^n m_t(\theta) \right)' \hat{\Upsilon}_n \left(\frac{1}{n} \sum_{t=1}^n m_t(\theta) \right) \right\}$$

for some stochastic positive semi-definite matrix $\hat{\Upsilon}_n \in \mathbb{R}^{q \times q}$, and $n \geq 1$ is the sample size. Under mild conditions $\hat{\theta}_g$ is asymptotically linear (e.g. Newey and McFadden 1994)

$$\sqrt{n}\left(\hat{\theta} - \theta^0\right) = A_n \times \frac{1}{\sqrt{n}} \sum_{t=1}^n m_t(\theta^0) + o_p(1) \text{ for some } A_n \in \mathbb{R}^{r \times q},$$

so asymptotics are grounded on $\sum_{t=1}^n m_t(\theta^0)$.

Finite variances $E[m_{i,t}^2(\theta^0)] < \infty$, along with standard regulatory conditions, ensures Gaussian asymptotics, but this requires ϵ_t and y_t to have finite 4^{th} and 8^{th} moments respectively. This rules out mildly heavy-tailed shocks, integrated random volatility (e.g. IGARCH), and much more. If over-identifying restrictions exist $q \geq 3$ with say $z_{3,t-1}$ $= |y_{t-1}|^{2+\delta/2}$ and $\delta > 0$ then y_t must have a finite $(8 + \delta)^{th}$ moment, a very tall order for financial time series. Models with heterogeneous estimating equations include the multifactor factor Capital Asset Pricing Model with high risk (e.g. oil futures), composite market returns (e.g. NYMEX), low risk asset returns (e.g. U.S. Treasury Bill), and factor premia (e.g. market capitalization and book-to-price ratio); VARX for causality modeling of financial and macroeconomic returns; and multivariate random volatility. See French and Fama (1996), Ding and Granger (1996), Mikosch and Stărică (2000), and Embrechts et al (2003).

Although GMM with a non-Gaussian limit is certainly achievable in the manner of M-estimators (e.g. Hannan and Kanter 1977, An and Chen 1982, Knight 1987, Cline 1989, Davis et al 1992), we seek an estimator that permits standard inference and is therefore simple to use. We propose asymptotically negligibly trimming $k_{1,i,n}$ left-tailed and $k_{2,i,n}$ right-tailed observations from each equation sample $\{m_{i,t}(\theta)\}_{t=1}^n$, where $k_{j,i,n} \rightarrow \infty$ and $k_{j,i,n}/n \rightarrow 0$

Define tail specific observations of $m_{i,t}(\theta)$ and sample order statistics:

$$m_{i,t}^{(-)}(\theta) := m_{i,t}(\theta) \times I\left(m_{i,t}(\theta) < 0\right) \quad \textit{and} \quad m_{i,(1)}^{(-)}(\theta) \leq \cdots \leq m_{i,(n)}^{(-)}(\theta) \leq 0$$

$$m_{i,t}^{(+)}(\theta) := m_{i,t}(\theta) \times I\left(m_{i,t}(\theta) > 0\right) \quad \textit{and} \quad m_{i,(1)}^{(+)}(\theta) \geq \cdots \geq m_{i,(n)}^{(+)}(\theta) \geq 0$$

$$m_{i,t}^{(a)}(\theta) := \left|m_{i,t}(\theta)\right| \quad \textit{and} \quad m_{i,(1)}^{(a)}(\theta) \geq \cdots \geq m_{i,(n)}^{(a)}(\theta) \geq 0.$$

Now trim any equation $m_{i,t}(\theta^0)$ that may have an infinite variance between its lower $k_{1,i,n}/n^{th}$ and upper $k_{2,i,n}/n^{th}$ sample quantiles:

$$\begin{aligned}\hat{m}^*_{i,n,t}(\theta) &:= m_{i,t}(\theta) \times I\left(m^{(-)}_{i,(k_{1,i,n})}(\theta) \leq m_{i,t}(\theta) \leq m^{(+)}_{i,(k_{2,i,n})}(\theta)\right) \\ &= m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta)\end{aligned} \tag{2}$$

$$\hat{m}^*_{n,t}(\theta) = \left[m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta)\right]_{i=1}^{q} \text{ where } \hat{I}_{j,n,t}(\theta) = 1 \text{ if equation } j \text{ is not trimmed,}$$

and $I(A) = 1$ is A is true, and 0 otherwise[1]. If the data generating process is symmetric and $m_{i,t}(\theta^0)$ is heavy-tailed (e.g. IGARCH) then symmetric trimming is appropriate: for $k_{i,n} \to \infty$ and $k_{i,n}/n \to 0$

$$\hat{m}^*_{i,n,t}(\theta) := m_{i,t}(\theta) \times I\left(|m_{i,t}(\theta)| \leq m^{(a)}_{i,(k_{i,n})}(\theta)\right). \tag{3}$$

The Generalized Method of Tail-Trimmed Moments [GMTTM] estimator solves

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\{ \left(\frac{1}{n}\sum_{t=1}^{n} \hat{m}^*_{n,t}(\theta)\right)' \times \hat{\Upsilon}_n \times \left(\frac{1}{n}\sum_{t=1}^{n} \hat{m}^*_{n,t}(\theta)\right) \right\}.$$

As long as $m_t(\theta^0)$ is an integrable martingale difference and standard smoothness conditions apply,

$$V_n^{1/2}\left(\hat{\theta}_n - \theta^0\right) \xrightarrow{d} N(0, I_r)$$

for some sequence of positive definite matrices $\{V_n\}$. The Gaussian limit holds for a host of heavy tailed time series, and simple rules of thumb can be applied to the selection of the trimming fractiles $k_{j,i,n}$.

Inference does not require knowledge of the rate of convergence since we self-normalize, and tail trimming equations with finite variance has no impact on the rate. In particular, $\hat{\theta}_n$ is $\sqrt{n}$-consistent if all equations have a finite variance (e.g. GARCH with finite eighth moment). Further, sub-, exact- *or* super-$\sqrt{n}$-convergence may arise in heavy tailed cases depending on the equation form (e.g. QML or least squares); relative tail thickness of error and regressor; and whether the error is iid (e.g. AR with iid shocks) or depends on the regressor through some form of feedback (e.g. AR with ARCH shocks). There is also a trade-off: the feasible rate of convergence is dampened due to trimming, but the damage can be nearly eradicated by a proper choice of fractiles $k_{j,i,n}$. See Section 3 for convergence rate derivation for dynamic linear regression, IV, ARCH and AR-ARCH models. See also Antoine and Renault (2008a) for broad GMM theory under variable coefficient estimator rates that are no greater than $\sqrt{n}$.

In Section 4 we verify the major assumptions for linear-in-parameters models, and show consistency and asymptotic normality are nearly primitive properties. We then perform a monte carlo study in Section 5 to demonstrate super-$\sqrt{n}$-convergence for an autoregression with iid shocks, and the superiority of GMTTM over GMM and QML for linear and nonlinear models including AR, GARCH, IGARCH, Quadratic-ARCH, and Threshold-ARCH with Gaussian or Paretian innovations.

Fixed quantile or central order trimming, by comparison, imposes $k_{j,i,n}/n \to \lambda_{j,i} \in (0,1)$ for each equation i and tail j. This is the standard in the robust M-estimation and

[1] Other criteria for trimming exist, including trimming according to the Euclidean norm $m^{(N)}_t(\theta) := ||m_t(\theta)||$. In this case $\hat{m}_{n,t}(\theta) = m_t(\theta)I(||m_t(\theta)|| \leq m^{(N)}_{(k_n)}(\theta))$ where $k_n \to \infty$ and $k_n/n \to 0$. Simulation work reveals the latter is massively dominated by component-wise trimming when $q > 1$, irrespective of distribution symmetry.

Method of Moments literatures where symmetry is imposed $\lambda_{1,i} = \lambda_{2,i}$ (see below). In this case without further information the data generating process must be symmetric to ensure identification of θ^0. Since key asymptotic arguments in this paper exploit negligibility and degeneracy properties under tail trimming, a direct extension to fixed quantile trimming is not evident.

Finally, we do not tackle the practical problem of fractile $k_{j,i,n}$ selection since that deviates from the central theme of Gaussian asymptotics under tail trimming. Nevertheless, we provide substantial detail on reasonable rates $k_{j,i,n} \to \infty$ for augmenting efficiency (Sections 3-5).

1.2 EXTANT METHODS

The best extant theory of Minimum Distance Estimation for time series covers M-estimators, in particular QML and LAD for GARCH models and Least Trimmed Squares in the robust estimation literature. Francq and Zakoïan (2004) prove the QMLE is asymptotically normal for strong-GARCH and ARMA-GARCH under $E(\epsilon_t^4) < \infty$. See, also, Hansen and Lee (1994), Lumsdaine (1996) and Jensen and Rahbek (2004) for results covering stationary and non-stationary cases. Hall and Yao (2003) characterize non-normal QMLE limit laws for linear GARCH models with possibly infinite variance errors and rate of convergence greater than $\sqrt{n}$. The GMTTM estimator with QML $m_t(\theta)$ is asymptotically normal for arbitrarily heavy-tailed y_t and $E(\epsilon_t^2) < \infty$, but sub-$\sqrt{n}$ rate of convergence due to trimming.

Linton et al (2010) prove asymptotic normality of the log-transformed LAD estimator for non-stationary GARCH provided $E(\epsilon_t^2) < \infty$ for martingale difference ϵ_t, and $E|\epsilon_t|^p < \infty$ for some $p > 0$ for iid ϵ_t. See also Peng and Yao (2003).

Although robust estimation has a substantial history (Huber 1964, Stigler 1973, Jurecková and Sen 1996), only a few results concern fully nonlinear models with heavy tails. Most regression treatments focus on breakdown point analysis for thin tailed data with outliers under contamination (e.g. Rousseeuw 1985, Basset 1991, He et al 1996, Čižek 2005, 2008, 2009); most concern M-estimator frameworks (e.g. Čižek 2008 and the citations therein); and when trimming, truncation or weighting are employed only non-tail data quantiles are considered. Almost all uses of tail trimming appear in the robust location and central limit theory literatures, and to our knowledge extremum estimation and tail-trimming have never been combined. See Hill (2009b) for a review, and see Horowitz (1998) for a tail-trimmed covariance matrix without supporting theory.

Let $s_t(\theta) \geq 0$ denote criterion equations, for example $s_t(\theta) = |y_t - \theta' x_t|$ for LAD. Ling (2005, 2007) symmetrically weighs LAD and QML equations $\sum_{t=1}^n w_t(c)s_t(\theta)$ where $w_t(c)$ is a smooth stochastic function on the data based on some threshold c. Since $w_t(c)$ is not a function of θ the threshold c is not with respect to the criterion $s_t(\theta)$. Linear autoregressive and GARCH models are separately covered allowing $E[\epsilon_t^2] < \infty$ and $E|y_t|^p < \infty$ for some $p > 0$. We easily match this setting in Section 3.4.

Hadi and Luceno (1997) characterize the Maximum Trimmed Likelihood [MTL] estimator but do not provide a formal theory. Čižek (2005, 2008) improves the breakdown point of M-estimators by trimming the $k_n \sim \lambda n$ largest $s_t(\theta)$. Nonlinear models and models with limited dependent variables are covered, the errors are assumed to be iid with a finite variance, and asymptotic variance estimation is neglected so inference is not available. Kan and Lewbel (2007) use trimming to solve bias problems in semiparametric least squares estimation for linear truncated regression models. The data are iid with thin tails, and trimming is based on a set distance of regressor observations to their sample maximum. See also Ruppert and Carroll (1980), Rousseeuw (1985), Stromberg (1993), and Agulló et al (2008) for LTS; Neykov and Neytchev (1990) for MTL; and Basset (1991) and Tableman (1994) for Least Trimmed Absolute Deviations; and see Chen et al (2001) for robust regression based on truncated observations.

The fundamental short-comings of trimming criterion equations $s_t(\theta)$ by a fixed quantile of itself are super-$\sqrt{n}$-convergence is impossible for stationary data; asymptotic normality cannot be achieved when regressors are heavy tailed; and asymmetric models are not covered. Adaptive weighting as in Ling (2005, 2007) neglects criterion and normal equation information. See Section 2.6 for direct comparisons of the LTS and MTL estimators with GMTTM.

A few results are couched in method of moments. Čižek (2009) trims a fixed quantile of $m_t(\theta)$ for thin-tailed cross-sections under data contamination, covering limited dependent and instrumental variables. Since the quantile is fixed identification must be assumed and an efficient criterion weight does not exist. Powell (1986) and Honoré (1992) construct least squares estimator couched in the method of trimmed moments for censored linear regressions models of iid data. Ronchetti and Trojani (2001) symmetrically truncate $m_t(\theta)$ and propose a method of simulated moments to overcome bias in asymmetric models. The error distribution must therefore be known and heavy-tailed cases are ignored.

Throughout $||x||_p := (\sum_{i,j} |x_{i,j}|^p)^{1/p}$ and $||\cdot|| = ||\cdot||_2$ the Euclidean matrix norm. $(z)_+ := \max\{0, z\}$. The L_p-norm is then $(E||x_t||_p^p)^{1/p} = (\sum_{i,j} E|x_{i,j}|^p)^{1/p}$. $K > 0$ is a finite constant whose value may change from line to line; $\iota, \delta > 0$ are arbitrarily tiny constants whose values may change; and N is an arbitrary positive integer. Denote by $\stackrel{p}{\rightarrow}$ and $\stackrel{d}{\rightarrow}$ convergence in probability and in distribution, and $\rightarrow$ denotes convergence in $||\cdot||$. I_d is a d-dimensional identity matrix and $A^{1/2}$ denotes the square-root matrix for positive definite A. $U^0(\delta)$ denotes a δ-neighborhood of θ^0. Throughout $\sup_\theta = \sup_{\theta\in\Theta}$ and $\inf_\theta = \inf_{\theta\in\Theta}$.

2. TAIL-TRIMMED GMM

In this section we develop a model-free theory of GMTTM based on primitive properties of $m_t(\theta)$. We treat specific models in Sections 3 and 4.

2.1 TAIL-TRIMMING

Denote by $L_i(\theta), U_i(\theta) \in [0, \infty]$ equation specific support bounds: $-L_i(\theta) \leq m_{i,t}(\theta) \leq U_i(\theta)$ *a.s.* The problem of interest are those $m_{i,t}(\theta^0)$ with unbounded support and infinite variance. Assume by convention the first $\underline{q} \in \{1, ..., q\}$ equations are trimmed:

$$\hat{m}^*_{n,t}(\theta) = \left[m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta)\right]_{i=1}^{q} = \left[\left\{m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta)\right\}_{i=1}^{\underline{q}\prime}, \{m_{i,t}(\theta)\}_{i=\underline{q}+1}^{q\prime}\right]'$$

and assume throughout $\underline{q} \geq 1$ since otherwise the following reduces to known results.

Let positive integer sequences $\{k_{1,i,n}, k_{2,i,n} : 1 \leq i \leq \underline{q}\}$ and positive sequences of threshold functions $\{l_{i,n}(\theta), u_{i,n}(\theta) : 1 \leq i \leq \underline{q}\}$ satisfy

$$k_{j,i,n} \rightarrow \infty,\ k_{j,i,n}/n \rightarrow 0,\ \ 1 \leq k_{1,i,n} + k_{2,i,n} < n$$

$$l_{i,n}(\theta) \rightarrow L_i(\theta) \ \text{ and } \ u_{i,n}(\theta) \rightarrow U_i(\theta) \ \text{ uniformly on compact } \Theta \subset \mathbb{R}^r,$$

and uniformly on Θ (e.g. Leadbetter et al 1983: Theorem 1.7.13)

$$\frac{n}{k_{1,i,n}} P\left(m_{i,t}(\theta) < -l_{i,n}(\theta)\right) \rightarrow 1 \ \text{ and } \ \frac{n}{k_{2,i,n}} P\left(m_{i,t}(\theta) > u_{i,n}(\theta)\right) \rightarrow 1. \tag{4}$$

Thus, $l_{i,n}(\theta)$ and $u_{i,n}(\theta)$ are asymptotically the equation specific lower $k_{1,i,n}/n^{th} \rightarrow 0$ and upper $k_{2,i,n}/n^{th} \rightarrow 0$ tail quantiles. The threshold sequences $\{l_{i,n}(\theta), u_{i,n}(\theta)\}$ are not unique for given fractiles $\{k_{1,i,n}, k_{2,i,n}\}$ since $\{l_{i,n}(\theta) \pm K_{l,n}, u_{i,n}(\theta) \pm K_{u,n}\}$ satisfy (4) for any sequences $K_{l,n} = o(l_{i,n}(\theta))$ and $K_{u,n} = o(u_{i,n}(\theta))$ uniformly on Θ.

The practice of GMTTM involves $\hat{m}^*_{n,t}(\theta)$ in (2), but theory centers around deterministic trimming:

$$m^*_{i,n,t}(\theta) := m_{i,t}(\theta) \times I\left(-l_{i,n}(\theta) \leq m_{i,t}(\theta) \leq u_{i,n}(\theta)\right) \tag{5}$$
$$= m_{i,t}(\theta) \times I_{i,n,t}(\theta) \quad : \quad 1 \leq i \leq \underline{q}$$

$$m^*_{n,t}(\theta) \quad = \left[m_{i,t}(\theta) \times I_{i,n,t}(\theta)\right]_{i=1}^{q} \text{ where } I_{j,n,t}(\theta) = 1 \text{ for } \underline{q}+1 \leq j \leq q.$$

Although $m_t(\theta)$ identifies θ^0, we can only assume $m^*_{n,t}(\theta)$ *eventually* identifies θ^0:

$$E\left[m^*_{n,t}(\theta^0)\right] \to 0.$$

This is easily guaranteed for arbitrarily many threshold sequences $\{l_{i,n}(\theta^0), u_{i,n}(\theta^0)\}$ since tail trimming is negligible. It is interesting to note "eventual identification" runs contrary to weak and nearly weak identification where information *vanishes* at some rate (e.g. Stock and Wright 2000, Antoine and Renault 2008b). Here, information *amasses* at some rate to be made precise below. If the DGP of $\{m_t(\theta^0)\}$ is symmetric then $E[m^*_{n,t}(\theta^0)]$ $= 0$ for any thresholds $l_{i,n}(\theta^0) = u_{i,n}(\theta^0)$ and fractiles $k_{1,i,n} = k_{2,i,n}$. This applies to linear-in-parameters models with symmetric shocks like autoregressions, GARCH, and so on.

Write compactly throughout

$$c_{i,n}(\theta) := \max\{l_{i,n}(\theta), u_{i,n}(\theta)\} \quad \text{and} \quad c_n(\theta) = \max_{1\leq i\leq \underline{q}}\{c_{i,n}(\theta)\}$$
$$k_{i,n} = \max\{k_{1,i,n}, k_{2,i,n}\} \quad \text{and} \quad k_n = \max_{1\leq i\leq \underline{q}}\{k_{i,n}\}$$
$$\{l_{i,n}, u_{i,n}, c_{i,n}\} = \left\{l_{i,n}\left(\theta^0\right), u_{i,n}\left(\theta^0\right), c_{i,n}\left(\theta^0\right)\right\}.$$

2.2 ASSUMPTIONS

Let $\{\Upsilon_n\}$ be a sequence of positive definite weight matrices $\Upsilon_n \in \mathbb{R}^{q\times q}$. The sample criterion is

$$\hat{Q}_n(\theta) := \hat{m}^*_n(\theta)' \times \hat{\Upsilon}_n \times \hat{m}^*_n(\theta), \text{ where } \hat{m}^*_n(\theta) := \frac{1}{n}\sum_{t=1}^{n}\hat{m}^*_{n,t}(\theta) \text{ and } \hat{\Upsilon}_n \in \mathbb{R}^{q\times q},$$

hence the GMTTME solves

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\text{argmin}}\{\hat{Q}_n(\theta)\}.$$

Under the identification and smoothness conditions detailed below, $\hat{\theta}_n$ exists and is unique.

Asymptotic arguments require the following constructions: the trimmed equation covariance matrix and moment envelope

$$\Sigma_n(\theta) := E\left[m^*_{n,t}(\theta)\, m^*_{n,t}(\theta)'\right], \;\; \Sigma_n := \Sigma_n(\theta^0), \text{ and } \mathfrak{m}_n = \sup_{\theta} E\left[\left\|m^*_{n,t}(\theta)\right\|\right];$$

the population and sample Jacobia

$$J_n(\theta) := \frac{\partial}{\partial\theta}E\left[m^*_{n,t}(\theta)\right] \in \mathbb{R}^{q\times r} \quad \text{and} \quad J_n = J_n(\theta^0)$$
$$J^*_{n,t}(\theta) := \left[\frac{\partial}{\partial\theta}m_{i,t}(\theta) \times I_{i,n,t}(\theta)\right]_{i=1}^{q} \quad \text{and} \quad J^*_n(\theta) := \frac{1}{n}\sum_{t=1}^{n}J^*_{n,t}(\theta)$$
$$\hat{J}^*_{n,t}(\theta) := \left[\frac{\partial}{\partial\theta}m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta)\right]_{i=1}^{q} \quad \text{and} \quad \hat{J}^*_n(\theta) := \frac{1}{n}\sum_{t=1}^{n}\hat{J}^*_{n,t}(\theta);$$

and the Hessian and scale

$$H_n(\theta) := J_n(\theta)'\Upsilon_n J_n(\theta) \in \mathbb{R}^{r\times r} \quad \text{and} \quad H_n := H_n(\theta^0)$$

$$V_n(\theta) := n \times H_n(\theta)\left[J_n'(\theta)\,\Upsilon_n\Sigma_n(\theta)\,\Upsilon_n J_n(\theta)\right]^{-1} H_n(\theta) \quad \text{and} \quad V_n := V_n(\theta^0).$$

Three sets of assumptions ensure identification for θ^0; $\hat{\theta}_n$ can be expressed as a linear function of $\sum_{t=1}^n \hat{m}^*_{n,t}(\theta^0)$; $\sum_{t=1}^n \hat{m}^*_{n,t}(\theta)$ is sufficiently close to $\sum_{t=1}^n m^*_{n,t}(\theta)$ uniformly on Θ; $\sum_{t=1}^n m^*_{n,t}(\theta^0)$ is asymptotically normal; and $\hat{J}^*_n(\tilde{\theta}_n)$ is consistent for J_n for any consistent plug-in $\tilde{\theta}_n$. Most are versions of standard regulatory conditions contoured to heavy tailed data under tail trimming. The remaining are easily verified. See Section 4.

Let $\{\Im_t\}$ be any sequence of increasing σ-fields adapted to $\{m_t(\theta)\}$, $\theta \in \Theta$, where $\{\Im_t\}$ itself does not depend on θ. The first set characterizes matrix norms, weight limits and covariance definiteness. Denote by $[\underline{\lambda}_n(\theta)]_{i=1}^q$ the minimum eigenvalue of $\Sigma_n(\theta)$ for each n and θ.

M1 (*weight*). *Υ_n is positive definite for each $n \geq N$; and $||\hat{\Upsilon}_n - \Upsilon_n|| \xrightarrow{p} 0$ and $||\Upsilon_n - \Upsilon_0|| \to 0$ for some positive definite Υ_0, $0 < ||\Upsilon_0|| < \infty$.*

M2 (*scale*). *$Kn^{1/2}||J_n|| \times ||\Sigma_n^{-1}||^{1/2} \geq ||V_n^{1/2}|| \to \infty$.*

M3 (*covariance*). *a. $\sup_\theta\{||\Sigma_n^{1/2}(\theta)||\} = o(n^{1/2})$; b. $\liminf_{n\geq N}\inf_\theta\{\underline{\lambda}_n(\theta)\} > 0$.*

Remark 3: M1 is standard. M2 is used solely to simplify bounding arguments and holds in the efficient weight case (see Section 2.3). M3.a ensures infinite variance equations $m_{i,t}(\theta^0)$ are trimmed since implicitly $||\Sigma_n(\theta)|| < \infty$ for each n, and trivially holds if $m_{i,t}(\theta^0)$ has regularly varying distribution tails. M3.b imposes positive definiteness for sufficiently large $n \geq N$ since trimming can technically render $\lambda_{n,i}(\theta) = 0$ for some i and finite n, and possibly all i (e.g. $\Sigma_n = 0$ a zero matrix for some finite n).

The second set promotes local identification of θ^0.

I1 (*identification by* $m_t(\theta)$). *$\{m_t(\theta), \Im_t\}$ forms an adapted martingale difference sequence* if and only if *$\theta = \theta^0$, a unique interior point of compact $\Theta \subset \mathbb{R}^r$.*

I2 (*identification by* $m^*_{n,t}(\theta)$). *$E[m^*_{n,t}(\theta^0)] = O(||\Sigma_n^{-1/2}||^{-1}/n^{1/2})$.*

I3 (*smoothness*). *$\inf_{n\geq N}\inf_{||\theta-\theta^0||>\delta}\{\mathfrak{m}_n^{-1}||E[m^*_{n,t}(\theta)]||\} > 0$ for tiny $\delta > 0$, and $\liminf_{n\geq N}\{\mathfrak{m}_n\} > 0$ for some $N \geq 1$.*

I4 (*orthogonality*). *$\max_{s\neq t}\{||E[m^*_{n,s}(\theta^0)m^*_{n,t}(\theta^0)']||\} = o(||\Sigma_n^{-1/2}||^{-2}/n)$.*

Remark 1: I1 is standard although hardly innocuous: $E[m_t(\theta^0)|\Im_{t-1}]$ exists if $m_t(\theta^0)$ is integrable. This necessarily limits allowed tail thickness depending on equation structure. See Sections 3 and 4 for examples.

Remark 2: Both $E[m^*_{n,t}(\theta^0)] \to 0$ and $E[m^*_{n,s}(\theta^0)m^*_{n,t}(\theta^0)'] \to 0$ are assured by I1 and Lebesgue's dominated convergence theorem. Conditions I2 and I4 merely ensure $m^*_{n,t}(\theta^0)$ behaves like a martingale difference *sufficiently fast* as $n \to \infty$. If the joint distributions of $\{m^*_{i,n,t}(\theta^0), m^*_{j,n,t}(\theta^0)\}$ are symmetric or independent then $E[m^*_{i,n,t}(\theta^0)] = 0$ and $E[m^*_{n,s}(\theta^0)m^*_{n,t}(\theta^0)'] = 0$ under I1 for any thresholds $l_{i,n}(\theta^0) = u_{i,n}(\theta^0)$. Otherwise by I2 and I4 $||\Sigma_n^{-1/2}||^{-1}/n^{1/2} \leq K||\Sigma_n^{1/2}||/n^{1/2} = o(1)$ under M3.a suffices. In turn $||\Sigma_n^{1/2}||/n^{1/2} = o(1)$ holds trivially in thin-tailed cases, and for equations with Paretian tails under intermediate order tail-trimming.

