

Generalized Method of Moments with Tail Trimming

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

Dec. 11, 2009

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} -consistency is achievable depending on regressor and error tail thickness and dependence; 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 Hansen (1982), Renault (1997) and Hall (2005).

1.1 TAIL TRIMMED ESTIMATING EQUATIONS

*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 Čížek, Feico Drost, Eric Ghysels, Shakeeb Khan, Denis Pelletier, Liang Peng, Bas Werker, and seminar participants at the Triangle Econometrics Conference, the Joint Statistical Meeting 2008, 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, and Center at the University of Tilburg for helpful comments.

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

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

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

$$E[m_t(\theta)] = 0 \text{ if and only if } \theta = \theta_0 \text{ for unique } \theta_0 \in \Theta.$$

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

$$y_t = h_t \epsilon_t, \quad h_t^2 = \alpha_0 + \beta_0 y_{t-1}^2, \quad \theta_0 = [\alpha_0, \beta_0]', \quad \epsilon_t \stackrel{iid}{\sim} (0, 1) \text{ and } \mathfrak{F}_t := \sigma(y_\tau : \tau \leq t) \quad (1)$$

with equations

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

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)

$$\sqrt{n}(\hat{\theta} - \theta_0) = A_n \times \frac{1}{\sqrt{n}} \sum_{t=1}^n m_t(\theta_0) + o_p(1) \text{ for some } A_n \in \mathbb{R}^{k \times q},$$

so asymptotics are grounded on $\sum_{t=1}^n m_t(\theta_0)$.

Clearly each $E[m_{i,t}^2(\theta_0)] < \infty$ which requires ϵ_t and y_t to have finite 4^{th} and 8^{th} moments, respectively, along with few additional assumptions, ensures Gaussian asymptotics. This rules out mildly heavy-tailed shocks, and 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 (e.g. NYMEX) and low risk (e.g. U.S. Treasury Bill) asset returns, and factor premia (e.g. market capitalization and book-to-price ratio); VARX for causality modeling of financial and macroeconomic returns; and random volatility with over identifying conditions. See French and Fama (1996), Ding and Granger (1996), Mikosch and Střaricř (2000), and Embrechts et al (2003).

Although GMM with a non-Gaussian limit is certainly achievable in the manner of least squares (e.g. Hannan and Kanter 1977, Knight 1987, Chan and Tran 1989, Cline 1989), 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:

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

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

$$m_{i,t}^{(a)}(\theta) := |m_{i,t}(\theta)| \quad \text{and} \quad m_{i,(1)}^{(a)}(\theta) \geq \dots \geq m_{i,(n)}^{(a)}(\theta) \geq 0.$$

Any equation $m_{i,t}(\theta_0)$ which may have an infinite variance, $m_{i,t}(\theta)$ is trimmed between its lower $k_{1,i,n}/n^{th}$ and upper $k_{2,i,n}/n^{th}$ sample quantiles:

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

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

and $I(A) = 1$ if A is true, and 0 otherwise¹. 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 $k_{i,n} \rightarrow \infty$ and $k_{i,n}/n \rightarrow 0$

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

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

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

As long as $\{m_t(\theta_0), \mathfrak{S}_t\}$ forms an adapted martingale difference sequence for some sequence of σ -fields $\{\mathfrak{S}_t\}$, and $E|m_{i,t}(\theta_0)|^p < \infty$ for some $p > 0$, standard smoothness conditions ensure

$$V_n^{1/2} \left(\hat{\theta}_n - \theta_0 \right) \xrightarrow{d} N(0, I_k)$$

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

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 asymptotics. In particular, $\hat{\theta}_n$ is \sqrt{n} -consistent if all equations have a finite variance (e.g. finite kurtosis GARCH), while sub-, exact- or super- \sqrt{n} -consistency may arise in heavy tailed cases depending on the 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 precisely due to trimming. 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} -consistency for an autoregression, 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 $k_{j,i,n}/n \rightarrow \lambda_{j,i} \in (0, 1)$ for each equation i and tail j . This is the standard in the robust M-estimation and Method of Moments literatures where symmetry is imposed $\lambda_{1,i} = \lambda_{2,i}$ (see below). In this

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

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 problem of fractile $\{k_{1,n}, k_{2,n}\}$ selection in practice since that deviates from the central theme of Gaussian asymptotics under tail-trimming. Nevertheless, we provide substantial detail on reasonable rates $k_{1,n} \rightarrow \infty$ and $k_{2,n} \rightarrow \infty$ 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 for GARCH models and Least Trimmed Squares in the robust estimation literature. Francq and Zakoian (2004) prove the QMLE is asymptotically normal for strong GARCH and ARMA-GARCH under $E(\epsilon_t^4) < \infty$. Extant results cover stationary and therefore IGARCH models under $E|\epsilon_t|^p < \infty$ for at least $p \geq 4$. See, also, Hansen and Lee (1994), Lumsdaine (1996) and Jensen and Rahbek (2004) for results covering stationary and non-stationary cases.

Linton et al (2008) 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), only a few results permit fully nonlinear models with heavy tails. Most focus on thin tailed environments with outliers under data contamination; most concern M-estimator frameworks; and when trimming, truncation or weighting are employed only non-tail data quantiles are considered. Evidently tail-trimming appears only in the robust location literature and has not been applied to extremum estimation.

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 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 $s_t(\theta)$, but instead the data y_t . Linear autoregressive and GARCH models are separately covered allowing $E[\epsilon_t^2] < \infty$ and $E|y_t|^p < \infty$ for some $p > 0$.

Hadi and Luceno (1997) characterize the Maximum Trimmed Likelihood [MTL] estimator but do not provide a formal theory. Čížek (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 covariance estimation is neglected. 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, but 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 Winsorizing observations.

Fundamental short-comings of trimming criterion equations $s_t(\theta)$ by a fixed quantile of $s_t(\theta)$ are super- \sqrt{n} -consistency is impossible for stationary data; asymptotic normality cannot be achieved when regressors are heavy tailed; and asymmetric models are not covered. See Section 2.6 for direct comparisons of the LTS and MTL estimators with GMTTM.

A few results are couched in method of moments. Čížek (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 couded in GMTM 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 \xrightarrow{p} and \xrightarrow{d} 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 . LLN = law of large numbers. 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 $m_t(\theta)$. See Sections 3 and 4 applications concerning specific models.

2.1 TAIL-TRIMMING

There is no impact on asymptotics if an equation $m_{i,t}(\theta_0)$ has a finite variance but is trimmed, due to the asymptotic negligibility of tail trimming. Thus, in order not to repeat ourselves and to reduce notation, we simply tail-trim *all* equations. The reader can feel safe to leave any equation $m_{i,t}(\theta_0)$ untrimmed if it is known to have a finite variance.

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

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

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

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

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

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 $\hat{m}_t^*(\theta)$ in (2) or (3), but theory centers around deterministically trimmed equations:

$$\begin{aligned} m_t^*(\theta) &:= [m_{i,t}(\theta) \times I(-l_{i,n}(\theta) \leq m_{i,t}(\theta) \leq u_{i,n}(\theta))]_{i=1}^q \\ &= [m_{i,t}(\theta) \times I_{i,t}(\theta)]_{i=1}^q. \end{aligned} \quad (5)$$

Although $m_t(\theta)$ identifies θ_0 , we can only assume $m_t^*(\theta)$ identifies some sample-size dependent $\theta_{n,0}$:

$$E[m_t^*(\theta)] = 0 \text{ if and only if } \theta = \theta_{n,0} \in \Theta.$$

Identification of θ_0 by the trimmed equations $m_t^*(\theta)$,

$$E [m_t^*(\theta_0)] \rightarrow 0,$$

however, is easily guaranteed for arbitrarily many threshold sequences $\{l_{i,n}(\theta_0), u_{i,n}(\theta_0)\}$ that satisfy (4) since tail trimming is asymptotically negligible. This 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, ensuring eventual identification of θ_0 . If the DGP of $\{m_t(\theta_0)\}$ is symmetric then $\theta_{n,0} = \theta_0$ for any thresholds $l_{i,n}(\theta) = u_{i,n}(\theta)$ and fractile $k_{1,i,n} = k_{2,i,n}$ under (4). This holds for linear-in-parameters models with symmetric shocks like autoregressions, GARCH, and so on.

Write compactly throughout

$$\begin{aligned} c_{i,n}(\theta) &:= \max\{l_{i,n}(\theta), u_{i,n}(\theta)\} \quad \text{and} \quad c_n(\theta) = \max_{1 \leq i \leq q} \{c_{i,n}(\theta)\} \\ k_{i,n} &= \max\{k_{1,i,n}, k_{2,i,n}\} \quad \text{and} \quad k_n = \max_{1 \leq i \leq q} \{k_{i,n}\} \\ \{l_{i,n}, u_{i,n}, c_{i,n}\} &= \{l_{i,n}(\theta_{n,0}), u_{i,n}(\theta_{n,0}), c_{i,n}(\theta_{n,0})\}. \end{aligned}$$

2.2 ASSUMPTIONS

Let $\{\Upsilon_n\}$ be a sequence of positive semi-definite matrices $\Upsilon_n \in \mathbb{R}^{q \times q}$. The population GMTTM criterion function is

$$Q_n(\theta) = E [m_t^*(\theta)]' \times \Upsilon_n \times E [m_t^*(\theta)]$$

with sample version

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

The GMTTME solves

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{arginf}} \{\hat{Q}_n(\theta)\}.$$

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 are

$$\Sigma_n(\theta) := E [m_t^*(\theta) m_t^*(\theta)'] \quad \text{and} \quad \Sigma_n := \Sigma_n(\theta_0)$$

$$\mathbf{m}_n = \sup_{\theta} E [\|m_t^*(\theta)\|];$$

the population and sample Jacobia are

$$\begin{aligned} J_n(\theta) &:= \frac{\partial}{\partial \theta} E [m_t^*(\theta)] \in \mathbb{R}^{q \times k} \quad \text{and} \quad J_n = J_n(\theta_0) \\ J_t^*(\theta) &:= \left[\frac{\partial}{\partial \theta} m_{i,t}(\theta) \times I_{i,t}(\theta) \right]_{i=1}^q \quad \text{and} \quad J_n^*(\theta) := \frac{1}{n} \sum_{t=1}^n J_t^*(\theta) \\ \hat{J}_t^*(\theta) &:= \left[\frac{\partial}{\partial \theta} m_{i,t}(\theta) \times \hat{I}_{i,t}(\theta) \right]_{i=1}^q \quad \text{and} \quad \hat{J}_n^*(\theta) := \frac{1}{n} \sum_{t=1}^n \hat{J}_t^*(\theta); \end{aligned}$$

and the Hessian and scale are

$$H_n(\theta) := J_n(\theta)' \Upsilon_n J_n(\theta) \in \mathbb{R}^{k \times k} \quad \text{and} \quad H_n := H_n(\theta_0)$$

$$V_n(\theta) := n \times H_n(\theta) [J_n'(\theta) \Upsilon_n \Sigma_n(\theta) \Upsilon_n J_n(\theta)]^{-1} H_n(\theta) \quad \text{and} \quad V_n := V_n(\theta_0).$$

Three sets of assumptions ensure identification for θ_0 ; $\hat{\theta}_n$ can be expressed as a linear function of $\sum_{t=1}^n \hat{m}_t^*(\theta_0)$; $\sum_{t=1}^n \hat{m}_t^*(\theta)$ is sufficiently close to $\sum_{t=1}^n m_t^*(\theta)$ uniformly on Θ ; $\sum_{t=1}^n m_t^*(\theta_0)$ is asymptotically normal; and $\hat{J}_n^*(\theta)$ is consistent for J_n for some plug-in for θ_0 . Most are versions of standard regulatory conditions contoured to heavy tailed data under tail trimming. The remaining are easily verified for linear-in-parameters models. See Section 4.

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

M1 (weight). Υ_n is positive semi-definite for each $n \geq N$; $\inf_{n \geq N} \{\|\Upsilon_n\|\} > 0$; $\sup_{n \geq 1} \{\|\Upsilon_n\|\} \leq K$; and $\|\hat{\Upsilon}_n - \Upsilon_n\| \xrightarrow{p} 0$ and $\|\Upsilon_n - \Upsilon_0\| \rightarrow 0$ for some positive semi-definite Υ_0 , $0 < \|\Upsilon_0\| < \infty$.

M2 (scale). $K n^{1/2} \|J_n\| \times \|\Sigma_n^{-1}\|^{1/2} \geq \|V_n^{1/2}\| \rightarrow \infty$.

M3 (covariance). There exists $N > 0$ such that Σ_n is positive definite for each $n \geq N$. Moreover $\liminf_{n \geq N} \inf_{\theta} \{\lambda_{n,i}(\theta)\} > 0$.

Remark 3: Weight property M1 is standard. Norm property M2 is used solely to simplify bounding arguments and holds in the efficient weight case (see Section 2.3). Positive definiteness M3 is imposed 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 for $\theta_{n,0}$). $\{m_t(\theta), \mathfrak{S}_t\}$ forms a adapted martingale difference sequence if and only if $\theta = \theta_0$, a unique interior point of compact $\Theta \subset \mathbb{R}^k$; $\{m_t^*(\theta_{n,0}), \mathfrak{S}_t\}$ forms for each n an adapted martingale difference array; $E[m_t^*(\theta)] = 0$ if and only if $\theta = \theta_{n,0}$.

I2 (identification for θ_0). $E[m_t^*(\theta_0)] = o(\|\Sigma_n^{-1/2}\|^{-1}/n^{1/2})$.

I3 (smoothness). $\inf_{n \geq N} \inf_{\|\theta - \theta_0\| > \delta} \{\mathfrak{m}_n^{-1} \|E[m_t^*(\theta)]\|\} > 0$ for tiny $\delta > 0$ and some $N \geq 1$.

Remark 1: The martingale difference component of I1 is a convenience for presenting central limit theory. In cross-sections far weaker conditions can be imposed.

Remark 2: If the DGP is symmetric then $E[m_{i,t}^*(\theta_0)] = 0$ for any thresholds $l_{i,n}(\theta_0) = u_{i,n}(\theta_0)$. Otherwise, as long as the trimmed equation covariance satisfies $\|\Sigma_n^{-1/2}\|^{-1}/n^{1/2} = o(1)$ we are assured $m_t^*(\theta)$ eventually identifies θ_0 sufficiently fast. In turn $\|\Sigma_n^{-1/2}\|^{-1}/n^{1/2} = o(1)$ holds trivially in thin-tailed cases, and heavy-tailed cases when the thresholds re bounded $c_n(\theta_0) = o(n^{1/2})$ as in D5 below. The latter is easily guaranteed by trimming sufficiently many tail observations: see Section 4.

Remark 3: Versions of smoothness I3 are standard (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. Thus, while $E[m_t^*(\theta)]$ need not

be well defined on Θ -a.e. asymptotically, $E[m_t^*(\theta)]/m_n$ is always well defined. This matters for a proof of consistency $\hat{\theta}_n \xrightarrow{P} \theta_0$ since consistency requires a uniform LLN $\sup_{\theta} \|1/n \sum_{t=1}^n \{m_t^*(\theta) - E[m_t^*(\theta)]\}\| = o_p(m_n)$. Consider an AR(1) $y_t = \theta_0 y_{t-1} + \epsilon_t$ with $|\theta_0| < 1$, $\mathfrak{S}_t = \sigma(y_\tau : \tau \leq t)$, martingale difference innovations $E[\epsilon_t | \mathfrak{S}_{t-1}] = 0$ with infinite variance $E[\epsilon_t^2] = \infty$ and one equation $m_t(\theta) = (y_t - \theta y_{t-1}) y_{t-1}$. Then $E[m_t(\theta_0) | \mathfrak{S}_{t-1}] = 0$ a.s. hence $E[m_t(\theta_0)] = 0$, but in general $m_t(\theta) = -(\theta - \theta_0) \times y_{t-1}^2$ is non-integrable for any coefficient $\theta \neq \theta_0$.

Remark 4: If $m_t(\theta)$ is uniformly integrable on Θ -a.e. then I3 reduces to $\inf_{n \geq N} \inf_{\|\theta - \theta_0\| > \delta} \{ \|E[m_t^*(\theta)]\| \} > 0$. An interesting case where $m_t(\theta)$ is both uniformly integrable on Θ -a.e. and not uniformly square integrable at θ_0 is a stationary AR(1) $y_t = \theta_0 y_{t-1} + \epsilon_t$ with ARCH(1) error $\epsilon_t = (\alpha + \beta \epsilon_{t-1}^2)^{1/2} u_t$, $u_t \stackrel{iid}{\sim} N(0, 1)$ 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 $m_t(\theta) = -(\theta - \theta_0) \times y_{t-1}^2$ is integrable on Θ -a.e. but $E[m_t^2(\theta_0)]$ does not exist.

The last set concerns properties of the equations $m_t(\theta)$ and the random Jacobian matrices J_t^* and \hat{J}_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 α -mixing: $\alpha_l := \sup_{A \subset \mathfrak{S}_t^{+\infty}, B \subset \mathfrak{S}_{t-l}^{-l}} |P(A \cap B) - P(A)P(B)| = o(\rho^l)$ 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_ν -bounded.*

D5 (thresholds and fractiles). *$k_{i,n} = O(n^\delta)$ for some $\delta \in (0, 1)$; $\sup_{\theta} |(n/k_{1,i,n})P(m_{i,t}(\theta) < -l_{i,n}(\theta)) - 1| = O(1/k_{1,i,n}^{1/2})$ and $\sup_{\theta} |(n/k_{2,i,n})P(m_{i,t}(\theta) > u_{i,n}(\theta)) - 1| = O(1/k_{2,i,n}^{1/2})$; $\inf_{\theta} \{c_n(\theta)\} \rightarrow \infty$ and $\sup_{\theta} \{c_n(\theta)\} = o(n^{1/2})$.*

D6 (Jacobia).

i. $\sup_{\theta} \|J_n(\theta)\| < \infty$ for each n ; $\{J_n(\theta), J_n^(\theta), E[J_t^*(\theta)], \hat{J}_n^*(\theta), E[\hat{J}_t^*(\theta)]\}$ 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 $\delta_n \rightarrow 0$.*

D7 (Indicator Class). *$\{I_{i,t}(\theta) : \theta \in \Theta\}$ form Vapnik-Chervonenkis [VC] classes of functions.*

Remark 1: Distribution continuity D1 and equation differentiability D2 reduce generality, but simplify key arguments showing consistency $Q_n(\hat{\theta}_n) \xrightarrow{P} 0$ and $\hat{m}_t^*(\theta)$ approximates $m_t^*(\theta)$ sufficiently fast.

Remark 2: Mixing D3 promotes uniform laws for $m_t^*(\theta)$ and $\hat{m}_t^*(\theta) - m_t^*(\theta)$. Geometric decay keeps notation simple, covering nonlinear AR-nonlinear GARCH (An and Huan 1996, Carrasco and Chen 2002, Meitz and Saikonen 2008), and can be relaxed to absolute regularity with long memory (Arcones and Yu 1994).

Remark 3: The D5 fractile bound $k_{i,n} = O(n^\delta)$ is used throughout the extreme value and tail-trimming literatures (Leadbetter et al 1983, Čížek 2008, Hill 2010a,b inter alia). It does not reduce generality for a large variety of tails that belong to the subexponential class, including at least Paretian, Weibull, log-logistic, and Fréchet (Hill 2010b). Since $(n/k_{1,i,n})P(|m_{i,t}(\theta)| < -l_{i,n}(\theta)) \rightarrow 1$ and $(n/k_{2,i,n})P(|m_{i,t}(\theta)| > u_{i,n}(\theta)) \rightarrow 1$ by construction, the D5 probability convergence orders merely sharpen the rates of

approximation. This is required for uniform asymptotics concerning the trimming indicators $I_{i,t}(\theta)$ derived from uniform laws for the tail arrays $\{1 - I_{i,t}(\theta)\}$. The rates hold for a large array of probability tails that satisfy second order regular variation or slow variation with remainder (e.g. Smith 1982, Leadbetter et al 1983, Haeusler and Teugels 1985, Goldie and Smith 1987, Hill 2009).

