Generalized Method of Moments with Tail Trimming

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

March. 2011

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, and it implies possibly heterogeneous convergence rates below, at or above $\sqrt{T}$. Further, it implies super-$\sqrt{T}$-consistency is achievable depending on regressor and error tail thickness and feedback, with a rate arbitrarily close to the largest possible rate amongst untrimmed minimum distance estimators for linear models with iid errors, and a faster rate than QML for heavy tailed GARCH. In the latter cases the optimal rate is achieved with the efficient GMM weight, and by using simple rules of thumb for choosing the number of trimmed equations. Simulation evidence shows the new estimator dominates GMM and QML when these estimators are not or have not been shown to be asymptotically normal, and for asymmetric GARCH models dominates a heavy tail robust weighted version of QML.

1. INTRODUCTION We develop a Generalized Method of Tail-Trimmed Moments estimator for possibly very heavy tailed nonlinear 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 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 (Hill and Shneyerov 2009, Chernozhukov 2010); and network traffic (Resnick 1997). Coupled with the necessity for

*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 Brendan Beare, Herman Bierens, Pavel Čižek, Robert de Jong, Eric Ghysels, Christian Gouriéroux, Alberto Holly, James Hamilton, Artem Prokhorov, Bas Werker and Hal White for helpful comments. We also thank conference participants at the Triangle Econometrics Conference 2008, the Joint Statistical Meetings 2008 and 2010, the First French Econometrics Conference 2009, the Brussels Conference on Latest Developments in Heavy-Tailed Distributions, and the Australian Meeting of the Econometric Society, and seminar participants at the University of Tilburg Dept. of Econometrics and O.R., CIREQ and Concordia University Dept. of Economics, Univeristy of Califormia at San Diego and Riverside Dept.'s of Economics, and Pennsylvania State University Dept. of Economics.

over-identifying restrictions in economic models, a robust GMM methodology will be useful to the analyst unwilling to impose ad hoc restrictions. See Hansen (1982), Renault (1997) and Hall (2005).

1.1 TAIL-TRIMMED ESTIMATING EQUATIONS

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

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

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

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

Consider a strong-ARCH(1) process $\{y_t\}$ as a simple example,

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

with least squares-type equations

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

The GMM estimator solves

$$\hat{\theta}_G = \underset{\theta \in \Theta}{\text{argmin}} \left\{ \left(\frac{1}{T} \sum_{t=1}^T m_t(\theta) \right)' \hat{\Upsilon}_T \left(\frac{1}{T} \sum_{t=1}^T m_t(\theta) \right) \right\}$$

for some positive definite matrix $\hat{\Upsilon}_T \in \mathbb{R}^{q \times q}$, and $T \geq 1$ is the sample size. Under standard regularity conditions $\hat{\theta}_G$ is asymptotically linear (e.g. Newey and McFadden 1994)

$$T^{1/2}\left(\hat{\theta} - \theta^0\right) = A_T \times \frac{1}{T^{1/2}} \sum_{t=1}^T m_t(\theta^0) + o_p(1) \text{ for some } A_T \in \mathbb{R}^{r \times q},$$

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

Finite variances $E[m_{i,t}^2(\theta^0)] < \infty$ expedites Gaussian asymptotics, but this requires ϵ_t and y_t to have finite 4^{th} and 8^{th} moments respectively. This rules out mildly heavy-tailed shocks, integrated random volatility (e.g. IGARCH), and much more. If over-identifying restrictions exist $q \geq 3$ with say $z_{3,t-1} = |y_{t-1}|^{2+\delta/2}$ and $\delta > 0$ then y_t must have a finite $(8+\delta)^{th}$ moment, a very tall order for financial time series (Embrechts et al 1997, Finkenstadt and Rootzén 2003). The problem carries over to a large class of random volatility models (cf. Meddahi and Renault 2004).

Models with heterogeneous estimating equations include the multifactor Capital Asset Pricing Model with high risk (e.g. oil futures), composite market returns (e.g. NYMEX), low risk asset returns (e.g. U.S. Treasury Bill), and factor premia (e.g. market capitalization and book-to-price ratio); VARX for causality modeling of financial and macroeconomic returns; and multivariate random volatility. See French and Fama (1996), Ding and Granger (1996), Mikosch and Stărică (2000), and Embrechts et al (2003).

QML equations by contrast allow for arbitrarily heavy tailed y_t as long as $E[\epsilon_t^4] < \infty$. But QML for GARCH with $E[y_t^4] = \infty$, like IGARCH, is well known for its poor finite sample properties (Lumsdaine 1995, 1996, Mikosch and Straumann 2006), and errors need not be thin tailed (Mikosch and Stărică 2000, Hall and Yao 2003, Ling 2007, Davis and Mikosch 2008, Linton et al 2010).

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

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

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

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

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

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

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

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

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

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

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

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

As long as $m_t(\theta^0)$ is integrable and satisfies a mixing condition, and standard smoothness conditions apply, then

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

for some sequence of positive definite matrices $\{V_T\}$. Negligible trimming ensures $\hat{m}_{T,t}^{*}(\theta)$ identifies θ^0 as $T \rightarrow \infty$ for symmetric or asymmetric processes, and the Gaussian limit holds for a host of heavy tailed time series. See Sections 2 and 4.

In Section 3 we present simple rules of thumb for selecting the rate $k_{j,i,T} \rightarrow \infty$ in order to optimize the rate $||V_T|| \rightarrow \infty$ for efficient GMTTM, where $||\cdot||$ is the spectral norm. Further, if $m_{i,t}(\theta^0)$ is symmetrically distributed and trimmed then any two-tailed $k_{i,T}$ ensure identification for each T. Otherwise, in Section 4 we present rules of thumb for relating $k_{1,i,T}$ and $k_{2,i,T}$ that optimize the rate of identification.

[1] Other criteria for trimming exist, including trimming according to a norm $||\cdot||$: $m_t^{(N)}(\theta) := ||m_t(\theta)||$. In this case $\hat{m}_{T,t}(\theta) = m_t(\theta)I(||m_t(\theta)|| \leq m_{(k_T)}^{(N)}(\theta))$ where $k_T \rightarrow \infty$ and $k_T/T \rightarrow 0$. Simulation work not presented here reveals the latter is massively dominated by component-wise trimming when $q > 1$, irrespective of distribution symmetry.

The components $\hat{\theta}_{i,T}$ may have heterogeneous rates, and linear combinations of subsets of $\hat{\theta}_T$ may help optimize the rates (Antoine and Renault 2010). In the sequel, where confusion is avoided, we simply say "rates of convergence" to refer variously to[2] $V_T^{1/2}$, $\text{diag}\{V_T^{1/2}\}$, or $||V_T||^{1/2}$.

Inference does not require knowledge of the rates of convergence since we self-normalize, and tail trimming equations with finite variance has no impact on the rate. In particular, $\hat{\theta}_T$ is $T^{1/2}$-consistent if all equations have a finite variance. In Section 3 we show sub-, exact- *or* super-$T^{1/2}$-convergence may arise in heavy tailed cases depending on the equation form (e.g. QML versus least squares-type equations for GARCH); 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 random volatility feedback (e.g. AR with ARCH shocks, or GARCH).

The feasible rates of convergence, however, is dampened in general precisely due to trimming, but the damage can be truly or nearly eradicated following simple rules of thumb for choosing $k_{j,i,T}$. GMTTM for linear regression models with infinite variance iid errors obtains the highest rate possible amongst M-estimators, identically $T^{1/\kappa}/L(T)$ for slowly varying $L(T) \to \infty$ where $\kappa < 2$ is the tail index (Davis et al 1992)[3]. In the case of AR-GARCH or GARCH feedback exists that depresses the rate to $o(T^{1/2})$. A simple rule of thumb for selecting $k_{j,i,T}$ optimizes the rate to $T^{1/2}/L(T)$ which dominates QML for GARCH with infinite kurtosis errors (Hall and Yao 2003). See Section 3 for convergence rate derivation for dynamic linear regression, IV, ARCH and AR-ARCH models. See also Antoine and Renault (2010) for broad GMM theory under variable coefficient estimator rates that are no greater than $T^{1/2}$.

In Section 5 we show consistency and asymptotic normality are primitive properties for linear-in-parameters models with innovations that have smooth distributions. Finally, perform a monte carlo study in Section 6 demonstrates 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 $k_{j,i,T}/T \to \lambda_{j,i} \in (0,1)$ 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 the review below. In this case without further information the equations $m_{i,t}(\theta^0)$ must be symmetrically distributed 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, under the lightest trimming case by extreme order sequences $k_{j,i,T} \to k_{j,i} \in \mathbb{N}$, too few equations are trimmed to ensure asymptotic normality.

1.2 EXTANT METHODS

The best extant theory of Minimum Distance Estimation for time series covers M-estimators, in particular QML and LAD for GARCH models and Least Trimmed Squares in the robust estimation literature. Francq and Zakoïan (2004) prove the QML estimator is asymptotically normal for strong-GARCH and ARMA-GARCH under $E(\epsilon_t^4) < \infty$. See, also, Hansen and Lee (1994), Lumsdaine (1996) and Jensen and Rahbek (2004) for results covering stationary and non-stationary cases. Hall and Yao (2003) characterize non-normal QML limit laws for linear GARCH models with possibly infinite variance errors.

[2] In t-tests of a single parameter restriction $V_{i,i,T}^{1/2}$ is the proper measure of *a rate* of convergence. The norm $||V_T||^{1/2}$ measures the maximum rate which is required for asymptotic arguments.

[3] A slowly varying function $L(T)$ satisfies $\lim_{T\to\infty} L(\lambda T)/L(T) = 1$ for all $\lambda > 0$, hence $L(T) = o(T^p)$ for all $p > 0$ (Resnick 1987). The natural log $\ln(T)$ is a classic example.

Davis et al (1992) characterize a general class of M-estimators with smooth criterion, and LAD, for stationary autoregressions with infinite variance iid errors.

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

Although robust estimation has a substantial history (Huber 1964, Stigler 1973, Jurecková and Sen 1996), only a few results concern fully nonlinear models with heavy tails. Most regression treatments focus on breakdown point analysis for thin tailed data with outliers under contamination (e.g. Rousseeuw 1985, Basset 1991, He et al 1996, Čižek 2005, 2008, 2009); most concern M-estimator frameworks (e.g. Čižek 2008 and the citations therein) or R-estimation (Koul and Saleh 1995, Andrews 2008); and when trimming, truncation or weighting are employed only non-tail data quantiles are considered. Almost all uses of tail trimming appear in the robust location and central limit theory literatures, and to our knowledge extremum estimation and tail-trimming have never been combined. See Horowitz (1998) for a tail-trimmed covariance matrix without supporting theory, and Berkes et al (2010) for a CUSUM test based on a tail-trimmed series.

There are few regression estimators that are asymptotically normal for heavy tailed data. Let $s_t(\theta) \geq 0$ denote criterion equations, for example $s_t(\theta) = |y_t - \theta' x_t|$ for LAD. Ling (2005, 2007) proposes Least Absolute Weighted Deviations [LAWD] and Quasi-Maximum Weighted Likelihood [QMWL] criteria $\sum_{t=1}^{T} w_t(c) s_t(\theta)$ where $w_t(c)$ is a symmetric smooth stochastic function of the data y_t based on some threshold c. Since $w_t(c)$ is not a function of θ, the threshold c is not with respect to the criterion $s_t(\theta)$. Linear autoregressive and GARCH models are separately covered allowing $E[\epsilon_t^2] < \infty$ and $E|y_t|^p < \infty$ for some $p > 0$ for consistency, but $E[\epsilon_t^4] < \infty$ must hold for asymptotic normality. The rate of convergence is $T^{1/2}$ for both AR and ARMA-GARCH models since restrictions imposed on $w_t(c)$ imply it operates like a smoothed fixed quantile trimming indicator. See also Pan et al (2007) for ARMA estimation.

Čižek (2005, 2008) improves the breakdown point of M-estimators by trimming the $k_T \sim \lambda T$ largest $s_t(\theta)$. Nonlinear models and models with limited dependent variables are covered, the errors are assumed to be iid with a finite variance, and asymptotic variance estimation is neglected so inference is not available. Khan and Lewbel (2007) use trimming to solve bias problems in semiparametric least squares estimation for linear truncated regression models of iid data with thin tails. The literature is too large for a fair review: see Čižek (2005, 2008) and his citations, and see Ruppert and Carroll (1980), Rousseeuw (1985), Basset (1991), Tableman (1994), Stromberg (1993), and Agulló et al (2008) for trimmed and truncated M-estimators.

The fundamental short-comings of trimming M-estimator criterion equations $s_t(\theta)$ by a fixed quantile of itself are super-$T^{1/2}$-convergence is impossible for stationary data under required moment restrictions; asymptotic normality cannot be achieved when regressors are heavy tailed; and in general asymmetric models and over-identifying restrictions are ignored. Adaptive weighting based on observable data as in Ling (2005, 2007) neglects criterion and normal equation information; and weights that act like fixed quantile trimming, and rank-orders can only deliver a $T^{1/2}$-consistent estimator. See Section 2.4 for direct comparisons of GMTTM with trimmed and weighted M-estimators.