Remark 3: Versions of smoothness I3 are standard for consistency (Huber 1967, Pakes and Pollard 1989, Newey and McFadden 1994). The envelope scale $\mathfrak{m}_n$ is required since $m_t(\theta)$ need not be integrable on Θ-*a.e.* in heavy-tailed cases, whereas $E[m^*_{n,t}(\theta)]/\mathfrak{m}_n$ is always well defined. Consider an AR(1) $y_t = \theta^0 y_{t-1} + \epsilon_t$ with $|\theta^0| < 1$, $\Im_t = \sigma(y_\tau : \tau \leq t)$, martingale difference innovations $E[\epsilon_t|\Im_{t-1}] = 0$ with infinite variance $E[\epsilon_t^2] = \infty$, and one equation $m_t(\theta) = (y_t - \theta y_{t-1})y_{t-1}$. Then $E[m_t(\theta^0)|\Im_{t-1}] = 0$ *a.s.* hence $E[m_t(\theta^0)] = 0$, but in general $m_t(\theta) = -(\theta - \theta^0) \times y^2_{t-1}$is non-integrable for any coefficient $\theta \neq \theta^0$. This matters for a proof of consistency $\hat{\theta}_n \xrightarrow{p} \theta^0$ since that requires a uniform law of large numbers for $m^*_{n,t}(\theta)$ by well known arguments (e.g. Pakes and Pollard (1989)[2].

Remark 4: If $m_t(\theta)$ is uniformly integrable on Θ-a.e. then I3 reduces to $\inf_{n\geq N} \inf_{||\theta-\theta^0||>\delta}\{||E[m^*_{n,t}(\theta)]||\} > 0$. Consider a stationary AR(1) $y_t = \theta^0 y_{t-1} + \epsilon_t$ with ARCH(1) error $\epsilon_t = (\alpha + \beta\epsilon^2_{t-1})^{1/2}u_t$, $u_t \overset{iid}{\sim} N(0,1)$ and one equation $m_t(\theta) = (y_t - \theta y_{t-1})y_{t-1}$. If ϵ_t has a finite variance and infinite kurtosis then $m_t(\theta) = -(\theta - \theta^0) \times y^2_{t-1}$ is integrable on Θ-*a.e.* but $E[m^2_t(\theta^0)]$ does not exist.

The last set concerns properties of the equations $m_t(\theta)$ and the random Jacobian matrices $J^*_{n,t}$ and $\hat{J}^*_{n,t}$.

D1 (*distribution continuity*). *The marginal distributions of* $m_t(\theta)$ *have support* $(-\infty, \infty)$ *and are absolutely continuous with respect to Lebesgue measure on* Θ.

D2 (*differentiability*). $m_t(\theta)$ *is continuous and differentiable on* Θ-*a.e.*

D3 (*mixing*). $\{m_t(\theta)\}$ *is strictly stationary, geometrically* β*-mixing (absolutely regular):* $\beta_l :=$
$\sup_{A\subset\Im^{+\infty}_{t+l}} E|P(A|\Im^t_{-\infty}) - P(A)| = o(\rho^l)$ *for some* $\rho \in (0,1)$.

D4 (*envelope bounds*). $\sup_\theta ||m_t(\theta)||$ *and* $\sup_\theta ||(\partial/\partial\theta)m_t(\theta)||\}$ *are* L_ι*-bounded.*

D5 (*thresholds and fractiles*).

i. $\sup_\theta |(n/k_{1,i,n})P(m_{i,t}(\theta) < -l_{i,n}(\theta)) - 1| = O(1/k^{1/2}_{1,i,n})$ *and* $\sup_\theta |(n/k_{2,i,n})P(m_{i,t}(\theta) > u_{i,n}(\theta)) - 1| = O(1/k^{1/2}_{2,i,n})$.

ii. $\inf_\theta\{c_n(\theta)\} \to \infty$ *and* $\sup_\theta c_n(\theta) = o(n^{1/2}\sup_\theta ||\Sigma_n^{-1/2}(\theta)||^{-1})$.

D6 (*Jacobia*).

i. $\sup_\theta ||J_n(\theta)|| < \infty$ *for each n;* $\{J_n(\theta),\ J^*_n(\theta),\ E[J^*_{n,t}(\theta)],\ \hat{J}^*_n(\theta),\ E[J^*_{n,t}(\theta)]\}$ *have full column rank for each* $n \geq N$.

ii. $\sup_{n\geq N} \inf_{\theta\in U^0(\delta_n)}\{J^*_n(\theta)\} > 0$ *and* $\sup_{\theta\in U^0(\delta_n)}\{||J^*_n(\theta) - J^*_n||\} = o_p(||J_n||)$ *for any* $\delta_n \to 0$.

D7 (*indicator class*). $\{I_{i,n,t}(\theta) : \theta \in \Theta\}$ *form Vapnik–Chervonenkis [VC] classes of functions.*

Remark 1: Distribution continuity D1 and equation differentiability D2 reduce generality, but simplify key uniform arguments since trimming adds substantial complexity. In regression models D1 requires at least idiosyncratic shocks to be continuously distributed.

Remark 2: Mixing D3 promotes uniform laws for $m^*_{n,t}(\theta)$ and $\hat{m}^*_{n,t}(\theta) - m^*_{n,t}(\theta)$, while geometric decay keeps notation simple. Nevertheless, many nonlinear AR-nonlinear

[2] Although the untrimmed equations $m_t(\theta)$ need not be integrable for arbitrary θ, we show $\sup_\theta ||1/n\sum^n_{t=1}\{m^*_{n,t}(\theta) - E[m^*_{n,t}(\theta)]\}|| = o_p(\mathfrak{m}_n)$ in Lemma C.3 in Appendix C, which suffices for consistency.

GARCH models are covered since geometric ergodicity implies β-mixing (An and Huan 1996, Carrasco and Chen 2002, Meitz and Saikonnen 2008).

Remark 3: Since by construction $(n/k_{1,i,n})P(|m_{i,t}(\theta) < -l_{i,n}(\theta)) \rightarrow 1$ and $(n/k_{2,i,n})P(|m_{i,t}(\theta) > u_{i,n}(\theta)) \rightarrow 1$ the D5.i probability orders merely sharpen the rates. This is required for uniform asymptotics concerning trimming indicators $I_{i,n,t}(\theta)$. The rates hold for a large array of probability tails that satisfy second order regular variation or slow variation with remainder (e.g. Haeusler and Teugels 1985, Goldie and Smith 1987, Hill 2009a).

Remark 4: The D5.ii threshold bound $c_n(\theta) = o(n^{1/2}||\Sigma_n^{-1/2}(\theta)||^{-1})$ ensures sufficiently many equations are trimmed for weak limit theory in the presence of heavy tails. The result $\max_{1\leq t\leq n}\{||\Sigma_n^{-1/2}(\theta)m_{n,t}^*(\theta^0)||\} = o_p(n^{1/2})$ matches the relative stability property of maxima of uniformly square integrable weakly dependent sequences, cf. Leadbetter et al (1983) and Naveau (2003), and aligns with a necessary and sufficient condition for the distribution limit of a sum of an iid array to be Gaussian (e.g. Kallenberg 2002: Theorem 5.15). The D5.ii bound is irrelevant if equation variances are finite, and holds for any fractiles $\{k_{1,i,n}, k_{2,i,n}\}$ if any equation has a Paretian tail. See Section 4.1.

Remark 5: The D4 moment bounds, D6 Jacobia properties and D7 indicator class help prove $1/n\sum_{t=1}^n\{\hat{m}_{n,t}^*(\theta) - m_{n,t}^*(\theta)\} = o_p(1)$ uniformly on Θ, required for consistency. See Vapnik and Chervonenkis (1971), Pollard (1984), Pakes and Pollard (1989), and van der Vaart and Wellner (1994) for a definition of the VC function class[3]. It suffices for $\{m_{i,t}(\theta) : \theta \in \Theta\}$ and $\{c_{i,n}(\theta) : \theta \in \Theta\}$ to form VC classes (van der Vaart and Wellner 1994: Lemma 2.6.18) which holds, for example, for finite dimensional functions (e.g. Pakes and Pollard 1989: Lemma 2.4), covering at least $m_t(\theta)$ polynomial in θ, hence dynamic linear regressions and ARCH.

2.3 MAIN RESULTS

The main results follow: $\hat{\theta}_n$ is consistent for θ^0 and asymptotically normal.

THEOREM 2.1 *Under D1-D7, I1-I3 and M1-M3* $\hat{\theta}_n \xrightarrow{p} \theta^0$.

The rate $V_n^{1/2}(\hat{\theta}_n - \theta^0) = O_p(I_r)$ can similarly be shown from first principles (Hill and Renault 2010: Theorem D.1). We do not present the result here since asymptotic normality follows from nearly the same set of assumptions.

THEOREM 2.2 *Under D1-D7, I1-I4 and M1-M3* $V_n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, I_r)$.

Remark 1: An "optimal" GMTTM weight sequence $\{\Upsilon_n\}$ in the sense of asymptotic efficiency is $\{\Sigma_n^{-1}/||\Sigma_n^{-1}||\}$ due to the quadratic form $V_n = nH_n(J_n'\Upsilon_n\Sigma_n\Upsilon_nJ_n)^{-1}H_n$ (Hansen 1982; Newey and MacFadden 1994: p. 2164). In this case

$$V_n = n\left(J_n'\Sigma_n^{-1}J_n\right)$$

hence scale bound M2 holds automatically. It is nevertheless not obvious that the trimming fractiles $\{k_{1,i,n}, k_{2,i,n}\}$ cannot be set to augment efficiency. We will see for linear-in-parameters models in Section 3 that, depending on the model, minimal trimming ($k_{j,i,n} \rightarrow \infty$ very slowly) or maximal trimming ($k_{1,i,n} \rightarrow \infty$ very quickly) is always optimal when the efficient weight $\Upsilon_n = \Sigma_n^{-1}/||\Sigma_n^{-1}||$ is use.

Remark 2: The existence of an efficient weight $\Upsilon_n = \Sigma_n^{-1}/||\Sigma_n^{-1}||$ is non-trivial since a symmetric variance form does not arise under *fixed* quantile trimming. In this case

[3] The VC class $\mathfrak{F}$ of functions $f \in \mathfrak{F}$ satisfies a uniform entropy or bracketing number bound required for $\mathfrak{F}$ to be P-Donsker (i.e. for empirical measures to satisfy a uniform central limit theorem on $\mathfrak{F}$). The entropy of a class $\mathfrak{F}$ quantifies smoothness. We refer the reader to Pollard (1984) and van der Vaart and Wellner (1994).

J_n has two components that enter V_n asymmetrically as $n \to \infty$, so an optimal weight does not exist (Čižek 2009). Under *tail* trimming, however, each $J_{i,j,n}$ also decomposes into two components $E[(\partial/\partial\theta_j)m_{i,t}(\theta)|_{\theta^0} \times I_{i,n,t}(\theta^0)] + (\partial/\partial\theta_j)E[m_{i,t}(\theta^0) \times I_{i,n,t}(\theta)]|_{\theta^0}$. The latter is asymptotically dominated by the former due to negligibility $I_{i,n,t}(\theta) \to 1$ *a.s.* so $J_{i,j,n} = E[(\partial/\partial\theta_j)m_{i,t}(\theta)|_{\theta^0} \times I_{i,n,t}(\theta^0)] \times (1 + o(1))$. See Lemma C.1 in Appendix C.

If the Jacobian and covariance are asymptotically bounded $J_n \to J$ and $\Sigma_n \to \Sigma$ then the rate is exactly $\sqrt{n}$ since $V_n^{1/2} \sim n^{1/2}V^{1/2}$ for some positive definite $V \in \mathbb{R}^{r\times r}$. This holds for any stationary DGP for which the conventional GMM estimator is asymptotically normal, so tail trimming is always a safe practice. Otherwise the rates need not be homogeneous over $\hat{\theta}_{i,n}$ and may be greater or less than $\sqrt{n}$. See Section 3.

LEMMA 2.3 *If* $\limsup_{n\geq 1} ||E[m^*_{n,t}(\theta^0)m^*_{n,t}(\theta^0)']|| < \infty$ *and* $\limsup_{n\geq 1} ||(\partial/\partial\theta)E[m^*_{n,t}(\theta)|_{\theta^0}]|| < \infty$ *then the rate of convergence of each* $\hat{\theta}_{i,n}$ *is* $\sqrt{n}$.

2.4 COVARIANCE AND JACOBIAN MATRIX ESTIMATION

In lieu of martingale difference I1 and trimmed equation orthogonality I4, a natural estimator of Σ_n is

$$\hat{\Sigma}_n(\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n} \hat{m}^*_{n,t}(\tilde{\theta}_n)\hat{m}^*_{n,t}(\tilde{\theta}_n)'$$

for some consistent plug-in $\tilde{\theta}_n$. Since $\hat{\Sigma}_n(\tilde{\theta}_n)$ may itself be used for GMTTM estimation, in practice $\tilde{\theta}_n$ need not be the final GMTTME $\hat{\theta}_n$. Candidate plug-ins include a one-step GMTTME (e.g. naïve $\hat{\Upsilon}_n = I_q$), but also untrimmed estimators like the LSE, GMME and QMLE since in general they converge faster than the GMTTME. See Section 3.

LEMMA 2.4 *Under D1-D6.i, D7, I1, I2, I4 and M2* $||\Sigma_n^{-1}\hat{\Sigma}_n(\tilde{\theta}_n) - I_q|| = o_p(1))$ *for any* $\tilde{\theta}_n = \theta^0 + O_p(||V_n^{1/2}||^{-1})$.

Tail trimming implies the Jacobian J_n is proportional to $E[J^*_{n,t}]$, cf. Lemma C.1 in Appendix C. Due to its simple form consistency $\hat{J}^*_n(\tilde{\theta}_n) = E[J^*_{n,t}] \times (1 + o_p(1))$ follows for any consistent $\tilde{\theta}_n$.

LEMMA 2.5 *Under D1-D7 and M1-M3* $J^*_n(\tilde{\theta}_n) = J_n \times (1 + o_p(1))$ *and* $\hat{J}^*_n(\tilde{\theta}_n) = J_n \times (1 + o_p(1))$ *for any* $||\tilde{\theta}_n - \theta^0|| \xrightarrow{p} 0$.

The covariance matrix V_n^{-1} is estimated by

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

where $\hat{H}_n(\theta) = \hat{J}^*_n(\theta)'\hat{\Sigma}_n^{-1}(\theta)\hat{J}^*_n(\theta)$. Since trivially $\hat{\theta}_n = \theta^0 + O_p(||V_n^{1/2}||^{-1})$ by Theorem 2.2 the scale is consistent by Lemmas 2.4 an 2.5.

THEOREM 2.6 *Under D1-D7, I1, I2, I4 and M1-M3* $\hat{V}_n(\hat{\theta}_n) = V_n \times (1 + o_p(1))$.

2.5 ROBUST M-ESTIMATORS

We now discuss why trimming M-estimator criterion equations may fail to promote asymptotic normality.

Least Trimmed Squares: Consider a linear model with least squares criterion

$$y_t = \theta^{0\prime} x_t + \epsilon_t \quad \text{with} \quad s_t(\theta) := \left(y_t - \theta' x_t\right)^2 .$$

Assume ϵ_t is zero-mean with distribution function $F_\epsilon(\epsilon) := P(\epsilon_t \leq \epsilon)$, and two-tailed inverse $F_{|\epsilon|}^{-1}(\lambda) := \inf\{\epsilon \geq 0 : P(|\epsilon_t| \leq \epsilon) \leq \lambda\}$. The fixed quantile LTSE is (Ruppert and Carrol 1980, Rousseeuw 1985, Čižek 2008)

$$\tilde{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}} \left\{ \frac{1}{n} \sum_{t=1}^{n} s_t(\theta) \times I\left(s_t(\theta) \leq s_{([n\lambda])}(\theta)\right) \right\},\ \lambda \in (0,1) .$$

If the distribution governing $s_t(\theta)$ is absolutely continuous on Θ-*a.e.*, $\{x_t, \epsilon_t\}$ have finite variance marginal distributions, $\{\epsilon_t, x_t\}$ are geometrically β-mixing, and $\tilde{J}(\lambda) := -E[x_t x_t' I(|\epsilon_t| \leq F_{|\epsilon|}^{-1}(\lambda))]$ is non-singular, then for the given linear DGP

$$\sqrt{n}\left(\tilde{\theta}_n - \theta^0\right) = \tilde{J}(\lambda)^{-1} \frac{1}{\sqrt{n}} \sum \epsilon_t x_t I\left(|\epsilon_t| \leq F_{|\epsilon|}^{-1}(\lambda)\right) + o_p(1) \xrightarrow{d} N\left(0, \tilde{V}^{-1}(\lambda)\right),$$

where $\tilde{V}(\lambda) = \tilde{J}(\lambda)'\tilde{\Sigma}^{-1}\tilde{J}(\lambda)$ and $\tilde{\Sigma}(\lambda) := E[\epsilon_t^2 x_t x_t' I(|\epsilon_t| \leq F_{|\epsilon|}^{-1}(\lambda))]$. See Čižek (2005, 2008). Clearly if the error ϵ_t is independent of x_t and any stochastic element of x_t has an infinite variance then $1/\sqrt{n}\sum \epsilon_t x_t I(\epsilon_t^2 \leq \epsilon_{([n\lambda])}^2)$ does not have a Gaussian limit *even under fixed quantile trimming*, and $\tilde{\Sigma}(\lambda)$ and $\tilde{J}(\lambda)$ do not exist. The same conclusion applies to Least Absolute Trimmed Deviations. Since the object that governs asymptotics is the gradient $(\partial/\partial\theta)s_t(\theta)|_{\theta^0}$, *its* components $\epsilon_t x_{i,t}$ must be trimmed to ensure asymptotic normality, not simply ϵ_t.

Quasi-Maximum Trimmed Likelihood: Consider an ARCH(1) $y_t = h_t\epsilon_t$, $\epsilon_t \overset{iid}{\sim} (0,1)$, $h_t^2(\theta) = \alpha + \beta y_{t-1}^2$, $(\alpha, \beta) \geq 0$ with QML criterion equations $s_t(\theta) = \ln h_t^2(\theta) + y_t^2/h_t^2(\theta)$. See Neykov and Neytchev (1990) and Čižek (2008) for Maximum Trimmed Likelihood of models of the conditional mean.

Since a standard question is whether a conditional heteroscedastic effect exists, suppose not for simplicity: $\beta^0 = 0$. If the distribution governing ϵ_t is absolutely continuous then by Lemma 2.1 of Čižek (2008) $g_{n,t}(\theta) := (\partial/\partial\theta)s_{n,t}(\theta) = (\partial/\partial\theta)s_t(\theta) \times I(s_t(\theta) \leq F_{s(\theta)}^{-1}(\lambda))$ *a.s.* on Θ-*a.e.* By direct computation it follows under $\beta^0 = 0$

$$g_{n,t}\left(\theta^0\right) = -\left(\epsilon_t^2 - 1\right)\left[1, y_{t-1}^2\right]' \times I\left(|\epsilon_t| \leq F_{|\epsilon|}^{-1}(\lambda)\right).$$

Now exploit independence to deduce

$$E\left[g_{2,n,t}^2\left(\theta^0\right)\right] = E\left[\left(\epsilon_t^2 - 1\right)^2 I\left(|\epsilon_t| \leq F_{|\epsilon|}^{-1}(\lambda)\right)\right] \times E\left[y_{t-1}^4\right].$$

Since $\beta^0 = 0$ we know y_t has an unbounded fourth moment $E[y_t^4] = \infty$ *if and only if* $E[\epsilon_t^4] = \infty$. The QMTL Jacobian is unbounded and by asymptotic linearity and independence between ϵ_t and y_{t-1}, the QMTLE is not asymptotically normal.

Adaptive M-Estimation: Ling's (2005, 2007) smooth symmetrically weighed LAD and QML criteria work like smoothed trimming. But theory is only delivered for symmetric DGP's, and only fixed quantiles of the data y_t are considered for the weight function. Although heavy-tails are allowed super-$\sqrt{n}$-convergence cannot be achieved due to the weight structure.

3. CONVERGENCE RATE FOR HEAVY-TAILED DATA Consider the efficient weight $\Upsilon_n = \Sigma_n^{-1}/||\Sigma_n^{-1}||$ for brevity. By positive definiteness and the Cauchy-Schwartz inequality we can define diagonal matrices $\Gamma_n \in \mathbb{R}^{q\times q}$ with

$$\Gamma_{i,i,n} = \Sigma_{i,i,n}^{-1/2} = \left(E[\left(m_{i,n,t}^{*}\left(\theta^0\right)\right)^2]\right)^{-1/2} \ : \ \Gamma_n^{-1}\Sigma_n\Gamma_n^{-1} \to \Sigma \text{ a positive definite matrix.}$$

Now write

$$V_n = n \times \left(\Gamma_n^{-1}J_n\right)' \times \Sigma^{-1} \times \left(\Gamma_n^{-1}J_n\right) \times (1 + o(1)) \quad \text{and} \quad \Sigma^{-1} = [\sigma^{i,j}]_{i,j=1}^{q}.$$

Assume $\sigma^{i,j} \neq 0 \ \forall i,j$, to simplify exposition. The component-wise rates n_{θ_i} are

$$n_{\theta_i} = V_{i,i,n}^{1/2} = Kn^{1/2} \times \left[\sum_{l_1,l_2=1}^{q} \sigma^{l_1,l_2}\Gamma_{l_1,l_1,n}^{-1}\Gamma_{l_2,l_2,n}^{-1}J_{l_1,i,n}J_{l_2,i,n}\right]^{1/2}. \tag{6}$$

Textbook intuition explains n_{θ_i}. Holding everything else constant, if $\Gamma_{i,i,n} = (E[(m_{i,n,t}^{*}\left(\theta^0\right))^2])^{1/2} \to \infty$ due to heavy-tailed errors then $n_{\theta_i} \to \infty$ slowly: sharp estimates are more difficult to obtain from models with disproportionately dispersive errors. If $|J_{l,i,n}| = |E[(\partial/\partial\theta_i)m_{l,t}(\theta)|_{\theta^0}I_{l,n,t}(\theta^0)]| \to \infty$ due to heavy tailed regressors then $n_{\theta_i} \to \infty$ quickly: sharpness improves with regressor dispersion and association. If both error and regressor are heavy-tailed and exhibit feedback then J_n may be overwhelmed by Γ_n. In this section we inspect the gamut of such cases.

In order to characterize Γ_n and J_n we consider dynamic linear regression and ARCH models where all equations are symmetrically trimmed with the same fractiles $k_{j,i,n} = k_n$ for simplicity. Throughout $\{\epsilon_t\}$ is iid L_p-bounded, $p > 0$, with an absolutely continuous distribution on $\mathbb{R}$-*a.e.*, symmetric about 0.

3.1 DYNAMIC REGRESSION WITH IID ERRORS

Consider a stationary dynamic linear regression with an intercept

$$y_t = \theta^{0\prime}x_t + \epsilon_t, \ x_{1,t} = 1, \ \ x_t \in \mathbb{R}^r \ \text{ with } \ m_t(\theta) = (y_t - \theta x_t')\, x_t,$$

where ϵ_t and x_t are mutually independent. Assume stochastic $x_{i,t}$ are measurable with $\mathbb{R}$-*a.e.* continuous, stationary, symmetric distributions. Independence rules out random volatility errors: see Sections 3.2-3.3 for this case.

Define the moment suprema of $z_t \in \{\epsilon_t, x_{i,t}\}$,

$$\kappa_z := \sup\left\{\alpha > 0 : E\left|z_t\right|^{\alpha} < \infty\right\} > 1,$$

where $\kappa_z = \infty$ is possible (e.g. uniform, normal, or bounded support). Identification requires integrability of $\epsilon_t x_{i,t}$ so we assume each $\kappa_z > 1$. If any $z_t \in \{\epsilon_t, x_{i,t}\}$ has an infinite variance $\kappa_z \in (1,2]$ then assume the distribution tail is Paretian (e.g. Resnick 1987):

$$P\left(|z_t| > z\right) = d_z z^{-\kappa_z}\left(1 + o\left(1\right)\right) \text{ with indices } \ \kappa_z \in \{\kappa_\epsilon, \kappa_i\} \in (1,2]. \tag{7}$$

Such z_t satisfy the Feller (1971: Theorem 1 in IX.8) property:

$$E\left[z_t^2 I\left(|z_t| \leq c\right)\right] \sim Kc^2 P\left(|z_t| > c\right) \text{ as } c \to \infty. \tag{FE}$$

Heavy-tailed convolutions $m_{i,t}(\theta^0) = \epsilon_t x_{i,t}$ also satisfy (7) with index $\kappa_{\epsilon,i} := \min\{\kappa_\epsilon, \kappa_i\}$ (Cline 1986), so the thresholds $c_{i,n} = K(n/k_n)^{1/\kappa_{\epsilon,i}}$ by the construction of $c_{i,n}$ and k_n.

Define

$$a^*_{\epsilon,(i)} := \min_{j \notin \{1,i\}} \{1/\kappa_{\epsilon,j} + (1 - 1/\kappa_j)\, \kappa_i/\kappa_{\epsilon,i}\}$$

$$A_n = (n/k_n)^{1/2+1/\kappa_{\epsilon,i}+\kappa_i/\kappa_{\epsilon,i}-2a^*_{\epsilon,(i)}}$$

$$B_n = \max_{j\neq i:\kappa_j\leq 2} \left\{(n/k_n)^{2/\kappa_j-1}\right\}$$

$$C_n = \max_{j\neq i:\kappa_j\leq 2} \left\{(n/k_n)^{2-2/\kappa_j-2/\kappa_i+2\kappa^{-1}_{\epsilon,i}(\kappa_i/\kappa_j+1-\kappa_i)}\right\}.$$

Note $\kappa_{\epsilon,1} = \kappa_\epsilon$ since $x_{1,t} = 1$, $a^*_{\epsilon,(i)} = \kappa_i/\kappa_{\epsilon,i}$ if there is only one stochastic regressor and $a^*_{\epsilon,(i)}$ is not defined if there is only an intercept.