Remark 4: The D5 threshold bound $\sup_{\theta} \{c_n(\theta)\} = o(n^{1/2})$ ensures sufficiently many equations are trimmed for asymptotic normality of $\sum_{t=1}^n m_t^*(\theta_0)$, and uniform laws for $\sum_{t=1}^n \{m_t^*(\theta) - E[m_t^*(\theta)]\}$ and $\sum_{t=1}^n \{\hat{m}_t^*(\theta) - m_t^*(\theta)\}$. Notice $\max_{1 \leq t \leq n} \{||m_t^*(\theta_0)||\} = o_p(n^{1/2})$ intuitively initiates the relative stability property of maxima of uniformly square integrable weakly dependent sequences, cf. Leadbetter et al (1983) and Naveau (2003).

Remark 5: The D4 moment bounds, D6 Jacobia properties and D7 indicator class help prove $1/n \sum_{t=1}^n \{\hat{m}_t^*(\theta) - m_t^*(\theta)\} = o_p(1)$ 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 class². 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 $\hat{\theta}_n \xrightarrow{p} \theta_0$.*

Remark: The proof reveals $||\hat{\theta}_n - \theta_{n,0}|| = o_p(1)$ for the sequence $\{\theta_{n,0}\}$ identified by $E[m_t^*(\theta)] = 0$ is an implication of D1-D7, I1, I3, and M1-M3. Identification I2 then crucially ensures correct identification of θ_0 : $||\theta_{n,0} - \theta_0|| = o(1)$.

The rate $V_n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(I_k)$ can similarly be shown from first principles. We present the result here for reference, but it is interesting to note that we do not actually require it for plug-in arguments. For example, the Jacobian $\hat{J}_n^*(\hat{\theta}_n)$ in Lemma 2.5 and in the proof of normality Theorem 2.3 only requires some $\tilde{\theta}_n$ to satisfy $||\tilde{\theta}_n - \theta_0|| = o_p(1)$.

THEOREM 2.2 *Under D1-D7, I1-I3 and M1-M3 $V_n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(I_k)$.*

The most important result of this paper follows: the GMTTME is asymptotically normal with rate characterized by $V_n^{1/2}$.

THEOREM 2.3 *Under D1-D7, I1-I3 and M1-M3 $V_n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_k)$.*

Remark 1: The scale $V_n^{1/2}$ implies the rate of convergence of element $\hat{\theta}_{i,n}$ is $V_{i,i,n}^{1/2}$ which need not be homogeneous over i . If the Jacobian and covariance as asymptotically bounded $J_n \rightarrow J$ and $\Sigma_n \rightarrow \Sigma$ then the GMTTME rate is exactly \sqrt{n} since $V_n^{1/2} \sim n^{1/2}V^{1/2}$ for some positive definite, bounded $V \in \mathbb{R}^{k \times k}$. This holds for any stationary DGP for which the conventional GMM estimator is asymptotically normal, so tail-trimming is always a safe practice. See Section 3 for rate derivation for heavy-tailed cases.

²The 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).

Remark 2: 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

$$V_n = n(J'_n\Sigma_n^{-1}J_n)$$

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, below, that in fact minimal trimming is always optimal for the sake of efficiency when $\Upsilon_n = \Sigma_n^{-1}/\|\Sigma_n^{-1}\|$.

Remark 3: 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 J_n has two components that enter V_n asymmetrically as $n \rightarrow \infty$, so an optimal weight does not exist (Čížek 2009). Under *tail* trimming, however, each $J_{i,j,n}$ also decomposes into two components $E[(\partial/\partial\theta_j)m_{i,t}(\theta)|_{\theta_0} \times I_{i,t}(\theta_0)] + (\partial/\partial\theta_j)E[m_{i,t}(\theta) \times I_{i,t}(\theta)]|_{\theta_0}$. The latter is asymptotically dominated by the former due to negligibility $I_{i,t}(\theta) \rightarrow 1$ *a.s.* so $J_{i,j,n} = E[(\partial/\partial\theta_j)m_{i,t}(\theta)|_{\theta_0} \times I_{i,t}(\theta_0)] \times (1 + o(1))$. See Lemma C.1 in Appendix C.

2.4 COVARIANCE AND JACOBIAN MATRIX ESTIMATION

A natural estimator of the trimmed equation covariance Σ_n is

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

for some consistent plug-in $\tilde{\theta}_n$. Since $\hat{\Sigma}_n(\tilde{\theta}_n)$ may itself be used for GMTTM estimation as in the asymptotically efficient weight $\hat{\Upsilon}_n = \hat{\Sigma}_n^{-1}(\tilde{\theta}_n)/\|\hat{\Sigma}_n^{-1}(\tilde{\theta}_n)\|$, in practice $\tilde{\theta}_n$ need not be the final GMTTM estimator $\hat{\theta}_n$. Candidate plug-ins include any consistent MDE including a single-step GMTTME (e.g. $\hat{\Upsilon}_n = I_q$), the LSE, GMME and QMLE provided they are consistent for θ_0 .

LEMMA 2.4 *Let $\|\tilde{\theta}_n - \theta_0\| = o_p(\min\{1, \|\Sigma_n^{-1}\|^{-1/2}\|J_n\|^{-1}\})$. Under D1-D6.i, D7 and I1 $\|\Sigma_n^{-1}\hat{\Sigma}_n(\tilde{\theta}_n) - I_q\| = o_p(1)$.*

Remark: In thin-tailed cases $\Sigma_n \sim \Sigma$ and $J_n \sim J$ the plug-in condition reduces to $\|\tilde{\theta}_n - \theta_0\| = o_p(1)$: $\tilde{\theta}_n$ need only be consistent. Otherwise, $\hat{m}_t^*(\tilde{\theta}_n)\hat{m}_t^*(\tilde{\theta}_n)'$ is sufficiently close to $\hat{m}_t^*(\theta_0)\hat{m}_t^*(\theta_0)'$ as $n \rightarrow \infty$ only when $\|\tilde{\theta}_n - \theta_0\| \xrightarrow{p} 0$ sufficiently fast. Nevertheless, since $\|\Sigma_n^{-1}\|^{-1/2}\|J_n\|^{-1} \leq n^{1/2}/\|V_n\|^{1/2}$ under the efficient weight $\Upsilon_n = \Sigma_n^{-1}/\|\Sigma_n^{-1}\|$, $\tilde{\theta}_n$ need not be $n^{1/2}$ - nor $\|V_n\|^{1/2}$ -consistent due to the negligibility of tail-trimming.

Tail trimming implies the Jacobian J_n is proportional to $E[J_t^*]$, cf. Lemma C.1 in Appendix C. Due to its simple form consistency $\hat{J}_n^*(\tilde{\theta}_n) = E[J_t^*] \times (1 + o_p(1))$ follows for any $\|\tilde{\theta}_n - \theta_0\| \xrightarrow{p} 0$.

LEMMA 2.5 *Under D1-D7, I1-I3 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 $\tilde{\theta}_n$ that satisfies $\|\tilde{\theta}_n - \theta_0\| \xrightarrow{p} 0$.*

The covariance matrix V_n^{-1} in general 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 $\hat{H}_n(\theta) = \hat{J}_n^*(\theta)' \hat{\Sigma}_n^{-1}(\theta) \hat{J}_n^*(\theta)$. The GMTTME satisfies $\|\hat{\theta}_n - \theta_0\| = O_p(\|V_n\|^{-1/2}) \leq o_p(\min\{1, \|\Sigma_n^{-1}\|^{-1/2}\|J_n\|^{-1}\})$ by Theorem 2.3 whenever $\|J_n\|/\|V_n\|^{1/2} = o(\|\Sigma_n^{-1}\|^{1/2})$. Since the latter bound holds under M2, Lemmas 2.4 and 2.5 imply $\hat{V}_n(\hat{\theta}_n)$ is consistent for V_n .

THEOREM 2.6 Under D1-D7, I1-I3 and M1-M3 $\hat{V}_n(\hat{\theta}_n) = V_n \times (1 + o_p(1))$.

2.5 ROBUST M-ESTIMATORS

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

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

$$y_t = \theta'_0 x_t + \epsilon_t \quad \text{with} \quad s_t(\theta) := (y_t - \theta' x_t)^2.$$

Assume ϵ_t is zero-mean with distribution function $F_\epsilon(\epsilon) := P(\epsilon_t \leq \epsilon)$, and inverse $F_{\epsilon^2}^{-1}(\lambda) := \inf\{\epsilon \geq 0 : P(\epsilon_t^2 \leq \epsilon) \leq \lambda\}$. The fixed quantile LTSE is (Ruppert and Carroll 1980, Rousseeuw 1985, Čížek 2008)

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

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 α -mixing, and $\tilde{J}(\lambda) := -E[x_t x_t' I(\epsilon_t^2 \leq F_{\epsilon^2}^{-1}(\lambda))]$ is non-singular, then for the given linear DGP

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = \tilde{J}(\lambda)^{-1} \frac{1}{\sqrt{n}} \sum \epsilon_t x_t I(\epsilon_t^2 \leq F_{\epsilon^2}^{-1}(\lambda)) + o_p(1) \xrightarrow{d} N(0, \tilde{V}^{-1}(\lambda)),$$

where $\tilde{V}(\lambda) = \tilde{J}(\lambda)' \tilde{\Sigma}^{-1} \tilde{J}(\lambda)$ and $\tilde{\Sigma}(\lambda) := E[\epsilon_t^2 x_t x_t' I(\epsilon_t^2 \leq F_{\epsilon^2}^{-1}(\lambda))]$. See Čížek (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 $1/\sqrt{n} \sum \epsilon_t x_t I(\epsilon_t^2 \leq F_{\epsilon^2}^{-1}(\lambda))$ does not have a Gaussian limit *even under fixed quantile trimming*, and $\tilde{\Sigma}(\lambda)$ and $\tilde{J}(\lambda)$ are unbounded. The same conclusion applies to Least Absolute Trimmed Deviations. Since the object that governs asymptotics is the gradient $(\partial/\partial\theta)s_t(\theta)$, its components $\epsilon_t x_t$ must be trimmed to ensure asymptotic normality, not simply ϵ_t^2 .

Quasi-Maximum Trimmed Likelihood: Consider an ARCH(1) $y_t = h_t \epsilon_t$, $\epsilon_t \stackrel{iid}{\sim} (0, 1)$, $h_t^2(\theta) = \alpha + \beta y_{t-1}^2$, $(\alpha, \beta) \geq 0$ 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 Čížek (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 Čížek (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 $\beta_0 = 0$

$$g_{n,t}(\theta_0) = -(\epsilon_t^2 - 1) [1, y_{t-1}^2]' \times I(\epsilon_t^2 \leq F_{\epsilon^2}^{-1}(\lambda)).$$

Now exploit independence to deduce

$$E[g_{2,n,t}^2(\theta_0)] = E\left[(\epsilon_t^2 - 1)^2 I(\epsilon_t^2 \leq F_{\epsilon^2}^{-1}(\lambda))\right] \times E[y_{t-1}^4].$$

Since $\beta_0 = 0$ we know y_t has an unbounded fourth moment $E[y_t^4] = \infty$ if and only if $E[\epsilon_t^4] = \infty$. In this case the QMTL Jacobian is unbounded and by asymptotic linearity and independence between ϵ_t and y_{t-1} , and the QMTLE is not asymptotically normal.

Adaptive M-Estimations: Ling's (2005, 2007) 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 the DGP must be symmetric and super- \sqrt{n} -consistency cannot be achieved.

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

$$\Gamma_{i,i,n} = \Sigma_{i,i,n}^{-1/2} = \left(E[(m_{i,t}^*(\theta_0))^2] \right)^{-1/2} : \Gamma_n^{-1} \Sigma_n \Gamma_n^{-1} \rightarrow \Sigma \text{ a positive definite matrix.}$$

Now write

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

We simplify exposition by assuming $\sigma^{i,j} \neq 0 \forall i, j$, although essential results carry over to the diagonal case $\sigma^{i,i} \neq 0 \forall i$. The component-wise rates n_{θ_i} are

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

Textbook intuition explains n_{θ_i} . If the trimmed equation standard deviation $\Gamma_{l,l,n} = (E[m_{l,n,t}^2(\theta_0)])^{1/2} \rightarrow \infty$ due to heavy-tailed errors then n_{θ_i} is small, ceteris paribus: sharp estimates are more difficult to obtain from models with disproportionately dispersive idiosyncratic shocks. If, however, the Jacobian $|J_{l,i,n}| = |E[(\partial/\partial\theta_i)m_{l,t}(\theta)|_{\theta_0} I_{l,n,t}(\theta_0)]| \rightarrow \infty$ due to heavy tailed regressors then n_{θ_i} is large, ceteris paribus: sharpness improves with regressor dispersion and association. If both error and regressor are heavy-tailed and exhibit feedback then the Jacobian $J_{l,i,n}$ may be overwhelmed by the standard deviation $\Gamma_{l,l,n}$.

We must therefore specify dependence and distribution tails in order to characterize Γ_n and J_n . Consider dynamic linear regression and ARCH models under symmetric trimming, with the same fractiles for all equations for simplicity: $k_{i,n} = k_n$.

Even then the cross-Jacobian $J_{i,j,n}$ can be difficult to formalize without more information. So, express n_{θ_i} as

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

By convention $\max_{j \neq i} \{ \cdot \} = 0$ if there is only one equation $q = 1$.

Let $\{\epsilon_t\}$ be an L_r -bounded, $r > 0$, iid innovations process 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

$$y_t = \theta_0' x_t + \epsilon_t, \quad x_{1,t} = 1, \quad x_t \in \mathbb{R}^k \quad \text{with} \quad m_t(\theta) = (y_t - \theta x_t') x_t,$$

where ϵ_t and x_t are mutually independent, symmetrically distributed and strictly stationary, and

$$E[m_t(\theta) | \mathfrak{S}_{t-1}] = 0 \text{ if and only if } \theta = \theta_0, \text{ where } \mathfrak{S}_t = \sigma(\{y_\tau, x_{\tau+1}\} : \tau \leq t).$$

In general x_t may contain lags of y_t or other random variables. Assume stochastic $x_{i,t}$ are measurable with \mathbb{R} -a.e. continuous distributions. Independence rules out random volatility errors: see Sections 3.2 and 3.3 for this case.

Assume each $z_t \in \{\epsilon_t, x_{i,t}\}$ has tail

$$P(|z_t| > z) = d_z z^{-\kappa_z} (1 + o(1)) \text{ with indices } \kappa_z \in \{\kappa_\epsilon, \kappa_i\}, \kappa_\epsilon > 0 \text{ and } \kappa_i \in (1, 2], \quad (6)$$

and define

$$\kappa_{\epsilon,i} := \min\{\kappa_\epsilon, \kappa_i\} \quad \text{and} \quad \kappa_{\epsilon,(i)}^* := \max_{j \neq i} \left\{ \frac{\kappa_j}{\kappa_{\epsilon,j}} \right\} \geq 1.$$

By convention $\kappa_{\epsilon,1} = \kappa_\epsilon$ since $x_{1,t} = 1$, and $\kappa_{\epsilon,(i)}^* = 1$ if there is only one regressor $k = 1$. Characterizing the Jacobian J_n is simplified by $E[x_{i,t}] = 0^3$.

LEMMA 3.1 *Let stochastic $x_{i,t}$ be mean-zero.*

i. Each $\Gamma_{i,i,n} = (n/k_n)^{1/\kappa_{\epsilon,i} - 1/2}$ and $J_{i,j,n} = -E[x_{i,t}x_{j,t}I(|\epsilon_t x_{j,t}| \leq c_{j,n})] \times (1 + o(1))$. In particular for stochastic $\{x_{i,t}, x_{j,t}\}$ the Jacobian $J_{i,i,n} \sim K(n/k_n)^{\kappa_{\epsilon,i}^{-1}(2-\kappa_i)}$, $J_{i,j,n} = O((n/k_n)^{\kappa_{\epsilon,j}^{-1}(\kappa_j/\kappa_i + 1 - \kappa_j)}) \forall i \neq j$ in general and $J_{i,j,n} = 0$ if $x_{i,t}$ is independent of $x_{j,t}$. Hence for each $i = 2, \dots, k$

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

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

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

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

Remark 2: Although we do not formally treat the generic cases $\kappa_i \gtrless 2$ and $\kappa_\epsilon \gtrless 2$ for the sake of brevity, the same qualitative relationships apply.

Exact slope rates are complicated by cross-Jacobia $J_{i,j,n}$ since their bounds are not sharp and depend on regressor dependence. A lower bound on the rate, however, is available.

EXAMPLE 1 (Slope Rate Lower Bound): In general $\liminf_{n \rightarrow \infty} n_{\theta_i} / [n^{1/2} (n/k_n)^{1/2 - \kappa_i/\kappa_{\epsilon,i} + 1/\kappa_{\epsilon,i}}] \geq K$ 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 ($\kappa_i \leq \kappa_\epsilon$): If $x_{i,t}$ is relatively heavy-tailed $\kappa_i \leq \kappa_\epsilon$ then the bound $n_{\theta_i} \geq Kn^{1/2} (n/k_n)^{1/\kappa_i - 1/2}$ only reflects the tails of $x_{i,t}$. Hence, super- \sqrt{n} -consistency is assured if $x_{i,t}$ has an infinite variance $\kappa_i < 2$.

Case 2 ($\kappa_i > \kappa_\epsilon$): If ϵ_t is more heavy-tailed $\kappa_\epsilon < \kappa_i$ then super- \sqrt{n} -consistency still arises as long as 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.

If all regressors are independent, or errors and regressors are tail-homogenous $\kappa_\epsilon = \kappa_i = \kappa$ (e.g. autoregressions), then simple solutions exist.

³If we set a convention $\infty - \infty = 0$ then $\kappa_i \leq 1$ for $x_{i,t}$ symmetric about 0 is allowed if we agree $E[x_{i,t}] = 0$.

EXAMPLE 2 (Independent Regressors): If stochastic $x_{i,t}$ are independent mean-zero random variables then $n_{\theta_i} \sim Kn^{1/2} (n/k_n)^{1/2 - \kappa_i/\kappa_{\epsilon,i} + 1/\kappa_{\epsilon,i}}$. In this case the inequality in Example 1 becomes equality asymptotically.

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

$$\begin{aligned} n_{\theta_i} &\sim Kn^{1/2} (n/k_n)^{1/\kappa - 1/2} \left[K + O\left((n/k_n)^{-2(1-1/\kappa) - 2(1/\kappa - 1)} \right) \right]^{1/2} \\ &\sim Kn^{1/2} (n/k_n)^{1/\kappa - 1/2}. \end{aligned}$$

Super- \sqrt{n} -consistency 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 examples provide more intuition as to *why* super- \sqrt{n} -consistency may or may not arise.

EXAMPLE 4 (Location): Consider estimating location

$$y_t = \theta_0 + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (6), \quad \kappa \in (1, 2],$$

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.

Remark 3: If $k_n = [n^{\lambda}]$ for $\lambda \in (0, 1)$ then $n_{\theta}/n^{1/2} = Kn^{-(1-\lambda)(1/\kappa - 1/2)}$. Conversely, under very slight trimming $k_n = [\lambda \ln(n)]$ it follows $n_{\theta}/n^{1/2} \geq Kn^{1/2 - 1/\kappa - \iota}$ for any $\lambda > 0$ and tiny $\iota > 0$.

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

$$y_t = \sum_{i=1}^k \theta_{0,i} y_{t-i} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (6), \quad \kappa \in (1, 2].$$

The AR process $\{y_t\}$ satisfies (6) with the same index κ (Cline 1989, Brockwell and Cline 1985). In this case $\kappa_{\epsilon} = \kappa_i = \kappa_{\epsilon,i} = \kappa$, so Example 3 applies: $n_{\theta_i}/n^{1/2} \sim K (n/k_n)^{1/\kappa - 1/2} \rightarrow \infty$.

Remark 1: Exact \sqrt{n} -consistency applies in the hairline infinite variance case $\kappa = 2$.

Remark 2: If $k_n = [n^{\lambda}]$ then $n_{\theta_i}/n^{1/2} \sim Kn^{(1-\lambda)(1/\kappa - 1/2)}$, and under slight trimming $k_n = [\lambda \ln(n)]$ then $n_{\theta_i} > Kn^{1/\kappa - \iota}$ for any $\lambda > 0$ and tiny $\iota > 0$.