A few results are couched in method of moments. Čižek (2009) trims a fixed quantile of $m_t(\theta)$ for thin-tailed cross-sections under data contamination, covering limited dependent and instrumental variables. Since the quantile is fixed identification must be assumed and an efficient criterion weight does not exist. Powell (1986) and Honoré (1992) construct least squares estimators couched in the method of trimmed moments for censored linear regressions models of iid data. Ronchetti and Trojani (2001a) symmetrically truncate $m_t(\theta)$ and propose a method of simulated moments to overcome bias in asymmetric models. See also Ronchetti and Trojani (2001b) and Dell'Aquila et al (2003). The error distribution must therefore be known and heavy-tailed cases are ignored.

Throughout $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the minimum and maximum eigenvalues of A. The L_p-norm is $||x||_p = (\sum_{i,j} E|x_{i,j}|^p)^{1/p}$, and the spectral (matrix) norm is $||A|| = (\lambda_{\max}(A'A))^{1/2}$. $(z)_+ := \max\{0, z\}$. K denotes a positive finite constant whose value may change from line to line; $\iota > 0$ is a tiny constant; N is a whole number. $\stackrel{p}{\rightarrow}$ and $\stackrel{d}{\rightarrow}$ denote probability and distribution convergence. I_d is a d-dimensional identity matrix and $A^{1/2}$ denotes the square-root matrix for positive definite A. $U^0(\delta_1, \delta_2) := \{\theta \in \Theta : \delta_1 \leq ||\theta - \theta^0|| \leq \delta_2\}$ for any $0 \leq \delta_1 < \delta_2$ and $U^0(\delta) = U^0(0, \delta)$. $\sup_\theta = \sup_{\theta \in \Theta}$ and $\inf_\theta = \inf_{\theta \in \Theta}$.

2. TAIL-TRIMMED GMM

In this section we develop a model-free theory of GMTTM based on primitive properties of $m_t(\theta)$. We treat specific models in Sections 3-5.

2.1 TAIL-TRIMMING

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

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

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

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

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

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

and for all $\theta \in \Theta$

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

Thus, $l_{i,T}(\theta)$ and $u_{i,T}(\theta)$ are identically the equation specific lower $k_{1,i,T}/T^{th} \rightarrow 0$ and upper $k_{2,i,T}/T^{th} \rightarrow 0$ tail quantiles (e.g. Leadbetter et al 1983: Theorem 1.7.13). We can always find threshold sequences $\{l_{i,T}(\theta), u_{i,T}(\theta)\}$ that satisfy (4) since we assume the marginal distributions of $m_{i,t}(\theta)$ are absolutely continuous. See Appendix A for all assumptions and related discussions, in particular condition D1.

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

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

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

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

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

This is easily guaranteed by Lebesgue's dominated convergence theorem for any fractiles $k_{j,i,T} \to \infty$ since $E[m_t(\theta^0)] = 0$ and tail trimming is negligible $k_{j,i,T} = o(T)$. It is interesting to note "eventual identification" runs contrary to weak and nearly weak identification where information *vanishes* at some rate (e.g. Stock and Wright 2000, Antoine and Renault 2009). Here, information *amasses* at some rate. If the DGP of $\{m_t(\theta^0)\}$ is symmetric then $E[m^*_{T,t}(\theta^0)] = 0$ for any thresholds $l_{i,T}(\theta^0) = u_{i,T}(\theta^0)$ and fractiles $k_{1,i,T}$ $= k_{2,i,T}$.

Since trimming is negligible and we assume $m_t(\theta)$ has absolutely continuos marginal distributions, the asymptotic covariance matrix of $\hat{\theta}_T$ is identical in form to the standard GMM estimator. We state the results here and develop the details in the appendices. Thus, we need standard covariance and Jacobian matrix constructions. The instantaneous and long run covariances are

$$\Sigma_T(\theta) := E\left[m^*_{T,t}(\theta)\, m^*_{T,t}(\theta)'\right] \quad \text{and} \quad \Sigma_T := \Sigma_T(\theta^0);$$

$$S_T(\theta) := \frac{1}{T}\sum_{s,t=1}^{T} E\left[m^*_{T,s}(\theta)\, m^*_{T,t}\theta'\right] \quad \text{and} \quad S_T := S_T(\theta^0);$$

population and sample Jacobia are

$$J_T(\theta) := \frac{\partial}{\partial\theta} E\left[m^*_{T,t}(\theta)\right] \in \mathbb{R}^{q\times r} \quad \text{and} \quad J_T = J_T(\theta^0)$$

$$J^*_{T,t}(\theta) := \left[\frac{\partial}{\partial\theta} m_{i,t}(\theta) \times I_{i,T,t}(\theta)\right]_{i=1}^{q} \quad \text{and} \quad J^*_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} J^*_{T,t}(\theta)$$

$$\hat{J}^*_{T,t}(\theta) := \left[\frac{\partial}{\partial\theta} m_{i,t}(\theta) \times \hat{I}_{i,T,t}(\theta)\right]_{i=1}^{q} \quad \text{and} \quad \hat{J}^*_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} \hat{J}^*_{T,t}(\theta);$$

and the GMTTM scale is

$$V_T(\theta) := T \times H_T(\theta)\left[J'_T(\theta)\,\Upsilon_T S_T(\theta)\,\Upsilon_T J_T(\theta)\right]^{-1} H_T(\theta) \quad \text{and} \quad V_T := V_T(\theta^0)$$

where

$$H_T(\theta) := J_T(\theta)'\Upsilon_T J_T(\theta) \in \mathbb{R}^{r\times r} \quad \text{and} \quad H_T := H_T(\theta^0).$$

We scale the covariance S_T with $1/T$ to help clarify the difference between $||V_T||^{1/2}$ and $T^{1/2}$. See Section 3.

2.2 MAIN RESULTS

The main results follow: $\hat{\theta}_T$ is consistent for θ^0 and asymptotically normal. See Appendix A for assumptions details concerning distribution properties D, identification properties I and matrix properties M.

THEOREM 2.1 *Under D1-D6, I1-I3 and M1-M2* $\hat{\theta}_T \xrightarrow{p} \theta^0$.

The rate $V_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(I_r)$ can similarly be shown under conditions D1-D6, I1-I3 and M1-M2, without proving asymptotic normality of $\hat{\theta}_T$, by using arguments in

Pakes and Pollard (1989) and Newey and McFadden (1994). We do not present the result here since asymptotic normality follows from the same set of assumptions[4].

THEOREM 2.2 *Under D1-D6, I1-I3 and M1-M2* $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{d} N(0, I_r)$.

Remark 1: An "optimal" GMTTM weight sequence $\{\Upsilon_T\}$ in the sense of asymptotic efficiency is $\{S_T^{-1}/||S_T^{-1}||\}$ for any fractiles $\{k_{j,i,T}\}$ due to the quadratic form $V_T = T \times H_T(J_T'\Upsilon_T S_T \Upsilon_T J_T)^{-1} H_T$ (Hansen 1982; Newey and MacFadden 1994: p. 2164). In this case

$$V_T = T\left(J_T' S_T^{-1} J_T\right).$$

Remark 2: Under β-mixing D3 the long-run covariance of the tail-trimmed equations S_T satisfies $||\Sigma_T^{-1} S_T|| \leq K$ and $||S_T^{-1}\Sigma_T|| \leq K$. See Lemma C.2 in Appendix C. The efficient weight therefore reduces to $\Upsilon_T = \Sigma_T^{-1}/||\Sigma_T^{-1}||$.

Remark 3: The existence of an efficient weight $\Upsilon_T = S_T^{-1}||S_T||$ is non-trivial since a symmetric variance form does not arise under *fixed* quantile trimming. In this case J_T has two components that enter V_T asymmetrically as $T \rightarrow \infty$, so an optimal weight does not exist (Čižek 2009). Under *tail* trimming, however, each $J_{i,j,T}$ also decomposes into two components, a mean Jacobian and a Jacobian of a mean:

$$J_{i,j,T} = E\left[\frac{\partial}{\partial\theta_j} m_{i,t}(\theta)|_{\theta^0} \times I_{i,T,t}(\theta^0)\right] + \frac{\partial}{\partial\theta_j} E\left[m_{i,t}(\theta^0) \times I_{i,T,t}(\theta)\right]|_{\theta^0}.$$

The latter is asymptotically dominated by the former due to negligibility $I_{i,T,t}(\theta) \rightarrow 1$ *a.s.* so

$$J_{i,j,T} = E\left[\frac{\partial}{\partial\theta_j} m_{i,t}(\theta)|_{\theta^0} \times I_{i,T,t}(\theta^0)\right] \times (1 + o(1)).$$

See Lemma C.4 in Appendix D.

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

LEMMA 2.3 *If* $\limsup_{T \geq 1} ||S_T|| < \infty$ *and* $\limsup_{T \geq 1} ||J_T]|| < \infty$ *then the rate of convergence of each* $\hat{\theta}_{i,T}$ *is* $T^{1/2}$.

2.3 COVARIANCE AND JACOBIAN MATRIX ESTIMATION

Unless specific dependence information is available for $m_t(\theta^0)$ we must estimate the long run covariance S_T non-parametrically. A now classic approach is a kernel HAC estimator

$$\hat{S}_T(\tilde{\theta}_T) = \frac{1}{T}\sum_{s,t=1}^{T} k\left((s-t)/\gamma_T\right) \hat{m}_{T,t}^*(\tilde{\theta}_T)\hat{m}_{T,t}^*(\tilde{\theta}_T)'$$

[4]A standard aproach to proving asymptotic normality of MDE's involves first demonstrating consistency and $T^{1/2}$-convergence (cf. Pakes and Pollard 1989, Newey and McFadden 1994). This pattern is not well suited to our problem since showing $V_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(I_r)$ follows from consistency is quite tedious in lieu of non-differentiability and the nonlinearity imbedded order statistics $\{m_{i,(k_{1,i,T})}^{(-)}(\theta), m_{i,(k_{2,i,T})}^{(+)}(\theta)\}$. Our proof of Theorem 2.2 relies on a version of the δ-method based on an asymptotic expansion that exploits non-degeneracy properties under tail-trimming. See Appendix D for background theory.

for some consistent plug-in $\tilde{\theta}_T$, where $k(\cdot)$ is a kernel function and $\gamma_T \to \infty$ is bandwidth (cf. Newey and West 1987). We restrict $k(\cdot)$ and $\gamma_T = o(T)$ in condition K1 of Appendix A based on theory developed in Davidson and de Jong (2000). Kernels covered include Bartlett, Parzen, Quadratic Spectral, Tukey-Hanning and others.

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

LEMMA 2.4 *Under D1-D6, I1-I2, K1 and M1* $||S_T^{-1}\hat{S}_T(\tilde{\theta}_T) - I_q|| = o_p(1)$ *for any* $||\tilde{\theta}_T - \theta^0|| = O_p(T^{-1/2}||S_T||^{1/2}||J_T||^{-1})$.

Remark 1: If the equations are sufficiently orthogonal that $S_T = \Sigma_T(1 + o(1))$ then the appropriate estimator of S_T is $\hat{\Sigma}_T(\tilde{\theta}_T) = 1/T\sum_{t=1}^T \hat{m}^*_{T,t}(\tilde{\theta}_T)\hat{m}^*_{T,t}(\tilde{\theta}_T)'$. Since kernel class K1 includes $k((s-t)/\gamma_T) = 0\ \forall s \neq t$ and $k((s-t)/\gamma_T) = 1\ \forall s \neq t$ we have $||S_T^{-1}\hat{\Sigma}_T(\tilde{\theta}_T) - I_q|| = o_p(1)$ by Lemma 2.4.

Remark 2: In principle $\tilde{\theta}_T$ may be a first-stage GMTTM estimator based on some weight Υ_T and fractile policy $\{\tilde{k}_{j,i,T}\}$. Notice irrespective of the choice of plug-in $\tilde{\theta}_T$, we must have the Jacobian J_T and covariance S_T reflect the policy $\{k_{j,i,T}\}$ of the *final* GMTTM estimator based on its weight and fractiles.

Tail trimming implies the Jacobian J_T is proportional to $E[J^*_{T,t}]$, cf. Lemma C.4 in Appendix D. Due to its simple form consistency $\hat{J}^*_T(\tilde{\theta}_T) = J_T \times (1 + o_p(1))$ follows for any consistent $\tilde{\theta}_T$.

LEMMA 2.5 *Under D1-D6, I2 and M1-M2* $J^*_T(\tilde{\theta}_T) = J_T \times (1 + o_p(1))$ *and* $\hat{J}^*_T(\tilde{\theta}_T) = J_T \times (1 + o_p(1))$ *for any* $||\tilde{\theta}_T - \theta^0|| \xrightarrow{p} 0$.