LEMMA 3.1 (ARX with IID Error)

a. Let $\max\{\kappa_\epsilon, \kappa_2, ..., \kappa_r\} \leq 2$. *Each* $\Gamma_{i,i,n} = (n/k_n)^{1/\kappa_{\epsilon,i}-1/2}$ *and* $J_{i,j,n} \sim -E[x_{i,t}x_{j,t}I(|\epsilon_t x_{j,t}| \leq c_{j,n})]$. *For stochastic* $\{x_{i,t}, x_{j,t}\}$ *in general* $J_{i,i,n} \sim K(n/k_n)^{\kappa^{-1}_{\epsilon,i}(2-\kappa_i)}$, $J_{i,j,n} = O((n/k_n)^{\kappa^{-1}_{\epsilon,j}(\kappa_j/\kappa_i+1-\kappa_j)})$ $\forall i \neq j$, *and* $J_{i,j,n} \sim K$ *if* $x_{i,t}$ *is independent of* $x_{j,t}$. *Hence the slope rates are*

$$n_{\theta_i} \sim Kn^{1/2}\,(n/k_n)^{1/2-\kappa_i/\kappa_{\epsilon,i}+1/\kappa_{\epsilon,i}}\,[K + O\,(A_n)]^{1/2}\,,\ i = 2, ..., r.$$

Further $J_{1,1,n} = -1+o(1)$ *and* $J_{i,1,n}, J_{1,i,n} = O(1) \times (1 + o(1))$, *hence the intercept rate is*

$$n_{\theta_1} = Kn^{1/2} \times K(k_n/n)^{1/\kappa_\epsilon-1/2}\,(1 + O(1))\,.$$

b. Let $\kappa_\epsilon > 2$. *If* $\kappa_i > 2$ *then* $n_{\theta_i} \sim n^{1/2} \times [K + O(B_n)]^{1/2}$, *and if* $\kappa_i \leq 2$ *then* $n_{\theta_i} \sim Kn^{1/2}(n/k_n)^{1/\kappa_i-1/2} \times [K + O(C_n)]^{1/2}$.

Remark 1: Notice $n_{\theta_1} = o(n^{1/2})$ when $\kappa_\epsilon < 2$. Irrespective of the other regressors, as long as the error ϵ_t is heavy tailed the intercept rate is sub-$\sqrt{n}$-consistent due to $k_n/n \to 0$ under tail trimming.

Remark 2: If the error has a finite variance then n_{θ_i} is governed entirely by the regressor tails, hence $\hat{\theta}_{i,n}$ is super-$\sqrt{n}$-consistent when $x_{i,t}$ has an infinite variance. If ϵ_t and $x_{i,t}$ have finite variances but some other regressor $x_{j,t}$ is heavy-tailed, then all we can say is $\hat{\theta}_{i,n}$ is at least $\sqrt{n}$-consistent since cross-Jacobia complexity precludes sharper results.

In the following assume all random variables are heavy tailed $\max\{\kappa_\epsilon, \kappa_2, ..., \kappa_r\} \leq 2$. General examples are similar.

EXAMPLE 1 (Slope Rate Lower Bound): Note $\liminf_{n\to\infty} n_{\theta_i}/[n^{1/2} (n/k_n)^{1/2-\kappa_i/\kappa_{\epsilon,i}+1/\kappa_{\epsilon,i}}] \geq K$ depends solely on the dispersion of ϵ_t and $x_{i,t}$. As long as $1/2 - \kappa_i/\kappa_{\epsilon,i} + 1/\kappa_{\epsilon,i} > 0$ then $\hat{\theta}_{i,n}$ is super-$\sqrt{n}$-consistent. There are two cases.

Case 1 ($\kappa_i \leq \kappa_\epsilon$): Here $n_{\theta_i} \geq Kn^{1/2}\,(n/k_n)^{1/\kappa_i-1/2}$ only reflects the tails of $x_{i,t}$. Simply choose a light fractile $k_n = [\lambda \ln n]$ for $\lambda > 0$ to obtain

$$n_{\theta_i} \geq Kn^{1/\kappa_i}/\ln(n).$$

The GMTTME rate is therefore arbitrarily close to the highest achievable rate n^{1/κ_i} for stationary data with Paretian tails by any estimation method, including untrimmed OLS, LAD, QML, smooth M-estimators, and Whittle and Yule-Walker estimation (e.g. Hannan and Kanter 1977, An and Chen 1982, Cline 1989, Davis et al 1992, Hall and Yao 2003). The latter estimators, however, have non-Gaussian limits, so the GMTTME offers a two-fold advantage: it obtains (nearly) the highest possible rate and is asymptotically normal. This case covers conventional time series models like ARMA: see Example 5, below.

Case 2 ($\kappa_i > \kappa_\epsilon$): If ϵ_t is more heavy-tailed than $x_{i,t}$ then super-$\sqrt{n}$-convergence still arises as long as $x_{i,t}$ has an infinite variance $\kappa_i < 2$ and the dispersion of ϵ_t is not too great, $\kappa_\epsilon > 2(\kappa_i - 1)$. If $\kappa_i = 1.5$, for example, then any $\kappa_\epsilon \geq 1$ applies.

EXAMPLE 2 (Independent Regressors): If stochastic $x_{i,t}$ are independent random variables then $J_{i,j,n} \sim K \; \forall i \neq j$. In this case it can be shown $A_n = o(1)$ hence $n_{\theta_i} \sim Kn^{1/2} \left(n/k_n\right)^{1/2-\kappa_i/\kappa_{\epsilon,i}+1/\kappa_{\epsilon,i}}$.

EXAMPLE 3 (Tail Homogeneity): If $\kappa_\epsilon = \kappa_i = \kappa$ for all i then $\kappa_{\epsilon,i} = \kappa$ and $a^*_{\epsilon,(i)} = 1$ hence

$$n_{\theta_i} \sim Kn^{1/2} \left(n/k_n\right)^{1/\kappa - 1/2}.$$

Super-$\sqrt{n}$-convergence arises *if and only if* variance is infinite $\kappa < 2$. In the hairline infinite variance case $\kappa = 2$ exact $\sqrt{n}$-convergence applies.

The following provide intuition as to *why* super-$\sqrt{n}$-convergence may or may not arise.

EXAMPLE 4 (Location): Consider estimating location

$$y_t = \theta^0 + \epsilon_t, \; \epsilon_t \overset{iid}{\sim} (7), \;\; \kappa \in (1, 2],$$

with one equation $m_t(\theta) = y_t - \theta$. The Jacobian is $J_n = -1 + o(1)$ and the covariance scale $\Gamma_n = \Sigma_n^{1/2} = n^{1/2}(n/k_n)^{1/\kappa - 1/2}$. Therefore $n_\theta = Kn^{1/2}(k_n/n)^{1/\kappa-1/2} = o(n^{1/2})$ under tail trimming, so the GMTTME is sub-$\sqrt{n}$-consistent when $\kappa < 2$.

Remark 1: The reason for the sluggish rate is given above: a model without stochastic regressors cannot provide explanatory leverage against a heavy tailed shock. In the hairline infinite variance case $\kappa = 2$, however, $n_\theta = Kn^{1/2}$.

Remark 2: It is straightforward to show over identifying restrictions involving lags of y_t have no impact on the sub-$\sqrt{n}$ rate since the added regressors y_{t-i} are independent.

EXAMPLE 5 (AR with iid error): Consider a stationary infinite variance autoregression

$$y_t = \sum_{i=1}^{r} \theta_i^0 y_{t-i} + \epsilon_t, \;\; \epsilon_t \overset{iid}{\sim} (7), \;\; \kappa \in (1, 2].$$

The AR process $\{y_t\}$ satisfies (7) with the same index κ (Cline 1989, Brockwell and Cline 1985). Since $\kappa_\epsilon = \kappa_i = \kappa_{\epsilon,i} = \kappa$, apply Example 3 to get $n_{\theta_i}/n^{1/2} \sim K\left(n/k_n\right)^{1/\kappa-1/2} \to \infty$.

Remark 1: Stochastic regressors always leverage toward increasing the rate, so minimal trimming is optimal (slow $k_n \to \infty$). If k_n is regularly varying $k_n = [n^\lambda]$ then $n_{\theta_i} \sim Kn^{1/\kappa - \lambda(1/\kappa-1/2)} > Kn^{1/\kappa-\iota}$ for any tiny $\lambda, \iota > 0$. Conversely if a slowly varying $k_n = [\lambda \ln(n)]$ then $n_{\theta_i} > Kn^{1/\kappa-\iota}$.

Remark 2: The AR(1) case is particularly revealing:

$$n_\theta = Kn^{1/2}\frac{|J_n|}{\Gamma_n} \sim Kn^{1/2}\frac{E\left[y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,n}\right)\right]}{\left(E\left[\epsilon_t^2 y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,n}\right)\right]\right)^{1/2}} \sim Kn^{1/2}\frac{\left(n/k_n\right)^{2/\kappa-1}}{\left[\left(n/k_n\right)^{2/\kappa-1}\right]^{1/2}}.$$

The numerator Jacobian $E[y_{t-1}^2 I(|\epsilon_t y_{t-1}| \leq c_{1,n})] \sim K(n/k_n)^{2/\kappa-1}$ works like a tail-trimmed *variance*. If sequences $\{c_{y,n}, k_n\}$ satisfy $(n/k_n)P(|y_t| > c_{y,n}) \to \infty$ then arguments in the proof of Lemma 3.1 reveal $E[y_{t-1}^2 I(|y_{t-1}| \leq c_{y,n})] \sim c_{y,n}^2 P(|y_{t-1}| > c_{y,n})$ $= K(n/k_n)^{2/\kappa-1}$. Trimming by $\epsilon_t y_{t-1}$ delivers the same rate because ϵ_t is independent of y_{t-1} and each have tail index $\kappa \leq 2$, hence $\epsilon_t y_{t-1}$ has index κ (Cline 1986). By comparison the denominator $(E[\epsilon_t^2 y_{t-1}^2] I(|\epsilon_t y_{t-1}| \leq c_{1,n}))^{1/2} = (n/k_n)^{1/\kappa-1/2}$ is a tail-trimmed *standard deviation* of an object with the *same* tail index κ. Therefore $n_\theta \sim Kn^{1/2}(n/k_n)^{1/\kappa-1/2}$ dominates $\sqrt{n}$ when ϵ_t has an infinite variance. If ϵ_t is not independent of y_{t-1} then the above arguments fails, and feedback can cause $\Gamma_n \to \infty$ so fast that super-$\sqrt{n}$-convergence cannot arise. See Sections 3.2 and 3.3.

EXAMPLE 6 (Instrumental Variables): It is tempting to use heavy-tailed instruments z_t to induce super-$\sqrt{n}$-convergence. Consider a simple scalar model for reference

$$y_t = \theta^0 x_t + \epsilon_t, \text{ where } \{x_t, \epsilon_t\} \overset{iid}{\sim} (0,1) \text{ and } m_t(\theta) = (y_t - \theta x_t) z_t \in \mathbb{R}.$$

Assume the instrument $z_t \in \mathbb{R}$ has tail (7) and index $\kappa_z < 2$, and is valid: it is independent of ϵ_t and $\inf_{n \geq N} |E[x_t z_t I(|\epsilon_t z_t| \leq c_n)]| > 0$ for large N. For example, we might use z_t $= x_t^2$ if x_t has a finite variance and infinite kurtosis. Since $\epsilon_t z_t$ has tail index κ_z (Cline 1986), the Cauchy-Schwartz inequality and arguments in the proof of Lemma 3.1 reveal

$$n_{\theta_i} = Kn^{1/2} \frac{E[x_t z_t I(|\epsilon_t z_t| \leq c_{1,n})]}{(E[\epsilon_t^2 z_t^2 I(|\epsilon_t z_t| \leq c_{1,n})])^{1/2}} \leq Kn^{1/2} \left(\frac{E[z_t^2 I(|\epsilon_t z_t| \leq c_{1,n})]}{E[\epsilon_t^2 z_t^2 I(|\epsilon_t z_t| \leq c_{1,n})]} \right)^{1/2} = Kn^{1/2}.$$

A thin-tailed regressor x_t handicaps the Jacobian rate irrespective of the instrument z_t.

3.2 ARCH

Consider a Strong-ARCH(p) model

$$y_t = h_t \epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0,1),\ h_t^2 = \alpha^0 + \sum_{i=1}^{p} \beta_i^0 y_{t-i}^2 = \theta^{0\prime} x_t,\ \alpha^0 > 0, \beta^0 \geq 0,\ \theta = [\alpha, \beta']'$$

$$m_t(\theta) = \left(y_t^2 - \theta' x_t\right) \times x_t,\ \ x_t = \left[1, y_{t-1}^2, ..., y_{t-p}^2\right]',$$

where the Lyapunov exponent associated with the stochastic recurrence equation form is negative. Then $\{y_t, h_t\}$ are stationary with tail (7) and index $\kappa_y > 0$ as long at least one as one $\beta_i^0 > 0$ (Basrak et al 2002: Theorem 3.1). If $\beta_i^0 > 0$ then $m_t(\theta^0) = (\epsilon_t^2 - 1)h_t^2 x_t$ have tails (7) with indices $\{\kappa_y/2, \kappa_y/4, ..., \kappa_y/4\}$.

Integrability of $m_t(\theta^0)$ requires $E[\epsilon_t^2] < \infty$ and $E|h_t^2 y_{i,t}^2| < \infty$, so assume

$$E\left[y_t^4\right] < \infty \text{ hence } \kappa_y > 4.$$

The moment requirement above is relaxed in a QML environment. See Section 3.4. If β^0 $= 0$ then each $m_{i,t}(\theta^0)$ has a finite variance and therefore all rates $n_\alpha, n_{\beta_i} \sim Kn^{1/2}$.

Since $(\partial/\partial\theta)m_t(\theta) = -x_t x_t'$ have indices in $\{\kappa_y/4, \kappa_y/2, \infty\}$ and $\kappa_y > 4$ each component is integrable, thus $J_n \sim J$. The standard deviations $\Gamma_{i,i,n}$ are almost as simple to compute since $\kappa_y > 4$ implies the intercept term $\Gamma_{1,1,n} \leq (E[(\epsilon_t^2 - 1)^2 h_t^4])^{1/2} = K$, and the remaining $\Gamma_{i,i,n} = (E[m_{i,t}^2(\theta^0) I_{i,n,t}(\theta^0)])^{1/2} \sim Kc_{i,n}(k_n/n)^{1/2} = K(n/k_n)^{4/\kappa_y - 1/2}$ by Feller property (FE) and $c_{i,n} = K(n/k_n)^{1/(\kappa_y/4)}$ for tail (7). Therefore $J_{1,1,n}/\Gamma_{1,1,n} \sim K$ and all other $J_{i,i,n}/\Gamma_{i,i,n} \sim K(n/k_n)^{-(4/\kappa_y - 1/2)} = o(1)$. Now use (6) to deduce $n_\alpha, n_{\beta_i} \sim$ $Kn^{1/2}$. This proves the next claim.

LEMMA 3.2 (Strong-ARCH) *Any stationary strong-ARCH process with negative Lyapunov exponent and* $\kappa_y > 4$ *has GMTTME rates* n_α, $n_{\beta_i} \sim Kn^{1/2}$.

Remark 1: The tails of ϵ_t do not play any role *per se*. Thicker tailed ϵ_t and/or larger slopes β^0 imply y_t is heavier tailed: *why* y_t is heavy tailed is irrelevant.

Remark 2: Strong-ARCH are AR in squares $y_t^2 = \theta' x_t + v_t$, where $E[v_t|\Im_{t-1}] = 0$. But stationary AR equations all have the same tail index κ when ϵ_t is iid with tail (7). In the ARCH case the "error" $y_t^2 - h_t^2 = (\epsilon_t^2 - 1)\theta^{0\prime} x_t$ depends on x_t, thus $m_{1,t}(\theta^0) = (\epsilon_t^2 - 1)h_t^2$ has tail index $\kappa/2$ and all other $m_{i,t}(\theta^0) = (\epsilon_t^2 - 1)h_t^2 y_{t-i+1}^2$ for $i \geq 2$ have index $\kappa/4$ due to feedback. The intuition from Section 3.1 suffices to explain the rate: models with disproportionately heavy-tailed "errors" render less sharp estimates (in this case, less than super-$\sqrt{n}$-consistent).

3.3 AUTOREGRESSIONS WITH ARCH-ERRORS

Consider an AR(1) with ARCH(1) error

$$y_t = \rho^0 y_{t-1} + u_t, \quad \left|\rho^0\right| < 1, \; u_t = h_t \epsilon_t, \; \epsilon_t \overset{iid}{\sim} N(0,1)$$

$$h_t^2 = \alpha^0 + \beta^0 u_{t-1}^2, \; \alpha^0 > 0, \beta^0 > 0, \; \theta = [\rho, \alpha, \beta]'$$

$$E\left[\ln\left|\rho^0 + \left(\beta^0\right)^{1/2} \epsilon_t\right|\right] < 0$$

and three equations used to estimate each $\theta = [\rho^0, \alpha^0, \beta^0]'$,

$$m_t(\theta) = \left[\begin{array}{l} (y_t - \rho y_{t-1}) y_{t-1} \\ (y_t - \rho y_{t-1})^2 - \alpha - \beta (y_{t-1} - \rho y_{t-2})^2 \\ \left((y_t - \rho y_{t-1})^2 - \alpha - \beta (y_{t-1} - \rho y_{t-2})^2\right) \times (y_{t-1} - \rho y_{t-2})^2 \end{array}\right].$$

Thus $\{y_t\}$ is stationary, geometrically β-mixing with regular varying tail (7) and index $\kappa_y > 0$ (Borkovec and Klüppelberg 2001: Theorems 1 and 3). Further, the equations $m_{i,t}(\theta^0)$ satisfy (7) with indices $\{\kappa_y, \kappa_y/2, \kappa_y/4\}$.

Integrability again requires $\kappa_y > 4$, hence $J_n \sim J$. If $\kappa_y > 8$ then $\Gamma_n = n^{1/2} I_3$, otherwise use $E[y_t^4] < \infty$ to deduce $E[u_t^4] < \infty$, $E[h_t^4] < \infty$ and $E[\epsilon_t^4] < \infty$ so $\Sigma_{1,1,n} \sim K$, $\Sigma_{2,2,n} \sim K$, and $\Sigma_{3,3,n} = E[(\epsilon_t^2 - 1)^2 h_t^4 y_{t-1}^4 I(|\epsilon_t^2 - 1| h_t^2 y_{t-1}^2 \leq c_{3,n})] \sim K(n/k_n)^{8/\kappa_y - 1}$. This gives trimmed standard deviations $\Gamma_{1,1,n} = \Gamma_{2,2,n} = 1$ and $\Gamma_{3,3,n} = (n/k_n)^{4/\kappa_y - 1/2}$. The same conclusion as the Strong-ARCH case is therefore reached by working through formula (6): all rates are $\sqrt{n}$ since feedback between u_t and y_{t-1} dulls the rate below super-$\sqrt{n}$-convergence

LEMMA 3.3 (AR-ARCH) *Any stationary AR(1)-ARCH(1) with* $\epsilon_t \overset{iid}{\sim} N(0,1)$, $E[\ln|\rho^0 + \left(\beta^0\right)^{1/2} \epsilon_t|] < 0$ *and* $\kappa_y > 4$ *has GMTTME rates* $n_\rho = n_\alpha = n_\beta \sim Kn^{1/2}$.

EXAMPLE 7 (AR with ARCH Error): Estimate only the AR slope ρ^0 in the above AR-ARCH:

$$m_t(\rho) = (y_t - \rho y_{t-1}) y_{t-1}.$$

Notice $m_t(\rho^0) = \epsilon_t h_t y_{t-1}$ has a tail index $\kappa_y/2$ due to feedback, half that in the iid case Example 5. But this implies $m_t(\rho^0)$ is integrable only if $\kappa_y > 2$. Arguments in the proof of Lemmas 3.1 can be used to deduce $n_\rho \sim Kn^{1/2}$ if $\kappa_y \geq 4$, and $n_\rho \sim$

$Kn^{1/2}(n/k_n)^{-(2/\kappa_y-1/2)} = o(n^{1/2})$ if $\kappa_y \in (2,4)$. Feedback between error u_t and regressor y_{t-1} substantially elevates estimating equation tail thickness relative to the Jacobian, hence the convergence rate n_ρ falls below $n^{1/2}$.

COROLLARY 3.4 (AR with ARCH Error) *Under the conditions of Lemma 3.3 if only ρ^0 is estimated with one equation $m_t(\rho) = (y_t - \rho y_{t-1})y_{t-1}$ then $n_\rho \sim Kn^{1/2}$ if $\kappa_y \geq 4$, and if $\kappa_y \in (2,4)$ then $n_\rho \sim Kn^{1/2}(n/k_n)^{-(2/\kappa_y-1/2)}$.*

3.4 TAIL-TRIMMED QML FOR GARCH

Consider an ARCH(1) with QML equations

$$y_t = h_t\epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0,1),\ \ h_t^2 = \alpha^0 + \beta^0 y_{t-1}^2 = \theta^{0\prime}x_t,\ \ \alpha^0 > 0, \beta^0 \geq 0$$

$$m_t(\theta) = \left(y_t^2 - \theta' x_t\right)\{\theta' x_t\}^{-2}x_t,$$

where $E[m_t(\theta)] = 0$ *if and only if* $\theta = \theta^0$. Assume at least one $\beta_i^0 > 0$ to highlight the impact of scaling.

If $E[\epsilon_t^4] = \infty$ assume ϵ_t has tail (7) with index $\kappa_\epsilon \in (2,4]$. Scaling $m_t(\theta^0) = (\epsilon_t^2 - 1)h_t^{-2}x_t$ with $\beta^0 > 0$ implies only the tails of ϵ_t matter, and ensures $J_n \sim J$. This will imply a diminished rate of convergence when $E[\epsilon_t^4] = \infty$ for the reasons given in Section 3.1.

LEMMA 3.5 *Assume $E|y_t|^p < \infty$ for some $p > 0$, $\beta^0 > 0$ and $E[\epsilon_t^2] < \infty$. If $\kappa_\epsilon \geq 4$ then $n_\alpha = n_\beta = Kn^{1/2}$. If $\kappa_\epsilon \in (2,4]$ then $n_\alpha, n_\beta \sim Kn^{1/2}(n/k_n)^{-(2/\kappa_\epsilon-1/2)}$.*

Remark 1: In lieu of scaling only trivial restrictions on the tails of y_t are required, *per se*, as long as $E[\epsilon_t^2] < \infty$.

Remark 2: Although QML equations permit far heavier tails, the rate of convergence suffers when ϵ_t has an infinite kurtosis since scaling eliminates any beneficial impact the Jacobian might have. As in all ARCH cases, above, maximal trimming augments the rate: $n_\alpha, n_\beta \nearrow Kn^{1/2}$ as $k_n \nearrow n$.

Remark 3: Suppose $\kappa_\epsilon \in (2,4]$. If $k_n = [n^\lambda]$ then $n_\alpha, n_\beta \sim Kn^{1/2-(1-\lambda)(2/\kappa_\epsilon-1/2)}$ and if $k_n = [n/\ln(n)]$ then $n_\alpha, n_\beta \geq Kn^{1/2-\iota}$ for all tiny $\iota > 0$.

Remark 4: The same basic conclusion is met for (nonlinear) GARCH $y_t = h_t(\theta^0)\epsilon_t$, $\epsilon_t \overset{iid}{\sim} (0,1)$, with QML equations $m_t(\theta) = (y_t^2 - h_t^2(\theta))h_t^{-4}(\theta) \times (\partial/\partial\theta)h_t(\theta)$ as long as $h_t^{-2}(\theta^0) \times (\partial/\partial\theta)h_t(\theta)|_{\theta^0}$ is integrable.

4. AUTOREGRESSIONS AND GARCH We now verify the major assumptions for heavy-tailed stationary AR and GARCH under symmetric trimming $m_{i,n,t}^*(\theta) = m_{i,t}(\theta)I(|m_{i,t}(\theta)| \leq c_{i,n}(\theta))$ where $(n/k_n)P(|m_{i,t}(\theta)| > c_{i,n}(\theta)) \to 1$, the same fractile k_n for each equation, and weight $\Upsilon_n = \Sigma_n^{-1}/||\Sigma_n^{-1}||$.

4.1 Autoregression

Consider a stationary AR(r) process with iid, heavy-tailed errors

$$y_t = \theta^{0\prime}x_t + \epsilon_t,\ \ x_t = [y_{t-1}, ..., y_{t-r}]',\ \ \epsilon_t \overset{iid}{\sim} (7) \text{ with } \kappa \in (1,2),\ E[\epsilon_t] = 0 \quad (8)$$

$$m_t(\theta) = \left(y_t - \theta' x_t\right)x_t.$$

Assume ϵ_t has an absolutely continuous marginal distribution symmetric at zero and positive $\mathbb{R}$-*a.e.* Then y_t is uniformly $L_{1+\iota}$-bounded geometrically β-mixing (An and Huang 1996: Theorem 3.1), and y_t and $m_{i,t}(\theta^0) = \epsilon_t y_{t-i}$ have tail (7) with the same index κ (Cline 1986, 1989).

Stationarity, linearity, distribution continuity, mixing and the efficient weight ensure D1, D2, D3, I1 and M2 are satisfied. The envelope bounds D4 are trivial given the linear form of $m_t(\theta)$, compactness of Θ, and L_p-boundedness.

Further I2 and I4 are trivial since $E[m^*_{n,t}(\theta^0)] = 0$ and $E[m^*_{n,s}(\theta^0)m^*_{n,t}(\theta^0)'] = 0$ under symmetry, integrability and orthogonality; M1 holds under I1 and D1-D7 given $\Upsilon_n = \Sigma_n^{-1}/||\Sigma_n^{-1}||$ and Lemma 2.4. Finally, D5.i merely defines the threshold sequence. What remains is envelope moment bounds D4, threshold bound D5.ii, Jacobia property D6, indicator property D7, smoothness I3, and covariance bound M3.