Remark 3: The AR(1) case is particularly revealing:

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

The numerator Jacobian term $E[y_{t-1}^2 I(|\epsilon_t y_{t-1}| \leq c_{1,n})] \sim K (n/k_n)^{2/\kappa - 1}$ works like a tail-trimmed *variance*. If sequences $\{c_{y,n}, k_n\}$ satisfy $(n/k_n)P(|y_t| > c_{y,n}) \rightarrow \infty$ 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 $\Gamma_n \rightarrow \infty$ very fast such that sub- \sqrt{n} -consistency arises. See Sections 3.2 and 3.3

Remark 4: The straight least squares estimator is asymptotically non-Gaussian with convergence rate $n^{1/\kappa} \geq n^{1/2}$ (e.g. Hannan and Kanter 1977, Knight 1987, Chan and Tran 1989, Cline 1989). By trimming we have ensured asymptotic normality, but at a cost of speed since

$$n_{\theta_i}/n^{1/\kappa} = n^{1/2-1/\kappa} (n/k_n)^{1/\kappa-1/2} = k_n^{1/2-1/\kappa} \rightarrow 0 \quad \forall \kappa < 2.$$

The rate spread $n_{\theta_i}/n^{1/\kappa}$ vanishes monotonically as the number of trimmed observations k_n decreases per sample n . In fact

$$k_n \sim \lambda \ln(n) \quad \text{implies} \quad n_{\theta_i}/n^{1/\kappa-\iota} \rightarrow \infty \quad \text{for any } \lambda > 0 \text{ and tiny } \iota > 0,$$

so the GMTTME rate can be made arbitrarily close to the maximum rate $n^{1/\kappa}$.

EXAMPLE 6 (Instrumental Variables): There are many variations on the above theme since $m_t(\theta)$ is quite general. It is, therefore, tempting to use heavy-tailed instruments z_t to induce super- \sqrt{n} -consistency. Consider a simple scalar model for reference

$$y_t = \theta_0 x_t + \epsilon_t, \quad \text{where } \{x_t, \epsilon_t\} \stackrel{iid}{\sim} (0, 1) \quad \text{and } m_t(\theta) = (y_t - \theta x_t) z_t \in \mathbb{R}.$$

Assume the instrument $z_t \in \mathbb{R}$ has tail (6) 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 an ARCH(p) model

$$y_t = h_t \epsilon_t, \quad \epsilon_t \text{ is } iid, \quad h_t^2 = \alpha_0 + \sum_{i=1}^p \beta_{0,i} y_{t-i}^2 = \theta'_0 x_t, \quad \alpha_0 > 0, \beta_0 \geq 0, \quad \theta = [\alpha, \beta']'$$

$$\beta_{0,i} > 0 \quad \text{for at least one } i \in \{1, \dots, p\}$$

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

We assume at least one $\beta_{0,i} > 0$ since otherwise the cross-Jacobia $J_{i,1,n}$ do not exist when ϵ_t has an infinite variance: by $\beta_0 = 0$ and independence $J_{i,1,n} \sim -E[y_{t-i}^2 I(K|\epsilon_t^2 - 1| \leq c_{1,n})] \sim -E[y_t^2] = -K \times E[\epsilon_t^2]$.

Assume $|y_t| \geq 1$ a.s. to simplify arguments, $E|\epsilon_t|^p < \infty$ for some $p > 0$, $E[\max\{0, \ln |\epsilon_t|\}] < \infty$, and the Lyapunov exponent associated with the stochastic recurrence equation form is negative⁴.

LEMMA 3.2 *In general $J_{i,j,n} = -Ex_{i,t}x_{j,t}I(|\epsilon_t^2 - 1|h_t^2x_{j,t}| \leq c_{j,n}) \times (1 + o(1))$. If $\kappa_y \geq 4$ then $n_\alpha = n_{\beta_i} = Kn^{1/2}$ for each $i = 1..p$, and if $\kappa_y \in (2, 4]$ then $n_\alpha = n_{\beta_i} = Kn^{1/2} (k_n/n)^{2/\kappa_y - 1/2} \times (1 + O(1))$.*

Remark 1: If y_t has a finite fourth moment $E[y_t^4] < \infty$ or exhibits the hairline infinite kurtosis case $\kappa = 4$, then $n_\alpha = n_\beta = Kn^{1/2}$. Otherwise, both GMTTME's $\hat{\alpha}_n$ and $\hat{\beta}_n$ are sub- \sqrt{n} -consistent. Notice the tails of ϵ_t do not play any role *per se* when there are ARCH affects. 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 $E[v_t|\mathfrak{F}_{t-1}] = 0$. But stationary AR equations all have the same tail index κ when ϵ_t is iid with tail (6). See the proof of Lemma 3.1, and Brockwell and Cline (1986). In the ARCH case, however, the error $y_t^2 - h_t^2 = (\epsilon_t^2 - 1)\theta_0'x_t$ depends on x_t . Specifically $m_{1,t}(\theta_0) = (\epsilon_t^2 - 1)h_t^2$ has tail index $\kappa/2$ and all other $m_{i,t}(\theta_0) = (\epsilon_t^2 - 1)h_t^2y_{t-i+1}^2$ for $i \geq 2$ have index $\kappa/4$ due to feedback. This implies the slope-components of the covariance Σ_n diverge *faster* relative to the regressor Jacobian J_n , compared to AR models with iid errors. The same intuition from Section 3.1 suffices: models with disproportionately heavy-tailed errors render less sharp estimates.

Remark 3: All results of this section follow intimately from properties of tail-trimmed variances. See especially the proof of Lemma 3.1. Thus, the above formula need not hold under central order trimming $k_n = [\lambda n]$, $\lambda \in (0, 1)$. This implies, for example, we cannot conclude $n_\alpha = n_{\beta_i} = Kn^{1/2} (k_n/n)^{2/\kappa_y - 1/2} \sim Kn^{1/2}$ is feasible when $\kappa_y < 4$.

3.3 AUTOREGRESSIONS WITH ARCH-ERRORS

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

$$y_t = \rho_0 y_{t-1} + u_t, \quad |\rho_0| < 1, \quad u_t = h_t \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} N(0, 1)$$

$$h_t^2 = \alpha_0 + \beta_0 u_{t-1}^2, \quad \alpha_0 > 0, \beta_0 > 0, \quad \theta = [\rho, \alpha, \beta]'$$

$$E \left[\ln \left| \rho_0 + \beta_0^{1/2} \epsilon_t \right| \right] < 0$$

and three equations used to estimate each $\theta = [\rho_0, \alpha_0, \beta_0]'$,

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

The Jacobian J_n is block diagonal asymptotically. But with ARCH affects the AR equation $m_{1,t}(\theta_0) = \epsilon_t h_t y_{t-1}$ is more heavy tailed than in a pure AR model, hence sub- \sqrt{n} -consistency arises.

LEMMA 3.3 *n_α and n_β are characterized by Lemma 3.2, and $n_\rho = n_\beta$.*

⁴See Basrak et al (2002: Theorem 3.1): in this setting the SRE form obtains a unique stationary solution and y_t has regularly varying tails.

Remark 1: The results of Sections 3.1 and 3.2 crucially rely on the properties of regularly varying tails. Gaussian iid errors and the Lyapunov condition $E[\ln |\rho_0 + \beta_0^{1/2} \epsilon_t|] < 0$ ensure $\{y_t\}$ is stationary geometrically α -mixing with regular varying tails, although more general environments are possible (Borkovec and Klüppelberg 2001: Theorems 1 and 3).

EXAMPLE 7 (AR with ARCH error): The impact a non-iid error has on the convergence rate can be seen by simply estimating the AR slope ρ_0 with one equation $m_t(\rho) = (y_t - \rho y_{t-1})y_{t-1}$. Notice $m_t(\rho_0) = \epsilon_t h_t y_{t-1}$ has a tail index $\kappa_y/2$ due to feedback, half that from the iid case Example 5. Arguments in the proofs of Lemmas 3.1 and 3.3 can be used to deduce $n_\rho \sim Kn^{1/2}$ if $\kappa_y \geq 4$, and

$$\begin{aligned} n_\rho &\sim Kn^{1/2} \frac{1}{(n/k_n)^{2/\kappa_y - 1/2}} = o(n^{1/2}) \text{ if } \kappa_y \in [2, 4) \\ &\sim Kn^{1/2} \frac{(n/k_n)^{2/\kappa_y - 1}}{(n/k_n)^{2/\kappa_y - 1/2}} = k_n^{1/2} = o(n^{1/2}) \text{ if } \kappa_y < 2. \end{aligned}$$

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 $n^{1/2}$.

EXAMPLE 8 (Optimal Trimming and Efficiency): The AR cases Examples 5 and 7 present diametrically opposing relationships between trimming k_n and efficiency n_θ . In the iid infinite variance case since large values of error and regressor are unassociated *minimal* trimming optimizes efficiency for the textbook reason given above: n_θ increases as k_n decreases. Conversely, error and regressor feedback in the ARCH case greatly increases equation tails and therefore diminishes efficiency. Since large regressors are therefore associated with large errors the optimal strategy is *maximal* trimming: n_θ increases as k_n increases. Optimal selection of the sequence $\{k_n\}$, however, is beyond the scope of the present paper.

3.4 TAIL TRIMMED QML FOR ARCH

Consider an ARCH(1) with QML equations

$$y_t = h_t \epsilon_t, \quad h_t^2 = \alpha_0 + \beta_0 y_{t-1}^2 = \theta_0' x_t, \quad \alpha_0 > 0, \beta_0 \geq 0$$

$$m_t(\theta) = (y_t^2 - \theta' x_t) \{\theta' x_t\}^{-2} x_t.$$

Since scaling $m_t(\theta_0) = (y_t - h_t \epsilon_t) h_t^{-4} x_t = (\epsilon_t^2 - 1) h_t^{-2} x_t$ implies only the tails of ϵ_t matter, we assume ϵ_t has regularly varying tails. Nevertheless the Jacobian is unbounded if there are no ARCH affects and ϵ_t has an infinite variance. Further, QML equations do not improve on the rate of convergence for the reasons above: feedback between error $y_t^2 - h_t^2$ and regressor $h_t^{-4} x_t$ implies the covariance dominates the Jacobian.

LEMMA 3.4 *Assume $\beta_0 > 0$ and $\epsilon_t \stackrel{iid}{\sim} (6)$ with index $\kappa_\epsilon > 0$. If $\kappa_\epsilon \geq 4$ then $n_\alpha = n_\beta = Kn^{1/2}$, and if $\kappa_\epsilon \in (2, 4]$ then $n_\alpha = n_\beta = Kn^{1/2} (k_n/n)^{2/\kappa_\epsilon - 1/2} \times (1 + O(1))$.*

4. AUTOREGRESSIONS AND ARCH We now verify the major assumptions for heavy-tailed stationary autoregression and ARCH under symmetric trimming $m_{i,t}^*(\theta) = m_{i,t}(\theta) I(|m_{i,t}(\theta)| \leq c_{i,n}(\theta))$ where $(n/k_n)P(|m_{i,t}(\theta)| > c_{i,n}(\theta)) \rightarrow 1$, 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(k) process with iid, heavy-tailed errors

$$y_t = \theta'_0 x_t + \epsilon_t, \quad x_t = [y_{t-1}, \dots, y_{t-k}]', \quad \epsilon_t \stackrel{iid}{\sim} (6) \text{ with } \kappa \in (1, 2), \quad E[\epsilon_t] = 0 \quad (7)$$

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

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 α -mixing (An and Huang 1996: Theorem 3.1), and y_t and $m_{i,t}(\theta_0) = \epsilon_t y_{t-i}$ have tail (6) with the same index κ (Cline 1986, 1989).

Stationarity, linearity, distribution continuity, and mixing ensure D1, D2, D3, I1 and M3 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 is trivial since $E[m_t^*(\theta_0)] = 0$ under symmetry and integrability; M1 holds under I1 and D1-D7 given $\Upsilon_n = \Sigma_n^{-1}/\|\Sigma_n^{-1}\|$ and Lemma 2.4; and M2 holds since $\Upsilon_n = \Sigma_n^{-1}/\|\Sigma_n^{-1}\|$ trivially implies $Kn^2\|J_n\|^2\|\Sigma_n^{-1}\| \geq \|V_n\| \rightarrow \infty$ given Lemma 3.1. Finally, D5 merely defines the threshold sequence. What remains is envelope moment bounds D4, Jacobia property D6, indicator property D7, and smoothness I3.

Although I2 is trivial under symmetry, note in asymmetric cases $E[m_t^*(\theta_0)] = o(\|\Sigma_n^{-1/2}\|^{-1}/n^{1/2}) = o(1)$ under the D5 implication $c_n(\theta_0) = o(n^{1/2})$. We therefore also discuss what $c_n(\theta_0) = o(n^{1/2})$ entails for tail-trimming.

D5 (threshold bound). Since $m_{i,t}(\theta_0)$ has tail (6) it is easy to show each $c_{i,n}(\theta_0) = K(n/k_n)^{1/\kappa}$. Thus $c_{i,n}(\theta_0) = o(n^{1/2})$ requires sufficiently many tail observations to be trimmed: $k_n/n^{1-\kappa/2} \rightarrow \infty$. This only matters in the heavy-tailed case $\kappa < 2$, since otherwise we are free to choose an intermediate order sequence $\{k_n\}$.

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 $\|J_n\| \rightarrow \infty$ by Lemma 3.1. Then stationarity, L_p -boundedness of y_t and the construction $J_{i,j,t}^*(\theta) = -y_{t-i}y_{t-j}I_{j,t}^*(\theta)$ imply $E[(\sup_{\theta \in U_0(\delta_n)} \{\|J_n^*(\theta) - J_n^*\|\})^r] \leq K$ for any $r \in (0, p/2)$. Therefore $\sup_{\theta \in U_0(\delta_n)} \{\|J_n^*(\theta) - J_n^*\|\} = o_p(\|J_n\|)$ follows by Markov's inequality and $\|J_n\| \rightarrow \infty$.

D7 (Indicator Class). Since $m_{i,t}(\theta)$ is a linear function of θ , we can without loss of generality assume $c_{i,n}(\theta)$ is also linear in θ . 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). Identification and the definition of a derivative imply $E[m_t^*(\theta)] = J_n(\theta - \theta_0) + o(\|J_n\| \times \|\theta - \theta_0\|)$, hence $\mathbf{m}_n := \sup_{\theta} \|E[m_t^*(\theta)]\| \leq K\|J_n\| \times (1 + o(1))$ given compactness of Θ . Therefore since Υ_n is bounded

$$\inf_{\|\theta - \theta_0\| > \delta} \{\mathbf{m}_n^{-2} Q_n(\theta)\} \geq \inf_{\|\theta - \theta_0\| > \delta} \left\{ (\theta - \theta_0)' \frac{J_n'}{\|J_n\|} \Upsilon_n \frac{J_n}{\|J_n\|} (\theta - \theta_0) \right\} \times (1 + o(1)) + o(1).$$

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.

Thus, the GMTTME of the slope θ_0 in a stationary AR with errors governed by regularly varying tails is consistent and asymptotically normal under only regulatory conditions D5 and M1 for the thresholds and weight. The above verification above and Lemma 3.1 suffice to prove the following claim by invoking Theorems 2.2, 2.3 and 2.6.

COROLLARY 4.1 Consider (7), let $\Upsilon_n = \Sigma_n^{-1}/\|\Sigma_n^{-1}\|$, and assume D5 and M1 hold. Then $\hat{\theta}_n = \theta_0 + O_p(\|V_n^{1/2}\|^{-1})$, $V_n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_k)$ 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 each i and some $V_i < \infty$.

Remark: Asymptotic normality easily generalizes to autoregressive distributed lags. This reveals a very useful result: in linear models with iid errors as long as the data are stationary and geometrically α -mixing then standard inference is available under heavy tails.

4.2 ARCH

Recall the ARCH model of Section 3.2 with finite variance and infinite kurtosis, assume at least one $\beta_i > 0$ and assume the distribution of $\epsilon_t \stackrel{iid}{\sim} (0, 1)$ does not have an atom at zero. Then $\{y_t\}$ is stationary, L_p -bounded, geometrically α -mixing, symmetrically distributed with tail (6) (e.g. Basrak et al 2002: Theorem 3.1). Further, the estimating equations $m_t(\theta) = (y_t^2 - \theta'x_t)x_t$ are differentiable and linear in θ with absolutely continuous marginal distributions, and integrable at θ_0 for all $\kappa_y > 2$. Thus D1-D4, D6, D7, I1-I3 and M2 and M3 are either trivial or are verifiable exactly as in Section 4.1.

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 $\kappa = \sup\{\alpha > 0 : E|y_t|^\alpha < \infty\}$. For each data generating process described in Table 1, 1000 samples of size $n = 1000$ are generated.

TABLE 1 - Data Generating Processes

Type	Subtype	Model	θ_k	iid errors ϵ_t	κ
LOCATION		$y_t = 1 + \epsilon_t$	1	$P_{1.5}, P_{2.5}$	1.5, 2.5
AR(1)		$y_t = .9 \times y_{t-1} + \epsilon_t$.9	$P_{1.5}, P_{2.5}$	1.5, 2.5
ARCH(1)		$y_t = h_t \epsilon_t$			
	ARCH	$h_t^2 = .3 + .6y_{t-1}^2$.6	$N_{0,1}$	3.81^5
GARCH(1,1)		$y_t = h_t \epsilon_t$			
	GARCH	$h_t^2 = .3 + .3y_{t-1}^2 + .6h_{t-1}^2$.6	$N_{0,1}$	2.67
	GARCH	$h_t^2 = .3 + .3y_{t-1}^2 + .6h_{t-1}^2$.6	$P_{2.5}$	0.81
	IGARCH	$h_t^2 = .3 + .4y_{t-1}^2 + .6h_{t-1}^2$.6	$N_{0,1}$	2.00
TARCH(1)		$y_t = h_t \epsilon_t$			
	TARCH	$h_t^2 = .3 + .6y_{t-1}^2 \times I(y_{t-1} < 0)$.6	$N_{0,1}$	5.24^6
	TARCH	$h_t^2 = .3 + .6y_{t-1}^2 \times I(y_{t-1} < 0)$.6	$P_{2.5}$	2.59
	TIARCH	$h_t^2 = .3 + y_{t-1}^2 \times I(y_{t-1} < 0)$	1	$N_{0,1}$	3.37
QARCH(1)		$y_t = h_t \epsilon_t$			
	QARCH	$h_t^2 = (.3 + .8y_{t-1})^2$.8	$N_{0,1}$	3.51^7
	QIARCH	$h_t^2 = (.3 + y_{t-1})^2$	1	$N_{0,1}$	2.00

⁵Basrak 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

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.

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

Model	$u_t(\theta) \in \mathbb{R}$	$\theta \in \mathbb{R}^k$	$z_t \in \mathbb{R}^q$
LOCATION	$y_t - \mu$	μ	$[1, y_{t-1}]'$
AR(1)	$y_t - \rho y_{t-1}$	ρ	$[y_{t-i}]_{i=1}^2$
ARCH(1)	$y_t^2 - \alpha - \beta y_{t-1}^2$	$[\alpha, \beta]$	$[1, \{y_{t-i}^2\}_{i=1}^2]'$
GARCH(1,1)	$y_t^2 - \alpha - \beta y_{t-1}^2 - \gamma h_{t-1}^2$	$[\alpha, \beta, \gamma]$	$[1, \{y_{t-i}^2, h_{t-i}^2\}_{i=1}^2]'$
TARCH(1)	$y_t^2 - \alpha - \beta y_{t-1}^2 I_{t-1}$	$[\alpha, \beta]$	$[1, y_{t-1}^2, y_{t-2}^2]'$
QARCH(1)	$y_t^2 - (\alpha + \beta y_{t-1}^2)^2$	$[\alpha, \beta]$	$[1, \{y_{t-i}, y_{t-i}^2\}_{i=1}^2]'$

Tail thickness for iid and AR(1) data with iid innovations is gauged by the Pareto innovation's index κ (Hannan and Kanter 1977, Cline 1986, 1989).

The collective GARCH group have heavy tails due to the innovations $\epsilon_t \stackrel{iid}{\sim} P_{2.5}$ or to the parametric structure under $\epsilon_t \stackrel{iid}{\sim} N_{0,1}$. The kurtosis of y_t is infinite in all cases except TARCH with Gaussian shocks, variance is infinite for IGARCH and GARCH with Pareto errors, and the absolute first moment is infinite for GARCH with Pareto $P_{2.5}$ errors. Thus, in all save one random volatility model 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 since⁸.