The efficient-weighted covariance matrix V_T^{-1} is estimated by

$$\hat{V}_T^{-1}(\theta) = T \times \left\{\hat{J}^*_T(\theta)'\hat{S}_T^{-1}(\theta)\,\hat{J}^*_T(\theta)\right\}^{-1}.$$

Since trivially the efficient estimator satisfies $||\hat{\theta}_T - \theta^0||_2 = O(||V_T||^{-1/2}) = O(T^{-1/2}||J_T||^{-1}||S_T||^{1/2})$ by Theorem 2.2, the scale estimator $\hat{V}_T^{-1}(\hat{\theta}_T)$ is consistent by Lemmas 2.4 and 2.5.

THEOREM 2.6 *Under D1-D6, I1, I2 and M1-M2* $\hat{V}_T(\hat{\theta}_T) = V_T \times (1 + o_p(1))$.

Remark: It is important to note we only show $||\hat{V}_T(\hat{\theta}_T) - V_T|| = o_p(||V_T||)$, and not $||\hat{V}_T(\hat{\theta}_T) - V_T|| \xrightarrow{p} 0$ since that requires $\hat{\theta}_T \xrightarrow{p} \theta^0$ far faster than $O_p(||V_T||^{-1/2})$. This is irrelevant, however, since any estimator $\hat{V}_T(\hat{\theta}_T) = V_T \times (1 + o_p(1)) = V_T + o_p(||V_T||)$ suffices for inference. Consider a t-ratio $\hat{V}^{1/2}_{i,i,T}(\hat{\theta}_T)\hat{\theta}_{i,T}$: if $\theta^0_i = 0$ then $\hat{V}^{1/2}_{i,i,T}(\hat{\theta}_T)\hat{\theta}_{i,T} = V^{-1/2}_{i,i,T}\hat{\theta}_{i,T} \times (1 + o_p(1)) \xrightarrow{d} N(0,1)$ by Theorems 2.2 and 2.6, and Cramér's Theorem.

2.4 ROBUST M-ESTIMATORS

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

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

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

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

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

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

$$T^{1/2}\left(\tilde{\theta}_T - \theta^0\right) = \tilde{J}(\lambda)^{-1} \frac{1}{T^{1/2}} \sum \epsilon_t x_t I\left(|\epsilon_t| \leq F_{|\epsilon|}^{-1}(\lambda)\right) + o_p(1) \xrightarrow{d} N\left(0, \tilde{V}^{-1}(\lambda)\right),$$

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

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

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

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

Now exploit independence to deduce

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

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

Adaptive M-Estimation: Ling's (2005, 2007) symmetrically weighed LAD and QML criteria work like smoothed fixed quantile trimmed criteria. Although heavy-tails are allowed the imposed weight structure rules out super-$T^{1/2}$-convergence. Further, theory is only delivered for symmetric DGP's, only fixed quantiles of the data y_t are considered for the weight function, and by construction over-identification is not allowed. The theory developed cannot be immediately extended to the problem of tail-trimming parametric equations $m_t(\theta)$.

3. FRACTILE SELECTION AND GMTTM CONVERGENCE RATES

We now characterize convergence rates for heavy tailed models, which provides rules of thumb for selecting the fractile rate $k_{j,i,T} \to \infty$. We treat asymmetric fractile selection separately in Section 4.

We specifically treat regression models with martingale difference equations $m_t(\theta^0)$. Let $h \geq 1$. In the spirit of Section 4 it can be shown $E[m_{i,T,s}^*(\theta^0)m_{j,T,t+h}^*(\theta^0)] = 0$ for $\{m_{i,T,s}^*(\theta^0), m_{j,T,t+h}^*(\theta^0)\}$ with symmetric joint distributions, and any two-tailed $\{k_{i,T}\}$ hence $S_T = \Sigma_T$. Similarly, $E[m_{i,T,s}^*(\theta^0)m_{j,T,t+h}^*(\theta^0)] \to 0$ sufficiently fast for infinitely many policies $\{k_{1,i,T}, k_{2,i,T}\}$ such that $S_T = \Sigma_T(1 + o(1))$. We therefore assume $S_T = \Sigma_T$ for clarity.

By positive definiteness and the Cauchy-Schwarz inequality we can define diagonal matrices $\Gamma_T \in \mathbb{R}^{q \times q}$ with elements

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

Now write

$$V_T = T \times \left(\Gamma_T^{-1}J_T\right)' \times \Sigma^{-1} \times \left(\Gamma_T^{-1}J_T\right) \times (1 + o(1)) \quad \text{and} \quad \Sigma^{-1} = [\sigma^{i,j}]_{i,j=1}^q.$$

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

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

Textbook intuition explains T_{θ_i}. Holding everything else constant, if $\Gamma_{i,i,T} = (E[(m_{i,T,t}^*\left(\theta^0\right))^2])^{1/2}$ $\to \infty$ due to heavy-tailed errors then $T_{\theta_i} \to \infty$ slowly: sharp estimates are more difficult to obtain from models with disproportionately dispersive errors (an "outlier" effect). Conversely, using Jacobian relation Lemma C.4 in Appendix C, if $|J_{l,i,T}| = |E[(\partial/\partial\theta_i)m_{l,t}(\theta)|_{\theta^0}I_{l,T,t}(\theta^0)]|$ $\times$ $(1 + o(1)) \to \infty$ due to heavy tailed regressors then $T_{\theta_i} \to \infty$ quickly: sharpness improves with regressor dispersion and association (a "leverage" effect.). If both error and regressor are heavy-tailed and exhibit feedback then Γ_T may overwhelm J_T. In this section we inspect the gamut of such cases.

In order to characterize Γ_T and J_T we consider dynamic linear regression, AR, ARCH and AR-ARCH models with least-squares and in some cases QML equations. Since we only care about convergence rates, initially all equations are symmetrically trimmed with two-tailed thresholds $c_{i,T}$ and the same fractiles $k_{j,i,T} = k_T$ to simplify notation. Hence

$$\frac{T}{k_T}P\left(\left|m_{i,t}\left(\theta^0\right)\right| > c_{i,T}\right) = 1.$$

This allows for a cleaner discussion of rules for selecting the rate $k_T \to \infty$.

Throughout $\{\epsilon_t\}$ is iid L_p-bounded, $p > 0$, with an absolutely continuous distribution on $\mathbb{R}$-*a.e.*, symmetric about 0. Assume covariance and Jacobian non-degeneracy properties D 5 and M2 throughout. See Appendix A. Let $L(T)$ be a slowly varying function,

$$L\left(T\right) \to \infty \text{ and } L\left(T\right) \leq T,$$

whose value or rate may change with the context.

3.1 DYNAMIC REGRESSION WITH IID ERRORS

We first examine a stationary dynamic linear regression with an intercept

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

where ϵ_t and x_t are mutually independent. Stochastic $x_{i,t}$ are measurable with $\mathbb{R}$-*a.e.* continuous, stationary, symmetric distributions. Independence rules out random volatility errors. See Sections 3.2-3.4 for this case.

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

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

where $\kappa_z = \infty$ is possible (e.g. uniform, normal, or bounded support). Identification requires integrability of $\epsilon_t x_{i,t}$ hence $\kappa_z > 1$. If any $z_t \in \{\epsilon_t, x_{i,t}\}$ has an infinite variance $\kappa_z \in (1,2]$ then assume the distribution tail decays as a power law:

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

By Karamata's Theorem z_t satisfies the following property for trimmed higher moments (e.g. Resnick 1987: Theorem 0.6; Problem 4.2.8; cf. Feller 1971: §IX.8):

$$\kappa_z < 2 : E\left[z_t^2 I(|z_t| \le c)\right] \sim Kc^2 P(|z_t| > c) \text{ as } c \to \infty \tag{TM}$$

$$\kappa_z = 2 : E\left[z_t^2 I(|z_t| \le c)\right] \sim L(c) \text{ as } c \to \infty \text{ where } L(c) \to \infty \text{ is slowly varying.}$$

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

Define

$$a^*_{\epsilon,(i)} := \min_{j \notin \{1,i\}} \{1/\kappa_{\epsilon,j} + (1 - 1/\kappa_j)\kappa_i/\kappa_{\epsilon,i}\}, \quad A_T := (T/k_T)^{1/2 + 1/\kappa_{\epsilon,i} + \kappa_i/\kappa_{\epsilon,i} - 2a^*_{\epsilon,(i)}}$$

$$B_T := \max_{j \notin \{1,i\}:\kappa_j \le 2} \left\{(T/k_T)^{2/\kappa_j - 1}\right\}, \quad C_T := \max_{j \notin \{1,i\}:\kappa_j \le 2} \left\{(T/k_T)^{2 - 2/\kappa_j - 2/\kappa_i + 2\kappa_{\epsilon,i}^{-1}(\kappa_i/\kappa_j + 1 - \kappa_i)}\right\}.$$

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

LEMMA 3.1 (ARX with IID Error)

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

$$T_{\theta_i} \sim KT^{1/2}(T/k_T)^{1/2 - \kappa_i/\kappa_{\epsilon,i} + 1/\kappa_{\epsilon,i}}[K + O(A_T)]^{1/2}, \; i = 2, ..., r.$$

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

$$T_{\theta_1} = KT^{1/2} \times K(k_T/T)^{1/\kappa_\epsilon - 1/2}(1 + O(1)).$$

b. Let $\kappa_\epsilon > 2$. *If* $\kappa_i > 2$ *then* $T_{\theta_i} \sim T^{1/2} \times [K + O(B_T)]^{1/2}$, *and if* $\kappa_i < 2$ *then* $T_{\theta_i} \sim KT^{1/2}(T/k_T)^{1/\kappa_i - 1/2} \times [K + O(C_T)]^{1/2}$.

Remark 1: Notice $T_{\theta_1} = o(T^{1/2})$ when $\kappa_\epsilon < 2$. Irrespective of the other regressors, as long as the error ϵ_t is heavy tailed the intercept rate is sub-$T^{1/2}$-consistent due an outlier effect: to $k_T/T \to 0$ under tail trimming allows $||\Sigma_T|| \to \infty$.

Remark 2: If the error has a finite variance then T_{θ_i} is governed entirely by the regressor tails, hence $\hat{\theta}_{i,T}$ is super-$T^{1/2}$-consistent when $x_{i,t}$ has an infinite variance due to a leverage effect: under tail-trimming $||J_T|| \to \infty$. If ϵ_t and $x_{i,t}$ have finite variances but some other regressor $x_{j,t}$ is heavy-tailed, then all we can say is $\hat{\theta}_{i,T}$ is at least $T^{1/2}$-consistent since cross-Jacobia complexity precludes sharper results.

Remark 3: We ignore the hairline infinite variance cases $\kappa_\epsilon = 2$ or $\kappa_i = 2$ to retain brevity. We inspect the case in specific models below which can be easily verified from (TM) and the arguments used to prove Lemma 3.1.

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

EXAMPLE 1 (Slope Rate Lower Bound): Assume $\max\{\kappa_\epsilon, \kappa_2, ..., \kappa_r\} < 2$, and note $\liminf_{T\to\infty} T_{\theta_i}/[T^{1/2}\,(T/k_T)^{1/2-\kappa_i/\kappa_{\epsilon,i}+1/\kappa_{\epsilon,i}}] \geq K$ 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 $\hat{\theta}_{i,T}$ is super-$T^{1/2}$-consistent. There are two cases.

Case 1 ($\kappa_i \leq \kappa_\epsilon$): In this case the leverage effect dominates, so $T_{\theta_i} \geq KT^{1/2}\,(T/k_T)^{1/\kappa_i-1/2}$ only reflects the tails of $x_{i,t}$. Simply choose a light fractile $k_T = [\lambda L(T)]$ for $\lambda > 0$ and slowly varying $L(T) \to \infty$ to obtain

$$T_{\theta_i} \geq KT^{1/\kappa_i}/L(T).$$

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

Case 2 ($\kappa_i > \kappa_\epsilon$): If ϵ_t is more heavy-tailed than $x_{i,t}$ then super-$T^{1/2}$-convergence still arises as long as $x_{i,t}$ has an infinite variance $\kappa_i < 2$ and the dispersion of ϵ_t is not too great, $\kappa_\epsilon > 2(\kappa_i - 1)$. If $\kappa_i = 1.5$, for example, then any $\kappa_\epsilon \geq 1$ allows a dominate leverage affect. The converse $\kappa_\epsilon \leq 2(\kappa_i - 1)$ naturally arises in random volatility models with AR-in-squares representations. See Sections 3.2-3.4.

EXAMPLE 2 (Independent Regressors): If stochastic $x_{i,t}$ are independent random variables then because they have finite means $J_{i,j,T} \sim K\ \forall i \neq j$. The leverage effect vanishes hence $T_{\theta_i} = o(T^{1/2})$ if the error has an infinite variance.

EXAMPLE 3 (Tail Homogeneity): If $\kappa_\epsilon = \kappa_i = \kappa < 2$ for all i, then $\kappa_{\epsilon,i} = \kappa$ and $a^*_{\epsilon,(i)} = 1$, hence $A_T = (k_T/T)^{1/2} = o(1)$ and the rate reduces to

$$T_{\theta_i} \sim KT^{1/2}\,(T/k_T)^{1/\kappa-1/2}.$$

Similarly, using property (TM) when $\kappa = 2$ it is easy to verify $T_{\theta_i} \sim KT^{1/2}L(T)$. Super-$T^{1/2}$-convergence arises *if and only if* variance is infinite $\kappa \leq 2$. In this case the leverage affect dominates the outlier effect.