D5.ii (*threshold bound*). Since linearity and power-law tails imply $c_{i,n}(\theta) = c_{i,n}g(\theta)$ for a finite dimensional function g (see D7, below) consider $c_{i,n} = c_{i,n}(\theta^0)$. Since $m_{i,t}(\theta^0)$ has tail (7) and $\kappa < 2$ it follows $c_{i,n} = K(n/k_n)^{1/\kappa}$ and from property (FE) $E[(m^*_{i,n,t}(\theta^0))^2] \sim Kc^2_{i,n}(k_n/n) = K(n/k_n)^{2/\kappa-1}$. Therefore $c_{i,n} = o(n^{1/2}||\Sigma_n^{-1/2}||^{-1}) = o(n^{1/2}||\Sigma_n^{1/2}||) = o(n^{1/2}c_n(k_n/n)^{1/2}) = o(c_n k_n^{1/2})$, so D5.ii is trivial for any thresholds $c_{i,n}$ and fractiles $k_n \to \infty$.

D6 (*Jacobia*).

D6.i. Each part is trivial given the linear data generating process and iid innovations with absolutely continuous marginal distribution.

D6.ii. The lower bound is trivial. Consider the upper bound and note $\kappa \in (1,2)$ implies $||J_n|| \to \infty$ by Lemma 3.1. Then stationarity, L_p-boundedness of y_t and the construction $J^*_{n,i,j,t}(\theta) = -y_{t-i}y_{t-j}I_{n,j,t}(\theta)$ imply $E[(\sup_{\theta\in U^0(\delta_n)}\{||J^*_n(\theta) - J^*_n||\})^s] \leq K$ for any $s \in (0, p/2)$. Therefore $\sup_{\theta\in U^0(\delta_n)}\{||J^*_n(\theta) - J^*_n||\} = o_p(||J_n||)$ follows by Markov's inequality and $||J_n|| \to \infty$.

D7 (*indicator class*). Since $m_{i,t}(\theta) = \epsilon_t x_t - (\theta - \theta^0)x_t y_{t-i}$ is a linear function of θ and $(\theta - \theta^0)x_t y_{t-i}$ dominates the tail structure for any $\theta \neq \theta^0$, we can without loss of generality assume $c_{i,n}(\theta) = c_{i,n}g(\theta)$ for some finite dimensional function $g : \Theta \to \mathbb{R}_+$. Therefore $\{m_{i,t}(\theta) : \theta \in \Theta\}$ and $\{c_{i,n}(\theta) : \theta \in \Theta\}$ form VC classes (Pakes and Pollard 1989: Lemma 2.4), hence $\{I(|m_{i,t}(\theta)| \leq c_{i,n}(\theta)) : \theta \in \Theta\}$ forms a VC class (van der Vaart and Wellner 1994: Lemma 2.6.18).

I3 (*smoothness*). Define $Q_n(\theta) := E[m^*_{n,t}(\theta)]' \times \Upsilon_n \times E[m^*_{n,t}(\theta)]$. Identification and the definition of a derivative imply $E[m^*_{n,t}(\theta)] = J_n(\theta - \theta^0) + o(||J_n|| \times ||\theta - \theta^0||)$, hence $\mathfrak{m}_n := \sup_\theta ||E[m^*_{n,t}(\theta)]|| \leq K||J_n|| \times (1 + o(1))$ given compactness of Θ. Therefore since Υ_n is bounded

$$\inf_{||\theta-\theta^0||>\delta}\left\{\mathfrak{m}_n^{-2}Q_n(\theta)\right\} \geq \inf_{||\theta-\theta^0||>\delta}\left\{\left(\theta-\theta^0\right)'\frac{J_n'}{\|J_n\|}\Upsilon_n\frac{J_n}{\|J_n\|}\left(\theta-\theta^0\right)\right\}\times(1+o(1))+o(1).$$

But boundedness and positive definiteness of Υ_n imply $J_n'\Upsilon_n J_n/||J_n||^2$ is positive definite for sufficiently large n, so I3 follows.

M3 (*covariance*). Similar to D5.ii, consider the simpler case $\Sigma_n^{1/2} = \Sigma_n^{1/2}(\theta^0)$. Use Feller property (FE) to deduce $||\Sigma_n^{1/2}|| \leq Kc_n(k_n/n)^{1/2} = K(n/k_n)^{1/\kappa-1/2}$ which is $o(n^{1/2})$ given $\kappa \in (1, 2]$.

Asymptotic normality is therefore a primitive property of the GMTTME for stationary AR data. The above verification, Theorems 2.2 and 2.6 and Lemma 3.1 suffice to prove the following claim.

COROLLARY 4.1 *Let y_t satisfy (8), let $\Upsilon_n = \Sigma_n^{-1}/||\Sigma_n^{-1}||$, and assume the thresholds satisfy D5.i. Then $V_n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, I_r)$ and $\hat{V}_n = V_n(1 + o_p(1))$. In particular, $n^{1/2}(n/k_n)^{1/\kappa - 1/2}(\hat{\theta}_{i,n} - \theta_i^0) \xrightarrow{d} N(0, V_i)$ for any $k_n \to \infty$ and $k_n = o(n)$ each i and some $V_i < \infty$, and $(n^{1/\kappa}/\ln(n))(\hat{\theta}_{i,n} - \theta_i^0) \xrightarrow{d} N(0, V_i)$.*

Remark: The result generalizes to autoregressive distributed lags, and AR-ARCH with adjustments to the rate.

4.2 Nonlinear AR-Nonlinear GARCH

Recall the ARCH model of Sections 3.2 and 3.5, assume at least one slope $\beta_i > 0$ and the distribution of $\epsilon_t \overset{iid}{\sim} (0,1)$ does not have an atom at zero. Then $\{y_t\}$ is stationary, L_p-bounded, geometrically β-mixing (Meitz and Saikkonen 2008) with tail (7) (Basrak et al 2002: Theorem 3.1). Least squares and QML type equations $m_t(\theta)$ are differentiable in θ with absolutely continuous marginal distributions, and integrable at θ^0 for all $\kappa_y > 2$. Similar to the AR model all conditions are either trivial (e.g. I2), regulatory (e.g. D5.i), or can be verified using arguments from Section 4.1 (e.g. M3).

The same basic arguments apply to a wide variety of smooth nonlinear AR-GARCH models, including Quadratic ARCH, smooth transition AR and GARCH, Asymmetric GARCH, and so on (An and Huang 1996, Borkovec and Klüppelberg 2001, Carrasco and Chen 2002, Cline 2007, Meitz and Saikkonen 2008).

5. SIMULATION STUDY In this section we compare one-step and two-step GMTTME's, denoted $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n^{(2)}$, to conventional GMM and QML estimators $\hat{\theta}_g$ and $\hat{\theta}_q$. Write $\hat{\theta}$ to denote any estimator. The models are LOCATION, AR(1), ARCH(1), GARCH(1,1), Threshold ARCH(1) and Quadratic ARCH(1), covering symmetric and asymmetric DGP's.

Let $N_{0,1}$ denote a standard normal law and P_γ a symmetric Pareto law with index $\gamma > 0$: if ϵ_t is governed by P_γ then $P(\epsilon_t > \epsilon) = P(\epsilon_t < -\epsilon) = (1/2) \times (1 + \epsilon)^{-\gamma}$. Define $\kappa_y = \sup\{\alpha > 0 : E|y_t|^\kappa < \infty\}$. See Table 1 for each DGP $y_t = f(\theta^0, \epsilon_t, x_t)$ where x_t may contain a constant, lags of y_t and so on, and for κ_y. We simulate 1000 samples of size $n = 1000$ for each model.

TABLE 1 - Data Generating Processes

Model Type	Functional Form: $f(\theta^0, \epsilon_t, x_t)$	θ_r^0	iid errors ϵ_t	κ_y
LOCATION	$1 + \epsilon_t$	1	$P_{1.5}$, $P_{2.5}$	1.50, 2.50
AR(1)	$.9 \times y_{t-1} + \epsilon_t$,	.9	$P_{1.5}$, $P_{2.5}$	1.50, 2.50
ARCH(1)	$\{.3 + .5y_{t-1}^2\}^{1/2}\epsilon_t$	.5	$N_{0,1}$	4.65[4]
ARCH(1)	$\{.3 + .6y_{t-1}^2\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	3.80
ARCH(1)	$\{.3 + .6y_{t-1}^2\}^{1/2}\epsilon_t$	.6	$P_{2.5}$	1.80
ARCH(1)	$\{.3 + .6y_{t-1}^2\}^{1/2}\epsilon_t$	.6	$P_{2.1}$	1.40
GARCH(1,1)	$\{.3 + .3y_{t-1}^2 + .6h_{t-1}^2\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	4.10
IGARCH(1,1)	$\{.3 + .4y_{t-1}^2 + .6h_{t-1}^2\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	2.00
GARCH(1,1)	$\{.3 + .3y_{t-1}^2 + .6h_{t-1}^2\}^{1/2}\epsilon_t$	.6	$P_{2.5}$	1.50
TARCH(1)	$\{.3 + .6y_{t-1}^2 \times I(y_{t-1} < 0)\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	5.25[5]
TIARCH(1)	$\{.3 + y_{t-1}^2 \times I(y_{t-1} < 0)\}^{1/2}\epsilon_t$	1	$N_{0,1}$	3.40
TARCH(1)	$\{.3 + .6y_{t-1}^2 \times I(y_{t-1} < 0)\}^{1/2}\epsilon_t$	.6	$P_{2.5}$	2.60
QARCH(1)	$\|.3 + .8y_{t-1}\|\,\epsilon_t$	.8	$N_{0,1}$	3.50[6]
QIARCH(1)	$\|.3 + y_{t-1}\|\,\epsilon_t$	1	$N_{0,1}$	2.00

The GMM estimating equations are $m_t(\theta) = u_t(\theta) \times z_t$ for some $u_t(\theta) \in \mathbb{R}$ and $z_t \in \mathbb{R}^q$ described in Table 2. In each location and AR model least squares $u_t(\theta) = y_t - \theta' x_t$ are used. Recall QML equations in GARCH cases allows for the minimal moment condition $E[\epsilon_t^2] < \infty$. We therefore use QML-type forms $u_t(\theta) = (y_t^2 - h_t^2(\theta))h_t^{-2}(\theta)$ for GARCH models with index $\kappa_y \leq 4$ and least squares-type $u_t(\theta) = y_t^2 - h_t^2(\theta)$ if $\kappa_y > 4$.

TABLE 2 - Estimating Equations $m_t(\theta) = u_t(\theta) \times z_t$

Model	iid errors ϵ_t	κ_y	$u_t(\theta) \in \mathbb{R}$	$z_t \in \mathbb{R}^q$
Location	$P_{1.5}$, $P_{2.5}$	1.5, 2.5	LS	$[1, y_{t-1}]'$
AR(1)	$P_{1.5}$, $P_{2.5}$	1.5, 2.5	LS	$[y_{t-i}]_{i=1}^2$
ARCH(1)	$N_{0,1}$	4.65	LS	$\left[1, \{y_{t-i}^2\}_{i=1}^2\right]'$
ARCH(1)	$N_{0,1}$, $P_{2.5}$, $P_{2.1}$	3.80, 1.80, 1.40	QML	$\left[1, \{y_{t-i}^2\}_{i=1}^2\right]'$
GARCH(1,1)	$N_{0,1}$	4.10	LS	$\left[1, \{y_{t-i}^2, h_{t-i}^2\}_{i=1}^2\right]'$
GARCH(1,1)	$N_{0,1}$, $P_{2.5}$	2.00, 1.50	QML	$\left[1, \{y_{t-i}^2, h_{t-i}^2\}_{i=1}^2\right]'$
TARCH(1)	$N_{0,1}$	5.25	LS	$\left[1, y_{t-1}^2, y_{t-2}^2\right]$
TARCH(1)	$N_{0,1}$, $P_{2.5}$	3.40, 2.60	QML	$\left[1, y_{t-1}^2, y_{t-2}^2\right]'$
QARCH(1)	$N_{0,1}$, $N_{0,1}$	3.50, 2.00	QML	$\left[1, \{y_{t-i}, y_{t-i}^2\}_{i=1}^2\right]'$

All processes in this study have regularly varying distribution tails (7) with index $\kappa_y > 0$ (Hannan and Kanter 1977, Cline 1986, 1989, Borkovec and Klüppelberg 2001, Cline 2007). In the iid and AR cases tail thickness is gauged by the Pareto innovations.

[4] Basrak et al (2002: eq. 2.10) show $E[(\beta\epsilon_t^2 + \gamma)^{\kappa/2}] = 1$ for GARCH(1,1) $y_t = h_t\epsilon_t$ with iid ϵ_t and $h_t^2 = \alpha + \beta y_{t-1}^2 + \gamma h_{t-1}^2$, provided the Lyapunov index is negative. The index κ is computed as $\hat{\kappa} = \arg\min_{\kappa \in K}\{|1/N \sum_{t=1}^N (\beta\epsilon_t^2 + \gamma)^{\kappa/2} - 1|\}$ over $K \in \{.01, .02, ..., 10\}$ based on $N = 100,000$ iid random draws ϵ_t from $N_{0,1}$, $P_{2.5}$ or $P_{2.1}$. The 1% bands are less than .001 in all cases.

[5] An ARCH affect exists only for the left-tail, so κ solves $\beta^{\kappa/2}E[|\epsilon_t|^\kappa I(\epsilon_t < 0)] = 1$ (Cline 2007: Lemma 2.1 and Example 3). But ϵ_t is symmetrically distributed about 0, hence $\beta^{\kappa/2}E[|\epsilon_t|^\kappa] = 2$. The monte carlo experiment described above allows for computation of κ.

[6] Since $y_t = |\alpha + \beta y_{t-1}|\epsilon_t$ use Lemma 2.1 of Cline (2007) to deduce $\beta^\kappa E[|\epsilon_t|^\kappa] = 1$.

The collective GARCH group have heavy tails due to the innovations $\epsilon_t \overset{iid}{\sim} P_{2.5}$ and/or the parametric structure. The kurtosis of y_t is infinite in most cases, and variance is infinite for IGARCH, ARCH and GARCH with Pareto errors and QIARCH. Thus, in all random volatility models the GMME is *not* asymptotically normal, and the QMLE has not been shown to be asymptotically normal when $E[\epsilon_t^4] = \infty$. Nevertheless, the QMLE is consistent in all cases[7].

5.1 Tail Fractile

Linear models (location, AR, ARCH, GARCH) are estimated with symmetric trimming and the same fractile for all equations:

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

Nonlinear models (TARCH, QARCH) demand asymmetric trimming with left- and right-tailed $k_{1,n}$ and $k_{2,n}$, where $k_{j,n} = [n^{\lambda_j}]$ for each $\lambda_j \in \{.01, .02, ..., .99\}$.

5.2 Evaluation

We analyze r^{th} parameter estimates $\hat{\theta}_r$ for brevity. Consult the third column of Table 1 for the true θ_r^0. Estimator performance is gauged by simulation means, mean-squared-errors, and Kolmogorov-Smirnov tests of standard normality. Let $\{\hat{\theta}_{j,r}\}_{j=1}^{1000}$ be the independently drawn sequence of estimates of θ_r^0. In a first experiment we fix the sample size $n = 1000$ and use the simulation mse $\hat{s}_{n,r}^2 = (1/1000)\sum_{j=1}^{1000}(\hat{\theta}_{j,r} - \theta_r^0)^2$ to generate an iid sequence of ratios $\{\hat{T}_{j,r}\}_{j=1}^{1000}$, $\hat{T}_{j,r} = \{\hat{\theta}_{j,r} - \theta_r^0\}/\hat{s}_{n,r}^2$. We report $(1/1000)\sum_{j=1}^{1000}\hat{\theta}_{j,r}$, $\hat{s}_{n,r}$, and the KS test based on $\{\hat{T}_{j,r}\}_{j=1}^{1000}$. In the case of GMTTM only the KS minimizing λ or pair (λ_1, λ_2) is used. See Tables 3 and 4.

5.3 GMTTM Weight

Let $\hat{\theta}_n^{(1)}$ be the one-step GMTTME based on the naïve weight $\hat{\Upsilon}_n = I_q$, and $\hat{\theta}_n^{(2)}(\tilde{\theta}_n)$ the two-step estimator with efficient weight $\hat{\Upsilon}_n = \hat{\Sigma}_n^{-1}(\tilde{\theta}_n)/||\hat{\Sigma}_n^{-1}||$ and plug-in $\tilde{\theta}_n$. In simulations not reported here we found $\hat{\theta}_n^{(1)}$ dominated $\hat{\theta}_n^{(2)}(\hat{\theta}_n^{(1)})$ across models and evaluation criteria due to the computational complexity of a multi-step algorithm under nonlinearity associated with trimming. Further, the two-step $\hat{\theta}_n^{(2)}(\hat{\theta}_q)$ with a QML plug-in dominated $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n^{(2)}(\hat{\theta}_g)$ with a GMM plug-in: $\hat{\theta}_n^{(2)}(\tilde{\theta}_n)$ is efficient for any consistent $\tilde{\theta}_n$, the QMLE $\hat{\theta}_q$ is more stable than the GMME $\hat{\theta}_g$ and one-step GMTTME $\hat{\theta}_n^{(1)}$ due to scaling and non-trimming, and $\hat{\theta}_q$ is consistent for each DGP in this study. The GMTTME is therefore computed as $\hat{\theta}_n = \hat{\theta}_n^{(2)}(\hat{\theta}_q)$.

In all cases the GMME is computed in two steps using the QMLE $\hat{\theta}_q$ as a plug-in.

5.4 Rate of Convergence

In a second experiment we analyze the rate of convergence for location $y_t = \theta^0 + \epsilon_t$, AR(1) $y_t = \theta^0 y_{t-1} + \epsilon_t$ and ARCH(1) $y_t = (\alpha^0 + \beta^0 y_{t-1}^2)^{1/2}\epsilon_t$. We estimate each model

[7] Let $\hat{Q}_q(\theta)$ denote the QML criterion. It can be easily verified for all models above $(\partial/\partial\theta)\hat{Q}_q(\theta) = \sum_{t=1}^n g_t(\theta)$ for some $L_{1+\iota}$-bounded martingale difference sequence $\{g_t(\theta^0), \Im_t\}$. Since an $L_{1+\iota}$-bounded mds trivially forms a uniformly integrable L_1-mixingale (McLeish 1975), Andrews' (1988: Theorem 1) law of large numbers applies: $1/n\sum_{t=1}^n g_t(\theta_0) \overset{p}{\to} 0$. Further, each $g_t(\theta)$ satisfies Andrew's (1992: W-LIP) Lipschitz condition given differentiability and the QML criterion form, so $\sup_\theta |1/n\sum_{t=1}^n \{g_t(\theta) - E[g_t(\theta)]\}| \overset{p}{\to} 0$ by Theorem 3 of Andrews (1992). Consistency of the QMLE is now a standard exercise (e.g. Pakes and Pollard 1989: Corollary 3.4). See also Hall and Yao (2003) for the GARCH case with infinite kurtosis errors.

by exactly identified GMTTM based on the Table 2 equation forms over sample sizes $n \in N = \{1000...10000\}$ with increments of 20. In the ARCH case we focus on the slope β^0. This produces 450 estimates of the simulation mse's $\{\hat{s}^2_{n,r}\}_{n \in N}$.

Since the rate satisfies $n_{\theta_r} \sim K\hat{s}^{-1}_{n,r}$, according to location and AR Examples 4 and 5 if y_t has an infinite variance $\kappa_y < 2$ then

$$\text{Location} : \hat{s}^{-1}_{n,r}/n^{1/2} \sim K n^{-(1-\lambda)(1/\kappa_y - 1/2)}$$

$$\text{AR} : \hat{s}^{-1}_{n,r}/n^{1/2} \sim K n^{(1-\lambda)(1/\kappa_y - 1/2)}.$$

By Lemma 3.5 if the ARCH error has index $\kappa_\epsilon \in (2, 4]$ then

$$\text{ARCH} : \hat{s}^{-1}_{n,r}/n^{1/2} \sim K n^{-(1-\lambda)(2/\kappa_\epsilon - 1/2)}$$

In all cases we may write $\ln(\hat{s}^{-1}_{n,r}/n^{1/2})$ as a log-linear trend in n:

$$\ln(\hat{s}^{-1}_{n,r}/n^{1/2}) = a + b \ln(n) + v_n \text{ for some } v_n.$$

In the infinite variance AR model, for example, $b = (1 - \lambda)(1/\kappa_y - 1/2)$.

The trend slope $b < 0$, $= 0$ and > 0 imply sub-, exact-, and super-$\sqrt{n}$-convergence. We can in principle select a KS minimizing λ for each sample size n, but this adds complexity since then b depends on n. We therefore use the KS minimizing λ based on the Section 5.2 experiment with $n = 1000$ and estimate b by least squares.

5.5 Summary of Results

Refer to Tables 3-6 for all results. Tail-trimming always delivers an approximately normal estimator. The GMTTME is roughly normal even for the profoundly heavy-tailed linear and nonlinear GARCH models. By comparison the standard GMME fails tests of normality in every heavy tailed case as expected, and the QMLE is non-normal in all cases where it is not asymptotically normal (infinite variance location and AR, GARCH with infinite kurtosis error), and in most cases where it has not been shown to be (GARCH with infinite skew errors; heavy-tailed QARCH). The most notable findings are summarized below.

i. In the presence of heavy-tails the GMME and QMLE strongly fail KS tests of normality, while the GMTTME passes roughly as well as in any other case. Models include infinite variance location and AR, and infinite kurtosis or infinite variance ARCH, GARCH, TARCH and QARCH.

ii. Asymmetric trimming for asymmetrically distributed equations is always optimal, where more observations are trimmed from the heavier tail. Consider the TIARCH model (see Table 6): left-tailed y_t have an infinite variance and right-tailed y_t are Gaussian. The optimal trimming pair $\{\lambda_1, \lambda_2\} = \{.40, .25\}$ translates to trimming $k_{1,n} = 16$ left-tailed and $k_{2,n} = 6$ right-tailed observations when $n = 1000$.

iii. Typically only a few tail observations need to be trimmed to ensure approximate normality. Examples include symmetric GARCH with Pareto errors: $k_n = [1000^{.35}] = 11$; and TIARCH with a heavier left-tail: $k_{1,n} + k_{2,n} = 16 + 6 = 22$.

iv. The QMLE fails normality tests in all GARCH cases where the errors have an infinite fourth moment. Further, even though the QMLE for IGARCH with Gaussian innovations is asymptotically normal (e.g. Lumsdaine 1996), for small samples it is demonstrably non-normal as shown elsewhere (e.g. Lumsdaine 1995).

5.6 Rate of Convergence

See Table 4 for the simulation 95% confidence bands of the rate log-trend slope b. In most cases simulation results closely match theory, and in all cases the correct sign is deduced. The finite variance location and AR the slope is $b = 0$ and the respective bands are $.031 \pm .051$ and $.027 \pm .039$.

Examples 4 and 5 reveal b for infinite variance location and AR are

$$\text{location}: \; b = -(1-\lambda)(1/\kappa_y - 1/2) = -(1-.22)(1/1.5 - 1/2) = -0.112$$

$$\text{AR}: b = (1-\lambda)(1/\kappa_y - 1/2) = (1-.32)(1/1.5 - 1/2) = .113.$$

The 95% band for location is $-.130 \pm .024$, and for AR is $.116 \pm .012$, both quite sharp and containing the true b. In the AR case the maximum rate of convergence amongst M-estimators is $n^{1/\kappa} = n^{.67} > n^{.5}$ while the GMTTME achieved on average $n^{1/2+b} = n^{.5+.116} = n^{.616} \in (n^{.5}, n^{.67})$, exactly as theory dictates.

In the ARCH case $b = 0$ if the kurtosis of y_t is finite with least squares equations, or the error kurtosis is finite with QML equations (Lemmas 3.2 and 3.5). Only the ARCH model with Pareto errors deviates since $\kappa_\epsilon = 2.5$ and $\kappa_y = 1.8$ or $\kappa_\epsilon = 2.1$ and $\kappa_y = 1.4$. Use Lemma 3.5 and Table 4 to deduce

$$\text{ARCH } (\kappa_\epsilon = 2.5)\text{: } b = -(1-\lambda)(2/\kappa_\epsilon - 1/2) = -(1-.4)(2/2.5 - 1/2) = -.18$$

$$\text{ARCH } (\kappa_\epsilon = 2.1)\text{: } b = -(1-\lambda)(2/\kappa_\epsilon - 1/2) = -(1-.38)(2/2.1 - 1/2) = -.28.$$

The 95% simulation bands are $-.387 \pm .306$ and respectively $-.412 \pm .393$. Feedback erodes efficiency substantially.

Finally, the KS minimizing $\lambda = .40$ or $.38$ are far smaller than the optimal setting $\lambda \approx 1$ for optimizing the rate $n_{\theta_r} \nearrow n^{1/2}$. This suggests a sharp trade-off exists between using the asymptotic Gaussian distribution for small sample inference and for improving efficiency. The analyst interested in efficiency may want to set λ near 1 and bootstrap standard errors, while the analyst interested in Gaussian inference may want to forego efficiency. All of these matters, however, are left for future consideration.

6. CONCLUSION This paper develops a robust GMM estimator for possibly very heavy tailed data commonly encountered in financial and macroeconomic applications. This is accomplished by trimming an asymptotically vanishing portion of the sample estimating equations. Our approach applies equally to asymmetric or symmetric data generating processes with thin or thick tails.

We prove trimming estimating equations themselves ensures asymptotic normality, while *tail*-trimming can promote super-$\sqrt{n}$-convergence for stationary data. Indeed, tail trimming provides a potentially massive lift in the convergence rate for heavy tailed linear models with more regressors than simply a constant term and without error-regressor feedback.

Simulation work demonstrates the new estimator is approximately normal for a variety of linear and nonlinear data generating processes with heavy tails; symmetric trimming leads to profoundly poor estimates for asymmetric data; and GMTTM dominates GMM and QML in heavy tailed cases where the latter estimators are not, or have not been shown to be, asymptotically normal. Perhaps the most important lessons to be drawn are very few observations need to be trimmed to induce consistency for θ^0 and approximate normality; and very slight trimming can lead to massive efficiency gains.