5.1 Tail Fractile

In symmetric data cases (iid, AR, GARCH) symmetric trimming is used with the same fractile for all equations

$$k_n = [n^{\lambda^*}],$$

where $\max\{\lambda, .01\} \leq \lambda^* \leq \max\{\lambda, .99\}$ for each λ in the set $\Lambda_n := \{1/n, 2/n, \dots, (n-1)/n\}$. Thus $\lambda^* \in [.01, .99]$. Asymmetric processes (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 $\max\{\lambda_j, .01\} \leq \lambda_j^* \leq \max\{\lambda_j, .99\}$ and $\lambda_j \in \Lambda_n$.

5.2 Evaluation

We analyze k^{th} parameter estimates $\hat{\theta}_k$ for brevity. Consult the fourth column of Table 1. Estimator performance is gauged by simulation means, standard deviations, tests of the null hypothesis $\theta_k = \theta_{0,k}$, and Kolmogorov-Smirnov tests of standard normality. Let $\{\hat{\theta}_{j,k}\}_{j=1}^{1000}$ be the independently drawn sequence of estimates of $\theta_{0,k}$. In a first experiment

⁵ $h_t^2 = \alpha + \beta y_{t-1}^2 + \gamma h_{t-1}^2$, 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 $K \in \{.01, .02, \dots, 10\}$ based on $N = 100,000$ iid random draws ϵ_t from $N_{0,1}$ or $P_{2.5}$, with resulting 1% bands less than .001 in all cases.

⁶An 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 $\beta^{\kappa/2} E[|\epsilon_t|^\kappa] = 2$.

⁷Since $y_t = |\alpha + \beta y_{t-1}| \epsilon_t$ use Lemma 2.1 of Cline (2007) to deduce $\beta^\kappa E[|\epsilon_t|^\kappa] = 1$.

⁸Let $\hat{Q}_q(\theta)$ 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 $\{g_t(\theta_0), \mathfrak{F}_t\}$. Since an $L_{1+\iota}$ -bounded martingale difference trivially forms a uniformly integrable L_1 -mixingale (McLeish 1975), Andrews' (1988: Theorem 1) law of large numbers shows $1/n \sum_{t=1}^n g_t(\theta_0) \xrightarrow{P} 0$. Further, each $g_t(\theta)$ satisfies Andrew's (1992: W-LIP) Lipschitz condition given differentiability and the QML criterion form, so $\sup_\theta |1/n \sum_{t=1}^n \{g_t(\theta) - E[g_t(\theta)]\}| \xrightarrow{P} 0$ by Theorem 3 of Andrews (1992). Consistency of the QMLE is now a standard exercise (e.g. Pakes and Pollard 1989: Corollary 3.4).

we fix the sample size $n = 1000$ and use the simulation variance $\hat{s}_{n,k}^2 = (1/999) \sum_{j=1}^{1000} (\hat{\theta}_{j,k} - \theta_{0,k})^2$ to generate a sequence of ratios $\{\hat{T}_{j,k}\}_{j=1}^{1000}$,

$$\hat{T}_{j,k} = \frac{\hat{\theta}_{j,k} - \theta_{0,k}}{\hat{s}_{n,k}^2}.$$

Both the rejection frequencies of t-tests of the hypothesis $\theta_k = \theta_{0,k}$ and the KS test of standard normality based on the iid sample $\{\hat{T}_{j,k}\}_{j=1}^{1000}$ are reported. 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

In simulations not reported here the one-step GMTTME $\hat{\theta}_n^{(1)}$ based on the naïve weight $\hat{\Upsilon}_n = I_q$ dominated across evaluation criteria the two-step $\hat{\theta}_n^{(2)}$ with efficient weight $\hat{\Upsilon}_n = \hat{\Sigma}_n^{-1}(\tilde{\theta}_n)/\|\hat{\Sigma}_n^{-1}\|$ and plug-in $\tilde{\theta}_n = \hat{\theta}_n^{(1)}$. The likely reason is the computational complexity of a multi-step algorithm under nonlinearity associated with trimming. We therefore compute the one-step GMTTME $\hat{\theta}_n^{(1)}$, and two-step GMTTME $\hat{\theta}_n^{(2)}$ with a second-step QMLE plug-in $\tilde{\theta}_n = \hat{\theta}_q$ since QML is in general more robust to heavy tails than GMM, and consistent for the DGP's in this study.

In all cases the GMME is computed in two steps using the QMLE as a second-step 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 by exactly identified one- and two-step GMTTM based on the Table 2 equations, fractiles $k_n = \lceil n^\lambda \rceil$ and $k_n = \lceil \lambda \ln(n) \rceil$, and over sample sizes $n \in N = \{1000 \dots 10000\}$ with increments of 20 observations. In the ARCH case we focus on β_0 . This produces $R = 450$ estimates of the simulation standard deviation $\{\hat{s}_{n,k}\}_{n \in N}$.

Since $\hat{s}_{n,k} \sim n_{\theta_k}^{-1} K$, according to Examples 4 and 5 if variance is infinite $\kappa < 2$ then for location

$$\begin{aligned} \text{Location : } \quad \hat{s}_{n,k}^{-1}/n^{1/2} &\sim K n^{-(1-\lambda)(1/\kappa-1/2)} \quad \text{if } k_n = \lceil n^\lambda \rceil \\ &\in K \left(n^{-(1/\kappa-1/2)-\iota}, n^{-(1/\kappa-1/2)} \right) \quad \text{if } k_n = \lceil \lambda \ln(n) \rceil, \end{aligned}$$

and

$$\begin{aligned} \text{AR : } \quad \hat{s}_{n,k}^{-1}/n^{1/2} &\sim K n^{(1-\lambda)(1/\kappa-1/2)} \quad \text{if } k_n = \lceil n^\lambda \rceil \\ &\in K \left(n^{1/\kappa-1/2-\iota}, n^{1/\kappa-1/2} \right) \quad \text{if } k_n = \lceil \lambda \ln(n) \rceil. \end{aligned}$$

According to Lemma 3.2 if variance is finite and kurtosis is infinite then

$$\begin{aligned} \text{ARCH : } \quad \hat{s}_{n,k}^{-1}/n^{1/2} &\sim K n^{-(1-\lambda)(2/\kappa-1/2)} \quad \text{if } k_n = \lceil n^\lambda \rceil \\ &\in K \left(n^{-(2/\kappa-1/2)-\iota}, n^{-(2/\kappa-1/2)} \right) \quad \text{if } k_n = \lceil \lambda \ln(n) \rceil. \end{aligned}$$

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

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

In the AR model, for example, $b(\lambda) = (1 - \lambda)(1/\kappa - 1/2)$ if the fractile is $k_n = [n^\lambda]$, and $b(\lambda) \approx 1/\kappa - 1/2$ when $k_n = [\lambda \ln(n)]$ for any $\lambda > 0$.

Notice $b(\lambda) < 0$, $= 0$ and > 0 imply sub-, exact-, and super- \sqrt{n} -consistency. We can in principle select a KS minimizing λ for each sample size n , but this adds complexity since then $b(\lambda)$ depends on n . We therefore use the KS minimizing λ for $k_n = [n^\lambda]$ and $n = 1000$ based on Section 5.1's results, and the KS minimizing $\lambda \in \{.1, .2, \dots, 10\}$ for $k_n = [\lambda \ln(n)]$ and $n = 1000$ based on experiments not reported here under $n = 1000$. Since the samples are independently drawn, and assuming $E[v_n|n] = 0$, the least squares estimator $\hat{b}(\lambda)$ is asymptotically normal and consistent for $b(\lambda)$.

5.5 Summary of Results

Refer to Tables 3 and 4. 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 iid, AR), 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 cases of substantially heavy tails the GMME and QMLE strongly fail the KS normality test, while the GMTTME passes roughly as well as in any other case.

ii. The asymptotically efficient two-step GMTTME $\hat{\theta}_n^{(2)}$ slightly dominates the one-step naïve $\hat{\theta}_n^{(1)}$ across criteria.

iii. Asymmetric trimming for asymmetrically distributed equations is always optimal, where more observations are trimmed from the heavier tail. Consider the TIARCH model: 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.

iv. 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: $k_{1,n} + k_{2,n} = 16 + 6 = 22$.

v. QMLE fails normality tests in all GARCH cases where the errors have an infinite third 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 3 for the simulation 95% confidence bands of $b = b(\lambda)$. Consider the fractile case $k_n = [n^\lambda]$. We only highlight results for two-step GMTTM. In each case simulation results closely match theory. The finite variance location and AR $b = 0$ and the respective bands are $.031 \pm .051$ and $.027 \pm .039$.

In the infinite variance cases Examples 4 and 5 reveal b for location is $-(1 - \lambda)(1/\kappa - 1/2) = -(1 - .22)(1/1.5 - 1/2) = -0.112$ and in the AR case $b = (1 - \lambda)(1/\kappa - 1/2) = (1 - .32)(1/1.5 - 1/2) = .113$. The 95% band for location is $-.130 \pm .024$, and for AR is $.116 \pm .012$, both quite sharp and containing the true b . Finally, for an ARCH(1) with finite variance and infinite kurtosis use Lemma 3.2 to deduce $b = -(1 - \lambda)(2/\kappa - 1/2) = -(1 - .22)(2/3.81 - 1/2) = -.019$. The 95% band $-.022 \pm .011$.

Finally, when $k_n = [\lambda \ln(n)]$ the AR case is particularly interesting since the GMTTME should obtain a rate arbitrarily close to $n^{1/\kappa}$: $n_\theta/n^{1/2} \in K(n^{1/\kappa-1/2-\epsilon}, n^{1/\kappa-1/2})$ when $\kappa < 2$. Thus, for any λ the true $b \approx 1/\kappa - 1/2 = 1/5 - 1/2 \approx .167$ and the 95% band

is $.189 \pm .043$. The GMTTME can indeed deliver a massive efficiency gain over extant asymptotically normal MDE's for stationary data.

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} -consistency even 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.

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 unbiasedness 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, adaptive methods for selecting the trimming fractiles $\{k_{1,n}, k_{2,n}\}$, and other criteria for trimming rather than simply large values of the equations (e.g. smooth trimming functions; or trimming by errors and regressors separately).

APPENDIX A: Proofs of Main Results

We repeatedly use the following properties under I1, I2, D2, D5, M1, and M2:

1. $E[m_t^*(\theta)] = 0$ if and only if $\theta = \theta_{n,0}$ (8)
2. $Q_n(\theta) = 0$ and $(\partial/\partial\theta)Q_n(\theta) = 0$ if and only if $\theta = \theta_{n,0}$
3. $Q_n(\theta_0) = o\left(\left\|\Sigma_n^{-1/2}\right\|^{-2}/n\right) = o\left(\|J_n\|^2 \times \|V_n\|^{-1}\right)$
4. $\sup_{\theta} \{\|\Sigma_n(\theta)\|\} \leq K \sup_{\theta} \{c_n^2(\theta)\} = o(n)$ and $\sup_{\theta} \left\{\left\|\Sigma_n^{-1/2}(\theta)\right\|\right\} = O(1)$ for $n \geq N$
5. $\sup_{\theta} \left\{\left\|\Sigma_n^{1/2}(\theta)\right\|\right\} / n^{1/2} = o(1)$ and $\sup_{\theta} \left\{\left\|\Sigma_n^{-1/2}\right\|\right\} / n^{1/2} = o(1)$.

(8.1) follows from identification I1; (8.2) from (8.1), weight boundedness M1 and differentiability D2; (8.3) by noting $Q_n(\theta_0) \leq K\|E[m_t^*(\theta_0)]\|^2 = o\|\Sigma_n^{-1/2}\|^{-1}/n = o(\|J_n\|^2/\|V_n\|)$ under identification property I2, and M1 and M2; (8.4) from I1, threshold bound D5 and covariance non-degeneracy M3; and (8.5) from (8.4), I1, D5 and standard matrix norm inequalities: $\|\Sigma_n^{-1/2}(\theta)\|^{-1}/n^{1/2} \leq \|\Sigma_n^{1/2}(\theta)\|/n^{1/2} = Kc_n(\theta)/n^{1/2} = o(1)$ uniformly on Θ .

The following proofs exploit criterion properties Lemmas B1-B2 and limit theory Lemmas C.1-C.9. See Appendices B and C respectively. It is understood n is sufficiently large so degeneracy under trimming is avoided.

Proof of Theorem 2.1. $\hat{m}_n^*(\theta) = 1/n \sum_{t=1}^n \hat{m}_t^*(\theta)$ and $m_n^*(\theta) = 1/n \sum_{t=1}^n m_t^*(\theta)$. 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 $\|\hat{\theta}_n - \theta_0\| \xrightarrow{p} 0$.

First note by Lemma B.1

$$\sup_{\theta} \left\{ \frac{\mathfrak{m}_n^{-2} \times \left| \hat{Q}_n(\theta) - Q_n(\theta) \right|}{1 + \mathfrak{m}_n^{-2} \times Q_n(\theta)} \right\} = o_p(1).$$

Therefore

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

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 $\hat{Q}_n(\hat{\theta}_n) \leq \hat{Q}_n(\theta_{n,0})$, where $\hat{Q}_n(\theta_0) \leq K \|\hat{m}_n^*(\theta_0)\|^2 \leq K \|m_n^*(\theta_0)\|^2 + o_p(\|\Sigma_n^{-1/2}\|^{-1}/n^{1/2}) = K \|m_n^*(\theta_0)\|^2 + o_p(1)$ by weight bound M1, the Lemma C.2 asymptotic approximation and covariance bound (8.5). Finally $K \|m_n^*(\theta_0)\| \leq K \|m_n^*(\theta_{n,0})\| + o_p(\|\Sigma_n^{-1/2}\|^{-1}/n^{1/2}) = o_p(1)$ by the Lemma C.1.b equation expansion between $\theta_{n,0}$ and θ_0 , the Lemma C.3 LLN, and (8.5). ■

Proof of Theorem 2.2. The proof follows arguments summarized in Pakes and Pollard (1989), Newey and McFadden (1994) and others, updated to allow for degeneracy under tail-trimming, and arbitrary rates of convergence. See Hill and Renault (2010). ■

Proof of Theorem 2.3. The claim follows from asymptotic linearity Lemma C.4:

$$V_n^{1/2} \left(\hat{\theta}_n - \theta_0 \right) = -V_n^{1/2} \left(H_n^{-1} J_n' \Upsilon_n \right) \frac{1}{n} \sum_{t=1}^n \hat{m}_t^*(\theta_0) \times (1 + o_p(1));$$

asymptotic approximation Lemma C.2 coupled with covariance property (8.5) and the construction of V_n :

$$V_n^{1/2} \left(\hat{\theta}_n - \theta_0 \right) = -V_n^{1/2} \left(H_n^{-1} J_n' \Upsilon_n \right) \frac{1}{n} \sum_{t=1}^n m_t^*(\theta_0) \times (1 + o_p(1)) + o_p(1);$$

and therefore by central limit theorem Lemma C.6

$$\begin{aligned} V_n^{1/2} \left(\hat{\theta}_n - \theta_0 \right) &= -V_n^{1/2} \left(\frac{1}{n^{1/2}} H_n^{-1} J_n' \Upsilon_n \Sigma_n^{1/2} \right) \frac{\Sigma_n^{-1/2}}{n^{1/2}} \sum_{t=1}^n m_t^*(\theta_0) \times (1 + o_p(1)) + o_p(1) \\ &\xrightarrow{d} N(0, I_k). \end{aligned}$$

■

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

$$\begin{aligned} & \left\| \|\Sigma_n^{-1}\| \times \left\| \frac{1}{n} \sum_{t=1}^n \left\{ \hat{m}_t^*(\tilde{\theta}_n) \hat{m}_t^*(\tilde{\theta}_n)' - m_t^*(\tilde{\theta}_n) m_t^*(\tilde{\theta}_n)' \right\} \right\| \right\| \\ & + \left\| \|\Sigma_n^{-1}\| \times \left\| \frac{1}{n} \sum_{t=1}^n \left\{ m_t^*(\tilde{\theta}_n) m_t^*(\tilde{\theta}_n)' - m_t^*(\theta_0) m_t^*(\theta_0)' \right\} \right\| \right\| \\ & + \left\| \|\Sigma_n^{-1}\| \times \left\| \frac{1}{n} \sum_{t=1}^n \left\{ m_t^*(\theta_0) m_t^*(\theta_0)' - m_t^*(\theta_{n,0}) m_t^*(\theta_{n,0})' \right\} \right\| \right\| \\ & + \left\| \|\Sigma_n^{-1}\| \times \left\| \frac{1}{n} \sum_{t=1}^n \left\{ m_t^*(\theta_{n,0}) m_t^*(\theta_{n,0})' - \Sigma_n(\theta_{n,0}) \right\} \right\| \right\| \\ & + \left\| \|\Sigma_n^{-1}\| \times \|\Sigma_n(\theta_{n,0}) - \Sigma_n\| \right\| = \sum_{i=1}^5 E_{i,n}. \end{aligned}$$

The first term $E_{1,n} = o_p(1)$ by an argument identical to the proof of uniform asymptotic approximation Lemma C.2. The second term $E_{2,n} = o_p(1)$ by cross-product expansion Lemma C.1.c since

$$\frac{1}{n} \sum_{t=1}^n m_t^*(\theta) m_t^*(\theta)' = \frac{1}{n} \sum_{t=1}^n m_t^*(\tilde{\theta}_n) m_t^*(\tilde{\theta}_n)' + o_p \left(\left\{ \|\Sigma_n^{-1/2}\|^{-1} / n^{1/2} + K \|J_n\| \right\}^2 \times \|\theta - \tilde{\theta}\|^2 \right).$$

Thus, properties of the covariance Σ_n and scale V_n in (8.4), (8.5) and M2, and the plug-in supposition $\tilde{\theta}_n = \theta_0 + o_p(\min\{1, \|\Sigma_n^{-1/2}\|^{-1} / \|J_n\|\})$ result in

$$\begin{aligned} E_{2,n} &= O_p \left(\|\Sigma_n^{-1}\| \times \|J_n\|^2 \times \|\tilde{\theta}_n - \theta_0\|^2 \right) \\ &+ O_p \left(\|\Sigma_n^{-1}\| \times o \left\{ \|\Sigma_n^{-1/2}\|^{-2} / n \right\} \times \|\tilde{\theta}_n - \theta_0\|^2 \right) \\ &+ O_p \left(\|\Sigma_n^{-1}\| \times \|J_n\| \times o \left\{ \|\Sigma_n^{-1/2}(\theta_{n,0})\|^{-1} / n^{1/2} \right\} \times \|\tilde{\theta}_n - \theta_0\|^2 \right) \\ &= o_p(1). \end{aligned}$$

Similarly, the third term $E_{3,n} = o_p(1)$ by the equation expansion Lemma C.1.b and covariane bound (8.5). The fourth term $E_{4,n} = o_p(1)$ by the martingale difference decomposition and LLN Lemmas C.7 and C.8. Finally, $E_{5,n} = o(1)$ follows from equation expansions Lemma C.1.b,c and dominated convergence. ■

Proof of Lemma 2.5. Recall $J_n = J_n(\theta_0) = (\partial/\partial\theta)E[m_t^*(\theta)]|_{\theta_0}$ and $\hat{m}_n^*(\theta) = 1/n \sum_{t=1}^n \hat{m}_t^*(\theta)$. We only prove $\hat{J}_n^*(\tilde{\theta}_n) = J_n \times (1 + o_p(1))$ since $\check{J}_n^*(\tilde{\theta}_n) = J_n \times (1 + o_p(1))$ is similar.