The following provide deeper intuition as to *why* super-$T^{1/2}$-convergence may or may not arise.

EXAMPLE 4 (Location): Consider estimating location

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

with one equation $m_t(\theta) = y_t - \theta$. The Jacobian is $J_T = -1 + o(1)$, and the covariance $\Gamma_T = \Sigma_T^{1/2} = T^{1/2}(T/k_T)^{1/\kappa-1/2}$ when $\kappa < 2$ or $\Gamma_T = T^{1/2}L(T)$ when $\kappa = 2$. Therefore $T_\theta = KT^{1/2}(k_T/T)^{1/\kappa-1/2} = o(T^{1/2})$ if $\kappa < 2$ and $T_\theta = KT^{1/2}/L(T) = o(T^{1/2})$ if $\kappa = 2$, so GMTTM is sub-$T^{1/2}$-consistent when $\kappa \leq 2$.

Remark 1: The reason for the sluggish rate is essentially the same as in Example 2: bounded regressors cannot provide explanatory leverage against a heavy tailed shock. In the above simple model there is only an outlier effect.

Remark 2: It is straightforward to show over identifying restrictions involving lags of y_t have no impact on the sub-$T^{1/2}$ rate since the added regressors y_{t-i} are independent.

Remark 3: Since there is only an outlier effect when $\kappa \leq 2$, maximal trimming optimizes the rate. Set $k_T = [T/L(T)]$ to achieve $T_\theta = KT^{1/2}/L(T)$.

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

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

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

Remark 1: The AR(1) case is particularly revealing: for slowly varying $L_0\left(T\right), L\left(T\right) \rightarrow \infty$

$$\kappa < 2 : T_\theta = KT^{1/2}\frac{|J_T|}{\Gamma_T} \sim KT^{1/2}\frac{E\left[y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,T}\right)\right]}{\left(E\left[\epsilon_t^2 y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,T}\right)\right]\right)^{1/2}} \sim KT^{1/2}\frac{\left(T/k_T\right)^{2/\kappa - 1}}{\left[\left(T/k_T\right)^{2/\kappa-1}\right]^{1/2}}$$

$$\kappa = 2 : T_\theta \sim KT^{1/2}\frac{E\left[y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,T}\right)\right]}{\left(E\left[\epsilon_t^2 y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,T}\right)\right]\right)^{1/2}} \sim KT^{1/2}\frac{L_0\left(T\right)}{\left[L_0\left(T\right)\right]^{1/2}} = T^{1/2}L\left(T\right).$$

The case $\kappa = 2$ can be verified from (TM) and the proof of Lemma 3.1. Independent errors mean y_{t-1}^2 and $\epsilon_t^2 y_{t-1}^2$ have the same tail shape up to a constant scale, so $L_0(T)$ are the same in the numerator and denominator (Cline 1986), hence $T_\theta \sim T^{1/2}L\left(T\right)$ for some $L\left(T\right)$.

Consider the case $\kappa < 2$. The numerator Jacobian

$$E\left[y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,T}\right)\right] \sim K\left(T/k_T\right)^{2/\kappa - 1}$$

works like a tail-trimmed *variance*. If sequences $\{c_{y,T}, k_T\}$ satisfy $(T/k_T)P(|y_t| > c_{y,T}) \rightarrow \infty$ then arguments in the proof of Lemma 3.1 reveal $E[y_{t-1}^2 I(|y_{t-1}| \leq c_{y,T})] \sim c_{y,T}^2 P(|y_{t-1}| > c_{y,T}) = K(T/k_T)^{2/\kappa-1}$. 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 $\kappa < 2$, hence $\epsilon_t y_{t-1}$ has index κ (Cline 1986). By comparison the denominator

$$\left(E\left[\epsilon_t^2 y_{t-1}^2 I\left(|\epsilon_t y_{t-1}| \leq c_{1,T}\right)\right]\right)^{1/2} = \left(T/k_T\right)^{1/\kappa - 1/2}$$

is a tail-trimmed *standard deviation* of an object with the *same* tail index κ. Therefore $T_\theta \sim KT^{1/2}(T/k_T)^{1/\kappa-1/2}$ dominates $T^{1/2}$ 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 $\Gamma_T \rightarrow \infty$ so fast that super-$T^{1/2}$-convergence cannot arise. See Sections 3.2 and 3.3.

Remark 2: Heavy-tailed regressors without error-regressor feedback promotes a pure leverage effect, so minimal trimming is optimal. If k_T is regularly varying $k_T = [T^\lambda]$ for $\lambda \in (0, 1)$ then $T_{\theta_i} \sim KT^{1/\kappa - \lambda(1/\kappa - 1/2)} > KT^{1/\kappa - \iota}$ for any tiny $\lambda, \iota > 0$. Conversely if slowly varying $k_T = [\lambda \ln(T)]$ then $T_{\theta_i} = KT^{1/\kappa}/L(T)$, arbitrarily close the maximum rate $T^{1/\kappa}$, cf. Davis et al (1992).

EXAMPLE 6 (Instrumental Variables): It is tempting to use heavy-tailed instruments z_t to induce super-$T^{1/2}$-convergence. Consider a simple scalar model for reference

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

Assume the instrument $z_t \in \mathbb{R}$ has tail (7) and index $\kappa_z \in (1,2)$, and is valid: it is independent of ϵ_t and $\inf_{T \geq N} |E[x_t z_t I(|\epsilon_t z_t| \leq c_T)]| > 0$ for large N. For example, we might use $z_t = x_t^2$ if x_t has a finite variance and infinite kurtosis. Since $m_t(\theta^0) = \epsilon_t z_t$ has tail index κ_z (Cline 1986), the Cauchy-Schwarz inequality and arguments in the proof of Lemma 3.1 reveal

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

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

3.2 ARCH WITH LEAST SQUARES EQUATIONS

Consider a Strong-ARCH(p) model with least squares-type equations

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

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

Assume the Lyapunov exponent associated with the stochastic recurrence equation form is negative[5]. Then $\{y_t, h_t\}$ are stationary with tail (7) and index $\kappa_y > 0$ as long at least one as one $\beta_i^0 > 0$ (e.g. Basrak et al 2002: Theorem 3.1). If $\beta_i^0 > 0$ then $m_t(\theta^0) = (\epsilon_t^2 - 1) h_t^2 x_t$ have tails (7) with indices $\{\kappa_y/2, \kappa_y/4, ..., \kappa_y/4\}$[6].

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

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

If $\beta^0 = 0$ or $\beta^0 > 0$ and $\kappa_y > 8$ then $m_{i,t}(\theta^0)$ have finite variances so all $T_\alpha, T_{\beta_i} \sim KT^{1/2}$. The harsh moment requirement is relaxed with QML equations due to scaling. See Section 3.4.

Since $(\partial/\partial\theta) m_t(\theta) = -x_t x_t'$ have indices in $\{\kappa_y/4, \kappa_y/2, \infty\}$ and $\kappa_y > 4$ each component is integrable, thus $J_T \sim J$. The standard deviations $\Gamma_{i,i,T}$ are almost as simple to compute since $\kappa_y \in (4,8)$ implies the intercept term $\Gamma_{1,1,T} \leq (E[(\epsilon_t^2 - 1)^2 h_t^4])^{1/2} = K$, and the remaining $\Gamma_{i,i,T} = (E[m_{i,t}^2(\theta^0) I_{i,T,t}(\theta^0)])^{1/2} \sim K c_{i,T} (k_T/T)^{1/2} = K(T/k_T)^{4/\kappa_y - 1/2}$ by trimmed moment property (TM) and $c_{i,T} = K(T/k_T)^{1/(\kappa_y/4)}$ for tail (7). Therefore $J_{1,1,T}/\Gamma_{1,1,T} \sim K$ and all other $J_{i,i,T}/\Gamma_{i,i,T} \sim K(T/k_T)^{-(4/\kappa_y - 1/2)} = o(1)$. Now use (6) to deduce $T_\alpha, T_{\beta_i} \sim KT^{1/2}$. If $\kappa_y = 8$ similar steps under (TM) reveal $T_\alpha, T_{\beta_i} \sim KT^{1/2}$. This proves the next claim.

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

Remark 1: Strong-ARCH are AR in squares $y_t^2 = \theta' x_t + v_t$, where $E[v_t|\Im_{t-1}] = 0$. But stationary AR equations all have the same tail index κ when ϵ_t is iid with tail (7).

---

[5] The Lyapunov condition can be replaced with $\sum_{i=1}^{p} \beta_i^0 < 1$, or $\sum_{i=1}^{p} \beta_i^0 = 1$ and the error distribution does not have an atom at zero (Bougerol and Picard 1992).

[6] In the ARCH(1) case κ_y solves $(E|\epsilon_t|^{\kappa_y})^{2/\kappa_y} \beta_1^0 = 1$: large β_1^0 implies small κ_y (de Haan et al 1989).

In the ARCH case the AR-in-squares error $v_t = y_t^2 - h_t^2 = (\epsilon_t^2 - 1)\theta^{0\prime}x_t$ depends on x_t, thus $m_{1,t}(\theta^0) = (\epsilon_t^2 - 1)h_t^2 = v_t$ has tail index $\kappa_y/2 > 2$ and all other $m_{i,t}(\theta^0) = (\epsilon_t^2 - 1)h_t^2 y_{t-i+1}^2$ for $i \geq 2$ have index $\kappa_y/4 > 1$ due to feedback. The intuition from Section 3.1 suffices to explain the rate: models with disproportionately heavy-tailed "errors" render less sharp estimates (in this case, less than super-$T^{1/2}$-consistent).

Remark 2: The tails of ϵ_t do not play any role *per se*. Thicker tailed ϵ_t and/or larger slopes β^0 imply y_t is heavier tailed: *why* y_t is heavy tailed is irrelevant.

3.3 AR-ARCH WITH LEAST SQUARES EQUATIONS

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

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

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

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

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

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

Thus $\{y_t\}$ is stationary, geometrically β-mixing with regular varying tail (7) and index $\kappa_y > 0$ (Borkovec and Klüppelberg 2001: Theorems 1 and 3), and $m_{i,t}(\theta^0)$ satisfy (7) with indices $\{\kappa_y, \kappa_y/2, \kappa_y/4\}$. Error normality can be replaced with sufficient distribution smoothness.

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

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

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

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

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

COROLLARY 3.4 (AR with ARCH Error) *Under the conditions of Lemma 3.3 if only* ρ^0 *is estimated with one equation* $m_t(\rho) = (y_t - \rho y_{t-1})y_{t-1}$ *then* $T_\rho \sim KT^{1/2}$ *if* $\kappa_y > 4$, $T_\rho \sim KT^{1/2}\left(T/k_T\right)^{-(2/\kappa_y - 1/2)}$ *if* $\kappa_y \in (2, 4)$, *and* $T_\rho \sim KT^{1/2}/L(T)$ *if* $\kappa_y = 4$.

Remark: Feedback between error u_t and regressor y_{t-1} substantially elevates estimating equation tail thickness relative to the Jacobian, hence the convergence rate T_ρ falls below $T^{1/2}$.

3.4 ARCH WITH QML EQUATIONS

Consider an ARCH(1) with QML equations

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

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

where $E[m_t(\theta)] = 0$ *if and only if* $\theta = \theta^0$. Integrability requires $E[\epsilon_t^2] < \infty$, and assume at least one $\beta_i^0 > 0$ to highlight the impact of scaling.

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

LEMMA 3.5 *Assume $E|y_t|^p < \infty$ for some $p > 0$, $\beta^0 > 0$ and $E[\epsilon_t^2] < \infty$. If $\kappa_\epsilon > 4$ then $T_\alpha = T_\beta = KT^{1/2}$, if $\kappa_\epsilon = 4$ then $T_\alpha, T_\beta \sim KT^{1/2}/L(T)$, and if $\kappa_\epsilon \in (2,4)$ then $T_\alpha, T_\beta \sim KT^{1/2}\,(T/k_T)^{-(2/\kappa_\epsilon - 1/2)}$.*

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

Remark 2: Although QML equations permit far heavier tails than least-squares equations, the rates of convergence suffer when ϵ_t has an infinite kurtosis since scaling eliminates a leverage effect. Nevertheless the rate can be made arbitrarily close to $T^{1/2}$ when $\kappa_\epsilon \in (2,4]$ by maximal trimming $k_T = [T/L(T)]$: $T_\alpha, T_\beta \sim KT^{1/2}/L(T)$. Compare this with the QML rate T^{1-2/κ_ϵ} for any $\kappa_\epsilon \in (2,4)$ (Hall and Yao 2003: Theorem 2.1). Since $1 - 2/\kappa_\epsilon < 1/2$ is strict for any $\kappa_\epsilon \in (2,4)$, then GMTTM beats QML in rate of convergence as long as the rule of thumb $k_T = [T/L(T)]$ is used.