Future work should involve details on the convergence rate for broader classes of linear and nonlinear processes, the trade-off between small sample distribution and efficiency, adaptive methods for selecting the trimming fractiles $\{k_{1,i,n}, k_{2,i,n}\}$, and other criteria and methods for trimming (e.g. smooth trimming functions; trimming by errors and regressors separately).

APPENDIX A: Proofs of Main Results

We repeatedly use the following properties under M3:

$$1. \; \sup_{\theta} \{\|\Sigma_n(\theta)\|\} = o(n) \;\; \text{and} \;\; 2. \; \sup_{\theta} \left\{ \left\| \Sigma_n^{-1/2} \right\|^{-1} \right\} / n^{1/2} = o(1). \tag{9}$$

Property (9.1) follows from M3 since $\sup_{\theta}\{||\Sigma_n(\theta)||\} \leq \sup_{\theta}\{||\Sigma_n^{1/2}(\theta)||^2\} = o(n)$; and (9.2) follows from (9.1): $||\Sigma_n^{-1/2}(\theta)||^{-1}/n^{1/2} \leq ||\Sigma_n^{1/2}(\theta)||/n^{1/2} = o(1)$ uniformly on Θ.

The following proofs exploit criterion properties Lemmas B1-B2 and limit theory Lemmas C.1-C.8. See Appendices B and C respectively. It is understood n is sufficiently large to avoid covariance and Jacobian degeneracy.

Proof of Theorem 2.1. Define $\hat{m}_n^*(\theta) := 1/n \sum_{t=1}^n \hat{m}_{n,t}^*(\theta)$, $m_n^*(\theta) := 1/n \sum_{t=1}^n m_{n,t}^*(\theta)$, and $Q_n(\theta) := E[m_{n,t}^*(\theta)]' \times \Upsilon_n \times E[m_{n,t}^*(\theta)]$. The following is similar to Pakes and Pollard's (1989: p. 1039) argument. Use smoothness I3 and weight boundedness M1 to define $\epsilon(\delta) := \inf_{n \geq N} \inf_{||\theta - \theta^0|| > \delta}\{\mathfrak{m}_n^{-2} \times Q_n(\theta)\} > 0$ for arbitrarily large N and tiny $\delta > 0$. Since $P(||\hat{\theta}_n - \theta^0|| > \delta) \leq P(\mathfrak{m}_n^{-2} Q_n(\hat{\theta}_n) > \epsilon(\delta))$ it suffices to show $Q_n(\hat{\theta}_n) = o_p(\mathfrak{m}_n^2)$ to prove $||\hat{\theta}_n - \theta^0|| \xrightarrow{p} 0$.

Use the Lemma B.1 uniform criterion probability bound to deduce

$$Q_n(\hat{\theta}_n) \leq \hat{Q}_n(\hat{\theta}_n) + \left|\hat{Q}_n(\hat{\theta}_n) - Q_n(\hat{\theta}_n)\right| \leq \hat{Q}_n(\hat{\theta}_n) + \left(\mathfrak{m}_n^2 + Q_n(\hat{\theta}_n)\right) \times o_p(1),$$

hence $Q_n(\hat{\theta}_n)(1 - o_p(1)) \leq \hat{Q}_n(\hat{\theta}_n) + o_p(\mathfrak{m}_n^2)$. By construction $\hat{Q}_n(\hat{\theta}_n) \leq \hat{Q}_n(\theta^0)$, and $\hat{Q}_n(\theta^0) \leq K||\hat{m}_n^*(\theta^0)||^2 \leq K||m_n^*(\theta^0)||^2 + o_p(||\Sigma_n^{-1/2}||^{-1}/n^{1/2}) = K||m_n^*(\theta^0)||^2 + o_p(1)$ in lieu of weight bound M1, the Lemma C.2 asymptotic approximation and covariance bound (9.2). Finally $||m_n^*(\theta^0)|| \leq K||m_n^*(\theta^0) - E[m_{n,t}^*(\theta^0)]|| + o(||\Sigma_n^{-1/2}||^{-1}/n^{1/2}) = o_p(1)$ by the triangular equality, identification I2, the Lemma C.3 law of large numbers, and (9.2). ■

Proof of Theorem 2.2. Asymptotic linearity Lemma C.5 states

$$V_n^{1/2}\left(\hat{\theta}_n - \theta^0\right) = -V_n^{1/2}\left(H_n^{-1} J_n' \Upsilon_n\right) \frac{1}{n} \sum_{t=1}^n \hat{m}_{n,t}^*(\theta^0) \times (1 + o_p(1)).$$

Invoke asymptotic approximation Lemma C.2 coupled with the construction of V_n to deduce

$$V_n^{1/2}\left(\hat{\theta}_n - \theta^0\right) = -V_n^{1/2}\left(H_n^{-1} J_n' \Upsilon_n\right) \frac{1}{n} \sum_{t=1}^n m_{n,t}^*(\theta^0) \times (1 + o_p(1)) + o_p(1).$$

Now invoke identification I2, central limit theorem Lemma C.6 and $V_n^{1/2}(n^{-1/2} H_n^{-1} J_n' \Upsilon_n \Sigma_n^{1/2}) \rightarrow I_r$ to conclude

$$\begin{aligned} V_n^{1/2}\left(\hat{\theta}_n - \theta^0\right) &= -\Sigma_n^{-1/2} n^{-1/2} \sum_{t=1}^n \left\{m_{n,t}^*(\theta^0) - E\left[m_{n,t}^*(\theta^0)\right]\right\} \times (1 + o_p(1)) + o_p(1) \\ &\xrightarrow{d} N(0, I_r). \end{aligned}$$

■

Proof of Lemma 2.4. The triangular inequality and the definitions of $\hat{\Sigma}_n(\tilde{\theta}_n)$ and Σ_n imply $||\Sigma_n^{-1}\hat{\Sigma}_n(\tilde{\theta}_n) - I_q|| \leq ||\Sigma_n^{-1}|| \times ||\hat{\Sigma}_n(\tilde{\theta}_n) - \Sigma_n||$ is bounded by

$$\left\|\Sigma_n^{-1}\right\| \times \left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\hat{m}_{n,t}^*(\tilde{\theta}_n)\hat{m}_{n,t}^*(\tilde{\theta}_n)' - m_{n,t}^*(\tilde{\theta}_n)m_{n,t}^*(\tilde{\theta}_n)'\right\}\right\| + \left\|\Sigma_n^{-1}\right\| \times \left\|\frac{1}{n}\sum_{t=1}^{n}\left\{m_{n,t}^*(\tilde{\theta}_n)m_{n,t}^*(\tilde{\theta}_n)' - m_{n,t}^*(\theta^0)m_{n,t}^*(\theta^0)'\right\}\right\| + \left\|\Sigma_n^{-1}\right\| \times \left\|\frac{1}{n}\sum_{t=1}^{n} m_{n,t}^*(\theta^0)m_{n,t}^*(\theta^0)' - \Sigma_n\right\|.$$

The first term is $o_p(1)$ by an argument identical to the proof of uniform asymptotic approximation Lemma C.2, and the third term is $o_p(1)$ by Lemma C.8.

Consider the second term, and use expansion Lemma C.1.b to deduce for some $||\tilde{\theta}_{n,*} - \theta^0|| \leq ||\tilde{\theta}_n - \theta^0||$ and $r_n \to 0$ arbitrarily fast

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} m_{n,t}^*(\tilde{\theta}_n)m_{n,t}^*(\tilde{\theta}_n)' &= \frac{1}{n}\sum_{t=1}^{n} m_{n,t}^*(\theta^0)m_{n,t}^*(\theta^0)' \\ &+ o_p\left(\left\|\frac{1}{n}\sum_{t=1}^{n} J_{n,t}^*(\tilde{\theta}_{n,*})\right\| \times \left\|\sum_{t=1}^{n} m_{n,t}^*(\theta^0)\right\|\right) \times \left\|\tilde{\theta}_n - \theta^0\right\| \\ &+ o\left(n\left\|\frac{1}{n}\sum_{t=1}^{n} J_{n,t}^*(\tilde{\theta}_{n,*})\right\|^2 \times \left\|\tilde{\theta}_n - \theta^0\right\|^2\right) + r_n^2 \times o_p(1) \\ &= \frac{1}{n}\sum_{t=1}^{n} m_{n,t}^*(\theta^0)m_{n,t}^*(\theta^0)' + o_p(1) + o\left(\left\|\Sigma_n^{-1}\right\|^{-1}\right) + r_n \times o_p(1)\end{aligned}$$

The second equality follows from the supposition $\tilde{\theta}_n = \theta^0 + O_p(||V_n^{1/2}||^{-1})$, central limit theorem Lemma C.6, and Jacobian consistency Lemma 2.5. Therefore $||\Sigma_n^{-1}|| \times ||1/n\sum_{t=1}^{n}\{m_{n,t}^*(\tilde{\theta}_n)m_{n,t}^*(\tilde{\theta}_n)' - m_{n,t}^*(\theta^0)m_{n,t}^*(\theta^0)'\}|| = o_p(1)$ since $r_n \to 0$ is arbitrary. ■

Proof of Lemma 2.5. Recall $J_n = J_n(\theta^0) = (\partial/\partial\theta)E[m_{n,t}^*(\theta)]|_{\theta^0}$ and write $\hat{m}_n^*(\theta) = 1/n\sum_{t=1}^{n}\hat{m}_{n,t}^*(\theta)$. We only prove $\hat{J}_n^*(\tilde{\theta}_n) = J_n \times (1 + o_p(1))$ since $J_n^*(\tilde{\theta}_n) = J_n \times (1 + o_p(1))$ is similar.

Denote by $e_i \in \mathbb{R}^r$ the unit vector (e.g. $e_2 = [0,1,0,...,0]'$), define a sequence of positive numbers $\{\varepsilon_n\}$ that satisfies $\varepsilon_n \to 0$, $\varepsilon_n||V_n^{1/2}|| \to \infty$ and $||\tilde{\theta}_n - \theta^0||/\varepsilon_n \xrightarrow{p} 0$, and define

$$\check{J}_{i,j,n}^*(\theta, \varepsilon_n) := \frac{1}{2\varepsilon_n} \times \frac{1}{n}\sum_{t=1}^{n}\left\{\hat{m}_{j,n,t}^*(\theta + e_i\varepsilon_n) - \hat{m}_{j,n,t}^*(\theta - e_i\varepsilon_n)\right\}.$$

Minkowski's inequality implies for arbitrary θ

$$\left\|\hat{J}_n^*(\tilde{\theta}_n) - J_n\right\| \leq \left\|\hat{J}_n^*(\tilde{\theta}_n) - \check{J}_n^*(\theta, \varepsilon_n)\right\| + \left\|\check{J}_n^*(\theta, \varepsilon_n) - J_n\right\|$$

Apply asymptotic expansion Lemma C.1.a to deduce for some $\tilde{\theta}_{n,*} \in \{\tilde{\theta}_n - e_i\varepsilon_n, \tilde{\theta}_n + e_i\varepsilon_n\}$

$$\hat{J}_n^*(\tilde{\theta}_n) = \check{J}_{i,j,n}^*(\tilde{\theta}_{n,*}, \varepsilon_n) + o_p(n^{-1/2}) + o_p\left(\|J_n\|\right) = \check{J}_{i,j,n}^*(\tilde{\theta}_{n,*}, \varepsilon_n) + o_p\left(\|J_n\|\right),$$

where $n^{-1/2} = o(||J_n||)$ under Jacobian non-degeneracy D6.i.

It remains to show $||\check{J}_n^*(\tilde{\theta}_n, \varepsilon_n) - J_n|| = o_p(||J_n||)$ for any $||\tilde{\theta}_n - \theta^0|| \xrightarrow{p} 0$. Stochastic differentiability Lemma C.7 and the properties of ε_n imply for any constant $a \in \mathbb{R}^r$

$$\left\|\left\{\hat{m}_n(\tilde{\theta}_n + a\varepsilon_n) - \hat{m}_n\left(\theta^0\right)\right\} - \left\{E\left[m_{n,t}^*\left(\tilde{\theta}_n + a\varepsilon_n\right)\right] - E\left[m_{n,t}^*\left(\theta^0\right)\right]\right\}\right\|$$

$$\leq \left(1 + \left\|V_n^{1/2}\right\| \times \left\|\tilde{\theta}_n + a\varepsilon_n - \theta^0\right\|\right) \times o_p\left(\|J_n\| \times \left\|V_n^{1/2}\right\|^{-1}\right)$$

$$\leq \left(\left\|V_n^{1/2}\right\|^{-1} + \left\|\tilde{\theta}_n - \theta^0\right\| + \|a\|\,\varepsilon_n\right) \times o_p\left(\|J_n\|\right) = o_p\left(\|J_n\|\,\varepsilon_n\right).$$

Similarly, by differentiability of $E[m_{n,t}^*(\theta)]$,

$$\left\|\frac{E\left[m_{n,t}^*(\tilde{\theta}_n + a\varepsilon_n)\right] - E\left[m_{n,t}^*\left(\theta^0\right)\right]}{\varepsilon_n} - aJ_n\right\|$$

$$= \left\|J_n\varepsilon_n^{-1}\left(\tilde{\theta}_n + a\varepsilon_n - \theta^0\right) - aJ_n + o_p\left(\|J_n\|\,\varepsilon_n^{-1}\left(\tilde{\theta}_n + \varepsilon_n - \theta^0\right)\right)\right\|$$

$$= \left\|J_n\varepsilon_n^{-1}\left(\tilde{\theta}_n - \theta^0\right)\right\| + o_p\left(\|J_n\|\right) = o_p\left(\|J_n\|\right).$$

Replace $\tilde{\theta}_n + a\varepsilon_n$ with $\tilde{\theta}_n - a\varepsilon_n$ to deduce the same bounds. Therefore

$$\left\|\check{J}_n^*(\tilde{\theta}_n, \varepsilon_n) - J_n\right\| = \left\|\frac{\hat{m}_n^*(\tilde{\theta}_n + \varepsilon_n) - \hat{m}_n^*(\tilde{\theta}_n - \varepsilon_n)}{2\varepsilon_n} - J_n\right\| = o_p\left(\|J_n\|\right).$$

■

Proof of Lemma 3.1. We treat the cases $\kappa_\epsilon \leq 2$ and $\kappa_\epsilon > 2$ separately.

Case 1 ($\max\{\kappa_\epsilon, \kappa_2, ..., \kappa_r\} \leq 2$): We require some properties of regularly varying tails. Since ϵ_t and stochastic $x_{i,t}$ are mutually independent with tails (7) and indices κ_ϵ and κ_i, the convolutions $\epsilon_t x_{i,t}$ satisfy (Cline 1986)

$$m_{i,t}(\theta^0) = \epsilon_t x_{i,t} \sim (7) \text{ with index } \kappa_{\epsilon,i} := \min\left\{\kappa_\epsilon, \kappa_i\right\}. \tag{10}$$

If $x_{1,t} = 1$ then (10) holds with $\kappa_{\epsilon,i} = \kappa_\epsilon$. Therefore by the construction of $c_{i,n}$ and k_n, and tail (7),

$$c_{i,n} = K\left(\frac{n}{k_n}\right)^{1/\kappa_{\epsilon,i}}. \tag{11}$$

Finally, processes z_t with regularly varying tails (7) and index $\kappa \in (0, 2]$ satisfy (see Feller 1971: Theorem 1 in IX.8):

$$E\left[z_t^2 I\left(|z_t| \leq c\right)\right] \sim Kc^2 \times P\left(|z_t| > c\right) \text{ as } c \to \infty. \tag{12}$$

Express n_{θ_i} as

$$n_{\theta_i} = Kn^{1/2}\frac{J_{i,i,n}}{\Gamma_{i,i,n}} \times \left[K + K\left(\frac{\max_{j\neq i}\left\{\Gamma_{j,j,n}^{-1}J_{j,i,n}\right\}}{\Gamma_{i,i,n}^{-1}J_{i,i,n}}\right)^2 \times \left(1 + O\left(1\right)\right)\right]^{1/2}. \tag{13}$$

Step 1 (Σ_n, Γ_n): Properties (10)-(12) imply

$$\Sigma_{i,i,n} = E\left[m^2_{i.n,t}(\theta^0)\right] \sim c^2_{i,n} P\left(\left|m_{i,t}(\theta^0)\right| > c_{i,n}\right) \sim c^2_{i,n}\frac{k_n}{n} = K\left(\frac{n}{k_n}\right)^{2/\kappa_{\epsilon,i}-1}.$$

Hence $\Gamma_{i,i,n} = (n/k_n)^{1/\kappa_{\epsilon,i}-1/2}$.

Step 2 (J_n): The Lemma C.1.c Jacobian approximation implies by the linear equation form $J_{i,j,n} = -E[x_{i,t}x_{j,t}I(|\epsilon_t x_{j,t}| \le c_{j,n})] \times (1 + o(1))$. Assume initially all regressors are stochastic. Since $\kappa_i, \kappa_\epsilon \le 2$ it follows $\kappa_i < \kappa_\epsilon + 2$. Thus, by independence, (10) and (12)

$$\begin{aligned} E\left[x^2_{i,t} I\left(|\epsilon_t x_{i,t}| \le c_{i,n}\right)\right] &= K\int\left[\frac{c^2_{i,n}}{\epsilon^2}P\left(|x_{i,t}| > \frac{c_{i,n}}{|\epsilon|}\right)\right]f_\epsilon(d\epsilon) \\ &\sim K\int E\left[\frac{c^2_{i,n}}{\epsilon^2}\left(\frac{c_{i,n}}{|\epsilon|}\right)^{-\kappa_i}\right]f_\epsilon(d\epsilon) \\ &= Kc^{2-\kappa_i}_{i,n}E\left[|\epsilon_t|^{\kappa_i-2}\right] = Kc^{2-\kappa_i}_{i,n} = \left(\frac{n}{k_n}\right)^{(2-\kappa_i)/\kappa_{\epsilon,i}}. \end{aligned}$$

The cross-products $E[x_{i,t}x_{j,t}I(|\epsilon_t x_{j,t}| \le c_{j,n})]$ are bounded by case. If $x_{i,t}$ and $x_{j,t}$ are independent then $E[x_{i,t}x_{j,t}I(|\epsilon_t x_{i,t}| \le c_{i,n})] \sim K$ given $\sup_{t\in\mathbb{Z}} E|x_{i,t}| < \infty \; \forall i$. If they are perfectly positively dependent with marginal tail (7) and $\kappa_i, \kappa_j \le 2$ it can only be the case that $x_{i,t} = sign(x_{j,t}) \times |x_{j,t}|^p$ where $p = \kappa_j/\kappa_i$. Therefore, since $\kappa_j - p - 1 < \kappa_\epsilon$ is easy to verify,

$$\begin{aligned} E\left[x_{i,t}x_{j,t}I\left(|\epsilon_t x_{j,t}| \le c_{j,n}\right)\right] &= \int E\left[|x_{j,t}|^{p+1} I\left(|x_{j,t}|^{(p+1)/2} \le \left(\frac{c_{j,n}}{|\epsilon|}\right)^{(p+1)/2}\right)\right]f_\epsilon(d\epsilon) \\ &= K\int\left[\left(\frac{c_{j,n}}{|\epsilon|}\right)^{p+1}P\left(|x_{j,t}| > \frac{c_{j,n}}{|\epsilon|}\right)\right]f_\epsilon(d\epsilon) \\ &\sim K\int E\left[\left(\frac{c_{j,n}}{|\epsilon|}\right)^{p+1}\left(\frac{c_{j,n}}{|\epsilon|}\right)^{-\kappa_j}\right]f_\epsilon(d\epsilon) \\ &= Kc^{p+1-\kappa_j}_{j,n}\int E\left[|\epsilon|^{\kappa_j-p-1}\right]f_\epsilon(d\epsilon) \\ &= Kc^{p+1-\kappa_j}_{j,n} = K\left(\frac{n}{k_n}\right)^{(\kappa_j/\kappa_i+1-\kappa_j)/\kappa_{\epsilon,j}}. \end{aligned}$$

The perfect negative dependence case is similar. Hence $J_{i,j,n} = O((n/k_n)^{\kappa^{-1}_{\epsilon,j}(\kappa_j/\kappa_i+1-\kappa_j)})$.

Finally, if $x_{1,t} = 1$ then $E[x^2_{1,t}I(|\epsilon_t x_{i,t}| \le c_{i,n})] \sim 1 - k_n/n$, $E[x_{i,t}x_{1,t}I(|\epsilon_t x_{1,t}| \le c_{i,n})] = E[x_{j,t}I(|\epsilon_t| \le c_{i,n})] = O(1)$ and $E[x_{1,t}x_{i,t}I(|\epsilon_t x_{i,t}| \le c_{i,n})] = E(E[x_{i,t}I(|x_{i,t}| \le c_{i,n}/|\epsilon_t|]|\epsilon_t) = O(1)$ since $\sup_{t\in\mathbb{Z}} E|x_{i,t}| < \infty \; \forall i$. Therefore $J_{1,1,n} = -1 + o(1)$ and $J_{i,1,n}, J_{1,i,n} = O(1) \times (1 + o(1))$.

Step 3 (n_{θ_i}): Consider the slope rates, the intercept rate being similar. The claim follows from (13) by noting

$$\Gamma^{-1}_{i,i,n}J_{i,i,n} = (n/k_n)^{1/2-1/\kappa_{\epsilon,i}}(n/k_n)^{(2-\kappa_i)/\kappa_{\epsilon,i}} = (n/k_n)^{1/2+(1-\kappa_i)/\kappa_{\epsilon,i}}$$

and

$$\begin{aligned}\max_{j\neq i}\left\{\Gamma_{j,j,n}^{-1}J_{j,i,n}\right\} &= O\left(\max_{j\neq i}\left\{(n/k_n)^{1/2-1/\kappa_{\epsilon,j}}\times(n/k_n)^{(\kappa_i/\kappa_j+1-\kappa_i)/\kappa_{\epsilon,i}}\right\}\right)\\ &= O\left((n/k_n)^{1/2+1/\kappa_{\epsilon,i}-\min_{j\neq i}\{1/\kappa_{\epsilon,j}+(1-1/\kappa_j)\kappa_i/\kappa_{\epsilon,i}\}}\right)\\ &= O\left((n/k_n)^{1/2+1/\kappa_{\epsilon,i}-a^*_{\epsilon,(i)}}\right).\end{aligned}$$

Case 2 ($\kappa_\epsilon > 2$): In this case $\Sigma_{1,1,n} = E\left[m^2_{1.n,t}(\theta^0)\right] < \infty$ hence $\Gamma_{1,1,n} = 1$; if any $\kappa_i > 2$ then $\Gamma_{i,i,n} = 1$; and if $\kappa_i \leq 2$ then arguments under Case 1 imply $\Gamma_{i,i,n} = (n/k_n)^{1/\kappa_i-1/2}$

The intercept Jacobia $J_{1,i,n}$ and $J_{i,1,n}$ are characterized by Case 1 since $\kappa_i > 1$. If both $\kappa_i, \kappa_j \leq 2$ then $J_{i,j,n}$ follow from Case 1. If $\kappa_i, \kappa_j > 2$ then $J_{i,j,n} = -K$.

If $\kappa_i > 2 \geq \kappa_j$ then by Case 1 and $\kappa_{\epsilon,j} = \kappa_j$ it follows $J_{j,j,n} = K(n/k_n)^{2/\kappa_j-1}$, $J_{i,j,n} = O((n/k_n)^{1/\kappa_i+1/\kappa_j-1})$, and $J_{j,i,n} = O((n/k_n)^{\kappa^{-1}_{\epsilon,i}(\kappa_j/\kappa_i+1-\kappa_i)})$. The rate can be deduced from (13) by noting if $\kappa_i > 2$ then

$$\Gamma^{-1}_{i,i,n}J_{i,i,n} = 1\times K, \quad \left|\max_{j\neq i:\kappa_j>2}\left\{\Gamma^{-1}_{j,j,n}J_{j,i,n}\right\}\right| = K, \quad \max_{j\neq i:\kappa_j\leq 2}\left\{\Gamma^{-1}_{j,j,n}J_{j,i,n}\right\} = O\left((n/k_n)^{1/\kappa_j-1/2}\right)$$

hence $n_{\theta_i} \sim n^{1/2} \times [K + O(\max_{j\neq i:\kappa_j\leq 2}\{(n/k_n)^{2/\kappa_j-1}\})]^{1/2}$.

The rate under $\kappa_i \leq 2$ follows similarly since $\Gamma^{-1}_{i,i,n}J_{i,i,n} = (n/k_n)^{1/\kappa_i-1/2}$ and $\max_{j\neq i:\kappa_j\leq 2}\{\Gamma^{-1}_{j,j,n}J_{j,i,n}\}$ $= O(\max_{j\neq i:\kappa_j\leq 2}\{(n/k_n)^{1/2-1/\kappa_j+\kappa^{-1}_{\epsilon,i}(\kappa_i/\kappa_j+1-\kappa_i)}\})$. ■

Proof of Lemma 3.5. See Hill and Renault (2010). ■

APPENDIX B : GMTTM Criterion Properties

LEMMA B.1 (*uniform criterion law*) *Under D1-D5 and D7*

$$\sup_\theta\left\{\frac{\mathfrak{m}_n^{-2}\times\left|\hat{Q}_n(\theta)-Q_n(\theta)\right|}{1+\mathfrak{m}_n^{-2}\times Q_n(\theta)}\right\} = o_p(1) \;\; \textit{where}\;\; \mathfrak{m}_n := \sup_\theta E\left[\left\|m^*_{n,t}(\theta)\right\|\right].$$

LEMMA B.2 (moment expansion) *Under D6.i* $E[m^*_{n,t}(\theta)] - E[m^*_{n,t}(\theta')] = J_n(\theta')(\theta - \theta') + o(||J_n(\theta')|| \times ||\theta - \theta'||)$ *for any* $\theta, \theta' \in \Theta$.