Denote by $e_i \in \mathbb{R}^k$ the unit vector (e.g. $e_2 = [0, 1, 0, \dots, 0]'$), define a sequence of positive numbers $\{\varepsilon_n\}$ that satisfies $\varepsilon_n \rightarrow 0$, $\varepsilon_n \|V_n^{1/2}\| \rightarrow \infty$ and $\|\tilde{\theta}_n - \theta_0\|/\varepsilon_n \xrightarrow{p} 0$, and define

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

Minkowski's inequality implies for arbitrary θ

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

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

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

where $\|\Sigma_n^{-1/2}\|^{-1} / n^{1/2} = o_p(1) = o(\|J_n\|)$ by scale bound M2 and 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.9 and the properties of ε_n imply for any constant $a \in \mathbb{R}^k$

$$\begin{aligned} & \left\| \left\{ \hat{m}_n(\tilde{\theta}_n + a\varepsilon_n) - \hat{m}_n(\theta_0) \right\} - \left\{ E \left[m_t^* \left(\tilde{\theta}_n + a\varepsilon_n \right) \right] - E \left[m_t^* \left(\theta_0 \right) \right] \right\} \right\| \\ & \leq \left(1 + \left\| V_n^{1/2} \right\| \times \left\| \tilde{\theta}_n + a\varepsilon_n - \theta_0 \right\| \right) \times o_p \left(\|J_n\| \times \left\| V_n^{1/2} \right\|^{-1} \right) \\ & \leq \left(\left\| V_n^{1/2} \right\|^{-1} + \left\| \tilde{\theta}_n - \theta_0 \right\| + \|a\| \varepsilon_n \right) \times o_p(\|J_n\|) = o_p(\|J_n\| \varepsilon_n). \end{aligned}$$

Similarly, by differentiability of $E[m_t^*(\theta)]$,

$$\begin{aligned} & \left\| \frac{E \left[m_t^*(\tilde{\theta}_n + a\varepsilon_n) \right] - E \left[m_t^*(\theta_0) \right]}{\varepsilon_n} - aJ_n \right\| \\ & = \left\| J_n \varepsilon_n^{-1} \left(\tilde{\theta}_n + a\varepsilon_n - \theta_0 \right) - aJ_n + o_p \left(\|J_n\| \varepsilon_n^{-1} \left(\tilde{\theta}_n + \varepsilon_n - \theta_0 \right) \right) \right\| \\ & = \left\| J_n \varepsilon_n^{-1} \left(\tilde{\theta}_n - \theta_0 \right) \right\| + o_p(\|J_n\|) = o_p(\|J_n\|). \end{aligned}$$

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

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

■

Proof of Lemma 3.1. We require some properties of regularly varying tails. Since ϵ_t and stochastic $x_{i,t}$ are mutually independent and have tail (6) with indices κ_ϵ and κ_i , the convolutions $\epsilon_t x_{i,t}$ satisfy (Cline 1986)

$$m_{i,t}(\theta_0) = \epsilon_t x_{i,t} \sim (6) \text{ with index } \kappa_{\epsilon,i} := \min \{ \kappa_\epsilon, \kappa_i \}. \quad (9)$$

If $x_{1,t} = 1$ then (9) holds with $\kappa_{\epsilon,i} = \kappa_\epsilon$. Therefore by the construction of $c_{i,n}$ and k_n , and tail (6),

$$c_{i,n} = K \left(\frac{n}{k_n} \right)^{1/\kappa_{\epsilon,i}}. \quad (10)$$

Finally, processes z_t with regularly varying tails (6) and index $\kappa \in (0, 2]$ satisfy (see Feller 1971):

$$E \left[z_t^2 I(|z_t| \leq c) \right] \sim K c^2 \times P(|z_t| > c) \text{ as } c \rightarrow \infty. \quad (11)$$

Step 1 (Σ_n, Γ_n): Properties (9)-(1) imply

$$\Sigma_{i,i,n} = E \left[m_{i,n,t}^2(\theta_0) \right] \sim c_{i,n}^2 P(|m_{i,t}(\theta_0)| > c_{i,n}) \sim c_{i,n}^2 \frac{k_n}{n} = K \left(\frac{n}{k_n} \right)^{2/\kappa_{\epsilon,i}-1}.$$

Hence $\Gamma_{i,i,n} = (n/k_n)^{1/\kappa_{\epsilon,i}-1/2}$, and by the Cauchy-Schwartz inequality $\Sigma_{i,j,n} = O((n/k_n)^{1/\kappa_{\epsilon,i}+1/\kappa_{\epsilon,j}-1})$ $\forall i \neq j$.

Step 2 (J_n): The Lemma C.1.d Jacobian approximation implies by the linear equation form $J_{i,j,n} = -E[x_{i,t}x_{j,t}I(|\epsilon_t x_{i,t}| \leq c_{i,n})] \times (1 + o(1))$. Assume initially all regressors are stochastic. Since $\kappa_i, \kappa_\epsilon \leq 2$ it follows $\kappa_i < \kappa_\epsilon + 2$. Thus, by independence, (9) and (11)

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

The cross-products $E[x_{i,t}x_{j,t}I(|\epsilon_t x_{j,t}| \leq c_{j,n})]$ are bounded by case. If $x_{i,t}$ and $x_{j,t}$ are independent then $E[x_{i,t}x_{j,t}I(|\epsilon_t x_{i,t}| \leq c_{i,n})] = 0$ given $E[x_t] = 0$. If they are perfectly positively dependent with marginal tail (6) and $\kappa_i, \kappa_j \leq 2$ it can only be the case that $x_{i,t} = \text{sign}(x_{j,t}) \times |x_{j,t}|^p$ where $p = \kappa_j/\kappa_i$. Therefore, since $\kappa_j - p - 1 < \kappa_\epsilon$ is easy to verify,

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

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

Finally, if $x_{1,t} = 1$ then $E[x_{1,t}^2 I(|\epsilon_t x_{1,t}| \leq c_{1,n})] = P(|\epsilon_t| \leq c_{1,n}) = 1 - k_n/n$, and $E[x_{i,t}x_{1,t}I(|\epsilon_t x_{1,t}| \leq c_{1,n})] = E[x_{j,t}I(|\epsilon_t| \leq c_{1,n})] = O(1)$ and $E[x_{1,t}x_{i,t}I(|\epsilon_t x_{i,t}| \leq c_{i,n})] = E(E[x_{i,t}I(|x_{i,t}| \leq c_{i,n}/|\epsilon_t|)|\epsilon_t]) = O(1)$ given $x_{i,t}$ has a zero mean. 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 by noting

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

and

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

■

Proof of Lemmas 3.2 - 3.4. See Hill and Renault (2010). ■

APPENDIX B : GMTTM Criterion Properties

LEMMA B.1 (uniform criterion law) Recall $\mathfrak{m}_n := \sup_{\theta} E[|m_t^*(\theta)|]$. Under D1-D5 and D7

$$\sup_{\theta} \left\{ \frac{\mathfrak{m}_n^{-2} \times \left| \hat{Q}_n(\theta) - Q_n(\theta) \right|}{1 + \mathfrak{m}_n^{-2} \times Q_n(\theta)} \right\} = o_p(1).$$

LEMMA B.2 (moment expansion) Under D6.i and I1 $E[m_t^*(\theta)] - E[m_t^*(\theta')] = J_n(\theta')(\theta - \theta') + o(\|J_n(\theta')\| \times \|\theta - \theta'\|)$ for any $\theta, \theta' \in \Theta$.

Proof of Lemma B.1. Recall $\hat{m}_n^*(\theta) := 1/n \sum_{t=1}^n \hat{m}_t^*(\theta)$ and $m_n^*(\theta) := 1/n \sum_{t=1}^n m_t^*(\theta)$. By weight property M1

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

Uniform approximation Lemma C.2, smoothness I3 and threshold bound D5 imply $A_{1,n}(\theta)$ satisfies

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

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

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

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

Proof of Lemma B.2. The claim follows Jacobian existence D6.i and the definition of a derivative. ■

APPENDIX C: Limit Theory for Trimmed Sums

The following appendix contains limit theory for tail and tail trimmed arrays. The first two results characterize differentiability and expansions of the trimmed equations, and the rate of approximation for $\hat{m}_t^*(\theta)$.

LEMMA C.1 (expansions) *Assume D1-D7, I1-I2 and M2-M3 hold. Write $m_t^*(\theta) = m_t(\theta) \times I_t(\theta)$ and $\hat{m}_t^*(\theta) = m_t(\theta) \times \hat{I}_t(\theta)$. Choose $\theta, \theta', \tilde{\theta} \in \Theta$ where $\|\theta - \tilde{\theta}\| \leq \delta$ for tiny $\delta > 0$, and define*

$$R_n(\theta, \theta', \tilde{\theta}) = \left(\left\| \Sigma_n^{-1/2}(\theta') \right\|^{-1} / n^{1/2} + K \|J_n\| \right) \times \left\| \theta - \tilde{\theta} \right\| \in \mathbb{R}^q.$$

Notice θ' is otherwise unrestricted. In the following $o_p(1)$ terms are not functions of θ . For some sequence $\{\theta_{n,}, \tilde{\theta}_{n,*}\}$ satisfying $\|\theta_{n,*} - \tilde{\theta}\| \leq \|\theta - \tilde{\theta}\|$ and $\|\tilde{\theta}_{n,*} - \tilde{\theta}\| \leq \|\theta - \tilde{\theta}\|$, as $n \rightarrow \infty$ and/or as $\delta \rightarrow 0$:*

- a. (equation expansions): i. $m_n^*(\theta) = m_n^*(\tilde{\theta}) + J_n^*(\theta_{n,*})(\theta - \tilde{\theta}) + R_n(\theta, \theta', \tilde{\theta}) \times o_p(1)$; and ii. $\hat{m}_n^*(\theta) = \hat{m}_n^*(\tilde{\theta}) + \hat{J}_n^*(\tilde{\theta}_{n,*})(\theta - \tilde{\theta}) + R_n(\theta, \theta', \tilde{\theta}) \times o_p(1)$;*
- b. (expansion between $\theta_{n,0}$ and θ_0): $\|\theta_{n,0} - \theta_0\| = o(\|\Sigma_n^{-1/2}\|^{-1} / [n \|J_n\|^2]) = o(\|V_n^{1/2}\|^{-1})$ and $\|m_n^*(\theta_0) - m_n^*(\theta_{n,0})\| = \{\|\Sigma_n^{-1/2}(\theta)\|^{-1} / n^{1/2}\} \times o_p(1)$ for any $\theta \in \Theta$;*
- c. (cross-product expansion): $m_t^*(\theta)m_t^*(\theta)' - m_t^*(\tilde{\theta})m_t^*(\tilde{\theta})' = R_n(\theta, \theta', \tilde{\theta})^2 \times o_p(1)$ for each $1 \leq t \leq n$;*
- d. (Jacobian): $J_n = E[J_t^*] \times (1 + o(1))$.*

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

Next, a uniform law for $m_t^*(\theta)$ and rates for the trimming components $m_{i,(k_i,n)}(\theta)$ and $I(|m_{i,t}(\theta)| > c_{i,n})$. Recall $\mathbf{m}_n := \sup_{\theta} |E[m_t^*(\theta)]|$.

LEMMA C.3 (uniform LLN) *Under D2-D5 and I1 $1/n \sum_{t=1}^n m_t^*(\theta_{n,0}) = o_p(1)$; under D2-D5 $\sup_{\theta} \{ \|1/n \sum_{t=1}^n (m_t^*(\theta) - E[m_t^*(\theta)])\| \} = o_p(\mathbf{m}_n)$.*

LEMMA C.4 (asymptotic linearity) *Under D1-D7, I1-I3 and M1-M3*

$$V_n^{1/2} \left(\hat{\theta}_n - \theta_0 \right) = A_n \sum_{t=1}^n \hat{m}_t^*(\theta_0) \times (1 + o_p(1)) + o_p(1) \text{ a.s.}$$

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

LEMMA C.5 (uniform indicator bounds) *Let D1-D5 and D7 hold and let $i \in \{1, \dots, q\}$ be arbitrary.*

- a. $P(\sup_{\theta} \sup_{\tilde{\theta}: \|\theta - \tilde{\theta}\| \leq \delta} |I_{i,t}(\theta) - I_{i,t}(\tilde{\theta})| = 1) \rightarrow 0$ as $n \rightarrow \infty$ and/or $\delta \rightarrow 0$.*
- b. Define $X_t^*(\theta) := ((n/k_n)^{1/2-\iota}) \{I_{i,t}(\theta) - E[I_{i,t}(\theta)]\}$ for tiny $\iota > 0$, and $X_t^*(\theta, \tilde{\theta}) := X_t^*(\theta) - X_t^*(\tilde{\theta})$. Then $E[(\sup_{\theta} \{ |n^{-1/2} \sum_{t=1}^n X_t^*(\theta) | \})^2] = O(1)$ and $E[(n^{-1/2} \sum_{t=1}^n X_t^*(\theta, \tilde{\theta}))^2] = O(1) \times \|\theta - \tilde{\theta}\|$ where $O(1)$ is not a function of θ .*

The next three ensure a multivariate asymptotic Gaussian law for $\sum_{t=1}^n m_t^*(\theta_0)$ based theory developed in Hill (2010b) for mixingale tail-trimmed arrays. We present them here for completeness and ease of reference. Let $\{A_n(\theta)\}$ be any $\mathbb{R}_n^{q \times q}$ -valued, non-stochastic sequences that satisfy $A_n(\theta)\{\sum_n(\theta)n\}A_n'(\theta) \rightarrow I_q$, and define for any $r \in \mathbb{R}^q$, $r'r = 1$,

$$z_t^*(r, \theta) := r'A_n(\theta)m_t^*(\theta) \in \mathbb{R}.$$

LEMMA C.6 (CLT) *Under D4, D5, I1 and I2 $\sum_{t=1}^n z_t^*(r, \theta_0) \xrightarrow{d} N(0, 1) \forall r'r = 1$.*

The proof of Lemma C.6 shows $\sum_{t=1}^n z_t^*(r, \theta_{n,0}) \xrightarrow{d} N(0, 1)$ and $\sum_{t=1}^n \{z_t^*(r, \theta_0) - z_t^*(r, \theta_{n,0})\} = o_p(1)$ since only $\{z_t^*(r, \theta_{n,0}), \mathfrak{F}_t\}$ is guaranteed to form a martingale difference array under I1. Nevertheless, if $m_t^*(\theta_{n,0})$ is sufficiently heavy-tailed as $n \rightarrow \infty$ extant martingale difference limit theory fails to apply (e.g. McLeish 1974), so we exploit a telescoping sum arguments. Define positive integer sequences $\{j_n, r_n, h_n\}$ satisfying $h_n, j_n \rightarrow \infty$, $1 \leq h_n, j_n, r_n \leq n$,

$$h_n = o(c_n^\iota) \text{ for tiny } \iota > 0, \quad (12)$$

$r_n = \lfloor n/h_n \rfloor$, and $j_n = o(h_n)$, and define $F_{n,i} := \sigma(\cup_{\tau \leq ih_n} \mathfrak{F}_\tau)$,

$$Z_{n,i} = \sum_{t=(i-1)h_n+j_n}^{ih_n} z_t^*(r, \theta_{n,0}) \text{ and } W_{n,i} = E[Z_{n,i}|F_{n,i}] - E[Z_{n,i}|F_{n,i-1}].$$

LEMMA C.7 (decomposition) *Under D4 and D5 $\sum_{t=1}^n z_t^*(r, \theta_{n,0}) = \sum_{i=1}^{r_n} W_{n,i} + o_p(1)$.*

LEMMA C.8 (asymptotic variance) *Under D4, D5, I1 and I2 $\sum_{i=1}^{r_n} W_{n,i}^2 \xrightarrow{p} 1$.*

Lastly, stochastic differentiability aids proving Jacobian estimator consistency.

LEMMA C.9 (stochastic differentiability) *Under D1-DD7, I1 and M2-M3, for all $\{\delta_n\}$, $\delta_n \rightarrow 0$, and all $\theta \in \Theta$,*

$$\sup_{\theta \in U_0(\delta_n)} \left\{ \frac{\left(\left\| V_n^{1/2} \right\| / \left\| J_n \right\| \right) \left\| \hat{m}_n^*(\theta) - \hat{m}_n^*(\theta_0) \right\| - \left\{ E[m_t^*(\theta)] - E[m_t^*(\theta_0)] \right\}}{1 + \left\| V_n^{1/2} \right\| \times \|\theta - \theta_0\|} \right\} \xrightarrow{p} 0.$$

Proof of Lemma C.1.

Claim (a): We prove (i) since (ii) is similar. Assume θ and $m_t(\theta)$ are scalar and $m_t(\theta)$ is symmetrically trimmed to simplify notation. Write $m_t^*(\theta) = m_t(\theta) \times I_t(\theta)$ where $I_t(\theta) = I(|m_t(\theta)| \leq c_n(\theta))$, and choose $\|\theta - \tilde{\theta}\| \leq \delta$. Use D2 to deduce by Taylor's theorem

$$m_t^*(\theta) = \left\{ m_t(\tilde{\theta}) + J_t^*(\theta_{n,\delta})(\theta - \theta_{n,0}) \right\} \times I_t(\theta)$$

where $\|\theta_{n,\delta} - \theta_{n,0}\| \leq \|\theta - \tilde{\theta}\|$, and $J_t^*(\theta) := (\partial/\partial\theta)m_t(\theta)$. Therefore

$$\begin{aligned} m_n^*(\theta) - m_n^*(\tilde{\theta}) &= J_n^*(\theta_{n,\delta}) \times (\theta - \tilde{\theta}) + \frac{1}{n} \sum_{t=1}^n m_t(\theta) \times \left\{ I_t(\theta) - I_t(\tilde{\theta}) \right\} \\ &\quad + \frac{1}{n} \sum_{t=1}^n J_t^*(\theta_{n,\delta}) \times \left\{ I_t(\theta) - I_t(\tilde{\theta}) \right\} \times (\theta - \theta_{n,0}). \end{aligned} \quad (13)$$

Consider the second term in (13). Since $I_t(\theta) - I_t(\tilde{\theta}) \in \{-1, 0, 1\}$ we can write

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

Step 1 ($A_n(\theta, \tilde{\theta})$): Threshold property D5 implies $E|I_t(\theta) - I_t(\tilde{\theta})| = O(k_n^{1/2}/n)$ where $O(\cdot)$ is not a function of θ . Now use stationarity D3, envelope bound D4, the Cauchy-Schwartz inequality and $\sup_{\theta} \|\Sigma_n^{-1/2}(\theta)\| = O(1)$ by D5, I1 and M3 given covariance bound (8.4) to deduce for tiny $\iota > 0$ and any $\theta' \in \Theta$

$$\begin{aligned} \left(E \left[A_n(\theta, \tilde{\theta})^\iota \right] \right)^{1/\iota} &\leq n^{1/2} \left[E \left| m_t(\theta) \{I_t(\theta) - I_t(\tilde{\theta})\} \right|^\iota \right]^{1/\iota} \\ &= O \left(n^{1/2} \left\| \Sigma_n^{-1/2}(\theta') \right\|^{-1} \times \left(k_n^{1/2}/n \right)^{1/\iota} \right). \end{aligned}$$

Since $\iota > 0$ is arbitrarily small use D5 to deduce $[(k_n^{1/2}/n)^{1/\iota} \leq 1/(nk_n^\iota)]$. Hence by Markov's inequality for $o_p(\cdot)$ not a function of θ

$$A_n(\theta, \tilde{\theta}) = o_p \left(n^{1/2} \left\| \Sigma_n^{-1/2}(\theta') \right\|^{-1} \left(k_n^{1/2}/n \right)^{1/\iota} \right) = o_p \left(\left\| \Sigma_n^{-1/2}(\theta') \right\|^{-1} \times \left[n^{1/2} k_n^\iota \right]^{-1} \right).$$

Step 2 ($B_n(\theta, \tilde{\theta})$): Subadditivity and $|I_t(\theta) - I_t(\tilde{\theta})| \in \{0, 1\}$ imply for any $\xi > 0$, tiny $\iota > 0$, and sufficiently large n and/or as $\|\theta - \tilde{\theta}\| \leq \delta \rightarrow 0$

$$\begin{aligned} P \left(\frac{1}{k_n^\iota} \left| \frac{1}{n^{1/2}} B_n(\theta, \tilde{\theta}) - \frac{1}{n} \sum_{i=1}^n |I_{(i)}(\theta) - I_{(i)}(\tilde{\theta})| \right| > 2\xi \right) \\ \leq \sum_{t=1}^n P \left(|I_t(\theta) - I_t(\tilde{\theta})| > k_n^\iota \xi \right) + \sum_{i=1}^n P \left(|I_{(i)}(\theta) - I_{(i)}(\tilde{\theta})| > k_n^\iota \xi \right) = 0. \end{aligned}$$

In order to bound $1/n \sum_{i=1}^n |I_{(i)}(\theta) - I_{(i)}(\tilde{\theta})|$ define $k_n^*(\theta) := \sum_{t=1}^n \{1 - I_t(\theta)\}$. By construction

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n |I_{(i)}(\theta) - I_{(i)}(\tilde{\theta})| &= \frac{1}{n} \sum_{i=k_n^*(\tilde{\theta})+1}^{k_n^*(\theta)} I_{(i)}(\tilde{\theta}) + \frac{1}{n} \sum_{i=k_n^*(\theta)+1}^{k_n^*(\tilde{\theta})} (1 - I_{(i)}(\tilde{\theta})) \\ &\leq 2 \frac{k_n}{n} \frac{1}{k_n} |k_n^*(\theta) - k_n^*(\tilde{\theta})| \leq 2 \frac{k_n^{1/2}}{n} \left\{ |X_n(\theta, \tilde{\theta})| + |P_n(\theta, \tilde{\theta})| \right\} \end{aligned}$$

where

$$\begin{aligned} X_n(\theta, \tilde{\theta}) &:= \frac{1}{k_n^{1/2}} \sum_{t=1}^n \left\{ \Delta I_t(\theta, \tilde{\theta}) - E \left[\Delta I_t(\theta, \tilde{\theta}) \right] \right\} \quad \text{and} \quad \Delta I_t(\theta, \tilde{\theta}) := I_t(\theta) - I_t(\tilde{\theta}) \\ P_n(\theta, \tilde{\theta}) &:= \frac{n}{k_n^{1/2}} \{P(|m_t(\theta)| > c_n(\theta)) - 1\} - \frac{n}{k_n^{1/2}} \{P(|m_t(\tilde{\theta})| > c_n(\tilde{\theta})) - 1\}. \end{aligned}$$

Apply uniform indicator bound Lemma C.5.b to deduce $|X_n(\theta, \tilde{\theta})| = O_p((n/k_n)^\iota) \times \|\theta - \tilde{\theta}\|$ where $O_p(\cdot)$ is not functions of θ . Further, absolute continuity D1 ensures $P_n(\theta, \tilde{\theta})$

is Θ -a.e. differentiable (Royden 1968), so threshold property D5 and $P_n(\theta, \theta) = 0$ imply $P_n(\theta, \tilde{\theta}) = O(1) \times \|\theta - \tilde{\theta}\|$. Coupled with covariance bound (8.5) we deduce as $n \rightarrow \infty$ and/or $\|\theta - \tilde{\theta}\| \leq \delta \rightarrow 0$

$$B_n(\theta, \tilde{\theta}) = O_p \left(k_n^t n^{1/2} \frac{k_n^{1/2}}{n} \left(\frac{n}{k_n} \right)^t \|\theta - \tilde{\theta}\| \right) = O_p \left(k_n^t (k_n/n)^{1/2-t} \right) \times \|\theta - \tilde{\theta}\|.$$

Step 3: Steps 1 and 2 with $k_n/n \rightarrow \infty$ deliver as $n \rightarrow \infty$ and/or $\|\theta - \tilde{\theta}\| \leq \delta \rightarrow 0$

$$\frac{1}{n} \sum_{t=1}^n m_t(\theta) \{I_t(\theta) - I_t(\tilde{\theta})\} = o_p \left(\left\{ \left\| \Sigma_n^{-1/2}(\theta') \right\|^{-1} / n^{1/2} \right\} \right) \times \|\theta - \tilde{\theta}\|.$$

Repeat the argument for the third term in (13) by invoking envelope bound D4 for $J_t^*(\theta)$.