Remark 3: Ling (2007) requires $E[\epsilon_t^4] < \infty$ for his weighted QML estimator to be asymptotically normal. This follows since Ling only weighs according to lagged y_t, and not ϵ_t. By conparison, GMTTM achieves asymptotic normality under only $E[\epsilon_t^2] < \infty$ because by trimming by $m_t(\theta^0) = (\epsilon_t^2 - 1)h_t^{-2}x_t$ we necessarily control for large errors ϵ_t. Indeed, $||h_t^{-2}x_t||$ is square integerable (Francq and Zakoïan 2010), so heavy tails are solely associated with ϵ_t (Cline 1986).

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

Remark 5: Although we do not tackle the general problem of nonlinear error-regressor dependence, it seems reasonable to suspect such feedback will always reduce the rate below $T^{1/2}$ when tails are thick enough.

4. FRACTILE SELECTION FOR ASYMMETRIC EQUATIONS

The preceding section presents rules for selecting how fast $k_{j,T} \to \infty$ to maximize the efficient GMTTM rates of convergence. In this section we use the rate of identification $E[m_{T,t}^*] \to 0$ as a criterion for selecting an asymmetric policy $\{k_{1,T}, k_{2,T}\}$. Let $m_t(\theta^0) \in \mathbb{R}$ denote a specific equation, and drop θ^0 everywhere for notational simplicity. The existence of asymmetric equations is hardly in question. An ARCH(1) $y_t = (\alpha^0 + \beta^0 y_{t-1}^2)^{1/2}\epsilon_t$, $\alpha^0, \beta^0 > 0$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$ has asymmetrically distributed least squares and QML equations.

Let $m_t = (\epsilon_t^2 - 1)z_t^2$ denote a particular scalar least squares equation $z_t \in \{1, y_{t-1}^2\}$. Then m_t has tail (7) with index $\kappa > 0$ (de Haan et al 1989), hence as $m \to \infty$

$$P\left(\left(\epsilon_t^2 - 1\right) z_t^2 > 1\right) \sim m^\kappa P\left(\left(\epsilon_t^2 - 1\right) z_t^2 > m\right) \to d_2$$

$$P\left(\left(\epsilon_t^2 - 1\right) z_t^2 < -1\right) \sim m^\kappa P\left(\left(\epsilon_t^2 - 1\right) z_t^2 < -m\right) \to d_1.$$

Since $\epsilon_t \overset{iid}{\sim} N(0,1)$ it follows $d_2 > d_1$.

Now let m_t be an arbitrary scalar equation. Simplify exposition further by sharpening condition D1.ii in Appendix A to an exact Pareto tail: for all $m \geq M$ and some $M \geq 1$

$$P\left(m_t < -m\right) = d_1 m^{-\kappa_1} \quad \text{and} \quad P\left(m_t > m\right) = d_2 m^{-\kappa_2},$$

where $d_i > 0$, and $\min\{\kappa_i\} > 1$ ensures integrability. It is a simple exercise to show for large T

$$E\left[m_{T,t}^*\right] = E\left[m_t I\left(-l_T \leq m_t \leq u_T\right)\right] = \left(\frac{d_1\kappa_1}{\kappa_1 - 1}\right)\frac{1}{l_T^{\kappa_1 - 1}} - \left(\frac{d_2\kappa_2}{\kappa_2 - 1}\right)\frac{1}{u_T^{\kappa_2 - 1}}.$$

Although identification $E[m_{T,t}^*] \to 0$ holds for any thresholds $\{l_T, u_T\} \to \infty$ hence any intermediate order sequences $\{k_{1,T}, k_{2,T}\}$, we can always relate $k_{1,T}$ to $k_{2,T}$ to promote $E[m_{T,t}^*] \to 0$ arbitrarily fast. In general, the heavier tail (e.g. $d_2 > d_1$ and/or $\kappa_2 < \kappa_1$) should be trimmed less (e.g. $u_T > l_T$ hence $k_{2,T} < k_{1,T}$). Simply consider symmetric trimming $l_T = u_T = c_T$ with $d_2 > d_1$ and $\kappa_1 = \kappa_2$. Then $E\left[m_{T,t}^*\right] < 0$ because a disproportionate number of large positive values are being trimmed. The is ameliorated for each T by lowering the number of trimmed right-tail equations $k_{2,T} < k_{1,T}$. This is strongly demonstrated by simulation in Section 6.

The thresholds $\{l_T, u_T\}$ for Pareto tails by construction (4) are $l_T = d_1^{1/\kappa_1}(T/k_{1,T})^{1/\kappa_1}$ and $u_T = d_2^{1/\kappa_2}(T/k_{2,T})^{1/\kappa_2}$. Use this to deduce $E[m_{T,t}^*] \approx 0$ for any T when $d_2^{1/\kappa_2}/(T/k_{2,T})^{1-1/\kappa_2}$ $\approx d_1^{1/\kappa_1}/(T/k_{1,T})^{1-1/\kappa_1}$ or $\{k_{1,T}, k_{2,T}\}$ satisfy

$$\frac{k_{2,T}^{1-1/\kappa_2}}{k_{1,T}^{1-1/\kappa_1}} = T^{1/\kappa_1 - 1/\kappa_2}\frac{d_1^{1/\kappa_1}}{d_2^{1/\kappa_2}}\frac{(1 - 1/\kappa_2)}{(1 - 1/\kappa_1)} = T^{1/\kappa_1 - 1/\kappa_2} \times \mathcal{D}\left(d, \kappa\right)$$

The ARCH model is a simple case since $\kappa_1 = \kappa_2 = \kappa$ for any particular equation, hence $k_{2,T}/k_{1,T} \sim (d_1/d_2)^{1/(\kappa-1)}$. For example, if $k_{j,T} \sim \delta_j T/\ln(T)$ then any policy $\{\delta_1, \delta_2\}$ must satisfy $\delta_2/\delta_1 \sim (d_1/d_2)^{1/(\kappa-1)}$. Similarly, if $k_{j,T} \sim \delta_j T^{\lambda_j}$ then $\delta_2/\delta_1 \sim (d_1/d_2)^{1/(\kappa-1)}$ and $\lambda_1 = \lambda_2$.

Table 1 presents three classes of $k_{1,T}$ based on the Section 3 rules of thumb for optimization GMTTM rates convergence, and solving for $k_{2,T}$.

Table 1 : General Policy Classes $\{k_{1,T}, k_{2,T}\}$

$k_{1,T}$ Class	$k_{2,T}$	$k_{2,T}$ Class
$L_1(T)$	$\mathcal{D}(d,\kappa)^{1/(1-1/\kappa_2)} L_1(T)^{(1-1/\kappa_1)/(1-1/\kappa_2)} T^{\kappa_1(1-\kappa_1/\kappa_2)/(1-1/\kappa_2)}$	$L_2(T) T^{\lambda_2}$
$\dfrac{T}{L_1(T)}$	$\mathcal{D}(d,\kappa)^{1/(1-1/\kappa_2)} \dfrac{T}{L_1(T)^{(1-1/\kappa_1)/(1-1/\kappa_2)}}$	$\dfrac{T}{L_2(T)}$
$\delta_1 T^{\lambda_1}$	$\left(\delta_1^{1-1/\kappa_1}\mathcal{D}(d,\kappa)\right)^{1/(1-1/\kappa_2)} T^{[\lambda_1(1-1/\kappa_1)+(1/\kappa_1-1/\kappa_2)]/(1-1/\kappa_2)}$	$\delta_2 T^{\lambda_2}$

In each case the rate of increase $k_{j,T} \to \infty$ should differ across tails only if tail indices differ. The first class $k_{1,T} \sim L_1(T)$ only makes sense if $\kappa_2 \geq \kappa_1$ since $\kappa_2 < \kappa_1$ implies $\lambda_2 < 0$ hence $k_{2,T} \to 0$. Otherwise fix $k_{2,T} \sim L_2(T)$ and solve for $k_{1,T}$. If $\kappa_2 > \kappa_1$ the right-tailed observations are trimmed exponentially faster than left-tailed observations. Only the heaviest tail can have a slowly varying fractile in this class. Based on Section 3 this class apparently makes sense solely for models *without* error-regressor feedback.

The remaining classes $T/L_j(T)$ and $\delta_j T^{\lambda_j}$ apply to both tails for any scales or indices. In the second class $T/L_j(T)$ the heavier tail (e.g. $\kappa_1 < \kappa_2$) receives heavier trimming (e.g. $k_{1,T}/k_{2,T} \to \infty$) aligning this class naturally to models with error-regressor feedback. In the third class $\delta_1 T^{\lambda_1}$ the heavier tail (e.g. $\kappa_1 < \kappa_2$) receives lighter trimming (e.g. $k_{1,T}/k_{2,T} \to 0$), aligning this class with models that do not exhibit error-regressor feedback.

GARCH models with innovations that have the same left- and right-tail index have least squares and QML equations that are symmetric in $\kappa_1 = \kappa_2 = \kappa$. In each class above it is easy to show

$$\text{if } \kappa_1 = \kappa_2 = \kappa \text{ then } k_{2,T} = (d_1/d_2)^{1/(\kappa-1)} k_{1,T},$$

so a thicker right tail $d_2 > d_1$ implies trimming fewer right tail equations $k_{2,T} < k_{1,T}$.

The nuisance tail parameters d_i and κ_i are easily estimated (e.g. Hill 1975, Hall 1982). See Hill (2010a) for proofs of consistency for mixing processes.

5. AUTOREGRESSIONS AND ARCH

We now verify the major assumptions for heavy-tailed stationary AR and ARCH with weight $\Upsilon_T = S_T^{-1}||S_T||$.

5.1 Autoregression

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

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

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

Assume ϵ_t has an absolutely continuous marginal distribution symmetric at zero, uniformly bounded and positive $\mathbb{R}$-*a.e.* Then y_t is uniformly $L_{1+\iota}$-bounded geometrically β-mixing (An and Huang 1996: Theorem 3.1), and y_t and $m_{i,t}(\theta^0) = \epsilon_t y_{t-i}$ have tail (7) with the same index κ while $m_{i,t}(\theta) = \epsilon_t y_{t-i} - (\theta^0 - \theta)' x_t y_{t-i}^2$ has tail (7) with index $\kappa/2$ if $\theta \neq \theta^0$ (Cline 1986, 1989, Davis and Resnick 1996).

Impose symmetric trimming $m^*_{i,T,t}(\theta) = m_{i,t}(\theta) I(|m_{i,t}(\theta)| \leq c_{i,T}(\theta))$ where $(T/k_T)P(|m_{i,t}(\theta)| > c_{i,T}(\theta)) \to 1$, with the same fractile k_T for each equation.

Error independence with Paretian tail (7), stationarity, linearity, distribution continuity, mixing and the efficient weight ensure continuity and power-law tails D1 (cf. Cline 1986), differentiability D2, mixing D3, identification I1 and covariance non-degeneracy M2. 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 $E[m^*_{T,t}(\theta^0)] = 0$ under symmetry, integrability and orthogonality; M1 holds under I1 and D1-D6 for $\hat{\Upsilon}_T = \hat{S}_T^{-1}||\hat{S}_T||$ given HAC consistency Lemma 2.4.

What remains is Jacobia property D5, indicator property D6, and smoothness I3.

D5 (*Jacobia*).

D5.i. Each part is trivial given the linear data generating process and iid innovations with absolutely continuous marginal distribution.

D5.ii. We must show $\sup_{\theta \in U^0(\delta_T)}\{||J_T(\theta) - J_T||\} = o(||J_T||)$. Stationarity, L_p-boundedness of y_t and the construction $J^*_{T,i,j,t}(\theta) = -y_{t-i}y_{t-j}I_{T,j,t}(\theta)$ for a linear AR imply $E[(\sup_{\theta \in U^0(\delta_T)}\{||J^*_T(\theta) - J^*_T||\})^s] \leq K$ for any $s \in (0, p/2)$, hence $\sup_{\theta \in U^0(\delta_T)}\{||J^*_T(\theta) - J^*_T||\} = o_p(||J_T||)$ by Markov's inequality and $||J_T|| \to \infty$. Hence D5.ii follows if we show $||J^*_T - J_T|| = o_p(||J_T||)$.