Proof of Lemma B.1. Write $\hat{m}^*_n(\theta) := 1/n\sum_{t=1}^n \hat{m}^*_{n,t}(\theta)$ and $m^*_n(\theta) := 1/n\sum_{t=1}^n m^*_{n,t}(\theta)$. By weight property M1 and the triangular inequality

$$\begin{aligned}\mathfrak{m}_n^{-2}\left|\hat{Q}_n(\theta)-Q_n(\theta)\right| &\leq \mathfrak{m}_n^{-2}\left\|\hat{m}^*_n(\theta)\right\|^2\times\left\|\hat{\Upsilon}_n-\Upsilon_n\right\| + \mathfrak{m}_n^{-2}\left|\hat{m}^*_n(\theta)'\Upsilon_n\hat{m}^*_n(\theta)-Q_n(\theta)\right|\\ &\leq \left\|\hat{m}^*_n(\theta)\right\|^2\times o_p\left(\mathfrak{m}_n^{-2}\right) + K\left\|\hat{m}^*_n(\theta)-m^*_n(\theta)\right\|^2\\ &\quad +K\mathfrak{m}_n^{-2}\left\|m^*_n(\theta)\right\|\times\left\|\hat{m}^*_n(\theta)-m^*_n(\theta)\right\| + \mathfrak{m}_n^{-2}\left|m^*_n(\theta)'\Upsilon_n m^*_n(\theta)-Q_n(\theta)\right|\\ &= A_{1,n}(\theta)+A_{2,n}(\theta)+A_{3,n}(\theta)+A_{4,n}(\theta).\end{aligned}$$

Uniform approximation Lemma C.2 and $\liminf_{n \geq N} \mathfrak{m}_n > 0$ under smoothness I3 imply

$$\begin{aligned}
\sup_\theta \{A_{1,n}(\theta)\}^{1/2} &= \sup_\theta \left\{ \frac{\|\hat{m}_n^*(\theta)\|}{\sup_\theta \|E[m_t^*(\theta)]\|} \right\} \times o_p(1) \\
&\leq \frac{\sup_\theta \|m_n^*(\theta)\|}{\sup_\theta \|E[m_t^*(\theta)]\|} \times o_p(1) + o_p(1) \\
&\leq \frac{\sup_\theta \|m_n^*(\theta) - E[m_t^*(\theta)]\|}{\sup_\theta \|E[m_t^*(\theta)]\|} \times o_p(1) + o_p(1).
\end{aligned}$$

Now apply the uniform LLN Lemma C.3 to deduce $\sup_\theta \{A_{1,n}(\theta)\} = o_p(1)$. Similar arguments based on approximation Lemma C.2 reveal $\sup_\theta \{A_{2,n}(\theta)\}$ and $\sup_\theta \{A_{3,n}(\theta)\}$ are $o_p(1)$.

Finally, under M1

$$\begin{aligned}
A_{4,n}(\theta) &= \mathfrak{m}_n^{-2} \left| m_n^*(\theta)' \Upsilon_n m_n^*(\theta) - E[m_t^*(\theta)]' \Upsilon_n E[m_t^*(\theta)'] \right| \\
&\leq K \mathfrak{m}_n^{-2} \|m_n^*(\theta) - E[m_t^*(\theta)]\|^2 + K \mathfrak{m}_n^{-2} \|E[m_t^*(\theta)]\| \times \|m_n^*(\theta) - E[m_t^*(\theta)]\|.
\end{aligned}$$

Lemma C.3 implies each term is uniformly $o_p(1)$. ■

Proof of Lemma B.2. Apply Jacobian existence D6.i and the definition of a derivative.

■

APPENDIX C: Limit Theory for Trimmed Sums

This appendix contains limit theory for tail and tail-trimmed arrays. Since $m_t(\theta)$ are differentiable under D2, if some $m_{i.t}(\theta)$ has a finite variance and is untrimmed such that $m_{n,i.t}(\theta) = m_{i.t}(\theta)$ and $\hat{m}_{n,i.t}(\theta) = \hat{m}_{i.t}(\theta)$, the following claims applied to $m_{n,i.t}(\theta)$ can be proven from well known arguments[8]. We therefore assume all equations are trimmed for clarity: $\underline{q} = q$.

The first two results characterize expansions and rate of approximations for the trimmed equations.

LEMMA C.1 (expansions) *Assume D1-D7 and M2-M3 hold. Choose $\theta, \tilde{\theta} \in \Theta$ where $||\theta - \tilde{\theta}|| \leq \delta$ for tiny $\delta > 0$. Let $\{r_n\}$ be a sequence of positive numbers $r_n \to 0$ arbitrarily fast. In the following $o_p(1)$ is not a function of $t \in \mathbb{Z}$* or *$\theta \in \Theta$. For some sequence $\{\theta_{n,*}, \tilde{\theta}_{n,*}\}$ satisfying $||\theta_{n,*} - \tilde{\theta}|| \leq ||\theta - \tilde{\theta}||$ and $||\tilde{\theta}_{n,*} - \tilde{\theta}|| \leq ||\theta - \tilde{\theta}||$:*

a. (equation expansions): i. $m_n^*(\theta) = m_n^*(\tilde{\theta}) + J_n^*(\theta_{n,*})(\theta - \tilde{\theta}) + r_n \times o_p(1)$; and ii. $\hat{m}_n^*(\theta) = \hat{m}_n^*(\tilde{\theta}) + \hat{J}_n^*(\tilde{\theta}_{n,*})(\theta - \tilde{\theta}) + r_n \times o_p(1)$;

b. (cross-product expansion): $\sum_{t=1}^n m_{n,t}^*(\theta) m_{n,t}^*(\theta)' = \sum_{t=1}^n m_{n,t}^*(\tilde{\theta}) m_{n,t}^*(\tilde{\theta})' + o_p(||\sum_{t=1}^n J_{n,t}^*(\theta_{n,*})|| \times ||\sum_{t=1}^n m_{n,t}^*(\tilde{\theta})||) \times ||\theta - \tilde{\theta}|| + o(||\sum_{t=1}^n J_{n,t}^*(\theta_{n,*})||^2 \times ||\theta - \tilde{\theta}||^2) + r_n \times o_p(1)$;

c. (Jacobian): $J_n = E[J_{n,t}^*] \times (1 + o(1))$.

[8] In this expansions, approximation and indicator bounds Lemmas C.1, C.2 and C.4 are trivial; the Lemma C.3 uniform law follows from Andrews (1992) in lieu of mixing D3; differentiability implies asymptotic linearity Lemma C.5; McLeish's (1974) central limit theory for martingale differences suffices for Lemma C.6; and so on.

LEMMA C.2 (approximation) *Under D1-D5 and D7* $||\sum_{t=1}^{n}\{\hat{m}_{n,t}^{*}(\theta)-m_{n,t}^{*}(\theta)\}|| = o_p(n^{1/2}||\Sigma_n^{-1/2}(\theta)||^{-1})$ *for any* $\theta \in \Theta$, *and* $\sup_{\theta}\{||1/n\sum_{t=1}^{n}\{\hat{m}_{n,t}^{*}(\theta)-m_{n,t}^{*}(\theta)\}||\} = o_p(1)$.

Next, uniform laws and bounds for $m_{n,t}^{*}(\theta)$ and $I_{i,n,t}(\theta)$.

LEMMA C.3 (uniform LLN) *Under D1-D5* $1/n\sum_{t=1}^{n}(m_{n,t}^{*}(\theta)-E\left[m_{n,t}^{*}(\theta)\right]) = o_p(1)$ *for each* $\theta\in\Theta$, *and if additionally I3 holds* $\sup_{\theta}\{||1/n\sum_{t=1}^{n}(m_{n,t}^{*}(\theta)-E\left[m_{n,t}^{*}(\theta)\right])||\} = o_p(\mathfrak{m}_n)$.

LEMMA C.4 (uniform indicator laws) *Let D1-D5 hold and let* $i\in\{1,...,q\}$ *be arbitrary.*

a. $P(\sup_{\theta}\sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta}|I_{i,n,t}(\theta)-I_{i,n,t}(\tilde{\theta})|=1)\to 0$ *as* $n\to\infty$ *and/or* $\delta\to 0$.

Let D7 additionally hold. Define $X_{n,t}^{*}(\theta) := ((n/k_n)^{1/2-\iota})\{I_{i,n,t}(\theta)-E[I_{i,n,t}(\theta)]\}$ *for tiny* $\iota>0$ *and let* $\{X(\theta):\theta\in\Theta\}$ *be a Gaussian process with uniformly bounded and uniformly continuous sample paths with respect to* L_2*-norm.*

b. $\{n^{-1/2}\sum_{t=1}^{n}X_{n,t}^{*}(\theta):\theta\in\Theta\}\Longrightarrow\{X(\theta):\theta\in\Theta\}$ *and* $E[(\sup_{\theta}\{|n^{-1/2}\sum_{t=1}^{n}X_{n,t}^{*}(\theta)|\})^2] = O(1)$ *where* $O(1)$ *is not a function of* θ, *and* $\Longrightarrow$ *denotes weak convergence on a Polish space.*

Asymptotic linearity of the GMTTME and asymptotic normality of the tail-trimmed equations follow.

LEMMA C.5 (asymptotic linearity) *Under D1-D7, I1-I3 and M1-M3*

$$V_n^{1/2}\left(\hat{\theta}_n-\theta^0\right) = A_n\sum_{t=1}^{n}\hat{m}_{n,t}^{*}(\theta^0)\times(1+o_p(1))+o_p(1)\ \ a.s.$$

where $A_n = -V_n^{1/2}(H_n^{-1}J_n'\Upsilon_n)n^{-1}\in\mathbb{R}^{r\times q}$.

LEMMA C.6 (CLT) *Under D3-D5 and I4* $\Sigma_n^{-1/2}n^{-1/2}\sum_{t=1}^{n}\{m_{n,t}^{*}(\theta^0)-E[m_{n,t}^{*}(\theta^0)]\} \stackrel{d}{\to} N(0,1)$.

Stochastic differentiability aids proving Jacobian estimator consistency.

LEMMA C.7 (stochastic differentiability) *Under D1-D7, and M2-M3, for all* $\{\delta_n\}$, $\delta_n\to 0$, *and all* $\theta\in\Theta$,

$$\sup_{\theta\in U^0(\delta_n)}\left\{\frac{\left(\left\|V_n^{1/2}\right\|/\left\|J_n\right\|\right)\left\|\left\{\hat{m}_n^{*}(\theta)-\hat{m}_n^{*}(\theta^0)\right\}-\left\{E\left[m_{n,t}^{*}(\theta)\right]-E\left[m_{n,t}^{*}(\theta^0)\right]\right\}\right\|}{1+\left\|V_n^{1/2}\right\|\times\left\|\theta-\theta^0\right\|}\right\}\stackrel{p}{\to}0.$$

LEMMA C.8 (covariance) *Under D3-D5* $||1/n\sum_{t=1}^{n}m_{n,t}^{*}(\theta^0)m_{n,t}^{*}(\theta^0)'-\Sigma_n|| = o(||\Sigma_n^{-1}||^{-1})$.

Proof of Lemma C.1.

Claim (a): We prove (i) since (ii) is similar. Assume θ and $m_t(\theta)$ are scalars and $m_t(\theta)$ is symmetrically trimmed to simplify notation. Write $m_{n,t}^{*}(\theta) = m_t(\theta)\times I_{n,t}(\theta)$ where $I_{n,t}(\theta) = I(|m_t(\theta)|\leq c_n(\theta))$, and choose $||\theta-\tilde{\theta}||\leq\delta$. Use D2 to deduce by Taylor's theorem

$$m_{n,t}^{*}(\theta) = \left\{m_t(\tilde{\theta})+J_t(\theta_{n,\delta})(\theta-\tilde{\theta})\right\}\times I_{n,t}(\theta)$$

where $||\theta_{n,\delta} - \tilde{\theta}|| \leq ||\theta - \tilde{\theta}||$, and $J_t(\theta) := (\partial/\partial\theta)m_t(\theta)$. Therefore

$$m_n^*(\theta) - m_n^*(\tilde{\theta}) = J_n^*(\theta_{n,\delta}) \times (\theta - \tilde{\theta}) + \frac{1}{n}\sum_{t=1}^{n} m_t(\theta) \times \left\{ I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \right\} \quad (14)$$
$$+ \frac{1}{n}\sum_{t=1}^{n} J_t(\theta_{n,\delta}) \times \{I_{n,t}(\theta) - I_{n,t}(\theta_{n,\delta})\} \times (\theta - \tilde{\theta}).$$

Consider the second term in (14) and use $I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \in \{-1, 0, 1\}$ to bound

$$\left| \frac{1}{n}\sum_{t=1}^{n} m_t(\theta) \left\{ I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \right\} \right| \leq \frac{1}{n^{1/2}}\sum_{t=1}^{n} \left| m_t(\theta) \left\{ I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \right\} \right|$$
$$\times \frac{1}{n^{1/2}}\sum_{t=1}^{n} \left| I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \right| = A_n(\theta, \tilde{\theta}) \times B_n(\theta, \tilde{\theta}).$$

Consider $A_n(\theta, \tilde{\theta})$. The D5 probability orders and triangular inequality imply $E|I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})| = O(k_n/n)$ where $O(\cdot)$ is not a function of θ. Now use stationarity D3, envelope bound D4 and the Cauchy-Schwartz inequality to deduce for tiny $\iota > 0$

$$\left(E\left[A_n(\theta, \tilde{\theta})^{\iota} \right] \right)^{1/\iota} \leq n^{1/2} \left[E\left| m_t(\theta) \left\{ I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \right\} \right|^{\iota} \right]^{1/\iota} = O\left(n^{1/2} (k_n/n)^{1/\iota} \right).$$

Since $\iota > 0$ can be chosen arbitrarily small and $k_n/n \to 0$, invoke Markov's inequality to conclude for some $r_n \to 0$ arbitrarily fast and $o_p(\cdot)$ not a function of θ

$$A_n(\theta, \tilde{\theta}) = o_p\left(n^{1/2} \left(k_n^{1/2}/n \right)^{1/\iota} \right) = r_n \times o_p(1).$$

Since $E|B_n(\theta, \tilde{\theta})| \leq n^{1/2}$ follows trivially from $|I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})| \in \{0, 1\}$ we have shown $A_n(\theta, \tilde{\theta}) \times B_n(\theta, \tilde{\theta}) = r_n \times o_p(1)$ for some $r_n \to 0$ arbitrarily fast. Repeat the argument for the third term in (14) by invoking envelope bound D4 for $J_t(\theta)$.

Claim (b): The proof imitates Claim (a): expand $m_{n,t}(\theta)$ using D2 differentiability of $m_t(\theta)$, then exploit D4 L_p-boundedness of $m_t(\theta)$ and $(\partial/\partial)m_t(\theta)$ and the D5 probability orders to deduce the claim by Marvok's inequality.

Claim (c): Claim (a) and bounded convergence imply for $r_n \to 0$ arbitrarily fast

$$\frac{E\left[m_{n,t}^*(\theta)\right] - E\left[m_{n,t}^*(\theta^0)\right]}{\left\|\theta - \theta^0\right\|} = E\left[J_{n,t}^*(\theta_{n,\delta})\right] \times \left(1 + o\left(\left\|\theta - \theta^0\right\|\right)\right) + o(r_n).$$

Further, moment expansion Lemma B.2 asserts

$$\frac{E\left[m_{n,t}^*(\theta)\right] - E\left[m_{n,t}^*(\theta^0)\right]}{\left\|\theta - \theta^0\right\|} = J_n \times \left(1 + o\left(\left\|\theta - \theta^0\right\|\right)\right) + o(\|J_n\|).$$

Invoke covariance bound (9.2) and take $||\theta - \theta^0|| \leq \delta \to 0$ to complete the proof. ■

The proof of approximation Lemma C.2 requires consistency of the intermediate order statistics $m_{i,(k_{1,i,n})}^{(-)}(\theta)$ and $m_{i,(k_{2,i,n})}^{(+)}(\theta)$. Simplify notation by considering the two-tailed $m_{i,(k_{i,n})}^{(a)}(\theta)$ which estimates $c_{i,n}(\theta)$.

LEMMA C.2.1 *Under D1-D5 and D7* $m^{(a)}_{i,(k_n)}(\theta)/c_{i,n}(\theta) = 1 + O_p(k_{i,n}^{-1/2})$ *for each* $\theta \in \Theta$, *and* $\sup_\theta |m^{(a)}_{i,(k_n)}(\theta)/c_{i,n}(\theta) - 1| = O_p(k_{i,n}^{-1/2} n^\iota)$ *for infinitessimal* $\iota > 0$.

Proof. The claim follows from the weak law Lemma C.4.b and an argument linking tail indicator limit behavior to order statistics (e.g. Hsing 1991: p. 1553). See Hill and Renault (2010: Lemma C.2.1). ■

Proof of Lemma C.2. Assume θ and $m_t(\theta)$ are scalars and $m_t(\theta)$ is symmetrically trimmed for notational convenience, and write $\bar{I}_{n,t}(\theta) := 1 - I_{n,t}(\theta)$.

Claim 1: Let $\theta \in \Theta$ be arbitrary, and write $m_t = m_t(\theta)$, $c_n = c_n(\theta)$, $\hat{m}^*_{n,t} = \hat{m}^*_{n,t}(\theta)$, $m^*_{n,t} = m^*_{n,t}(\theta)$, $\bar{I}_{n,t} = 1 - I_{n,t}(\theta)$, $\hat{I}_{n,t} = \hat{I}_{n,t}(\theta)$, and $\tilde{\sigma}_n := ||\Sigma_n^{-1/2}(\theta)||^{-1}$. First bound

$$\left\| \sum_{t=1}^{n} \left\{ \hat{m}^*_{n,t} - m^*_{n,t} \right\} \right\| \le \max_{1 \le t \le n} \left\{ \left\| m_t \left\{ \hat{I}_{n,t} - I_{n,t} \right\} \right\| \right\} \times \sum_{t=1}^{n} \left\| \hat{I}_{n,t} - I_{n,t} \right\|.$$

By construction $||m_t\{\hat{I}_{n,t} - I_{n,t}\}|| \le 2||m^{(a)}_{(k_n)} - c_n||$. The intermediate order statistic is consistent $m^{(a)}_{(k_n)}/c_n = 1 + O_p(k_n^{-1/2})$ by Lemma C.2.1. Now use threshold bound D5.ii to deduce

$$\max_{1 \le t \le n} \left\{ \left\| m_t \left\{ \hat{I}_{n,t} - I_{n,t} \right\} \right\| \right\} \le 2 \left\| m^{(a)}_{(k_n)} - c_n \right\| = 2c_n \left\| m^{(a)}_{(k_n)}/c_n - 1 \right\| = o_p\left(\tilde{\sigma}_n \left(n/k_n \right)^{1/2} \right).$$

Next, by construction and the triangular inequality

$$\sum_{t=1}^{n} \left\| \hat{I}_{n,t} - I_{n,t} \right\| \le k_n^{1/2} \left\| \frac{1}{k_n^{1/2}} \sum_{t=1}^{n} \left\{ \bar{I}_{n,t} - E\left[\bar{I}_{n,t} \right] \right\} \right\| + k_n^{1/2} \left\| k_n^{1/2} \left(\frac{n}{k_n} E\left[\bar{I}_{n,t} \right] - 1 \right) \right\|$$

which is $O_p(k_n^{1/2})$ by D5 and an application of Lemma C.4.b. Therefore $\sum_{t=1}^{n}\{\hat{m}^*_{n,t} - m^*_{n,t}\} = o_p(\tilde{\sigma}_n (n/k_n)^{1/2} k_n^{1/2}) = o_p(\tilde{\sigma}_n n^{1/2})$.

Claim 2: Define $M^*_n := \max_{1 \le t \le n}\{\sup_\theta ||m_t(\theta)\{\hat{I}_{n,t}(\theta) - I_{n,t}(\theta)\}||\}$ and $\tilde{\sigma}_n := \sup_\theta\{||\Sigma_n^{-1/2}(\theta)||^{-1}\}$ and repeat the above argument to reach

$$\begin{aligned}\sup_\theta \left\| \frac{1}{n} \sum_{t=1}^{n} \left\{ \hat{m}^*_{n,t}(\theta) - m^*_{n,t}(\theta) \right\} \right\| \quad &\le \quad M^*_n \times \frac{k_n^{1/2}}{n} \sup_\theta \left\| \frac{1}{k_n^{1/2}} \sum_{t=1}^{n} \left\{ \bar{I}_{n,t}(\theta) - E\left[\bar{I}_{n,t}(\theta) \right] \right\} \right\| \\ &\quad + M^*_n \times \frac{k_n^{1/2}}{n} \sup_\theta \left\| k_n^{1/2} \left(\frac{n}{k_n} E\left[\bar{I}_{n,t}(\theta) \right] - 1 \right) \right\|.\end{aligned}$$

Uniform indicator law Lemma C.4.b and uniform probability orders D5 imply the right-hand-side is $O_p(M^*_n k_n^{1/2}/n)$.

Finally, since $|m_t(\theta)\{\hat{I}_{n,t}(\theta) - I_{n,t}(\theta)\}| \le 2c_n(\theta)||m^{(a)}_{(k_n)}(\theta)/c_n(\theta) - 1||$ it follows $M^*_n = O_p(n^{1/2+\iota}\tilde{\sigma}_n/k_n^{1/2}) = o_p(n/k_n^{1/2})$ from Lemma C.2.1, D5, and covariance bound (9.2) coupled with a continuity argument for infinitessimal $\iota > 0$. This proves $\sup_\theta ||1/n \sum_{t=1}^{n}\{\hat{m}^*_{n,t}(\theta) - m^*_{n,t}(\theta)\}|| = O_p(M^*_n k_n^{1/2}/n) = o_p(1)$ as claimed. ■

The uniform law in Lemma C.3 requires two supporting results concerning stochastic equicontinuity and term-wise stochastic equicontinuity [TSE]. See Andrews (1988) for definitions.

LEMMA C.3.1 *Under D1-D5 if* $\lim_{n\to\infty} E[\sup_\theta ||m^*_{n,t}(\theta)||] < \infty$ *then Andrews' (1988) Assumption* TSE *holds.*

Proof. An argument similar to Čižek's (2008: eq. (20)) proof of TSE suffices: we need only prove for arbitrary $\epsilon > 0$ and $\delta_1 > 0$ it follows $\forall\delta < \delta_1$ and some $\kappa > 0$

$$P\left(\sup_\theta \sup_{\tilde\theta:||\theta-\tilde\theta||\le\delta} \left|m_{i,t}(\theta) \times \left\{I_{i,n,t}(\theta) - I_{i,n,t}(\tilde\theta)\right\}\right| > \kappa\right) \le \epsilon, \tag{15}$$

$$P\left(\sup_\theta \sup_{\tilde\theta:||\theta-\tilde\theta||\le\delta} \left|m_{i,t}(\theta) - m_{i,t}(\tilde\theta)\right| \times I_{i,n,t}(\tilde\theta) > \kappa\right) \le \epsilon. \tag{16}$$

Since

$$\begin{aligned}&\sup_\theta \sup_{\tilde\theta:||\theta-\tilde\theta||\le\delta} \left|m_{i,t}(\theta)\left\{I_{i,n,t}(\theta) - I_{i,n,t}(\tilde\theta)\right\}\right| \\ &\qquad\le \sup_\theta |m_{i,t}(\theta)I_{i,n,t}(\theta)| \times \sup_\theta \sup_{\tilde\theta:||\theta-\tilde\theta||\le\delta} \left|I_{i,n,t}(\theta) - I_{i,n,t}(\tilde\theta)\right|\end{aligned}$$

use $\lim_{n\to\infty} E[\sup_\theta ||m^*_{n,t}(\theta)||] < \infty$ and indicator probability limit Lemma C.4.a to deduce (15). Next, differentiability D2 and moment bounds D4 imply for given $\kappa > 0$ and sufficiently tiny $\iota > 0$

$$\begin{aligned}P\left(\sup_\theta \sup_{\tilde\theta:||\theta-\tilde\theta||\le\delta} \left|m_{i,t}(\theta) - m_{i,t}(\tilde\theta)\right| \times \left|I_{i,n,t}(\tilde\theta)\right| > \kappa\right) &\le \frac{1}{\kappa^\iota} E\left[\sup_\theta \left\|\frac{\partial}{\partial\theta} m_{i,t}(\theta)\right\|^\iota\right] \times \left\|\theta - \tilde\theta\right\|^\iota \\ &\le K \times \delta^\iota \to 0\end{aligned}$$

as $\delta \to 0$, hence (16) holds. ■

LEMMA C.3.2 *If for arbitrary* $\epsilon > 0$ *and some sequence of positive real numbers* $\{\zeta_n\}$, $\zeta_n \to \infty$,

$$\lim_{n\to\infty} E\left[\sup_\theta \left|M^*_{n,t}(\theta)\right| \times I\left(\sup_\theta \left|M^*_{n,t}(\theta)\right| > \zeta_n\right)\right] = 0 \tag{17}$$

and

$$\lim_{\delta\to 0} \lim_{n\to\infty} \left\{\zeta_n \times P\left(\sup_\theta \sup_{\tilde\theta:||\theta-\tilde\theta||\le\delta} \left|M^*_{n,t}(\theta) - M^*_{n,t}(\tilde\theta)\right| > \epsilon\right)\right\} = 0 \tag{18}$$

then $\forall\epsilon > 0$ *there exists* $\delta > 0$ *such that* $X_n(\theta)$ *is stochastically equicontinuous.*

Proof. The proof is similar to Andrews (1992: Lemma 3) except we bypass uniform integrability with the weaker (17). Define $s^*_{n,t} := \sup_\theta |M^*_{n,t}(\theta)|$ and $\check M^*_{n,t}(\delta) := \sup_{\tilde\theta:||\theta-\tilde\theta||\le\delta} |M^*_{n,t}(\theta) - M^*_{n,t}(\tilde\theta)|$. For given $\epsilon > 0$ by supposition there exists $\delta > 0$ such that, uniformly in n, $E[2s^*_{n,t}I(2s^*_{n,t} > \zeta_n)] \le \epsilon^2/6$ and $P(\check M^*_{n,t}(\delta) > \epsilon^2/6) \le \epsilon^2/[6\zeta_n]$. Then by stationarity, Marvok's inequality and term-wise stochastic equicontinuity (18), for sufficiently large n

(e.g. Andrews 1992: p. 250)

$$P\left(\sup_{\theta}\sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta}\left|X_n^*(\theta)-X_n^*(\tilde{\theta})\right|>\epsilon\right)$$

$$\leq\frac{2}{\epsilon}\times E\left[\check{M}_{n,t}^*(\delta)\right]$$

$$\leq\frac{2}{\epsilon}\times E\left[\check{M}_{n,t}^*(\delta)I\left(\check{M}_{n,t}^*(\delta)\leq\frac{\epsilon^2}{6}\right)\right]+\frac{2}{\epsilon}\times E\left[\check{M}_{n,t}^*(\delta)I\left(\frac{\epsilon^2}{6}<\check{M}_{n,t}^*(\delta)\leq\zeta_n\right)\right]$$
$$+\frac{2}{\epsilon}\times E\left[\check{M}_{n,t}^*(\delta)I\left(\zeta_n<\check{M}_{n,t}^*(\delta)\right)\right]$$

$$\leq\frac{2}{\epsilon}\times\left\{\frac{\epsilon^2}{6}+\zeta_n\times P\left(\check{M}_{n,t}^*(\delta)>\frac{\epsilon^2}{6}\right)+E\left[s_{n,t}^*I\left(2s_{n,t}^*>\zeta_n\right)\right]\right\}\leq\epsilon.$$

This proves the claim. ■

Proof of Lemma C.3. Recall $\mathfrak{m}_n := \sup_\theta E[||m_{n,t}^*(\theta)||]$, let $\{a_n\}$ be an arbitrary sequence of positive real numbers, $\liminf_{n\geq N}\{a_n\}>0$, and define $M_{n,t}^*(\theta,a_n):=a_n^{-1}m_{i,n,t}^*(\theta)$, $X_{n,t}^*(\theta,a_n):=M_{n,t}^*(\theta,a_n)-E[M_{n,t}^*(\theta,a_n)]\}$ for arbitrary $i\in\{1,..,q\}$, and write $X_n^*(\theta,a_n)$ $:=1/n\sum_{t=1}^n X_{n,t}^*(\theta,a_n)$.