Claim (b): Moment expansion Lemma B.2 and $E[m_t^*(\theta_{n,0})] = 0$ given I1 and (8.1) imply $E[m_t^*(\theta_0)] = J_n(\theta_0 - \theta_{n,0}) + o(\|J_n\| \times \|\theta_0 - \theta_{n,0}\|)$, but $E[m_t^*(\theta_0)] = o(\|\Sigma_n^{-1/2}\|^{-1}/n^{1/2})$ by I2. Therefore $\|\theta_{n,0} - \theta_0\| = o(\|\Sigma_n^{-1/2}\|^{-1}/[n\|J_n\|^2])$ which is $o(\|V_n^{1/2}\|^{-1})$ under scale bound M2.

Now invoke Claim (a) and Jacobian consistency Lemma 2.5 to deduce for any $\theta \in \Theta$

$$\begin{aligned} |m_n(\theta_0) - m_n(\theta_{n,0})| &\leq \|J_n\| \times o \left(\left\| \Sigma_n^{-1/2}(\theta) \right\|^{-1} / [n^{1/2} \|J_n\|] \right) \\ &\quad + \left(\left\| \Sigma_n^{-1/2}(\theta) \right\|^{-1} / n^{1/2} + K \|J_n\| \right) \times o_p \left(\left\| \Sigma_n^{-1/2} \right\|^{-1} / [n^{1/2} \|J_n\|] \right) \\ &= o_p(1) \times \left\| \Sigma_n^{-1/2}(\theta) \right\|^{-1} / n^{1/2} \end{aligned}$$

where $o_p(1)$ is not a function of θ .

Claim (c): The proof simply initiates the Claim (a) argument.

Claim (d): Claim (a) and bounded convergence imply as $n \rightarrow \infty$ and/or $\|\theta - \theta_{n,\delta}\| \leq \|\theta - \theta_0\| \leq \delta \rightarrow 0$

$$\frac{E[m_t^*(\theta)] - E[m_t^*(\theta_0)]}{\|\theta - \theta_0\|} = E[J_t^*(\theta_{n,\delta})] \times (1 + o(\|\theta - \theta_0\|)) + o \left(\left\| \Sigma_n^{-1/2} \right\|^{-1} / n^{1/2} + \|J_n\| \right).$$

Further, moment expansion Lemma B.2 asserts

$$\frac{E[m_t^*(\theta)] - E[m_t^*(\theta_0)]}{\|\theta - \theta_0\|} = J_n \times (1 + o(\|\theta - \theta_0\|)) + o(\|J_n\|).$$

Invoke covariance bound (8.5) and take $\delta \rightarrow 0$ to complete the proof. ■

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

Claim 1: Let $\theta \in \Theta$ be arbitrary, and write $m_t = m_t(\theta)$, $c_n = c_n(\theta)$, $\hat{m}_t^* = \hat{m}_t^*(\theta)$, $m_t^* = m_t^*(\theta)$, $\bar{I}_t = 1 - I_t(\theta)$, and $\hat{I}_t = \hat{I}_t(\theta)$. First bound

$$\left\| \frac{1}{n} \sum_{t=1}^n \{\hat{m}_t^* - m_t^*\} \right\| \leq \max_{1 \leq t \leq n} \left\{ \left\| m_t \{ \hat{I}_t - I_t \} \right\| \right\} \times \frac{1}{n} \sum_{t=1}^n \left\| \hat{I}_t - I_t \right\|.$$

By construction $\|m_t\{\hat{I}_t - I_t\}\| \leq 2\|m_{(k_n)}^{(a)} - c_n\|$. Now, the intermediate order statistic is consistent $m_{(k_n)}^{(a)}/c_n = 1 + O_p(k_n^{-1/2})$ by Lemma C.2.1 below. Now use threshold bound D5 to deduce

$$\max_{1 \leq t \leq n} \left\{ \|m_t\{\hat{I}_t - I_t\}\| \right\} \leq 2\|m_{(k_n)}^{(a)} - c_n\| = 2c_n \left\| m_{(k_n)}^{(a)}/c_n - 1 \right\| = o_p\left((n/k_n)^{1/2}\right).$$

Next, by construction and the triangular inequality

$$\frac{1}{n} \sum_{t=1}^n \|\hat{I}_t - I_t\| \leq \frac{k_n^{1/2}}{n} \left\| \frac{1}{k_n^{1/2}} \sum_{t=1}^n \{\bar{I}_t - E[\bar{I}_t]\} \right\| + \frac{k_n^{1/2}}{n} \left\| k_n^{1/2} \left(\frac{n}{k_n} E[\bar{I}_t] - 1 \right) \right\|$$

which is $O_p(k_n^{1/2}/n)$ by D5 and an application of Lemma C.5.b. Together with covariance bound (8.5) $\sum_{t=1}^n \{\hat{m}_t^* - m_t^*\} = o_p((n/k_n)^{1/2} k_n^{1/2}) = o_p(n^{1/2})$.

Claim 2: Define $M_n^* := \max_{1 \leq t \leq n} \{\sup_{\theta} \|m_t(\theta)\{\hat{I}_t(\theta) - I_t(\theta)\}\|\}$ and repeat the above argument to reach

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

Uniform indicator bound Lemma C.5.b and uniform threshold property D5 imply the right-hand-side is $O_p(M_n^* k_n^{1/2}/n)$.

It remains to show $M_n^* = o_p(n/k_n^{1/2})$. First, since $\|m_t(\theta)\{\hat{I}_t(\theta) - I_t(\theta)\}\| \leq 2\|m_{(k_n)}^{(a)}(\theta) - c_n(\theta)\|$ for any t use D5 to deduce M_n is bounded by

$$2 \sup_{\theta} \left\{ \left\| m_{(k_n)}^{(a)}(\theta) - c_n(\theta) \right\| \right\} = o_p \left(\left(\frac{n}{k_n} \right)^{1/2} \sup_{\theta} \left\{ k_n^{1/2} \left\| m_{(k_n)}^{(a)}(\theta)/c_n(\theta) - 1 \right\| \right\} \right).$$

Property D5 and uniform bound Lemma C.5.b suffice to conclude $\sup_{\theta} \|1/n \sum_{t=1}^n \{\hat{m}_t^* - m_t^*\}\| = o_p(M_n^* k_n^{1/2}/n)$. Second, uniform L_L -boundedness D4 and Markov's inequality imply for some $\kappa > 0$ sufficiently small

$$P(|m_t(\theta)| > z) = K \times z^{-\kappa} \times \mathbf{O}(1) \text{ as } z \rightarrow \infty$$

$$P(|m_t(\theta)| > z) = K \times z^{-\kappa} \times \mathbf{O}(1) \times (1 + \mathbf{O}(g(z))) \text{ as } z \rightarrow \infty$$

where $\mathbf{O}(\cdot)$ is a contraction mapping, $\mathbf{O}(z) \in [0, z]$, $z > 0$, and $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is any measurable mapping that satisfies $k_n^{1/2} g(c_n) \rightarrow 0$. Now apply Lemma C.5.b and Hsing's (1991: p. 1553) argument to reach $\sup_{\theta} \{k_n^{1/2} \|m_{(k_n)}^{(a)}(\theta)/c_n(\theta) - 1\|\} = O_p(1)^9$. Therefore $M_n^* \leq 2 \sup_{\theta} \{\|m_{(k_n)}^{(a)}(\theta) - c_n(\theta)\|\} = o_p((n/k_n)^{1/2}) = o_p(n/k_n^{1/2})$ as required. ■

LEMMA C.2.1 Under D1, D3 and D4 $m_{(k_n)}^{(a)}(\theta)/c_n(\theta) = 1 + O_p(k_n^{-1/2})$ for every $\theta \in \Theta$.

⁹See especially the proof of Theorem 2.4 of Hsing (1991). Hsing exploits regular variation and does not deliver uniform results. His argument, however, trivially extends to the uniform case and to tails bounded by a regularly varying function. See also Lemma 3.1.1 of Hill (2010b).

Proof. See Lemma 3.1.1 of Hill (2010b) for a nearly identical result. In brief, marginal distribution support \mathbb{R} under D1, geometric α -mixing D3 and measurability ensure the tail array $\{\bar{I}_{i,t}(\theta)\} := \{1 - I_{i,t}(\theta)\}$ satisfies a pointwise central limit theorem (Hill 2009: Theorem 2.1, Lemma 3.1): $1/k_n^{1/2} \sum \{\bar{I}_{i,t}(\theta) - E[\bar{I}_{i,t}(\theta)]\} \xrightarrow{d} N(0, w_1^2(\theta))$ for some $w_1^2(\theta) < \infty$. It is then straightforward to prove an asymptotic identity between $k_n^{1/2} \ln(m_{(k_n)}^{(a)}(\theta)/c_n(\theta))$ and $1/k_n^{1/2} \sum \{\bar{I}_{i,t}(\theta) - E[\bar{I}_{i,t}(\theta)]\}$ based solely on L_p -boundedness D4 so that $k_n^{1/2} \ln(m_{(k_n)}^{(a)}(\theta)/c_n(\theta)) \xrightarrow{d} N(0, w_2^2(\theta))$ for some $w_2^2(\theta) < \infty$. The claim then follows from the mean-value-theorem. ■

Proof of Lemma C.3. The first claim follows from I1 and covariance bound (8.5):

$$\left\| E \left[\left(\frac{1}{n} \sum_{t=1}^n m_t^*(\theta_{n,0}) \right) \left(\frac{1}{n} \sum_{t=1}^n m_t^*(\theta_{n,0}) \right)' \right] \right\| = \|\Sigma_n(\theta_{n,0})\|/n = o(1).$$

Now invoke Chebyshev's inequality.

Consider the second claim, recall $\mathbf{m}_n := \sup_{\theta} E[|m_t^*(\theta)|]$ and define $M_t^*(\theta) := \mathbf{m}_n^{-1} m_{i,t}^*(\theta)$ and $X_t^*(\theta) := M_t^*(\theta) - E[M_t^*(\theta)]$ for arbitrary $i \in \{1, \dots, q\}$, and write $X_n^*(\theta) := 1/n \sum_{t=1}^n X_t^*(\theta)$.

Step 1: $X_t^*(\theta)$ is for any θ geometrically α -mixing under D3. Therefore stationarity, Ibragimov's (1962) bound, and $|m_{i,t}^*(\theta)| \leq c_n$ imply for some $\rho \in (0, 1)$ and tiny $\iota > 0$

$$\begin{aligned} E(X_n^*(\theta))^2 &\leq \frac{1}{n} E[(X_t^*(\theta))^2] + \frac{1}{n} \sum_{t=2}^{\infty} |E[X_1^*(\theta)X_t^*(\theta)]| \\ &\leq K \frac{1}{n} c_n^2(\theta) + \frac{1}{n} \sum_{t=2}^{\infty} \rho^t \left(E|X_t^*(\theta)|^{2+\iota} \right)^{1-\iota} \leq K \left(\frac{1}{n} c_n^2(\theta) + \frac{1}{n} c_n^{2+\iota}(\theta) \right). \end{aligned}$$

The right-hand-side is $o(1)$ by threshold bound D5 and a continuity argument, hence point-wise convergence follows by Chebyshev's inequality.

Step 2: Compactness of Θ and pointwise convergence imply uniform convergence follows by Theorem 1 of Andrews (1992) if we demonstrate stochastic equicontinuity: $\forall \epsilon > 0$ there exists $\delta > 0$ such that $\lim_{n \rightarrow \infty} P(\sup_{\theta} \sup_{\tilde{\theta}: \|\tilde{\theta} - \theta\| \leq \delta} |X_n(\theta) - X_n(\tilde{\theta})| > \epsilon) < \epsilon$.

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

Consider two possibly over-lapping cases, and write $\sup_{\tilde{\theta}(\delta)}$ for $\sup_{\tilde{\theta}: \|\tilde{\theta} - \theta\| \leq \delta}$. First, if $\lim_{n \rightarrow \infty} E[\sup_{\theta} |m_t^*(\theta)|] < \infty$ then $\lim_{n \rightarrow \infty} E[\sup_{\theta} |M_t^*(\theta)| \times I(\sup_{\theta} |M_t^*(\theta)| > \zeta)] \leq K(\zeta)$ for all $\zeta > 0$ and some mapping $K(\zeta) \searrow 0$ as $\zeta \rightarrow \infty$. Further, Andrews' (1992) Assumption TSE [*termwise stochastic equicontinuity*] holds by Lemma C.3.1, below. Therefore stochastic equicontinuity follows from Lemma 3 of Andrews (1992).

Second, if $\mathbf{m}_n := \sup_{\theta} E[|m_t^*(\theta)|] \rightarrow \infty$ then we need only verify conditions (16) and (17) of Lemma C.3.2. Define for arbitrary $\epsilon > 0$

$$\zeta_n := P \left(\sup_{\theta} \sup_{\tilde{\theta}(\delta)} |M_t^*(\theta) - M_t^*(\tilde{\theta})| > \epsilon \right)^{-1+\iota}.$$

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

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

Condition (16) is trivial since since $E[\sup_{\theta} |M_t^*(\theta)|] \leq 1$ and $\zeta_n \rightarrow \infty$. Condition (17) follows by the construction of M_n :

$$\begin{aligned} \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \left\{ \zeta_n \times P \left(\sup_{\theta} \sup_{\tilde{\theta}(\delta)} |X_t^*(\theta) - X_t^*(\tilde{\theta})| > \epsilon \right) \right\} \\ = \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \left\{ P \left(\sup_{\theta} \sup_{\tilde{\theta}(\delta)} |X_t^*(\theta) - X_t^*(\tilde{\theta})| > \epsilon \right)^{\iota} \right\} = 0. \end{aligned}$$

■

LEMMA C.3.1 *If $\lim_{n \rightarrow \infty} E[\sup_{\theta} \|m_t^*(\theta)\|] < \infty$ then TSE holds.*

Proof. An argument similar to Čížek'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 $\kappa > 0$

$$P \left(\sup_{\theta} \sup_{\tilde{\theta}(\delta)} |m_{i,t}(\theta) \times \{I_{i,t}(\theta) - I_{i,t}(\tilde{\theta})\}| > \kappa \right) \leq \epsilon, \quad (14)$$

$$P \left(\sup_{\theta} \sup_{\tilde{\theta}(\delta)} |m_{i,t}(\theta) - m_{i,t}(\tilde{\theta})| \times I_{i,t}(\tilde{\theta}) > \kappa \right) \leq \epsilon. \quad (15)$$

Since

$$\sup_{\theta} \sup_{\tilde{\theta}(\delta)} |m_{i,t}(\theta) \{I_{i,t}(\theta) - I_{i,t}(\tilde{\theta})\}| \leq \sup_{\theta} |m_{i,t}(\theta) I_{i,t}(\theta)| \times \sup_{\theta} \sup_{\tilde{\theta}(\delta)} |I_{i,t}(\theta) - I_{i,t}(\tilde{\theta})|$$

use $\lim_{n \rightarrow \infty} E[\sup_{\theta} \|m_t^*(\theta)\|] < \infty$ and indicator bound Lemma C.5.a to deduce (14). Next, differentiability D2 and moment bounds D4 imply for given $\kappa > 0$ and sufficiently tiny $\iota > 0$

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

as $\delta \rightarrow 0$, hence (15) holds. ■

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

$$\zeta_n \rightarrow \infty,$$

$$\lim_{n \rightarrow \infty} E \left[\sup_{\theta} |M_t^*(\theta)| \times I \left(\sup_{\theta} |M_t^*(\theta)| > \zeta_n \right) \right] = 0 \quad (16)$$

and

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \left\{ \zeta_n \times P \left(\sup_{\theta} \sup_{\tilde{\theta}(\delta)} |M_t^*(\theta) - M_t^*(\tilde{\theta})| > \epsilon \right) \right\} = 0 \quad (17)$$

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 (16). Define $s_t^* := \sup_{\theta} |M_t^*(\theta)|$ and $\tilde{M}_t^*(\delta) := \sup_{\tilde{\theta}(\delta)} |M_t^*(\theta) - M_t^*(\tilde{\theta})|$. For given $\epsilon > 0$ by supposition there exists $\delta > 0$ such that, uniformly in

n , $E[2s_t^* I(2s_t^* > \zeta_n)] \leq \epsilon^2/6$ and $P(\check{M}_t^*(\delta) > \epsilon^2/6) \leq \epsilon^2/[6\zeta_n]$. Then by stationarity, Marvok's inequality and term-wise stochastic equicontinuity for sufficiently large n

$$\begin{aligned}
& P\left(\sup_{\theta} \sup_{\tilde{\theta}(\delta)} \left|X_n^*(\theta) - X_n^*(\tilde{\theta})\right| > \epsilon\right) \\
& \leq P\left(\check{M}_t^*(\delta) + E[\check{M}_t^*(\delta)] > \epsilon\right) \leq \frac{2}{\epsilon} \times E[\check{M}_t^*(\delta)] \\
& \leq \frac{2}{\epsilon} E\left[\check{M}_t^*(\delta) I\left(\check{M}_t^*(\delta) \leq \frac{\epsilon^2}{6}\right)\right] + \frac{2}{\epsilon} \times E\left[\check{M}_t^*(\delta) I\left(\frac{\epsilon^2}{6} < \check{M}_t^*(\delta) \leq \zeta_n\right)\right] \\
& \quad + \frac{2}{\epsilon} \times E\left[\check{M}_t^*(\delta) I(\zeta_n < \check{M}_t^*(\delta))\right] \\
& \leq \frac{2}{\epsilon} \times \left\{\frac{\epsilon^2}{6} + \zeta_n \times P\left(\check{M}_t^*(\delta) > \frac{\epsilon^2}{6}\right) + 2 \times E[s_t^* I(2s_t^* > \zeta_n)]\right\} \leq \epsilon.
\end{aligned}$$

■

Proof of Lemma C.4. Under D1 and D2 $\hat{Q}_n(\theta)$ is continuous on Θ , and twice differentiable at $\hat{\theta}_n$ by Čížek's (2008: Lemma 2.1) argument. Therefore $\hat{Q}_n(\hat{\theta}_n) \leq \hat{Q}_n(\theta) \forall \theta \in \Theta$ implies

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

Consistency $\|\hat{\theta}_n - \theta_0\| = o_p(1)$ by Theorem 2.1 and the Lemma C.1.a asymptotic expansion for $1/n \sum_{t=1}^n \hat{m}_t^*(\hat{\theta}_n)$ imply we may write

$$\begin{aligned}
& \hat{J}_n^*(\hat{\theta}_n)' \hat{\Upsilon}_n \left\{ \hat{J}_n^*(\theta_{n,*})' (\hat{\theta}_n - \theta_0) + \frac{1}{n} \sum_{t=1}^n \hat{m}_t^*(\theta_0) \right\} \\
& \quad + \hat{J}_n^*(\hat{\theta}_n)' \hat{\Upsilon}_n \times o_p\left(\left\{\left\|\Sigma_n^{-1/2}\right\|^{-1} / n^{1/2} + K \|J_n\|\right\} \times \|\hat{\theta}_n - \theta_0\|\right) = 0 \text{ a.s.}
\end{aligned}$$

where $\|\theta_{n,*} - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$.