In lieu of the general Jacobian property $E[J^*_{T,t}] = J_T(1 + o(1))$ by Lemma C.4 in Appendix C, it suffices to show

$$\frac{1}{T}\sum_{t=1}^{T} \frac{y_{t-i}y_{t-j}I_{T,j,t}(\theta^0) - E\left[y_{t-i}y_{t-j}I_{T,j,t}(\theta^0)\right]}{||J_T||} := \frac{1}{T}\sum_{t=1}^{T}\mathcal{Y}_{i,j,T} = o_p(1).$$

Since geometric β-mixing ensures $\{y_{t-i}y_{t-j}I_{T,j,t}(\theta^0), \Im_t\}$ forms a geometric L_2-mixingale, apply stationarity and the Theorem 2.1 partial sum bound in Hill (2010b) to deduce $E(1/T\sum_{t=1}^{T}\mathcal{Y}_{i,j,T})^2 \leq T^{-1}E[\mathcal{Y}^2_{i,j,T}]$. Lemma 3.1 dictates $||J_T|| \sim K(T/k_T)^{2/\kappa - 1}$, and arguments in the line of proof of Lemma 3.1 can be used to show $E[y^2_{t-i}y^2_{t-j}I_{T,j,t}(\theta^0)] \sim K\,(T/k_T)^{4/\kappa - 1}$. Therefore

$$\frac{1}{T}E\left[\mathcal{Y}^2_{i,j,T}\right] = O\left(\frac{1}{T}\frac{(T/k_T)^{4/\kappa-1}}{(T/k_T)^{4/\kappa-2}}\right) = O\left(\frac{(T/k_T)}{T}\right) = o(1).$$

The law now follows by Chebyshev's inequality.

D6 (*indicator class*). The fdd's of $m_{i,t}(\theta)$ are absolutely continuous so we may assume without any loss of generality the thresholds $c_{i,T}(\theta)$ are continuous. Further, fdd's of $m_{i,t}(\theta)$ have bounded densities uniformly on Θ. Therefore $I_{i,T,t}(\theta)$ is L_2-Lipschitz: $E[(I_{i,T,t}(\theta) - I_{i,T,t}(\tilde{\theta}))^2] \leq K||\theta - \tilde{\theta}||$. Proving the L_2-bracketing numbers satisfy D6 is then a classic exercise. See Giné and Zinn (1984), Pollard (1984, 1989, 2002) and van der Vaart and Wellner (1996).

I3 (*smoothness*). Define $Q_T(\theta) := E[m^*_{T,t}(\theta)]' \times \Upsilon_T \times E[m^*_{T,t}(\theta)]$. Identification at interior point $\theta^0 \in \Theta$ and the definition of a derivative imply $E[m^*_{T,t}(\theta)] = J_T(\theta - \theta^0) + o(||J_T|| \times ||\theta - \theta^0||)$, hence $\mathfrak{m}_T := \sup_\theta ||E[m^*_{T,t}(\theta)]|| \leq K||J_T|| \times (1 + o(1))$ given compactness of Θ. Therefore since Υ_T is bounded

$$\inf_{||\theta-\theta^0||>\delta}\left\{\mathfrak{m}_T^{-2}Q_T(\theta)\right\} \geq \inf_{||\theta-\theta^0||>\delta}\left\{\left(\theta-\theta^0\right)'\frac{J'_T}{||J_T||}\Upsilon_T\frac{J_T}{||J_T||}\left(\theta-\theta^0\right)\right\}\times(1+o(1))+o(1).$$

But boundedness and positive definiteness of Υ_T imply $J_T' \Upsilon_T J_T/||J_T||^2$ is positive definite for sufficiently large T, so I3 follows.

Asymptotic normality is therefore a primitive property of GMTTM for stationary AR data with a sufficiently smooth error distribution. The above verification with Theorems 2.2 and 2.6 and Lemma 3.1 suffice to prove the following claim.

COROLLARY 5.1 *Let y_t satisfy (8) and let $\Upsilon_T = S_T^{-1}/||S_T||^{-1}$. Then $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{d} N(0, I_r)$ and $\hat{V}_T = V_T(1 + o_p(1))$. In particular, $T^{1/2}(T/k_T)^{1/\kappa - 1/2}(\hat{\theta}_{i,T} - \theta_i^0) \xrightarrow{d} N(0, V_i)$ for any $k_T \to \infty$ and $k_T = o(T)$ each i and some $V_i < \infty$. Further, $(T^{1/\kappa}/L(T))(\hat{\theta}_{i,T} - \theta_i^0) \xrightarrow{d} N(0, V_i)$ for some $V_i < \infty$ and slowly varying $L(T) \to \infty$ when $k_T \sim L(T)$.*

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

5.2 ARCH, GARCH, Nonlinear-GARCH

The ARCH(p) process $\{y_t\}$ from section 3.2 is stationary, L_p-bounded, and geometrically β-mixing with tail (7). See Basrak et al (2002: Theorem 3.1) and Meitz and Saikkonen (2008: Proposition 1). The Lyapunov condition can be replaced with $\sum_{i=1}^{p} \alpha_i^0 < 1$, or $\sum_{i=1}^{p} \alpha_i^0 = 1$ and the error distribution does not have an atom at zero (Bougerol and Picard 1992).

Least squares and QML equations $m_t(\theta)$ are differentiable in θ with absolutely continuous marginal distributions, and integrable at θ^0 for all $\kappa_y > 4$ (least squares) or all $\kappa_y > 2$ (QML). All conditions can be verified using arguments from Sections 4.1 and 5.1.

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

6. SIMULATION STUDY In this section we compare one-step and two-step GMTTM's, denoted $\hat{\theta}_T^{(1)}$ and $\hat{\theta}_T^{(2)}$, to a variety of estimators detailed below. 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 $P(\epsilon_t > \epsilon) = P(\epsilon_t < -\epsilon) = (1/2) \times (1 + \epsilon)^{-\gamma}$. Random draws from P_γ with $\gamma > 2$ are standardized to ensure $\epsilon_t \overset{iid}{\sim} (0, 1)$. Define the moment supremum $\kappa_y = \sup\{\alpha > 0 : E|y_t|^\kappa < \infty\}$. See Table 2 for each DGP $y_t = f(\theta^0, \epsilon_t, x_t)$ where x_t may contain a constant and lags of y_t. We simulate 1000 samples of size $T = 1000$ for each model.

TABLE 2 - Data Generating Processes

Model Type	Functional Form: $f(\theta^0, \epsilon_t, x_t)$	θ_r^0	iid errors ϵ_t	κ_y
LOCATION	$1 + \epsilon_t$	1	$P_{1.5}$, $P_{2.5}$	1.50, 2.50
AR(1)	$.9 \times y_{t-1} + \epsilon_t$,	.9	$P_{1.5}$, $P_{2.5}$	1.50, 2.50
ARCH(1)	$\{.3 + .5y_{t-1}^2\}^{1/2}\epsilon_t$	.5	$N_{0,1}$	4.65[7]
ARCH(1)	$\{.3 + .6y_{t-1}^2\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	3.80
ARCH(1)	$\{.3 + .6y_{t-1}^2\}^{1/2}\epsilon_t$	.6	$P_{2.5}$	1.80
GARCH(1,1)	$\{.3 + .3y_{t-1}^2 + .6h_{t-1}^2\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	4.10
IGARCH(1,1)	$\{.3 + .4y_{t-1}^2 + .6h_{t-1}^2\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	2.00
GARCH(1,1)	$\{.3 + .3y_{t-1}^2 + .6h_{t-1}^2\}^{1/2}\epsilon_t$	.6	$P_{2.5}$	1.50
TARCH(1)	$\{.3 + .6y_{t-1}^2 \times I(y_{t-1} < 0)\}^{1/2}\epsilon_t$	.6	$N_{0,1}$	5.25[8]
TARCH(1)	$\{.3 + .6y_{t-1}^2 \times I(y_{t-1} < 0)\}^{1/2}\epsilon_t$	.6	$P_{2.5}$	2.60
QARCH(1)	$\lvert .3 + .8y_{t-1}\rvert\, \epsilon_t$	.8	$N_{0,1}$	3.50[9]
QIARCH(1)	$\lvert .3 + y_{t-1}\rvert\, \epsilon_t$	1	$N_{0,1}$	2.00

All processes in this study have regularly varying distribution tails (7) with index κ_y > 0 (Hannan and Kanter 1977, Cline 1986, 1989, Borkovec and Klüppelberg 2001, Cline 2007). In the iid and AR cases tail thickness is gauged by the Pareto innovations, and all equations $m_{i,t}(\theta^0)$ have symmetric tail indices.

We compare GMTTM with Ling's (2005) and Pan et al's (2007) Least Absolute Weighted Deviations [LAWD] $\hat{\theta}_{LW}$, untrimmed GMM $\hat{\theta}_G$, QML $\hat{\theta}_{QL}$ and Ling's (2007) Quasi-Maximum Weighted Likelihood [QMWL] $\hat{\theta}_{QW}$. Similar results for GARCH models are obtained with Peng and Yao's (2003) Log-LAD.

Write $\hat{\theta}$ to denote any estimator. All model are estimated by GMTTM, GMM, and QML (QML is just OLS for LOCATION and AR models); we use LAWD for LOCATION and AR models; and QMWL for all GARCH models. Estimator descriptions are detailed below. Based on our data generating processes OLS is consistent, and LAWD and QMWL are $T^{1/2}$-convergent and asymptotically normal respectively for each LOCATION and linear AR, and each linear GARCH model. See Davis et al (1992) and Ling (2005, 2007).

The collective GARCH group have heavy tails due to the innovations $\epsilon_t \stackrel{iid}{\sim} P_{2.5}$ and/or the parametric structure. The kurtosis of y_t is infinite in most cases, and variance is infinite for IGARCH and QIARCH, and ARCH and GARCH with Pareto errors. Thus, in all random volatility models the GMM estimator is *not* asymptotically normal, and QML has not been shown to deliver an asymptotically normal estimator when $E[\epsilon_t^4] = \infty$. Nevertheless, QML is consistent in all cases[10].

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

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

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

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

6.1 Evaluation

We analyze r^{th} parameter estimates $\hat{\theta}_{T,r}$ for brevity. Consult the third column of Table 2 for the true θ_r^0. Estimator performance is gauged by simulation means, mean-squared-errors, and Kolmogorov-Smirnov tests of standard normality. Let $\{\hat{\theta}_{T,j,r}\}_{j=1}^{1000}$ be the independently drawn sequence of estimates of θ_r^0.

We use the simulation mse $\hat{s}_{T,r}^2 = (1/1000)\sum_{j=1}^{1000}(\hat{\theta}_{T,j,r} - \theta_r^0)^2$ to generate an iid sequence of ratios $\{\hat{t}_{j,r}\}_{j=1}^{1000}$, $\hat{t}_{j,r} = \{\hat{\theta}_{T,j,r} - \theta_r^0\}/\hat{s}_{T,r}^2$. We report $(1/1000)\sum_{j=1}^{1000}\hat{\theta}_{T,j,r}$, $\hat{s}_{T,r}$, and the KS test is based on $\{\hat{t}_{j,r}\}_{j=1}^{1000}$.

6.2 Estimating Equations

The GMM and GMTTM estimating equations are $m_t(\theta) = u_t(\theta) \times z_t(\theta)$ for some "error: $u_t(\theta) \in \mathbb{R}$ and "regressor" $z_t(\theta) \in \mathbb{R}^q$ described in Table 3. In each location and AR model least squares-type equations are used, $u_t(\theta) = y_t - \theta' x_t$. Recall QML-type equations in GARCH cases allows for the minimal moment condition $E[\epsilon_t^2] < \infty$. We therefore use QML-type $u_t(\theta) = (y_t^2 - h_t^2(\theta))h_t^{-4}(\theta)$ for GARCH models with index $\kappa_y \leq 4$ and least squares-type $u_t(\theta) = y_t^2 - h_t^2(\theta)$ if $\kappa_y > 4$.

We consider both exact and over-identification cases concerning choice of $z_t(\theta)$. Exact identification allows for direct comparisons with least squares and QML. The two cases result in qualitatively similar results, so we only present output for exact identification See Hill and Renault (2010) for omitted results.

TABLE 3 - Estimating Equations $m_t(\theta) = u_t(\theta) \times z_t(\theta)$

Model	iid errors ϵ_t	κ_y	$u_t(\theta) \in \mathbb{R}$	$z_t(\theta) \in \mathbb{R}^q$: $q = 1$ or 2^a
Location	$P_{1.5}$, $P_{2.5}$	1.5, 2.5	LS, LS	1 or $[1, y_{t-1}]'$
AR(1)	$P_{1.5}$, $P_{2.5}$	1.5, 2.5	LS, LS	$[y_{t-i}]_{i=1}^q$
ARCH(1)	$N_{0,1}$	4.65	LS	$\left[1, \{y_{t-i}^2\}_{i=1}^q\right]'$
ARCH(1)	$N_{0,1}$, $P_{2.5}$	3.80, 1.80	QML, QML	$\left[1, \{y_{t-i}^2\}_{i=1}^q\right]'$
GARCH(1,1)^b	$N_{0,1}$	4.10	LS	$\left[1, \{y_{t-i}^2, h_{t-i}^2(\theta)\}_{i=1}^q\right]'$
GARCH(1,1)	$N_{0,1}$, $P_{2.5}$	2.00, 1.50	QML, QML	$\left[1, \{y_{t-i}^2, h_{t-i}^2(\theta)\}_{i=1}^q\right]' + \beta\frac{\partial}{\partial\theta}h_{t-1}^2(\theta)$
TARCH(1)^c	$N_{0,1}, P_{2.5}$	5.25, 2.60	LS, QML	$\left[1, y_{t-1}^2 I_{t-1}, ..., y_{t-q}^q I_{t-q}\right]'$
QARCH(1)	$N_{0,1}$, $N_{0,1}$	3.50, 2.00	QML, QML	$h_t(\theta) \times [1, \{y_{t-i}\}_{i=1}^q]'$

Notes: a. In all cases q = 1 or 2. Exact identification corresponds to q = 1.
b. $h_t^2(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2(\theta)$; c. $I_t := I(y_t < 0)$.