Step 1: $X_{n,t}^*(\theta,a_n)$ is for any θ and $\{a_n\}$ stationary geometrically β-mixing under D3, hence it is geometrically α-mixing (e.g. Doukhan 1994). Apply Ibragimov's (1962) bound and $|m_{i,n,t}^*(\theta)|\leq c_n$ to deduce for some $\rho\in(0,1)$ and tiny $\iota>0$

$$\begin{aligned}E\left(X_n^*(\theta,a_n)\right)^2 &\leq \frac{1}{n}E\left[\left(X_{n,t}^*(\theta,a_n)\right)^2\right]+2\frac{1}{n}\sum_{t=2}^{\infty}\left|E\left[X_{n,1}^*(\theta,a_n)X_{n,t}^*(\theta,a_n)\right]\right|\\ &\leq K\left\{\frac{1}{n}E\left[\left(X_{n,t}^*(\theta,a_n)\right)^2\right]+\frac{1}{n}\sum_{t=2}^{\infty}\rho^{(t-1)\times\iota}\left(E\left|X_{n,t}^*(\theta,a_n)\right|^{2+\iota}\right)^{1-\iota}\right\}\\ &\leq K\left\{\frac{1}{n}E\left[\left(m_{i,n,t}^*(\theta)\right)^2\right]+\frac{1}{n}\left(E\left|m_{i,n,t}^*(\theta)\right|^{2+\iota}\right)^{1-\iota}\right\}.\end{aligned}$$

The first term on the right-hand-side is $o(1)$ by covariance bound (9.1). Further, (9.1), $E[(m_{i,n,t}^*(\theta))^2]\leq(E[(m_{i,n,t}^*(\theta))^{2+\iota}])^{1-\iota}$ by Lyapunov's inequality and absolute continuity D1 ensure we can find $\iota>0$ so small to ensure $(E[(m_{i,n,t}^*(\theta))^{2+\iota}])^{1-\iota}=o(n)$.

Thus by Chebyshev's inequality $X_n^*(\theta,1)\xrightarrow{p}0$, and since $\liminf_{n\geq N}\{\mathfrak{m}_n\}>0$ under I3 also $X_n^*(\theta,\mathfrak{m}_n)\xrightarrow{p}0$.

Step 2: Write $M_{n,t}^*(\theta)=M_{n,t}^*(\theta,\mathfrak{m}_n)$, $X_{n,t}^*(\theta):=X_{n,t}^*(\theta,\mathfrak{m}_n)$ and $X_n^*(\theta)=X_n^*(\theta,\mathfrak{m}_n)$. Compactness of Θ and pointwise convergence Step 1 imply uniform convergence follows by Theorem 1 of Andrews (1992) if we demonstrate stochastic equicontinuity: $\forall\epsilon>0$ there exists $\delta>0$ such that $\lim_{n\to\infty}P(\sup_\theta\sup_{\tilde{\theta}:||\tilde{\theta}-\theta||\leq\delta}|X_n^*(\theta)-X_n^*(\tilde{\theta})|>\epsilon)<\epsilon$.

Consider two possibly over-lapping cases. First, if $\lim_{n\to\infty}E[\sup_\theta||m_{n,t}^*(\theta)||]<\infty$ then $\lim_{n\to\infty}E[\sup_\theta|M_{n,t}^*(\theta)|\times I(\sup_\theta|M_{n,t}^*(\theta)|>\zeta)]\leq K(\zeta)$ for all $\zeta>0$ and some mapping $K(\zeta)\searrow 0$ as $\zeta\to\infty$. Further, Andrews' (1992) Assumption TSE holds by Lemma C.3.1. Therefore stochastic equicontinuity follows from Lemma 3 of Andrews (1992).

Second, if $\mathfrak{m}_n := \sup_\theta E[||m^*_{n,t}(\theta)||] \to \infty$ then we need only verify conditions (17) and (18) of Lemma C.3.2. Define for arbitrary $\epsilon > 0$

$$\zeta_n := P\left(\sup_\theta \sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta} \left|M^*_{n,t}(\theta) - M^*_{n,t}(\tilde{\theta})\right| > \epsilon\right)^{-1+\iota}.$$

We can always find sufficiently small $\iota > 0$ such that $\zeta_n \to \infty$ since by subadditivity, envelope bound D4 and Markov's inequality

$$P\left(\sup_\theta \sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta} \left|M^*_{n,t}(\theta) - M^*_{n,t}(\tilde{\theta})\right| > \epsilon\right) \leq K\frac{E\left(\sup_\theta \left|m^*_{i,n,t}(\theta)\right|^\iota\right)}{\mathfrak{m}_n^\iota} \leq K\mathfrak{m}_n^{-\iota} = o(1).$$

Condition (17) is trivial since since $E[\sup_\theta |M^*_{n,t}(\theta)|] \leq 1$ and $\zeta_n \to \infty$. Condition (18) follows by the construction of M_n:

$$\lim_{\delta\to 0}\lim_{n\to\infty}\left\{\zeta_n \times P\left(\sup_\theta \sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta} \left|M^*_{n,t}(\theta) - M^*_{n,t}(\tilde{\theta})\right| > \epsilon\right)\right\}$$
$$= \lim_{\delta\to 0}\lim_{n\to\infty}\left\{P\left(\sup_\theta \sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta} \left|M^*_{n,t}(\theta) - M^*_{n,t}(\tilde{\theta})\right| > \epsilon\right)^\iota\right\} = 0.$$

■

Proof of Lemma C.4. Assume for clarity θ and $m_t(\theta)$ are scalars and $m_t(\theta)$ is symmetrically distributed with two-tailed thresholds and fractiles $\{c_n(\theta), k_n\}$.

Claim (a): Distribution and equation continuity D1 and D2 imply we may assume the thresholds $c_n(\theta)$ are continuous on Θ without loss of generality: for any $\epsilon > 0$ such that $|c_n(\theta)/c_n(\tilde{\theta}) - 1| \leq \epsilon$ we can find $\delta > 0$ such that $||\tilde{\theta} - \theta|| \leq \delta$. Further, by D2 and envelope bound D4 there exists $\delta > 0$ such that for any $\kappa > 0$ and sufficiently tiny $\iota > 0$

$$P\left(\sup_\theta \sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta} \left|m_t(\theta) - m_t(\tilde{\theta})\right| > \epsilon\right) \leq K \times E\left[\sup_\theta \left|\frac{\partial}{\partial\theta}m_t(\theta)\right|^\iota\right] \times \delta^\iota \leq \kappa.$$

Therefore, by sub-additivity for arbitrary $\epsilon > 0$ and $\delta_1 > 0$ and each $\delta < \delta_1$

$$P_n(\delta) := P\left(\sup_\theta \sup_{\tilde{\theta}:||\theta-\tilde{\theta}||\leq\delta} \left|I(|m_t(\theta)| \leq c_n(\theta)) - I\left(\left|m_t(\tilde{\theta})\right| \leq c_n(\tilde{\theta})\right)\right| = 1\right)$$
$$\leq P\left(\exists\theta, \tilde{\theta} \in \Theta,\ ||\theta - \tilde{\theta}|| \ \leq \delta : |m_t(\theta)| \leq c_n(\theta) \text{ and } c_n(\tilde{\theta}) < |m_t(\tilde{\theta})|\right)$$
$$+ P\left(\exists\theta, \tilde{\theta} \in \Theta,\ ||\theta - \tilde{\theta}|| \ \leq \delta : \left|m_t(\tilde{\theta})\right| \leq c_n(\tilde{\theta}) \text{ and } c_n(\theta) < |m_t(\theta)|\right)$$
$$\leq 2 \times P\left(\exists\theta \in \Theta : |m_t(\theta)| > c_n(\theta)\left\{[1-\epsilon] - \frac{K\delta}{\inf_\theta\{c_n(\theta)\}}\right\}\right)$$
$$- 2 \times P\left(\exists\theta \in \Theta : |m_t(\theta)| \geq c_n(\theta)\left\{[1+\epsilon] + \frac{K\delta}{\sup_\theta\{c_n(\theta)\}}\right\}\right).$$

But this implies $P_n(\delta) = O(k_n/n) = o(1) \leq \epsilon$ for each ϵ and sufficiently large n by D5, non-uniqueness of the thresholds $\{c_n(\theta)\}$ and $k_n/n \to 0$. Since ϵ is arbitrary the proof is complete.

Claim (b): Recall $X^*_{n,t}(\theta) := ((n/k_n)^{1/2-\iota}))\{I_{i,n,t}(\theta) - E[I_{i,n,t}(\theta)]\}$ for tiny $\iota > 0$. $\{X^*_{n,t}(\theta) : \theta \in \Theta\}$ forms a VC class under D7, while mixing D3 and probability orders D5 ensure by measurability $\{X^*_{n,t}(\theta)\}$ is geometrically β-mixing and $L_{2+\iota}$-bounded uniformly on $1 \leq t \leq n$, $n \geq 1$ and $\theta \in \Theta$. It is now straightforward to extend Arcones and Yu's (1994: Theorem 2.1) uniform central limit theorem for stationary β-mixing sequences to geometrically β-mixing triangular arrays $\{X^*_{n,t}(\theta) : 1 \leq t \leq n\}_{n\geq 1}$ stationary over t: $\{n^{-1/2}\sum_{t=1}^{n} X^*_{n,t}(\theta) : \theta \in \Theta\} \Longrightarrow \{X(\theta) : \theta \in \Theta\}$ a Gaussian process with uniformly bounded and uniformly continuous sample paths with respect to L_2-norm.

Therefore $E[(\sup_\theta\{|n^{-1/2}\sum_{t=1}^{n} X^*_{n,t}(\theta)|\})^2] = O(1)$ by a straightforward generalization of Doukhan et al's (1995: Theorem 2) maximal inequality in lieu of mixing D2 and VC class D7[9]. ■

Proof of Lemma C.5. Apply Čižek's (2008: Lemma 2.1) argument to deduce absolute continuity of the equation distributions D1 and equation differentiability D2 ensures $\hat{Q}_n(\theta)$ is continuous and differentiable at $\hat{\theta}_n$. Therefore $\hat{Q}_n(\hat{\theta}_n) \leq \hat{Q}_n(\theta)\ \forall\theta \in \Theta$ implies

$$\hat{J}^*_n(\hat{\theta}_n)'\Upsilon_n\frac{1}{n}\sum_{t=1}^{n}\hat{m}^*_{n,t}(\hat{\theta}_n) = 0\ a.s.$$

The Lemma C.1.a asymptotic expansion for $1/n\sum_{t=1}^{n}\hat{m}^*_{n,t}(\hat{\theta}_n)$ implies we may write

$$\hat{J}^*_n(\hat{\theta}_n)'\hat{\Upsilon}_n\left\{\hat{J}^*_n(\theta_{n,*})'\left(\hat{\theta}_n - \theta^0\right) + \frac{1}{n}\sum_{t=1}^{n}\hat{m}^*_{n,t}(\theta^0)\right\} + \hat{J}^*_n(\hat{\theta}_n)'\hat{\Upsilon}_n \times o_p(r_n) = 0\ a.s.$$

for some $||\theta_{n,*} - \theta^0|| \leq ||\hat{\theta}_n - \theta^0||$ and $r_n \to 0$ arbitrarily fast.

Consistency $||\hat{\theta}_n - \theta^0|| \xrightarrow{p} 0$ under Theorem 2.1 and Jacobian consistency Lemma 2.5 ensure both $\hat{J}^*_n(\hat{\theta}_n) = J_n(1 + o_p(1))$ and $\hat{J}^*_n(\theta_{n,*}) = J_n(1 + o_p(1))$. Further $H_n^{-1} := (J_n'\Upsilon_n J_n)^{-1}$ exists given weight and Jacobian properties M1 and D6.i. Re-arrange terms and exploit the construction of V_n and $r_n \to 0$ arbitrarily fast to deduce

$$\begin{aligned} V_n^{1/2}\left(\hat{\theta}_n - \theta^0\right) &= -\left\{V_n^{1/2}H_n^{-1}J_n'\Upsilon_n\right\}\frac{1}{n}\sum_{t=1}^{n}\hat{m}^*_{n,t}(\theta^0) \times (1 + o_p(1)) + o_p(1) \\ &= A_n\sum_{t=1}^{n}\hat{m}^*_{n,t}(\theta^0) \times (1 + o_p(1)) + o_p(1). \end{aligned}$$

■

In order to prove central limit theorem Lemma C.6 we need to define a mixingale property for tail-trimmed arrays. Define for arbitrary $\xi \in \mathbb{R}^q$, $\xi'\xi = 1$:

$$z^*_{n,t}(\xi,\theta) = \xi' n^{-1/2}\Sigma_n^{-1/2}m_{n,t}(\theta)$$

Orthogonality I4 and $||\Sigma_n^{-1/2}||^{-2} \leq K||\Sigma_n||$ ensure

$$E\left[\sum_{t=1}^{n}m_{n,t}(\theta^0)\sum_{t=1}^{n}m_{n,t}(\theta^0)'\right] = n\Sigma_n + \sum_{s\neq t}E\left[m_{n,s}(\theta^0)m_{n,t}(\theta^0)'\right] = n\Sigma_n\left(1 + o\left(1\right)\right)$$

hence $E[(z^*_{n,t}(\xi,\theta^0))^2] = n^{-1/2}$ an $E[\sum_{t=1}^{n} z^*_{n,t}(\xi,\theta^0)\sum_{t=1}^{n} z^*_{n,t}(\xi,\theta^0)'] = 1 + o(1)$.

[9] Equation VC class D7 and β-mixing D3 imply bracketing number and mixing coefficient summability conditions in Doukhan et al (1995: Theorem 2) are satisfied. See eq. (2.17) in Doukhan et al (1995).

Although $\{z^*_{n,t}(\xi,\theta^0)\}$ is L_4-bounded geometrically β-mixing due to trimming and mixing D3, it is not necessarily uniformly square integrable since that requires uniformity over n. We therefore exploit central limit theory for L_2-mixingale arrays $\{x_{n,t},\Im_t\}$ with size λ: there exists an array of positive constants $\{e_{n,t}\}$ and a sequence of positive coefficients $\{\zeta_q\}$, $\zeta_q = o(q^{-\lambda})$ such that

$$\left(E\left[\left(x_{n,t} - E\left[x_{n,t}|\Im_{-\infty}^{t+q}\right]\right)^2\right]\right)^{1/2} \leq e_{n,t} \times \zeta_{q+1} = e_{n,t} \times o(q^{-\lambda}) \qquad \text{(MA)}$$
$$\left(E\left[\left(E\left[x_{n,t}\right] - E\left[x_{n,t}|\Im_{-\infty}^{t-q}\right]\right)^2\right]\right)^{1/2} \leq e_{n,t} \times \zeta_q = e_{n,t} \times o(q^{-\lambda}).$$

Since β-mixing implies α-mixing, invoke D3 and Lemma 2.1 of McLeish 1975) to deduce $\{z^*_{n,t}(\xi,\theta^0),\Im_t\}$ satisfies (MA) with constants $e_{n,t} = Kn^{-1/2}||\Sigma_n^{-1/2}||\times(E||m^*_{n,t}||^{2+\delta}_{2+\delta})^{1/(2+\delta)}$ for tiny $\delta > 0$ and arbitrary size $\lambda > 0$.

Our proof exploits a telescoping sum argument. Define positive integer sequences $\{d_n, h_n, j_n\}$ satisfying $h_n, j_n \to \infty$, $1 \leq h_n, j_n, d_n \leq n$,

$$h_n = o(c_n^{\iota}) \text{ for tiny } \iota > 0, \qquad (19)$$

$d_n = [n/h_n]$, and $j_n = o(h_n)$, and define $\digamma_{n,i} := \sigma\left(\cup_{\tau \leq ih_n}\Im_\tau\right)$ and blocks

$$Z_{n,i}(\theta) = \sum_{t=(i-1)h_n+j_n}^{ih_n} z^*_{n,t}(\xi,\theta) \text{ and } W_{n,i}(\theta) = E[Z_{n,i}(\theta)|\digamma_{n,i}] - E[Z_{n,i}(\theta)|\digamma_{n,i-1}]\ .$$

We require a generalization of McLeish'ss (1975: Lemma 2.1) maximal inequality. See Hill and Renault (2010).

LEMMA C.6.1 *If $\{x_{n,t},\Im_t\}$ forms an L_2-mixingale array with size $\lambda = 1/2$ then* $E[\max_{1\leq j\leq n}(\sum_{t=1}^{j}\{y_{n,t} - E[y_{n,t}]\})^2] = O(\sum_{t=1}^{n} E[y^2_{n,t}])$.

Remark: Notice the upper bound is based on $E[y^2_{n,t}]$ rather the constants $e^2_{n,t}$ as in McLeish (1975).

Define

$$\tilde{\sigma}^2_n := n \times \left\|\Sigma_n^{-1/2}\right\|^{-2},\ \ u^2_{n,i} := \tilde{\sigma}^2_n W^2_{n,i}, \text{ and } K_n \sim Kc_n^{2+\iota},$$

and a truncation function

$$\tilde{u}^2_{n,i}(K) := u^2_{n,i} I\left(u^2_{n,i} \leq K\right) = \tilde{\sigma}^2_n W^2_{n,i} I\left(\tilde{\sigma}^2_n W^2_{n,i} \leq K\right) \text{ and } \tilde{u}^2_{n,i} = \tilde{u}^2_{n,i}(K_n).$$

Lastly some supporting results based on the mixingale property. See Hill and Renault (2010).

LEMMA C.6.2 $\tilde{\sigma}_n^{-2}\sum_{i=1}^{d_n} E|\tilde{u}^2_{n,i} - u^2_{n,i}| = o(1)$ *and* $\tilde{\sigma}_n^{-2}\sum_{i=1}^{d_n}(\tilde{u}^2_{n,i} - u^2_{n,i}) = o_p(1)$.

LEMMA C.6.3 $\tilde{\sigma}_n^{-2}\sum_{i=1}^{d_n}(\tilde{u}^2_{n,i} - E[\tilde{u}^2_{n,i}]) = o_p(1)$.

LEMMA C.6.4 $||\sum_{i=1}^{d_n}(W^2_{n,i} - Z^2_{n,i})||_2 = o(1)$.

LEMMA C.6.5 $\sum_{i=1}^{d_n} E[Z^2_{n,i}] \stackrel{p}{\to} 1$.

Proof of Lemma C.6. We will show $\sum_{t=1}^{n} z_{n,t}^{*}(\xi,\theta^{0}) = \sum_{i=1}^{d_n} W_{n,i}(\theta^{0}) + o_p(1)$ and $\sum_{i=1}^{d_n} W_{n,i}(\theta^{0}) \xrightarrow{d} N(0,1)$. The claim then follows by the Cramér-Wold theorem. Drop the arguments ξ and θ^0.

Step 1 (decomposition): Decompose

$$\sum_{t=1}^{n} z_{n,t}^{*} = \sum_{i=1}^{d_n} W_{n,i} + \sum_{i=1}^{d_n} (Z_{n,i} - E[Z_{n,i}|\digamma_{n,i}]) + \sum_{i=1}^{d_n} E[Z_{n,i}|\digamma_{n,i-1}] + \sum_{i=1}^{d_n} \sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_{n,t}^{*} + \sum_{t=d_n h_n+1}^{n} z_{n,t}^{*} = \sum_{i=1}^{d_n} W_{n,i} + E_n.$$

$\{z_{n,t}^{*}, \Im_t\}$ satisfies (MA) with constants $e_{n,t} = Kn^{-1/2}||\Sigma_n^{-1/2}|| \times (E||m_{n,t}^{*}||_{2+\delta}^{2+\delta})^{1/(2+\delta)}$ and coefficients $\psi_q = o(q^{-\lambda})$ for any $\lambda > 0$. Define the index set $B_{n,t} = \{t : t \in \cup_{i=1}^{d_n}[(i - 1)h_n + j_n + 1, ..., ih_n]\}$. It can similarly be shown $\{z_{n,t}^{*} - E[z_{n,t}^{*}|\digamma_{n,i}], \Im_t\}_{t\in B_{n,t}}$ and $\{E[z_{n,t}^{*}|\digamma_{n,i-1}], \Im_t\}_{t\in B_{n,t}}$ satisfy (MA) with constants $e_{n,t}\psi_{j_n}^{\iota}$ and coefficients $\psi_{j_n}^{1-\iota}$ of arbitrary size. (de Jong 1997: A.7-A.12).

Maximal inequality Lemma C.6.1, stationarity under D3 and $E[(z_{n,t}^{*})^2] = n^{-1/2}$ imply

$$E\left[\left(\sum_{t=d_n h_n+1}^{n} z_{n,t}^{*}\right)^2\right] = O\left(\sum_{t=d_n h_n+1}^{n} E\left[\left(z_{n,t}^{*}\right)^2\right]\right) = o\left(1\right)$$

since $1 - [n/h_n](h_n/n) \rightarrow 0$ arbitrarily fast. Further, $E[(\sum_{i=1}^{d_n}\sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_{n,t}^{*})^2] = O(\sum_{i=1}^{d_n}\sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} E[(z_{n,t}^{*})^2]) = O(d_n j_n/n) = O(j_n/h_n) = o(1)$. Similarly, $E[(\sum_{i=1}^{d_n}\{Z_{n,i} - E[Z_{n,i}|\digamma_{n,i}]\})^2] = O(\sum_{i=1}^{d_n}\sum_{t=(i-1)h_n+j_n}^{ih_n} e_{n,t}^2\psi_{j_n}^{2\iota}) = O\left(j_n^{-2\iota\lambda}\right) = o\left(1\right)$, and $E[(\sum_{i=1}^{d_n} E[Z_{n,i}|\digamma_{n,i-1}])^2] = o(1)$. Therefore $E_n = o_p(1)$ by Chebyshev's inequality.

Step 2 (central limit theorem): $\sum_{i=1}^{d_n} W_{n,i} \xrightarrow{d} N(0,1)$ follows from Corollary 2.8 and equation (2.4) of McLeish (1974) given Steps 2.1 and 2.2 below.

Step 2.1 (LLN $\sum_{i=1}^{d_n} W_{n,i}^2 \xrightarrow{p} 1$): By the triangular inequality $|\sum_{i=1}^{d_n} W_{n,i}^2 - 1| \leq \sum_{i=1}^{5} E_{n,i}$ where

$$E_{n,1} = \left|\frac{1}{\tilde{\sigma}_n^2}\sum_{i=1}^{d_n}\left(\tilde{u}_{n,i}^2 - u_{n,i}^2\right)\right| \text{ and } E_{n,2} = \left|\frac{1}{\tilde{\sigma}_n^2}\sum_{i=1}^{d_n}\left(\tilde{u}_{n,i}^2 - E\left[\tilde{u}_{n,i}^2\right]\right)\right|$$

$$E_{n,3} = \frac{1}{\tilde{\sigma}_n^2}\sum_{i=1}^{d_n}\left|E\left[\tilde{u}_{n,i}^2\right] - E\left[u_{n,i}^2\right]\right| \text{ and } E_{n,4} = \left|\sum_{i=1}^{d_n}\left(E\left[W_{n,i}^2\right] - E\left[Z_{n,i}^2\right]\right)\right|$$

$$E_{n,5} = \left|\sum_{i=1}^{d_n} E\left[Z_{n,i}^2\right] - 1\right|.$$

Lemmas C.6.2 and C.6.3 imply $E_{n,1}$, $E_{n,2}$ and $E_{n,3}$ are $o_p(1)$, and Lemma C.6.4 and Lyapunov's inequality imply $E_{n,4} \leq ||\sum_{i=1}^{d_n}(W_{n,i}^2 - Z_{n,i}^2)||_2 = o(1)$. Finally, Lemma C.6.5 asserts $E_{n,5} = o_p(1)$.