Since $\|\hat{\theta}_n - \theta_0\| \xrightarrow{p} 0$ the Jacobian limit Lemma 2.5 implies 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, weight and Jacobian properties M1 and D6.i imply $H_n^{-1} := (J_n' \Upsilon_n J_n)^{-1}$ exists. Now re-arrange terms and exploit the construction of V_n and property M2 to deduce

$$\hat{\theta}_n - \theta_0 = -H_n^{-1} J_n' \Upsilon_n \frac{1}{n} \sum_{t=1}^n \hat{m}_t^*(\theta_0) + o_p\left(\|\hat{\theta}_n - \theta_0\|\right).$$

We may therefore write

$$\begin{aligned}
V_n^{1/2} (\hat{\theta}_n - \theta_0) & = -\left\{V_n^{1/2} H_n^{-1} J_n' \Upsilon_n\right\} \frac{1}{n} \sum_{t=1}^n \hat{m}_t^*(\theta_0) \times (1 + o_p(1)) + o_p(1) \\
& = A_n \sum_{t=1}^n \hat{m}_t^*(\theta_0) \times (1 + o_p(1)) + o_p(1).
\end{aligned}$$

■

Proof of Lemma C.5. Assume for clarity θ and $m_t(\theta)$ are scalars ($q = k = 1$) and $m_t(\theta)$ is symmetrically distributed, hence $(n/k_n)P(|m_t(\theta)| > c_n(\theta)) \rightarrow 1$.

Claim (a): Distribution and equation continuity D1 and D2 and threshold property D5 imply we may assume $c_n(\theta)$ is 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 $\iota > 0$

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

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

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

But this implies $P_n(\delta) = O(k_n^{1/2}/n) = o(1) \leq \epsilon$ for sufficiently large n by D5 and the non-uniqueness of the thresholds $\{c_n(\theta)\}$. Since ϵ is arbitrary this completes the proof.

Claim (b): $\{X_t^*(\theta) : \theta \in \Theta\}$ forms a VC class under D7, and D3 and D5 ensure by measurability $X_t^*(\theta)$ is geometrically α -mixing and L_{2+} -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 α -mixing triangular arrays $\{X_t^*(\theta) : 1 \leq t \leq n\}_{n \geq 1}$ stationary over t : $\{n^{-1/2} \sum_{t=1}^n X_t^*(\theta) : \theta \in \Theta\}$ converges to a Gaussian process $\{X(\theta) : \theta \in \Theta\}$ with uniformly bounded and uniformly continuous sample paths with respect to L_2 -norm. Therefore $E[(\sup_{\theta} \{|n^{-1/2} \sum_{t=1}^n X_t^*(\theta)\}|)^2] = O(1)$.

Finally, since $X_t^*(\theta, \tilde{\theta})$ is geometrically α -mixing under D3, use Ibragimov's bound and the Cauchy-Schwartz inequality to deduce

$$\begin{aligned} E\left[\left(\frac{1}{n^{1/2}} \sum_{t=1}^n X_t^*(\theta, \tilde{\theta})\right)^2\right] &\leq \sum_{q=1}^n E\left[\left(X_1^*(\theta) - X_1^*(\tilde{\theta})\right)\left(X_q^*(\theta) - X_q^*(\tilde{\theta})\right)\right] \\ &\leq K \times E\left(X_t^*(\theta) - X_t^*(\tilde{\theta})\right)^2 \\ &\leq \left(\frac{k_n}{n}\right)^{2\iota} \frac{n}{k_n} \left\{\left[\bar{P}_{i,t}^*(\theta)^{1/2} - \bar{P}_{i,t}^*(\tilde{\theta})^{1/2}\right]^2 - \left[\bar{P}_{i,t}^*(\theta) - \bar{P}_{i,t}^*(\tilde{\theta})\right]^2\right\} \\ &= \left(\frac{k_n}{n}\right)^{2\iota} \frac{n}{k_n} \bar{\mathcal{P}}_{i,t}^*(\theta, \tilde{\theta}) \end{aligned}$$

where $\bar{P}_{i,t}^*(\theta) := 1 - E[I_{i,t}^*(\theta)]$. Absolute continuity D1 ensures $\bar{\mathcal{P}}_{i,t}^*(\theta, \tilde{\theta})$ is Θ -a.e. differentiable (Royden 1968), $\bar{\mathcal{P}}_{i,t}^*(\theta, \theta) = 0$ by construction, and threshold property D5 implies by

a simple continuity argument $(n/k_n)^{1-2\iota} \bar{\mathcal{P}}_{i,t}^*(\theta, \tilde{\theta}) = o(1)$. Therefore $(n/k_n)^{1-2\iota} \bar{\mathcal{P}}_{i,t}^*(\theta, \tilde{\theta}) = O(1) \times \|\theta - \tilde{\theta}\|$ where $O(1)$ does not depend on θ . This completes the proof. ■

Proof of Lemma C.6. Define

$$\tilde{\sigma}_n^2 = n \times \left\| \Sigma_n^{-1/2} \right\|^{-2}.$$

The Lemma C.7 decomposition $\sum_{t=1}^n z_t^*(r, \theta_{n,0}) = \sum_{i=1}^{r_n} W_{n,i} + o_p(1)$ implies we need only prove $\sum_{i=1}^{r_n} W_{n,i} \xrightarrow{d} N(0, 1)$. Since $\{W_{n,i}, F_{n,i}\}$ forms a martingale difference array the required limit follows from Corollary 2.8 of McLeish (1974) provided $\sum_{i=1}^{r_n} W_{n,i}^2 \xrightarrow{p} 1$ and $\sum_{i=1}^{r_n} E[W_{n,i}^2 I(|W_{n,i}| > \epsilon)] \rightarrow 0 \forall \epsilon > 0$ (cf. McLeish 1974: eq. 2.4). The former is Lemma C.8. The latter Lindeberg condition follows from Lemma C.6.1, below, stationarity, $r_n h_n \sim n$, and covariance property (8.4):

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

Since (8.4) and D5 together imply $\tilde{\sigma}_n^2 / c_n^{2+\iota} \rightarrow \infty$ for sufficiently tiny $\iota > 0$ the right-hand-side is 0 for large n .

It is now a simple exercise to prove $\sum_{t=1}^n \{z_t^*(r, \theta_0) - z_t^*(r, \theta_{n,0})\} = o_p(1)$ by invoking equation expansions Lemma C.1.b,c. ■

LEMMA C.6.1 *Under D4, D5 and I1 $E[W_{n,i}^2 \tilde{\sigma}_n^2 I(W_{n,i}^2 \tilde{\sigma}_n^2 > M)] \leq K h_n^{\kappa+1} \int_M^{K c_n^{2+\iota}} u^{-\kappa/2} du$ for any $M > 0$, tiny $\iota > 0$ and some $\kappa \in (0, 2]$.*

Proof. Use subadditivity and the triangular inequality to deduce

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

say. Moment bound D4, Markov's inequality, the construction of z_t^* , and subadditivity imply for some $\kappa \in (0, 2]$

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

Consider $B_{n,i}$ and note trimming implies $|W_{n,i} - Z_{n,i}| \tilde{\sigma}_n \leq K h_n c_n$. By Markov's inequality

$$\begin{aligned} \int_{c_n^{2+\iota}}^\infty P(|W_{n,i} - Z_{n,i}| \tilde{\sigma}_n > (u/2)^{1/2}) du &= \int_{c_n^{2+\iota}}^{K h_n^2 c_n^2} P(|W_{n,i} - Z_{n,i}| \tilde{\sigma}_n > (u/2)^{1/2}) du \\ &\leq K \times E|W_{n,i} - Z_{n,i}| \tilde{\sigma}_n \int_{c_n^{2+\iota}}^{K h_n^2 c_n^2} u^{-1/2} du. \end{aligned}$$

Since $h_n = o(c_n^t)$ from (12) the right-hand-side is 0 for large n . ■

Proof of Lemma C.7. Define $z_t^* = z_t^*(r, \theta_{n,0})$ and decompose

$$\begin{aligned} \sum_{t=1}^n z_t^* &= \sum_{i=1}^{r_n} W_{n,i} + \sum_{i=1}^{r_n} (Z_{n,i} - E[Z_{n,i}|F_{n,i}]) + \sum_{i=1}^{r_n} E[Z_{n,i}|F_{n,i-1}] \\ &\quad + \sum_{i=1}^{r_n} \sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_t^* + \sum_{t=r_n h_n+1}^n z_t^* = \sum_{i=1}^{r_n} W_{n,i} + E_n. \end{aligned}$$

We need only show $E_n = o_p(1)$. $\{z_t^*, \mathfrak{F}_t\}$ forms an adapted martingale difference array under II, hence an L_2 -mixingale array with trivial constant $e_{n,t} = 0$ and coefficients $\psi_q = o(q^{-\lambda})$ for any size $\lambda > 0$ and integer $q \in \mathbb{N}$ since (e.g. McLeish 1975, Davidson 1994)

$$\|z_t^* - E[z_t^*|\mathfrak{F}_{-\infty}^{t+q}]\|_2 = \|z_t^* - z_t^*\|_2 = 0 = e_{n,t} \times o((q+1)^{-\lambda})$$

$$\|E[z_t^*] - E[z_t^*|\mathfrak{F}_{-\infty}^{t-q}]\|_2 = \left\| 0 - E\left[E[z_t^*|\mathfrak{F}_{-\infty}^{1-q}]|\mathfrak{F}_{-\infty}^{t-q}\right]\right\|_2 = 0 \leq e_{n,t} \times o(q^{-\lambda}).$$

Apply Lemmas C.7.1 and C.7.2 and the fact that the mixingale constants are trivial, to deduce $E[(\sum_{t=r_n h_n+1}^n z_t^*)^2] = o(1)$, $E[(\sum_{i=1}^{r_n} \sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_t^*)^2] = o(1)$, $E[(\sum_{i=1}^{r_n} (Z_{n,i} - E[Z_{n,i}|F_{n,i}]))^2] = o(1)$ and $E[(\sum_{i=1}^{r_n} E[Z_{n,i}|F_{n,i-1}])^2] = o(1)$. Therefore $E_n = o_p(1)$ by Chebyshev's inequality. ■

We require a generalization of McLeish's (1975) seminal maximal inequality to mixingale arrays, and to displacement sequences $\{q_n\}$ for use in the sequel. See Hall and Heyde (1980) and Andrews (1988) for related concepts, and see Hill (2010: Theorem 2.1) for a proof.

LEMMA C.7.1 *Let a uniformly L_2 -bounded triangular array $\{y_{n,t}\}$ and σ -fields $\{\mathfrak{F}_t\}$ satisfy $\|E[y_{n,t}] - E[y_{n,t}|\mathfrak{F}_{-\infty}^{t-q_n}]\|_2 \leq e_{n,t} \times o(q_n^{-1/2})$ and $\|y_{n,t} - E[y_{n,t}|\mathfrak{F}_{-\infty}^{t+q_n}]\|_2 \leq e_{n,t} \times o((q_n + 1)^{-1/2})$ for some positive deterministic array $\{e_{n,t}\}$ and some sequence of positive integers $\{q_n\}$. If $q_n = q \times g(n)$ for $q \in \mathbb{N}$ and some mapping $g: \mathbb{N} \rightarrow \mathbb{N}$, $g(n) \rightarrow 1$ or $g(n) \rightarrow \infty$ as $n \rightarrow \infty$, then $E[(\sum_{t=1}^n y_{n,t})^2] = O(\sum_{t=1}^n e_{n,t}^2)$.*

LEMMA C.7.2 *Define the index set $B_{n,t} = \{t : t \in \cup_{i=1}^{r_n} [(i-1)h_n + j_n + 1, \dots, ih_n]\}$. Under II $\{Z_{n,i} - E[Z_{n,i}|F_{n,i}], \mathfrak{F}_t\}_{t \in B_{n,t}}$ and $\{E[Z_{n,i}|F_{n,i-1}], \mathfrak{F}_t\}_{t \in B_{n,t}}$ form an L_2 -mixingale arrays with trivial constants and arbitrary size.*

Proof. We prove the claim for $\{E[Z_{n,i}|F_{n,i-1}], \mathfrak{F}_t\}_{t \in B_{n,t}}$, the second claim being similar. Since $\{z_t^*, \mathfrak{F}_t\}$ forms an adapted martingale difference array under II, by the construction $F_{n,i} := \sigma(\cup_{\tau \leq ih_n} \mathfrak{F}_\tau)$ it follows $E[Z_{n,i}|F_{n,i-1}] = \sum_{t=(i-1)h_n+1}^{ih_n} E[z_t^*|F_{n,i-1}] = 0$. Therefore $\|E[E[Z_{n,i}|F_{n,i-1}]|\mathfrak{F}_{-\infty}^{t-q}]\|_2 = 0$ and $\|E[Z_{n,i}|F_{n,i-1}] - E[E[Z_{n,i}|F_{n,i-1}]|\mathfrak{F}_{-\infty}^{t+q}]\|_2 = 0$. ■

Proof of Lemma C.8. Define

$$\tilde{\sigma}_n^2 := n \times \left\| \Sigma_n^{-1/2} \right\|^{-2}, \quad u_{n,i}^2 := \tilde{\sigma}_n^2 W_{n,i}^2, \quad \text{and} \quad K_n \sim K c_n^{2+t},$$

and a truncation function

$$\hat{u}_{n,i}^2(K) := u_{n,i}^2 I(u_{n,i}^2 \leq K) = \tilde{\sigma}_n^2 W_{n,i}^2 I(\tilde{\sigma}_n^2 W_{n,i}^2 \leq K) \quad \text{and} \quad \hat{u}_{n,i}^2 = \hat{u}_{n,i}^2(K_n).$$

By the triangular inequality $|\sum_{i=1}^{r_n} W_{n,i}^2 - 1| \leq \sum_{i=1}^5 E_{n,i}$ where

$$\begin{aligned} E_{n,1} &= \left| \frac{1}{\tilde{\sigma}_n^2} \sum_{i=1}^{r_n} (\tilde{u}_{n,i}^2 - u_{n,i}^2) \right| \quad \text{and} \quad E_{n,2} = \left| \frac{1}{\tilde{\sigma}_n^2} \sum_{i=1}^{r_n} (\tilde{u}_{n,i}^2 - E[\tilde{u}_{n,i}^2]) \right| \\ E_{n,3} &= \frac{1}{\tilde{\sigma}_n^2} \sum_{i=1}^{r_n} |E[\tilde{u}_{n,i}^2] - E[u_{n,i}^2]| \quad \text{and} \quad E_{n,4} = \left| \sum_{i=1}^{r_n} (E[W_{n,i}^2] - E[Z_{n,i}^2]) \right| \\ E_{n,5} &= \left| \sum_{i=1}^{r_n} E[Z_{n,i}^2] - 1 \right|. \end{aligned}$$

Lemmas C.8.1 and C.8.2 imply $E_{n,1}$, $E_{n,2}$ and $E_{n,3}$ are $o_p(1)$, and Lemma C.8.3 and Lyapunov's inequality imply $E_{n,4} \leq \|\sum_{i=1}^{r_n} (W_{n,i}^2 - Z_{n,i}^2)\|_2 = o(1)$. Finally, Lemma C.8.4 asserts $E_{n,5} = o_p(1)$. ■

LEMMA C.8.1 *Under D5 and I1 $\tilde{\sigma}_n^{-2} \sum_{i=1}^{r_n} E|\tilde{u}_{n,i}^2 - u_{n,i}^2| = o(1)$ and $\tilde{\sigma}_n^{-2} \sum_{i=1}^{r_n} (\tilde{u}_{n,i}^2 - u_{n,i}^2) = o_p(1)$.*

Proof. By construction $\tilde{\sigma}_n^2 W_{n,i}^2 \leq K h_n^2 c_n^2$. Now use $r_n h_n \sim n$, $h_n = O(c_n^\iota)$, $n/\tilde{\sigma}_n^2 = O(1)$, $K_n \sim K c_n^{2+\iota}$ for sufficiently large K and stationarity to deduce

$$\frac{1}{\tilde{\sigma}_n^2} \sum_{i=1}^{r_n} E|\tilde{u}_{n,i}^2 - u_{n,i}^2| = \frac{1}{\tilde{\sigma}_n^2} \sum_{i=1}^{r_n} E[u_{n,i}^2 I(u_{n,i}^2 > K_n)] \leq \frac{r_n}{\tilde{\sigma}_n^2} \int_{K c_n^{2+\iota}}^{K c_n^{2+\iota}} P(\tilde{\sigma}_n^2 W_{n,i}^2 > u) du = o(1).$$

The second claim now follows by the Markov and triangular inequalities. ■

LEMMA C.8.2 *Under D3, D5 and I1 $\tilde{\sigma}_n^{-2} \sum_{i=1}^{r_n} (\tilde{u}_{n,i}^2 - E[\tilde{u}_{n,i}^2]) = o_p(1)$.*

Proof. Geometric α -mixing D3 and measurability of $\tilde{u}_{n,i}^2$ imply $\{\tilde{u}_{n,i}^2/\tilde{\sigma}_n^2, F_{n,i}\}$ forms an L_2 -mixingale array with size $1/2$ and constants $K\tilde{\sigma}_n^{-1}$ (e.g. McLeish 1975: Theorem 2.1): for any sequence of positive integers $\{q_n\}$ and some $\rho \in (0, 1)$ and $\delta > 0$

$$\begin{aligned} \left(E \left(\frac{\tilde{u}_{n,i}^2}{\tilde{\sigma}_n^2} - E \left[\frac{\tilde{u}_{n,i}^2}{\tilde{\sigma}_n^2} \middle| F_{n,i+q_n} \right] \right)^2 \right)^{1/2} &\leq K \frac{1}{\tilde{\sigma}_n} (E[\tilde{u}_{n,i}^4])^{1/2} \rho^{q_n} \\ &\leq K \frac{1}{\tilde{\sigma}_n} \times \left\{ (E[\tilde{u}_{n,i}^4])^{1/2} \frac{1}{\tilde{\sigma}_n q_n^\delta} \right\} \times o((q_n + 1)^{-1/2}). \end{aligned}$$

Simply choose $q_n \rightarrow \infty$ as $n \rightarrow \infty$ and sufficiently large $\delta > 0$ to ensure

$$\left(E \left(\frac{\tilde{u}_{n,i}^2}{\tilde{\sigma}_n^2} - E \left[\frac{\tilde{u}_{n,i}^2}{\tilde{\sigma}_n^2} \middle| F_{n,i+q_n} \right] \right)^2 \right)^{1/2} \leq K \frac{1}{\tilde{\sigma}_n} \times o((q_n + 1)^{-1/2}).$$

An identical argument reveals

$$\left(E \left(E \left[\frac{\tilde{u}_{n,i}^2}{\tilde{\sigma}_n^2} \right] - E \left[\frac{\tilde{u}_{n,i}^2}{\tilde{\sigma}_n^2} \middle| F_{n,i-q_n} \right] \right)^2 \right)^{1/2} \leq K \frac{1}{\tilde{\sigma}_n} \times o(q_n^{-1/2}).$$

But this implies by Lemma C.7.1 $E(\tilde{\sigma}_n^{-2} \sum_{i=1}^{r_n} \{\tilde{u}_{n,i}^2 - E[\tilde{u}_{n,i}^2]\})^2 = O(\sum_{i=1}^{r_n} \tilde{\sigma}_n^{-2}) = O(r_n/\tilde{\sigma}_n^2) = o(1)$ since $r_n = o(n)$ and $n/\tilde{\sigma}_n^2 = O(1)$. ■

The last two follow directly the trivial mixingale property. See Hill and Renault (2010).