6.3 GMTTM Fractile Selection

Equations $m_{i,t}(\theta^0)$ in each model are either all symmetric or all asymmetric, *except* QARCH. In the latter case the constant term equation is skewed right,

$$\text{QARCH: } m_{1,t}(\theta^0) = \left(y_t^2 - h_t^2\right)h_t^{-3} = \left(\epsilon_t^2 - 1\right)h_t^{-1},$$

hence asymmetrically trimmed. The remaining QARCH equations are symmetric,

$$\text{QARCH: } m_{i,t}(\theta^0) = \left(y_t^2 - h_t^2\right)h_t^{-3}y_{t-i} = \left(\epsilon_t^2 - 1\right)h_t^{-1}y_{t-i} = \left(\epsilon_t^2 - 1\right)h_t^{-1}h_{t-i}\epsilon_{t-i},$$

because ϵ_t is iid symmetrically distributed about zero. Hence $m_{i,t}(\theta^0)$ are symmetrically trimmed. We therefore discuss QARCH separately.

In all models except QARCH we use the same fractiles $\{k_{1,T}, k_{2,T}\}$ for each equation since evaluation focuses on just $\hat{\theta}_{T,r}$. In the first experiment we use regularly varying trimming fractiles $k_{j,T} = [T^{\lambda_j}]$ where $\lambda_j \in \{.01, .02, ..., .99\}$. Models with symmetrically distributed equations (LOCATION, AR) are trimmed symmetrically: $\lambda_1 = \lambda_2$. All GARCH models except QARCH demand asymmetric trimming for each equation. These results are summarized in Tables 4-6.

The class T^{λ_j} is best aligned with symmetric equations without error-regressor feedback, as in LOCATION and AR. Still, this policy does not optimize the convergence rates in any model treated here. In a second experiment we use rate optimizing policies: $k_T = [\delta \ln(T)]$ for AR; $k_T = [\delta T/\ln(T)]$ for LOCATION; and $k_{j,T} = [\delta_j T/\ln(T)]$ for GARCH[11]. In all cases except QARCH we minimize the KS statistic over a one *or* two dimensional grid of $\delta_i \in \{.01, .02, ..., 3.0\}$. See Table 7.

In the QARCH case we use asymmetric fractiles for the first equation $\{k_{1,1,T}, k_{2,1,T}\}$ and identical left and right fractiles k_T for all remaining equations. A three dimensional grid search is therefore performed over $\{k_T, k_{1,1,T}, k_{2,1,T}\} = \{T^{\lambda}, T^{\lambda_1}, T^{\lambda_2}\}$ in the first experiment and $\{k_T, k_{1,1,T}, k_{2,1,T}\} = \{[\delta T/\ln(T)], [\delta_1 T/\ln(T)], [\delta_2 T/\ln(T)]\}$ in the second experiment.

6.4 GMTTM Weight

Let $\hat{\theta}_T^{(1)}$ be the one-step GMTTM estimator based on the naïve weight $\hat{\Upsilon}_T = I_q$, and $\hat{\theta}_T^{(2)}(\tilde{\theta}_T)$ the two-step estimator with efficient weight $\hat{\Upsilon}_T = \hat{\Sigma}_T^{-1}(\tilde{\theta}_T) \times ||\hat{\Sigma}_T(\tilde{\theta}_T)||$ and plug-in $\tilde{\theta}_T$.

In simulations not reported here $\hat{\theta}_T^{(1)}$ dominated $\hat{\theta}_T^{(2)}(\hat{\theta}_T^{(1)})$ across models and evaluation criteria due to the computational complexity of a multi-step algorithm under nonlinearity associated with trimming. Further, the two-step $\hat{\theta}_T^{(2)}(\hat{\theta}_Q)$ with a QML plug-in dominated $\hat{\theta}_T^{(1)}$ and $\hat{\theta}_T^{(2)}(\hat{\theta}_G)$ with a GMM plug-in. Since QML is more stable than GMM and one-step GMTTM due to non-trimming and scaling (for GARCH), and $\hat{\theta}_Q$ is consistent for each DGP in this study, we compute $\hat{\theta}_T = \hat{\theta}_T^{(2)}(\hat{\theta}_Q)$ in all cases. In all cases the GMM estimator is computed in two steps using a QML first-step plug-in.

6.5 LAWD and QMWL

Ling (2005,2007) proposes weighted versions of LAD and QML respectively for symmetric heavy tailed AR and GARCH models. In this study use LAWD (Ling 2005) for LOCATION and AR models, and QMWL for all GARCH models.

The LWAD estimator for AR models solves $\text{argmin}_{\theta\in\Theta}\{\sum_{t=2}^{T} w_t|y_t - \theta' x_t|\}$. Ling's (2005) suggested weight w_t is inspired by arguments in Huber (1964): $w_t = 1$ if $a_t = 0$ and $w_t = (y_{([.05T])}^{(a)})^3/a_t^3$ if $a_t \neq 0$, where $a_t := |y_{t-1}|I(|y_{t-1}| \geq y_{([.05T])}^{(a)})$. Thus, values above the 5^{th} two-tailed percentile are given monotonically less weight.

The QMWL estimator for GARCH models solves $\text{argmin}_{\theta\in\Theta}\{\sum_{t=2}^{T}(\ln h_t^2(\theta) + y_t^2/h_t^2(\theta))\}$ with Ling's only suggested weight $w_t = 1$ if $\mathcal{Y}_t = 0$ and $w_t = (y_{([.05T])}^{(a)})^4/a_t^4$ if $a_t \neq 0$, where $\mathcal{Y}_t := \sum_{i=1}^{R} i^{-9}|y_{t-i}|I(|y_{t-i}| \geq y_{([.05T])}^{(a)})$. Ling requires $R = \infty$ but does not suggest how to choose R in practice, so we simply use $R = [T^{1/2}]$. Notice trimming is symmetric which is appropriate for LOCATION, AR, ARCH and GARCH, but apparently not for asymmetric GARCH models, a topic beyond our present scope.

6.6 Summary of Results

[11] Convergence rate optimization has only been verified for LOCATION, AR and ARCH in Section 3. We must leave the study of the convergence rate for nonlinear models and GARCH models to future research.

Refer to Tables 4-7 for all results. Tail-trimming always delivers an approximately normal estimator. The GMTTME is roughly normal even for the profoundly heavy-tailed linear and nonlinear GARCH models. By comparison the GMME fails tests of normality in every heavy tailed case as expected, and the QMLE is non-normal in all cases where it is not asymptotically normal (infinite variance location and AR, GARCH with infinite kurtosis error).

As expected LAWD leads to a sharp estimate for heavy tailed LOCATION and AR models since those data generating processes are symmetric. However, the symmetrically weighted QMWL criterion generates biased estimates for asymmetric TARCH and QARCH models. The most notable findings are summarized below.

i. In the presence of heavy-tails the GMME and QMLE strongly fail KS tests of normality. By comparison the GMTTME passes roughly as well as in any other case.

ii. LAWD and QMWL are designed for heavy tailed symmetric models, although in principle can be extended to asymmetric models. Following the methods described in Ling (2005, 2007) the estimators are sharp for heavy tailed linear models, while QMWL in general leads to biased estimates for heavy tailed asymmetric GARCH. The distortion is particularly acute for heavy tailed TARCH.

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

v. Asymmetric trimming for asymmetrically distributed equations is always optimal. Approximate normality for a small sample is promoted by trimming more observations from the thinner tail, with the intuition discussed in Section 4.

v. Lighter trimming ensures approximate normality for location and AR models, where heavier tailed equations require monotonically less trimming. Conversely, GARCH models with heavier tailed equations requires heavier trimming. Both outcomes match theory predictions from Section 3. This is verified by optimizing the KS statistic over fractile classes $[T^{\lambda_j}]$, $[\delta_j \ln(T)]$ or $[\delta_j T/\ln(T)]$ and comparing λ_j and δ_j. In general the heavier the tails in location and AR the smaller is the KS minimizing λ and δ; and the heavier the tails in GARCH the larger is λ and δ.

vi. The GMTTME works equally well if $k_{j,T} = [T^{\lambda_j}]$ or the convergence rate optimizing variety $k_{j,T} = [\delta_j \ln(T)]$ or $[\delta_j T/\ln(T)]$ is used. This is rather trivial since a grid search renders the KS minimizing values $\{k_{2,T}, k_{2,T}\}$ essentially identical.

vii. A remarkably few number of trimmed large equation observations renders an approximately normal GMM estimator, and corrects for bias and efficiency affects due to heavy tails. The largest fractiles in this study occurred with the infinite variance AR model with least squares equations and the infinite variance ARCH model with QML equations ($k_T = 28$) and the GARCH model with infinite variance ($k_T = 30$). In the ARCH model trimming fewer than 3% of total equations drops the KS statistic from .175 (p-value $< .01$) to .075 (p-value $> .05$), and in the GARCH model the drop is .195 to .063.

7. 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 processes with thin or thick tails.

We prove trimming estimating equations themselves ensures asymptotic normality, while *tail*-trimming can promote consistency for θ^0, super-$T^{1/2}$-convergence for models without error-regressor feedback, and a rate that beats QML for GARCH models with infinite kurtosis errors. Although Ling's (2007) weighted QML for GARCH with heavy-tailed errors leads to a slightly higher rate of convergence than GMTTM, by construction over-identifying conditions are ignored and to date nonlinear GARCH is left untreated. GMTTM allows both by construction.

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; GMTTM dominates GMM and QML in heavy tailed cases; and dominates QMWL for asymmetric GARCH models with heavy tailed errors.

Future work should tackle rates of convergence for nonlinear processes; the trade-off between small sample distribution and efficiency; adaptive methods for selecting the trimming fractiles $\{k_{1,i,T}, k_{2,i,T}\}$; and other criteria for trimming.

APPENDIX A: Assumptions

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

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

hence the GMTTME solves

$$\hat{\theta}_T = \underset{\theta\in\Theta}{\operatorname{argmin}}\{\hat{Q}_T(\theta)\}.$$

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

Asymptotic arguments require the following matrix constructions, some of which are already defined above. The trimmed equation instantaneous and long run covariance matrices are

$$\Sigma_T(\theta) := E\left[m_{T,t}^*(\theta)\, m_{T,t}^*(\theta)'\right] \text{ and } \Sigma_T := \Sigma_T(\theta^0)$$

$$S_T(\theta) := \frac{1}{T}\sum_{s,t}^{T} E\left[m_{T,s}^*(\theta)\, m_{T,t}^*(\theta)'\right] \text{ and } S_T := S_T(\theta^0);$$

the tail-trimmed moment envelope is

$$\mathfrak{m}_T = \sup_{\theta\in\Theta}\left\|E\left[m_{T,t}^*(\theta)\right]\right\|;$$

population and sample Jacobia are

$$J_T(\theta) := \frac{\partial}{\partial\theta}E\left[m_{T,t}^*(\theta)\right] \in \mathbb{R}^{q\times r} \text{ and } J_T = J_T(\theta^0)$$

$$J_{T,t}^*(\theta) := \left[\frac{\partial}{\partial\theta}m_{i,t}(\theta) \times I_{i,T,t}(\theta)\right]_{i=1}^{q} \text{ and } J_T^*(\theta) := \frac{1}{T}\sum_{t=1}^{T} J_{T,t}^*(\theta)$$

$$\hat{J}_{T,t}^*(\theta) := \left[\frac{\partial}{\partial\theta}m_{i,t}(\theta) \times \hat{I}_{i,T,t}(\theta)\right]_{i=1}^{q} \text{ and } \hat{J}_T^*(\theta) := \frac{1}{T}\sum_{t=1}^{T} \hat{J}_{T,t}^*(\theta);$$

and the Hessian and scale are

$$H_T(\theta) := J_T(\theta)'\Upsilon_T J_T(\theta) \in \mathbb{R}^{r\times r} \quad \text{and} \quad H_T := H_T(\theta^0)$$

$$V_T\left(\theta\right) := T \times H_T\left(\theta\right)\left[J_T'\left(\theta\right)\Upsilon_T S_T\left(\theta\right)\Upsilon_T J_T\left(\theta\right)\right]^{-1} H_T\left(\theta\right) \quad \text{and} \quad V_T := V_T(\theta^0).$$

Write compactly throughout

$$c_{i,T}\left(\theta\right) := \max\left\{l_{i,T}\left(\theta\right), u_{i,T}\left(\theta\right)\right\} \quad \text{and} \quad c_T\left(\theta\right) = \max_{1\leq i\leq q}\left\{c_{i,T}\left(\theta\right)\right\}$$

$$k_{i,T} = \max\left\{k_{1,i,T}, k_{2,i,T}\right\} \quad \text{and} \quad k_T = \max_{1\leq i\leq q}\left\{k_{i,T}\right\}.$$

Four sets of assumptions ensure identification for θ^0; $\hat{\theta}_T$ can be expressed asymptotically as a linear function of $\sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0)$; $\sum_{t=1}^T \hat{m}_{T,t}^*(\theta)$ is sufficiently close to $\sum_{t=1}^T m_{T,t}^*(\theta)$ uniformly on Θ; $S_T^{-1/2}\sum_{t=1}^T m_{T,t}^*(\theta^0)$ is asymptotically normal; and $\hat{J}_T^*(\tilde{\theta}_T)$ and $\hat{S}_T(\tilde{\theta}_T)$ are consistent for consistent $\tilde{\theta}_T$. Most are versions of standard regularity conditions contoured to heavy tailed data under tail trimming. The remaining are easily verified for linear-in-parameters models. See Section 4 for dynamic linear regression and ARCH models.