Step 2.2 (Lindeberg $\sum_{i=1}^{d_n} E[W_{n,i}^2 I(|W_{n,i} > \epsilon)] \rightarrow 0$): We first show $E[W_{n,i}^2\tilde{\sigma}_n^2 I(W_{n,i}^2\tilde{\sigma}_n^2 > M)] \leq Kh_n^{\kappa+1}\int_M^{Kc_n^{2+\iota}} u^{-\kappa/2}du$ for any $M > 0$, tiny $\iota > 0$ and some $\kappa \in (0,2]$. Use subadditivity and the triangular inequality to deduce

$$E\left[W_{n,i}^2\tilde{\sigma}_n^2 I\left(W_{n,i}^2\tilde{\sigma}_n^2 > M\right)\right] = \int_M^\infty P\left(W_{n,i}^2\tilde{\sigma}_n^2 > u\right) du$$
$$\leq \int_M^{c_n^{2+\iota}} P\left(|Z_{n,i}|\,\tilde{\sigma}_n > (u/2)^{1/2}\right) du$$
$$+ \int_{c_n^{2+\iota}}^\infty P\left(|W_{n,i} - Z_{n,i}|\,\tilde{\sigma}_n > (u/2)^{1/2}\right) du = A_{n,i} + B_{n,i},$$

say. Moment bound D4, Markov's inequality, the construction of $z_{n,t}^*$, and subadditivity imply for some $\kappa \in (0, 2]$

$$A_{n,i} \leq \sum_{t=(i-1)h_n+j_n}^{ih_n} \int_M^{c_n^{2+\iota}} P\left(\left|z_{n,t}^*\right|\tilde{\sigma}_n > \frac{1}{h_n}(u/2)^{1/2}\right) du$$
$$\leq \sum_{t=(i-1)h_n+j_n}^{ih_n} \int_M^{c_n^{2+\iota}} P\left(K\left\|m_t\left(\theta^0\right)\right\| > \frac{1}{h_n}(u/2)^{1/2}\right) du \leq h_n^{\kappa+1}\int_M^{c_n^{2+\iota}} u^{-\kappa/2}du.$$

Next, trimming implies $|W_{n,i} - Z_{n,i}|\tilde{\sigma}_n \leq Kh_nc_n$. By Markov's inequality

$$B_{n,i} = \int_{c_n^{2+\iota}}^{Kh_n^2c_n^2} P\left(|W_{n,i} - Z_{n,i}|\,\tilde{\sigma}_n > (u/2)^{1/2}\right) du \leq K \times E\,|W_{n,i} - Z_{n,i}|\,\tilde{\sigma}_n \int_{c_n^{2+\iota}}^{Kh_n^2c_n^2} u^{-1/2}du.$$

Since $h_n = o(c_n^\iota)$ from (19) the right-hand-side is 0 for large n

Now use stationarity, $d_nh_n \sim n$, and covariance property (9.1) to deduce

$$\sum_{i=1}^{d_n} E\left[W_{n,i}^2 I\left(|W_{n,i}| > \epsilon\right)\right] = \frac{1}{\tilde{\sigma}_n^2}\sum_{i=1}^{d_n} E\left[W_{n,i}^2\tilde{\sigma}_n^2 I\left(|W_{n,i}\tilde{\sigma}_n| > \epsilon\tilde{\sigma}_n\right)\right]$$
$$\leq K\frac{d_n}{\tilde{\sigma}_n^2}h_n^{\kappa+1}\int_{\epsilon^2\tilde{\sigma}_n^2}^{Kc_n^{2+\iota}} u^{-\kappa/2}du \leq Kh_n^\kappa \int_{\epsilon^2\tilde{\sigma}_n^2}^{Kc_n^{2+\iota}} u^{-\kappa/2}du.$$

Since (9.1) and D5 together imply $\tilde{\sigma}_n^2/c_n^{2+\iota} \to \infty$ for sufficiently tiny $\iota > 0$, the right-hand-side is 0 for large n. ■

Proof of Lemma C.7. Minkowski's inequality, the M2 bound $||V_n^{1/2}|| \leq Kn^{1/2}||J_n|| \times ||\Sigma_n^{-1}||^{1/2}$, the matrix norm bound $||\Sigma_n^{-1}||^{1/2} \leq K||\Sigma_n^{-1/2}||$, $||\Sigma_n^{-1/2}||^{-1}/n^{1/2} = o(1)$ under (9.2), and the Lemma C.2 uniform approximation imply

$$\sup_{\theta\in U^0(\delta_n)}\left\{\frac{\left(\left\|V_n^{1/2}\right\|/\left\|J_n\right\|\right)\left\|\left\{\hat{m}_n^*(\theta) - \hat{m}_n^*(\theta^0)\right\} - \left\{E\left[m_{n,t}^*(\theta)\right] - E\left[m_{n,t}^*(\theta^0)\right]\right\}\right\|}{1 + \left\|V_n^{1/2}\right\| \times \left\|\theta - \theta^0\right\|}\right\}$$
$$\leq \sup_{\theta\in U^0(\delta_n)}\left\{\frac{\left(\left\|V_n^{1/2}\right\|/\left\|J_n\right\|\right)\left\|\left\{m_n^*(\theta) - m_n^*(\theta^0)\right\} - \left\{E\left[m_{n,t}^*(\theta)\right] - E\left[m_{n,t}^*(\theta^0)\right]\right\}\right\|}{1 + \left\|V_n^{1/2}\right\| \times \left\|\theta - \theta^0\right\|}\right\} + o_p(1).$$

Now apply moment expansion Lemma B.2, equation expansion Lemma C.1.a, covariance bound (9.2) and scale bound M2, to deduce

$$\sup_{\theta \in U^0(\delta_n)} \left\{ \frac{\left(\left\|V_n^{1/2}\right\| / \|J_n\|\right) \left\|\{m_n^*(\theta) - m_n^*(\theta^0)\} - \left\{E\left[m_{n,t}^*(\theta)\right] - E\left[m_{n,t}^*(\theta^0)\right]\right\}\right\|}{1 + \left\|V_n^{1/2}\right\| \times \left\|\theta - \theta^0\right\|} \right\}$$

$$\leq \sup_{\theta \in U^0(\delta_n)} \left\{ \frac{\|J_n^*(\theta) - J_n\|}{\|J_n\|} \right\} + o_p(1)$$

$$\leq \sup_{\theta \in U^0(\delta_n)} \left\{ \frac{\|J_n^*(\theta) - J_n^*\|}{\|J_n\|} \right\} + \left\{ \frac{\|J_n^* - J_n\|}{\|J_n\|} \right\} + o_p(1).$$

The first term is $o_p(1)$ by supposition D6.ii and the second term is $o_p(1)$ by Lemma 2.5. ■

Proof of Lemma C.8. The claim follows by invoking the martingale difference decomposition and law of large numbers in Steps 2 and 3 of the proof of Lemma C.6. ■

REFERENCES

[1] Agulló, J., C. Croux, S. Van Aelst (2008). The Multivariate Least-Trimmed Squares Estimator, Journal of Multivariate Analysis 99, 311-338.

[2] An, H.Z. and Z.G. Chen (1982). On Convergence of LAD Estimates in Autoregression with Infinite Variance, Journal of Multivariate Analysis 12, 335-345.

[3] An, H.Z., and F.C. Huang (1996). The Geometrical Ergodicity of Nonlinear Autoregressive Models, Statistica Sinica 6, 943-956.

[4] Andrews, D.W.K (1988). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables, Econometric Theory 4, 458-467.

[5] Andrews, D.W.K (1992). Generic Uniform Convergence, Econometric Theory 8, 241-257.

[6] Antoine, B. and E. Renault (2008a). Efficient Minimum Distance Estimation with Multiple Rates of Convergence, Working Paper, Dept. of Economics, University of North Carolina - Chapel Hill.

[7] Antoine, B. and E. Renault (2008b). Efficient GMM with Nearly-Weak Instruments, Working Paper, Dept. of Economics, University of North Carolina - Chapel Hill.

[8] Arcones, M.A. and B. Yu (1994). Central Limit Theorems for Empirical and U-Processes of Stationary Mixing Sequences, Journal of Theoretical Probability 7, 47-71 .

[9] Bassett, G.W. (1991). Equivariant, Monotonic, 50% Breakdown Estimators, American Statistician 45, 135–137.

[10] Basrak, B., R.A. Davis, and T, Mikosch (2002). Regular Variation of GARCH Processes, Stochastic Processes and their Applications 12, 908-920.

[11] Borkovec, M. and C. Klüppelberg (2001). The Tail of the Stationary Distribution of an Autoregressive Process with ARCH(1) Errors, Annals of Applied Probability 11, 1220–124..

[12] Brockwell, P.J. and D.B.H. Cline (1985). Linear Prediction of ARMA Processes with Infinite Variance, Stochastic Processes and their Applications 19, 281-296.

[13] Campbell, J. and L. Hentschel (1992). No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns, Journal of Financial Economics 31, 281-318.

[14] Carrasco M., and X. Chen (2002). Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, Econometric Theory 18, 17-39.

[15] Chen, L.-A., A. H. Welsh and W. Chan (2001). Estimators for the Linear Regression Model Based on Winsorized Observations, Statistica Sinica 11, 147-172.

[16] Chernozhukov, V. (2005). Inference for Extreme Conditional Quantile Models, with an Application to Birthweight, Working Paper, Dept. of Economics, Massachusetts Institute of Technology.

[17] Čižek, P. (2005). Least Trimmed Squares in Nonlinear Regression Under Dependence, Journal of Statistical Planning and Inference 136, 3967–3988.
[18] Čižek, P. (2008). General Trimmed Estimation: Robust Approach to Nonlinear and Limited Dependent Variable Models, Econometric Theory 24 , 1500-1529.
[19] Čižek, P. (2009). Generalized Method of Trimmed Moments, Working Paper, CentER, University of Tilburg.
[20] Cline, D.B.H. (1986). Convolution Tails, Product Tails and Domains of Attraction, Probability Theory and Related Fields 72, 529–557.
[21] Cline, D.B.H. (1989). Consistency for Least Squares Regression Estimators with Infinite Variance Data, Journal of Statistical Planning and Inference 23, 163-179.
[22] Cline, D.B.H. (2007). Regular Variation of Order 1 Nonlinear AR-ARCH models, Stochastic Processes and Theiry Applications 117, 840-861.
[23] Davidson, J. (1994). Stochastic Limit Theory. Oxford University Press: Oxford.
[24] Davis, R.A. and S.I. Resnick (1986). Limit Theory for Sample for the Sample Covariance and Correlation Functions of Moving Averages, Annals of Statistics 14, 533-558.
[25] Davis, R.A., K. Knight and J. Liu (1992). M-Estimation for Autoregressions with Infinite Variance, Stochatic Processes and Their Applications 40, 145-180.
[26] Ding, Z. and C.W.J. Granger (1996). Modeling Volatility Persistence of Speculative Returns: A New Approach, Journal of Econometrics 73, 185-215.
[27] Doukhan, P. (1994). Mixing: Properties and Examples, Lecture Notes in Statistics 85. Springer: Berlin.
[28] Doukhan, P., P. Massart, and E. Rio (1995). Invariance Principles for Absolutely Regullar Empirical Processes, Annales de l'Institut Henri Poincaré 31, 393-427.
[29] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag: Frankfurt.
[30] Engle, R. amd V. Ng (1993). Measuring and Testing the Impact of News On Volatility, Journal of Finance 48, 1749-1778.
[31] Fama, E.F., and K.R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds, Journal of Financial Economics 33, 3–5.
[32] Feller, W. (1971). An Introduction to Probability Theory and its Applications, 2 ed., Vol. 2. Wiley: New York.
[33] Finkenstadt, B., and H. Rootzén (2003). Extreme Values in Finance, Telecommunications and the Environment. Chapman and Hall: New York.
[34] Francq, C. and J.M. Zakoïan (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes, Bernoulli 10, 605–637.
[35] Goldie, C.M., and R.L. Smith (1987). Slow Variation with Remainder: Theory and Applications, Quarterly Journal of Mathematics 38, 45-71.
[36] Griffin, P.S. and W.E. Pruitt (1987). The Central Limit Problem for Trimmed Sums, Mathematical Proceedings of the Cambridge Philosophical Society 102, 329-349.
[37] Haeusler, E., and J. L. Teugels (1985). On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation, Annals of Statistics 13, 743-756.
[38] Hadi, A. and A. Luceno (1997). Maximum Trimmed Likelihood Estimators: A Unified Approach, Examples and Algorithms, Computational Statistics & Data Analysis 25, 251–272.
[39] Hall, A. (2005). Generalized Method of Moments. Oxford University Press: Oxford.
[40] Hall, P. and C.C. Heyde (1980). Martingale Limit Theory and Its Applications, Academic Press : New York.
[41] Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH Models with Heavy-Tailed Errors, Econometrica 71, 285-317.
[42] Hannan, E.J. and M. Kanter (1977). Autoregressive Processes with Infinite Variance, Journal of Applied Probability 14, 411-415.
[43] Hansen, L.P. (1982). Large Sample Properties of Method of Moments Estimators, Econo-

metrica 50, 1029-1054.
[44] Hansen, B. and S.W. Lee (1994). Asymptotic Theory for the Garch (1,1) Quasi-Maximum Likelihood Estimator, Econometric Theory 10, 29-52.
[45] He, X., J. Jurecková, R. Koenker and S. Portnoy (1990). Tail Behavior of Regression Estimators and their Breakdown Points, Econometrica 58, 1195-1214.
[46] Hill, J.B. (2009a). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation, Journal of Statistical Planning and Inference 139, 2091-2110.
[47] Hill, J.B. (2009b). Central Limit Theory for Kernel-Self Normalized Tail-Trimmed Sums of Dependent, Heterogeneous Data with Applications, Working Paper, Dept. of Economics, University of North Carolina - Chapel Hill.
[48] Hill, J.B. (2010). On Tail Index Estimation for Dependent, Heterogeneous Data, Econometic Theory 26: in press.
[49] Hill, J.B. and E. Renault (2010). Technical Appendix for "Generalized Method of Moments with Tail Trimming", University of North Carolina - Chapel Hill; www.unc.edu/∼jbhill/gmttm_tech_append.pdf.
[50] Hill, J.B. and A. Shneyerov (2009). Are There Common Values on BC Timber Sales? A Tail-Index Nonparametric Test, Working Paper, Dept. of Economics, University of North Carolina - Chapel Hill.
[51] Honoré, B. (1992). Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects, Econometrica 60, 533-565.
[52] Horowitz, J. L. (1998). Bootstrap Methods for Covariance Structures, Journal of Human Resourses 33, 39-61.
[53] Hsing, T. (1991). On Tail Index Estimation Using Dependent Data, Annals of Statistics 19, 1547-1569.
[54] Huber, P.J. (1964). Robust Estimation of Location Parameter, Annals of Mathematical Statistics 35 73-101.
[55] Ibragimov, I. A. and Y. V. Linnik (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhof: Berlin.
[56] Jensen, S.T. and A. Rahbek (2004). Asymptotic Normality of the QML Estimator of ARCH in the Nonstationary Case, Econometrica 72, 641-646.
[57] Jurecková, J. and P.K. Sen (1996). Robust Statistical Procedures: Asymptotics and Interrelations. Wiley: New York.
[58] Kan., S., and A. Lewbel (2007). Weighted and Two-Stage Least Squares Estimations of Semiparamertic Truncated Regression Models, Econometric Theory 23, 309-347.
[59] Knight, K. (1987). Rate of Convergence of Centred Estimates of Autoregressive Parameters for Infinite Variance Autoregressions, Journal of Time Series Analysis 8, 51-60.
[60] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag: New York.
[61] Ling, S. (2005). Self-Weighted LAD Estimation for Infinite Variance Autoregressive Models, Journal of the Royal Statistical Society Series B 67, 381–393.
[62] Ling, S. (2007). Self-Weighted and Local Quasi-Maximum Likelihood Estimators for ARMA-GARCH/IGARCH Models, Journal of Econometrics 150, 849-873.
[63] Linton, O., J. Pan and H. Wang (2010). Estimation for a Non-stationary Semi-Strong GARCH(1,1) Model with Heavy-Tailed Errors, Econometric Theory 26, 1-28.
[64] Lumsdaine, R. (1995). Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation, Journal of Business and Economic Statistics. 13, 1-10.
[65] Lumsdaine, R. (1996). Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, Econometrica 64, 575-596.

[66] Mandelbrot, B. (1963). The Variation of Certain Speculative Prices, Journal of Business 36, 394-419.
[67] McFadden D.L. (1989). A Method of Simulated Moments for Estimation of Multinomial Discrete Response Models Without Numerical Integration, Econometrica 57, 995-1026.
[68] McLeish, D.L. (1974). Dependent Central Limit Theorems and Invariance Principles, Annals of Probability 2, 620-628.
[69] McLeish, D.L. (1975). Maximal Inequality and Dependent Strong Laws, Annals of Probability 3, 829-839.
[70] Meitz M. and P. Saikkonen (2008). Stability of Nonlinear AR-GARCH Models, Journal of Time Series Analysis 29, 453-475.
[71] Mikosch, T. and C. Stărică (2000). Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process, Annals of Statistics 28, 1427-1451 .
[72] Naveau, P. (2003). Almost Sure Relative Stability of the Maximum of a Stationary Sequence, Advances in Applied Probability 35, 721-736.
[73] Neykov, N.M. and P.N. Neytchev (1990). A Robust Alternative of the Maximum Likelihood Estimator, Short Communications of Compstat, Dubrovnik 1990, 99–100.
[74] Newey, W.K., and D.L. McFadden (1994). Large Sample Estimation and Hypothesis Testing, in: R. Engle and D. McFadden, eds., Handbook of Econometrics, Vol. 4. Amsterdam: North-Holland.
[75] Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators, Econometrica 57, 1027-1057.
[76] Peng, L. and Q. Yao (2003). Least Absolute Deviations Estimation for ARCH and GARCH Models, Biometrika 90, 967-975.
[77] Pollard, D. (1984). Convergence of Stochastic Processes. Springer-Verlang New York.
[78] Powell, J.L. (1986). Symmetrically Trimmed Least Squares Estimation for Tobit Models, Econometrica 54, 1435-1460.
[79] Renault, E. (1997). Économétrie de la Finance: La Méthode des Moments Généralisés, in Encyclopédie des Marchés Financiers, edited by Y. Simon, Economica, Paris, 1997, 330-407.
[80] Resnick, S.I. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag: New York.
[81] Resnick, S.I. (1997). Heavy Tail Modeling and Traffic Data, Annals of Statistics 25, 1805-1849.
[82] Ronchetti, E. and F. Trojani (2001). Robust Inference with GMM Estimators, Journal of Econometrics 101, 37-69.
[83] Rousseeuw, P.J. (1985). Multivariate Estimation with High Breakdown Point, In W. Grossman, G. Pflug, I. Vincze, and W. Wertz (eds.), Mathematical Statistics and Applications, volume B, 283-297, Reidel: Berlin.
[84] Ruppert, D. and J. Carroll (1980). Trimmed Least Squares Estimation in the Linear Model, Journal of the American Statistical Association 75, 828-838.
[85] Stock, J.H. and J.H. Wright (2000). GMM with Weak Identification, Econometrica 68, 1055-1096.
[86] Tableman, M. (1994). The Asymptotics of the Least Trimmed Absolute Deviations (LTAD) Estimator, Statistics and Probability Letters 19, 329-337.
[87] van der Vaart, A.W., and J.A. Wellner (1996). Weak Convergence and Empirical Processes, Springer-Verlang: New York.
[88] Vapnik, V. N., and A. YA. Cervonenkis (1971). On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities, Theory of Probability and Its Application 16, 264-280.

TABLE 3 : Location, AR

Location				AR			
$\epsilon_t \sim P_{2.5}$, $\kappa_y = 2.5^a$, $\theta_r^0 = 1^b$				$\epsilon_t \sim P_{2.5}$, $\kappa_y = 1.5$, $\theta_r^0 = .9$			
$\hat{\theta}$	Mean[c]	λ[d]	KS[e]	$\hat{\theta}$	Mean	λ	KS
GMTTM[f]	.994 (.03)	.47	.101	GMTTM	.901 (.01)	.48	.051
GMM	1.00 (.08)	-	.076	GMM	.898 (.01)	-	.086
QML	.999 (.04)	-	.121	QML	.899 (.01)	-	.132

Location				AR			
$\epsilon_t \sim P_{1.5}$, $\kappa_y = 1.5$, $\theta_r^0 = 1$				$\epsilon_t \sim P_{1.5}$, $\kappa_y = 1.5$, $\theta_r^0 = .9$			
$\hat{\theta}$	Mean	λ	KS	$\hat{\theta}$	Mean	λ	KS
GMTTM	.994 (.13)	.22	.051	GMTTM	.900 (.01)	.32	.069
GMM	.939 (.21)	-	.172	GMM	.900 (.01)	$-$	.187
QML	1.01 (.15)	-	.272	QML	.889 (.01)		.253

a. True parameter value.
b. Error distribution ($P_{1.5}$ or $N_{0,1}$), and moment supremum or tail index of y_t: $E|y_t|^{\kappa_y} = \infty$ and $E|y_t|^{\kappa_y - \iota} < \infty$.
c. Simulation mean of parameter estimation (square root of mse in parentheses).
d. Kolmogorov-Smirnov test p-value. In the case of GMTTM, KS p-values are evaluated at that $k_n = [n^{\lambda}]$ which minimizes KS. 1%, 5%, 10% critical values: .136, .122, .107.
e. KS minimizing λ in the trimming fractile $k_n = [n^{\lambda}]$, $\lambda \in \{.01, .02, ..., .99\}$.
f. GMTTM and GMM are performed in two steps with the efficient weight and QMLE plug-in.

TABLE 4 :ARCH

$\epsilon_t \sim N_{0,1}$, $\kappa_y = 4.65$, $\theta_r^0 = .5$					$\epsilon_t \sim N_{0,1}$, $\kappa_y = 3.8$, $\theta_r^0 = .6$			
$\hat{\theta}$	Mean	λ	KS		$\hat{\theta}$	Mean	λ	KS
GMTTM[e]	.525 (.32)	.76	.082		GMTTM	.605 (.26)	.40	.075
GMM	.457 (.13)	-	.228		GMM	.600 (.17)	-	.162
QML	.492 (.07)	-	.051		QML	.599 (.07)	-	.082

$\epsilon_t \sim P_{2.5}$, $\kappa_y = 1.8$, $\theta_r^0 = .6$					$\epsilon_t \sim P_{2.1}$, $\kappa_y = 1.4$, $\theta_r^0 = .6$			
$\hat{\theta}$	Mean	λ	KS		$\hat{\theta}$	Mean	λ	KS
GMTTM	.602 (.38)	.50	.093		GMTTM	.582 (.34)	.38	.105
GMM	.547 (.22)	-	.289		GMM	.542 (.26)	-	.213
QML	.620 (.27)	-	.110		QML	.683 (.18)	-	.414

TABLE 5 : GMTTM Convergence Rate

Model	ϵ_t	θ_r^0	κ_ϵ	κ_y	$m_t(\theta)$	b	$\hat{b}$
LOCATION	$P_{2.5}$	1	2.50[a]	2.50	LS[b]	.000[c]	.031±.051[d]
LOCATION	$P_{1.5}$	1	1.50	1.50	LS	-.129	-.130±.024
AR	$P_{2.5}$	.9	2.50	2.50	LS	.000	.027±.039
AR	$P_{1.5}$	.9	1.50	1.50	LS	.112	.116±.012
ARCH	$N_{0,1}$	.5	∞	4.65	LS	.000	-.094±.162
ARCH	$N_{0,1}$	.6	∞	3.80	QML	.000	.195±.422
ARCH	$P_{2,1}$	.6	2.5	1.80	QML	-.180	-.387±.306
ARCH	$P_{2,1}$	.6	2.5	1.40	QML	-.280	-.412±.393

a. Tail indices (moment suprema) of ϵ_t and y_t.
b. Equation type. Infinite kurtosis ARCH are estimated with QML equations.
c. Slope parameter b in the rate of convergence regression $\ln(\hat{s}_{n,r}^{-1}/n^{1/2}) \sim a + b\ln(n)$, where $\hat{s}_{n,r}^2$ is computed by GMTTM with the KS minimizing λ when $k_n = [n^\lambda]$ and $n = 1000$. The true values are: Location: $b = -(1-\lambda)(1/\kappa_y - 1/2)$; AR:: $b = (1-\lambda)(1/\kappa_y - 1/2)$; ARCH with QML equations and error with index $\kappa_\epsilon \in (2,4]$: $b = -(1-\lambda)(2/\kappa_\epsilon - 1/2)$.
d. Last squares asymptotic 95% band.

TABLE 6 : GARCH, TARCH, QARCH

GARCH, $N_{0,1}$, $\kappa_y = 4.1$, $\theta_r^0 = .6$				TARCH, $N_{0,1}$, $\kappa_y = 5.25$, $\theta_r^0 = .6$			
$\hat{\theta}$	Mean	KS[a]	λ[b]	$\hat{\theta}$	Mean: .6	KS	λ
GMTTM[c]	.595 (.17)	.064	.26	GMTTM	.617 (.16)	.083	.29,.16
GMM	.577 (.20)	.177	-	GMM	.536 (.15)	.269	-
QML	.597 (.08)	.098	-	QML	.595 (.09)	.073	-
IGARCH, $N_{0,1}$, $\kappa_y = 2.0$, $\theta_r^0 = .6$				TIARCH, $N_{0,1}$, $\kappa_y = 3.4,$, $\theta_r^0 = 1$			
$\hat{\theta}$	Mean	KS	λ	$\hat{\theta}$	Mean	KS	λ
GMTTM	.599 (.17)	.078	.31	GMTTM	.998 (.24)	.064	.41,.24
GMM	.628 (.23)	.273	-	GMM	.839 (.39)	.284	-
QML	.586 (.20)	.262	-	QML	1.00 (.13)	.082	-
GARCH, $P_{2.5}$, $\kappa_y = 1.5$, $\theta_r^0 = .6$				TARCH, $P_{2.5}$, $\kappa_y = 2.6$, $\theta_r^0 = .5$			
$\hat{\theta}$	Mean	KS	λ	$\hat{\theta}$	Mean	KS	λ
GMTTM	.594 (.23)	.094	.35	GMTTM	.619 (.29)	.104	.67,.09
GMM	.264 (.35)	.448	-	GMM	.270 (.20)	.532	-
QML	.605 (.18)	.569	-	QML	.516 (.38)	.323	-
QARCH, $N_{0,1}$, $\kappa_y = 3.5$, $\theta_r^0 = .8$				QIARCH, $N_{0,1}$, $\kappa_y = 2.0$, $\theta_r^0 = 1$			
$\hat{\theta}$	Mean	KS	λ	$\hat{\theta}$	Mean	KS	λ
GMTTM	.798 (.09)	.066	.32,.10	GMTTM	1.01 (.09)	.050	.57,.37
GMM	.673 (.09)	.454	-	GMM	.721 (.09)	.643	-
QML	.896 (.66)	.389	-	QML	.997 (.64)	.287	-

a. KS critical values for test size 1%, 5%, 10% are .136, .122, .107.
b. The KS minimizing λ for symmetric data generating processes (GARCH), or the KS minimizing pair $\{\lambda_1, \lambda_2\}$ for asymmetric processes (TARCH, QARCH).
c. GMTTM and GMM are performed in two steps with an efficient weight and QMLE plug-in.