LEMMA C.8.3 Under D5 and I1 $\|\sum_{i=1}^{r_n}(W_{n,i}^2 - Z_{n,i}^2)\|_2 = o(1)$.

LEMMA C.8.4 Under D5 and I1 $\sum_{i=1}^{r_n} E[Z_{n,i}^2] \xrightarrow{p} 1$.

Proof of Lemma C.9. 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 (8.5), and the Lemma C.2 uniform approximation imply

$$\begin{aligned} & \sup_{\theta \in U_0(\delta_n)} \left\{ \frac{\left(\|V_n^{1/2}\| / \|J_n\| \right) \|\{\hat{m}_n^*(\theta) - \hat{m}_n^*(\theta_0)\} - \{E[m_t^*(\theta)] - E[m_t^*(\theta_0)]\}\|}{1 + \|V_n^{1/2}\| \times \|\theta - \theta_0\|} \right\} \\ & \leq \sup_{\theta \in U_0(\delta_n)} \left\{ \frac{\left(\|V_n^{1/2}\| / \|J_n\| \right) \|\{m_n^*(\theta) - m_n^*(\theta_0)\} - \{E[m_t^*(\theta)] - E[m_t^*(\theta_0)]\}\|}{1 + \|V_n^{1/2}\| \times \|\theta - \theta_0\|} \right\} + o_p(1). \end{aligned}$$

Now apply moment expansion Lemma B.2, equation expansion Lemma C.1.a, covariance bound (8.5) and scale bound M2, to deduce

$$\begin{aligned} & \sup_{\theta \in U_0(\delta_n)} \left\{ \frac{\left(\|V_n^{1/2}\| / \|J_n\| \right) \|\{m_n^*(\theta) - m_n^*(\theta_0)\} - \{E[m_t^*(\theta)] - E[m_t^*(\theta_0)]\}\|}{1 + \|V_n^{1/2}\| \times \|\theta - \theta_0\|} \right\} \\ & \leq \sup_{\theta \in U_0(\delta_n)} \left\{ \frac{\|J_n^*(\theta) - J_n\|}{\|J_n\|} \right\} + o_p(1) \\ & \leq \sup_{\theta \in U_0(\delta_n)} \left\{ \frac{\|J_n^*(\theta) - J_n^*\|}{\|J_n\|} \right\} + \left\{ \frac{\|J_n^* - J_n\|}{\|J_n\|} \right\} + o_p(1). \end{aligned}$$

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

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 F.C. Huang (1996). The Geometrical Ergodicity of Nonlinear Autoregressive Models, *Statistica Sinica*, 6, p. 943-956.
- [3] Andrews, D.W.K (1988). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables, *Econometric Theory* 4, 458-467.
- [4] Andrews, D.W.K (1992). Generic Uniform Convergence, *Econometric Theory* 8, 241-257.
- [5] 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.
- [6] Antoine, B. and E. Renault (2008b). Efficient GMM with Nearly-Weak Instruments, working paper, Dept. of Economics, University of North Carolina - Chapel Hill.
- [7] 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 .
- [8] Bassett, G.W. (1991). Equivariant, Monotonic, 50% Breakdown Estimators, *American Statistician* 45, 135-137.
- [9] Basrak, B., R.A. Davis, and T. Mikosch (2002). Regular Variation of GARCH Processes, *Stochastic Processes and their Applications* 12, 908-920.

- [10] 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..
- [11] 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.
- [12] 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.
- [13] Carrasco M., and X. Chen (2002). Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, *Econometric Theory* 18, 17-39.
- [14] Chan, N.H. and L.T. Tran (1989). On the First-Order Autoregressive Process with Infinite Variance, *Econometric Theory* 5, 354–362.
- [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, MIT.
- [17] Čížek, P. (2005). Least Trimmed Squares in Nonlinear Regression Under Dependence, *Journal of Statistical Planning and Inference* 136, 3967–3988.
- [18] Čížek, P. (2008). General Trimmed Estimation: Robust Approach to Nonlinear and Limited Dependent Variable Models, *Econometric Theory* 24 , 1500-1529.
- [19] Čížek, 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, *Stoch. Proc. Appl.* 117, 840-861.
- [23] Davidson, J. (1994). *Stochastic Limit Theory*, Oxford Univ. Press: Oxford.
- [24] Ding, Z. and C.W.J. Granger (1996). Modeling Volatility Persistence of Speculative Returns: A New Approach, *Journal of Econometrics* 73, 185-215.
- [25] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag: Frankfurt.
- [26] Engle, R. and V. Ng (1993). Measuring and Testing the Impact of News On Volatility. *Journal of Finance* 48, 1749-1778.
- [27] 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.
- [28] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, 2 ed., Vol. 2. Wiley: New York.
- [29] Finkenstadt, B., and H. Rootzén (2003). *Extreme Values in Finance, Telecommunications and the Environment*. Chapman and Hall: New York.
- [30] Francq, C. and J.M. Zakořan (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli* 10, 605–637.
- [31] Goldie, C.M., and R.L. Smith (1987). Slow Variation with Remainder: Theory and Applications, *Quarterly Journal of Mathematics* 38, 45-71.
- [32] 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.
- [33] 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.
- [34] Hall, A. (2005). *Generalized Method of Moments*. Oxford University Press: Oxford.
- [35] Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Applications*, Academic Press : New York.
- [36] Hannan, E.J. and M. Kanter (1977). Autoregressive Processes with Infinite Variance, Jour-

- nal of Applied Probability 14, 411-415.
- [37] Hansen, L.P. (1982). Large Sample Properties of Method of Moments Estimators, *Econometrica* 50, 1029-1054.
- [38] Hansen, L.P., J. Heaton and R. Jagannathan (1992). Econometric Evaluation of Intertemporal Asset Pricing Models Using Volatility Bounds, mimeo, University of Chicago.
- [39] Hansen, B. and S.W. Lee (1994). Asymptotic Theory for the Garch (1,1) Quasi-Maximum Likelihood Estimator, *Econometric Theory* 10, 29-52.
- [40] Hill, J.B. (2009). 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.
- [41] Hill, J.B. (2010a). On Tail Index Estimation for Dependent, Heterogeneous Data, *Econometric Theory* 26: in press.
- [42] Hill, J.B. (2010b). 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.
- [43] 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/gmmtm_tech_append.pdf.
- [44] 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, www.unc.edu/~jbhill/cv_test.pdf.
- [45] Honoré, B. (1992). Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects, *Econometrica* 60, 533-565.
- [46] Hsing, T. (1991). On Tail Index Estimation Using Dependent Data, *Annals of Statistics* 19, 1547-1569.
- [47] Huber, P.J. (1964). Robust Estimation of Location Parameter, *Annals of Mathematical Statistics* 35 73-101.
- [48] Ibragimov, I. A. and Y. V. Linnik (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhof: Berlin.
- [49] Jensen, S.T. and A. Rahbek (2004). Asymptotic Normality of the QML Estimator of ARCH in the Nonstationary Case, *Econometrica* 72, 641-646.
- [50] Kan., S., and A. Lewbel (2007). Weighted and Two-Stage Least Squares Estimations of Semiparametric Truncated Regression Models, *Econometric Theory* 23, 309-347.
- [51] Knight, K. (1987). Rate of Convergence of Centred Estimates of Autoregressive Parameters for Infinite Variance Autoregressions, *Journal of Time Series Analysis* 8, 51-60.
- [52] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag: New York.
- [53] Ling, S. (2005). Self-Weighted LAD Estimation for Infinite Variance Autoregressive Models, *Journal of the Royal Statistical Society Series B* 67, 381-393.
- [54] Ling, S. (2007). Self-Weighted and Local Quasi-Maximum Likelihood Estimators for ARMA-GARCH/IGARCH Models, *Journal of Econometrics* 150, 849-873.
- [55] Linton, O., J. Pan and H. Wang (2008). Estimation for a Non-stationary Semi-Strong GARCH(1,1) Model with Heavy-Tailed Errors, mimeo, London School of Economics.
- [56] 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.
- [57] 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.
- [58] Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business* 36, 394-419.

- [59] McFadden D.L. (1989). A Method of Simulated Moments for Estimation of Multinomial Discrete Response Models Without Numerical Integration, *Econometrica* 57, 995-1026.
- [60] McLeish, D.L. (1974). Dependent Central Limit Theorems and Invariance Principles, *Annals of Probability* 2, 620-628.
- [61] McLeish, D.L. (1975). Maximal Inequality and Dependent Strong Laws, *Annals of Probability* 3, 829-839.
- [62] Meitz M. and P. Saikkonen (2008). Stability of Nonlinear AR-GARCH Models, *Journal of Time Series Analysis* 29, 453-475.
- [63] 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 .
- [64] Naveau, P. (2003). Almost Sure Relative Stability of the Maximum of a Stationary Sequence, *Advances in Applied Probability* 35, 721-736.
- [65] Nelson, D. (1991). Conditional Heteroscedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.
- [67] Neykov, N.M. and P.N. Neytchev (1990). A Robust Alternative of the Maximum Likelihood Estimator, *Short Communications of Compstat, Dubrovnik 1990*, 99–100.
- [68] 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.
- [69] Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, 57, 1027-1057.
- [70] Peng, L. and Q. Yao (2003). Least Absolute Deviations Estimation for ARCH and GARCH Models, *Biometrika* 90, 967-975.
- [71] Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlang New York.
- [72] Powell, J.L. (1986). Symmetrically Trimmed Least Squares Estimation for Tobit Models, *Econometrica* 54, 1435-1460.
- [73] Renault, E. (1997). *Econométrie de la Finance: La Méthode des Moments Généralisés*, in *Encyclopédie des Marchés Financiers*, edited by Y. Simon, Economica, Paris, 1997, 330-407.
- [74] Resnick, S.I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag: New York.
- [75] Resnick, S.I. (1997). Heavy Tail Modeling and Traffic Data, *Annals of Statistics* 25, 1805-1849.
- [76] Ronchetti, E. and F. Trojani (2001). Robust Inference with GMM Estimators, *Journal of Econometrics* 101, 37-69.
- [77] 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.
- [78] Royden, H.L. (1968). *Real Analysis*. Macmillan: New York.
- [79] Ruppert, D. and J. Carroll (1980). Trimmed Least Squares Estimation in the Linear Model, *Journal of the American Statistical Association* 75, 828-838.
- [80] Smith, R.L. (1982). Uniform Rates of Convergence in Extreme-Value Theory, *Advances in Applied Probability* 14, 600-622.
- [81] Stock, J.H. and J.H. Wright (2000). GMM with Weak Identification, *Econometrica* 68, 1055-1096.
- [82] van der Vaart, A.W., and J.A. Wellner (1996). *Weak Convergence and Empirical Processes*, Springer-Verlang: New York.
- [83] 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, ARCH

LOCATION, $\epsilon_t \sim P_{1.5}, \theta = 1.0^a$									
	Mean	StDev	t-test ^b	KS ^c	λ^d	$b(\lambda)^e$	$\hat{b}(\lambda)^f$	b^g	\hat{b}
GMTTM ^{(1)h}	.992	.139	.02,.06,.10	.052	.23	-.129	-.131±.026	-.167	-.211±.065
GMTTM ⁽²⁾	.994	.126	.02,.05,.11	.051	.22	-.129	-.130±.024	-.167	-.203±.059
GMM	.939	.212	.01,.08,.13	.172	-	-	-	-	-
QML	1.01	.151	.02,.07,.15	.272	-	-	-	-	-
LOCATION, $\epsilon_t \sim P_{2.5}, \theta = 1.0$									
	Mean	StDev	t-test	KS	λ	$b(\lambda)$	\hat{b}	b	\hat{b}
GMTTM ⁽¹⁾	.994	.038	.01,.06,.11	.108	.49	.000	.047±.059	.000	.064±.044
GMTTM ⁽²⁾	.994	.033	.01,.05,.11	.101	.47	.000	.031±.051	.000	.058±.039
GMM	1.00	.075	.01,.03,.10	.076	-	-	-	-	-
QML	.999	.037	.01,.04,.10	.121	-	-	-	-	-
AR, $\epsilon_t \sim P_{1.5}, \theta = .90$									
	Mean	StDev	t-test	KS	λ	$b(\lambda)$	$\hat{b}(\lambda)$	b	\hat{b}
GMTTM ⁽¹⁾	.900	.007	.01,.04,.12	.075	.33	.112	.119±.015	.167	.194±.039
GMTTM ⁽²⁾	.900	.005	.01,.05,.12	.069	.32	.112	.116±.012	.167	.189±.033
GMM	.900	.012	.02,.07,.15	.187	-	-	-	-	-
QML	.889	.013	.02,.04,.08	.253	-	-	-	-	-
AR, $\epsilon_t \sim P_{2.5}, \theta = .90$									
	Mean	StDev	t-test	KS	λ	$b(\lambda)$	$\hat{b}(\lambda)$	b	\hat{b}
GMTTM ⁽¹⁾	.899	.011	.02,.05,.10	.052	.49	.000	.032±.044	.000	.042±.077
GMTTM ⁽²⁾	.901	.010	.01,.05,.10	.051	.48	.000	.027±.039	.000	.048±.081
GMM	.898	.013	.01,.06,.09	.086	-	-	-	-	-
QML	.899	.014	.02,.04,.09	.132	-	-	-	-	-
ARCH, $\epsilon_t \sim N_{0,1}, \theta = .60$									
	Mean	StDev	t-test	KS	λ	$b(\lambda)$	$\hat{b}(\lambda)$	b	\hat{b}
GMTTM ⁽¹⁾	.602	.152	.01,.06,.11	.065	.24	-.019	-.032±.022	-.024	-.048±.037
GMTTM ⁽²⁾	.601	.139	.01,.05,.11	.063	.22	-.019	-.030±.021	-.024	-.050±.037
GMM	.600	.176	.01,.06,.10	.162	-	-	-	-	-
QML	.599	.067	.01,.05,.11	.091	-	-	-	-	-

- a. $N_{0,1} = N(0,1)$; $P_\kappa =$ Pareto with index κ .
- b. t-test rejection frequency based on standard normal 1%, 5%, and 10% critical values. t-tests and KS tests are based on $\hat{T}_{j,k} = (\hat{\theta}_{j,k} - \theta_0) / \hat{s}_{n,k}^2$ where $\hat{s}_{n,k}^2$ is the simulation standard deviation of $\hat{\theta}_{j,k}$.
- c. Kolmogorov-Smirnov test p-value. Norm trimming and symmetric trimming uses one fractile $k_n = [n^\lambda]$. In the case of GMTTM, KS p-values are evaluated at that λ which minimizes KS. The 1%, 5%, 10% KS critical values are .136, .122, .107.
- d. The KS minimizing two-tailed GMTTME symmetric trimming parameter, $k_n = [n^\lambda]$.
- e. The slope parameter $b(\lambda)$ in the rate of convergence regression $\ln(\hat{s}_{n,k}^{-1}/n^{1/2}) \sim a + b(\lambda) \ln(n)$, where $\hat{s}_{n,k}^2$ is computed by GMTTM with the KS minimizing λ when $k_n = [n^\lambda]$ and $n = 1000$. See Sections 3 and 5.6 for verification of the true b , and Table 1 for κ . In all cases $b = 0$ if $\kappa \geq 2$. Location: $b = -(1 - \lambda)(1/\kappa - 1/2)$; AR: $b = (1 - \lambda)(1/\kappa - 1/2)$; ARCH: $b = -(1 - \lambda)(2/\kappa - 1/2)$.
- f. The least squares asymptotic 95% band when $k_n = [n^\lambda]$ and $n = 1000$.
- g. The same regression parameter and band, when $k_n = [\lambda \ln(n)]$ with the KS minimizing λ for $n = 1000$. Location: $b \approx 1/2 - 1/\kappa$; AR: $b \approx (1/\kappa - 1/2)$; ARCH: $b \approx 1/2 - 2/\kappa$.
- h. GMTTM⁽¹⁾ is the one-step GMTTME; GMTTM⁽²⁾ is the two-step GMTTME with a QMLE plug-in.

TABLE 4 : GARCH, TARCH, QARCH

GARCH, $\epsilon_t \sim N_{0,1}, \theta = .60$						TARCH, $\epsilon_t \sim N_{0,1}, \theta = .60$				
	Mean	StDev	Z-test	KS ^a	λ^b	Mean	StDev	Z-test	KS	λ
GMTTM ⁽¹⁾	.594	.174	.02,.05,.09	.067	.27	.625	.165	.01,.06,.11	.084	.29,.17
GMTTM ⁽²⁾	.595	.168	.02,.05,.10	.064	.26	.617	.163	.02,.06,.10	.083	.29,.16
GMM	.577	.195	.02,.04,.09	.177	-	.536	.147	.00,.05,.10	.269	-
QML	.597	.076	.02,.02,.05	.098	-	.595	.094	.01,.04,.10	.073	-
GARCH, $\epsilon_t \sim P_{2.5}, \theta = .60$						TARCH, $\epsilon_t \sim P_{2.5}, \theta = .50$				
	Mean	StDev	Z-test	KS	λ	Mean	StDev	Z-test	KS	λ
GMTTM ⁽¹⁾	.593	.231	.00,.06,.11	.096	.09	.622	.297	.00,.02,.08	.106	.68,.09
GMTTM ⁽²⁾	.594	.228	.01,.06,.11	.094	.09	.619	.289	.00,.02,.09	.104	.67,.09
GMM	.264	.346	.00,.00,.08	.448	-	.270	.202	.00,.00,.00	.532	-
QML	.605	.184	.00,.00,.00	.569	-	.516	.381	.00,.00,.00	.323	-
IGARCH, $\epsilon_t \sim N_{0,1}, \theta = .60$						TIARCH, $\epsilon_t \sim N_{0,1}, \theta = 1.0$				
	Mean	StDev	Z-test	KS	λ	Mean	StDev	Z-test	KS	λ
GMTTM ⁽¹⁾	.602	.177	.01,.06,.11	.083	.32	1.01	.242	.01,.06,.10	.066	.40,.25
GMTTM ⁽²⁾	.599	.171	.01,.05,.11	.078	.31	.998	.238	.01,.06,.10	.064	.41,.24
GMM	.628	.225	.01,.02,.05	.273	-	.839	.393	.00,.03,.06	.284	-
QML	.586	.198	.03,.14,.17	.262	-	1.00	.132	.02,.05,.09	.082	-
QARCH, $\epsilon_t \sim N_{0,1}, \theta = .80$						QIARCH, $\epsilon_t \sim N_{0,1}, \theta = 1.0$				
	Mean	StDev	Z-test	KS	λ	Mean	StDev	Z-test	KS	λ
GMTTM ⁽¹⁾	.807	.094	.02,.06,.09	.068	.34,.11	1.00	.088	.01,.05,.10	.051	.58,.39
GMTTM ⁽²⁾	.798	.092	.02,.06,.10	.066	.32,.10	1.01	.085	.01,.06,.10	.050	.57,.37
GMM	.673	.093	.00,.05,.08	.454	-	.721	.092	.00,.00,.03	.643	-
QML	.896	.659	.00,.00,.15	.389	-	.997	.638	.00,.00,.01	.287	-

a. KS critical values for test size 1%, 5%, 10% are .136, .122, .107.

b. The KS minimizing λ for the GMTTME for symmetric data generating processes (GARCH), or the KS minimizing pair $\{\lambda_1, \lambda_2\}$ for asymmetric processes (TARCH, QARCH).

c. GMTTM⁽¹⁾ is the one-step GMTTME; GMTTM⁽²⁾ is the two-step GMTTME with a QMLE plug-in.