Let $\{\Im_t\}$ be any sequence of increasing σ-fields adapted to $\{m_t(\theta)\}$, $\theta \in \Theta$, where $\{\Im_t\}$ itself does not depend on θ. The first set characterizes matrix norms, weight limits and covariance definiteness[12].

M1 (*weight*). *Υ_T is positive definite for every $T \geq N$ and sufficiently large $N \geq 1$; and $||\hat{\Upsilon}_T - \Upsilon_T|| \xrightarrow{p} 0$ and $||\Upsilon_T - \Upsilon_0|| \to 0$ for some positive definite Υ_0, $0 < ||\Upsilon_0|| < \infty$.*

M2 (*covariance non-degeneracy*). *Each $A_T(\theta) \in \{\Sigma_T(\theta), S_T(\theta)\}$ satisfies $\liminf_{T\geq N}\inf_\theta\{\lambda_{\min}(A_T(\theta))\} > 0$.*

Remark 3: M1 is standard. M2 imposes positive definiteness for sufficiently large $T \geq N$ since trimming can technically render a zero matrix, hence, e.g. $\lambda_{\min}(\Sigma_T(\theta)) = 0$ for some T.

The second set promotes local identification of θ^0.

I1 (*identification by* $m_t(\theta)$). $E[m_t(\theta)] = 0$ if and only if $\theta = \theta^0$, *a unique interior point of compact $\Theta \subset \mathbb{R}^r$.*

I2 (*identification by* $m_{T,t}^*(\theta)$). *$E[m_{T,t}^*(\theta^0)] = o(||S_T||^{1/2}/T^{1/2})$ for $T \geq N$ and sufficiently large $N \geq 1$.*

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

Remark 1: Identification $E[m_{T,t}^*(\theta^0)] \to 0$ is assured by I1 and Lebesgue's dominated convergence. In lieu of the GMM quadratic criterion, I2 states $m_{T,t}^*(\theta)$ identifies θ^0 *sufficiently fast* as information accumulates $T \to \infty$. If the marginal distributions $m_t(\theta^0)$

[12] We simplify notation by ignoring measurability issues that arise when taking a supremum over an index set when the index itself is a function. We implicitly assume all functions in this paper satisfy Pollard's (1984) permissibility criteria, the measure space that governs all random variables is complete, and therefore all majorants are measurable. See also Dudley (1978) for an appeal to Suslin measurability or image admissible Suslin. Thus, probability statements for majorants are with respect to *outer probability*, and expectations over majorants are *outer expectations*.

are symmetric then $E[m_{i,T,t}^*(\theta^0)] = 0$ by I1 for any fractiles $k_{1,i,T} = k_{2,i,T}$ or thresholds $l_{i,T}(\theta^0) = u_{i,T}(\theta^0)$. Further, we show in Section 4 if $m_{i,T,t}^*(\theta^0)$ has exact Pareto tails then $E[m_{i,T,t}^*(\theta^0)] \approx 0$ arbitrarily close for each T for infinitely many sequences $\{k_{1,i,T}, k_{2,i,T}\}$. Otherwise $S_T = o(T)$ is trivial in thin-tailed cases by the Cauchy-Schwarz inequality, and holds for any threshold sequences $\{l_{i,T}(\theta), u_{i,T}(\theta)\}$ for equations with power-law tail decay under intermediate order tail-trimming. See Lemma C.2.

Remark 2: Versions of smoothness I3 are standard for consistency (Huber 1967, Pakes and Pollard 1989, Newey and McFadden 1994). The envelope scale $\mathfrak{m}_T$ is required since $m_t(\theta)$ need not be integrable on Θ-*a.e.* in heavy-tailed cases, whereas $E[m_{T,t}^*(\theta)]/\mathfrak{m}_T$ is always well defined. Consider an AR(1) $y_t = \theta^0 y_{t-1} + \epsilon_t$ with $|\theta^0| < 1$, $\Im_t = \sigma(y_\tau : \tau \leq t)$, martingale difference innovations $E[\epsilon_t|\Im_{t-1}] = 0$ with infinite variance $E[\epsilon_t^2] = \infty$, and one equation $m_t(\theta) = (y_t - \theta y_{t-1})y_{t-1}$. Then $E[m_t(\theta^0)|\Im_{t-1}] = 0$ *a.s.* hence $E[m_t(\theta^0)] = 0$, but in general $m_t(\theta) = -(\theta - \theta^0) \times y_{t-1}^2$is non-integrable for any coefficient $\theta \neq \theta^0$. This matters for a proof of consistency $\hat{\theta}_T \xrightarrow{p} \theta^0$ since that requires a ULLN for $m_{T,t}^*(\theta)$ by well known arguments (e.g. Pakes and Pollard (1989)[13].

The next set concerns properties of the equations $m_t(\theta)$, trimming indicators $I_{i,T,t}(\theta)$ and the random Jacobian matrices $J_{T,t}^*(\theta)$ and $\hat{J}_{T,t}^*(\theta)$. Let $\kappa_i(\theta) \in (0, \infty]$ denote the moment supremum of $m_{i,t}(\theta)$: $E|m_{i,t}(\theta)|^p < \infty$ *if and only if* $p < \kappa_i(\theta)$, where $\kappa_i(\theta) = \infty$ is possible (e.g. bounded support, exponential tail decay). Similarly $\Theta_{2,i} \subseteq \Theta$ is the set of all θ such that $\kappa_i(\theta) \leq 2$.

D1 (*distribution*).

i. The finite dimensional distributions of $m_t(\theta)$ are strictly stationary and absolutely continuous with respect to Lebesgue measure on Θ.

ii. Let $\inf_\theta \kappa_i(\theta) > 0$ and $\kappa_i(\theta^0) > 1$. If $\kappa_i(\theta) \leq 2$ then $P(|m_{i,t}(\theta)| > m)\} = d_i(\theta)m^{-\kappa_i(\theta)}(1 + o(1))$. In particular $\sup_{\theta \in \Theta_{2,i}} |m^{\kappa_i(\theta)}P(|m_{i,t}(\theta)| > m) - d_i(\theta)| \to 0$ where $\inf_{\theta \in \Theta_{2,i}} d_i(\theta) > 0$.

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

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

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

D5 (*Jacobia*).

i. $J_T(\theta)$ exists on Θ-a.e.; $\liminf_{T \geq N} ||J_{i,i,T}|| > 0$ for each $i \in \{1, ..., q\}$; $\sup_\theta ||J_T(\theta)|| < \infty$ for each T; $\{J_T(\theta),\ J_T^*(\theta),\ E[J_{T,t}^*(\theta)],\ \hat{J}_T^*(\theta)\}$ have full column rank for each $T \geq N$ and each $\theta \in \Theta$.

ii. $\sup_{\theta \in U^0(\delta_T)}\{||J_T(\theta) - J_T\} = o(||J_T||)$ for any $\delta_T \to 0$.

D6 (*indicator class*). *$\{I_{i,T,t}(\theta) : \theta \in \Theta\}$ satisfies metric entropy with L_2-bracketing $\mathcal{H}_{[\,]}(\varepsilon, \Theta, ||\cdot||_2) = O(\ln(\varepsilon))$, $\varepsilon \in (0, 1)$.*

Remark 1: Distribution continuity D1.i and equation differentiability D2 reduce generality, but simplify key uniform arguments since trimming adds substantial complexity. In regression models D1.i requires at least idiosyncratic shocks to have a density. Power-law tails D1.ii in the infinite variance case permit elegant representaitons of tail

[13] Although the untrimmed equations $m_t(\theta)$ need not be integrable for arbitrary θ, we prove $\sup_\theta ||1/T \sum_{t=1}^T \{m_{T,t}^*(\theta) - E[m_{T,t}^*(\theta)]\}|| = o_p(\mathfrak{m}_T)$ in Lemma D.3 in Appendix D, which suffices for consistency.

trimmed vectors. The assumption is very mild since distribution tails are bounded by regularly varying function by Markov's inequality, and power-laws govern those infinite variances processes that satisfy a central limit theorem (Ibragimov and Linnik 1997, Leadbetter et al 1983, Resnick 1987)[14].

Remark 2: Mixing D3 and indicator metric entropy D6 promote uniform laws for $m^*_{T,t}(\theta)$ and $\hat{m}^*_{T,t}(\theta) - m^*_{T,t}(\theta)$, while geometric decay keeps notation simple. Nevertheless, many nonlinear AR-nonlinear GARCH models are covered (An and Huan 1996, Carrasco and Chen 2002, Meitz and Saikonnen 2008).

Remark 3: Although we do explicitly bound the rate $c_T(\theta) \to \infty$ it is implicitly bounded under power-law tail decay D1.ii since for any intermediate order sequences $\{k_{1,i,T}, k_{2,i,T}\}$

$$\sup_{\theta} \left\{ c_T(\theta) / \left\| \Sigma_T(\theta) \right\|^{1/2} \right\} = o(T^{1/2}).$$

See Lemma C.1 in Appendix C. Therefore power-law tails ensures $\Sigma_T^{-1/2}(\theta) m^*_{T,t}(\theta)$ is uniformly relatively stable

$$\max_{1 \leq t \leq T} \left\{ \sup_{\theta} \left\| \Sigma_T^{-1/2}(\theta) m^*_{T,t}(\theta) \right\| \right\} = o_p\left(T^{1/2}\right),$$

a property held by finite variance processes that are stationary with weakly dependent maxima (Naveau 2003), or are merely identically distributed (Bonnal and Renault 2004: Lemma A.1; Kitamura et al 2004: Lemma D.2)[15].

Remark 4: Jacobian non-degeneracy D5.i is standard. Smoothness D5.ii ensures if $||J_T(\theta)|| \to \infty$ due to heavy tails then the rate is proportional to $||J_T||$ for θ "close" to θ^0. The assumption automatically holds for linear models. See Section 4.

Remark 5: The D4 moment bounds, D5 Jacobia properties and D6 indicator class are used to prove $1/T \sum_{t=1}^{T} \{\hat{m}^*_{T,t}(\theta) - m^*_{T,t}(\theta)\} = o_p(1)$ uniformly on Θ, required for consistency. Uniformity under non-differentiability of $I_{i,T,t}(\theta)$ is expedited if $\{I_{i,T}(\theta) : \theta \in \Theta\}$ exhibits good metric entropy properties, where metric entropy with L_2-bracketing D6 delivers a required uniform CLT and uniform maximal inequality[16]. We must exploit a uniform maximal inequality, and the simplest set of sufficient conditions with the greatest payoff appears to be due to Doukhan et al (1995) under D3 and D6.

Finally, the kernel class for the HAC estimator $\hat{S}_T(\theta)$.

K1 (*kernel*). *$k(\cdot)$ is a member of class K, where*

$$\begin{aligned} K = \{k : \mathbb{R} \to [-1,1] \mid k(0) = 1,\ k(x) = k(-x)\ \forall x \in \mathbb{R}, \\ \int_{-\infty}^{\infty} |k(x)| dx < \infty,\ \int_{-\infty}^{\infty} |\varpi(\xi)| d\xi < \infty, \\ k(\cdot) \textit{ is continuous at } 0 \textit{ and all but a finite number of points}\}, \end{aligned}$$

[14] We assume a Paretian tail $P(|m_{i,t}(\theta)| > m)\} = d_i(\theta) m^{-\kappa_i(\theta)}(1 + o(1))$ to simplify notation. It is straightforward to generalized D1.ii to $P(|m_{i,t}(\theta)| > m)\} = m^{-\kappa_i(\theta)} L(\theta, m)$ for slowly varying $L(\cdot, m)$ and bounded $L(\theta, \cdot)$. See, e.g., Resnick (1987).

[15] Technically D1.ii only requires the tails to be identically distributed. Relative stability aligns with a necessary and sufficient condition for the distribution limit of a sum of an iid array to be Gaussian (e.g. Kallenberg 2002: Theorem 5.15).

[16] The brackets $\{l, u\}$ of an index function class $\mathcal{F}$ satisfy $l \leq f \leq u$ for every member $f \in \mathcal{F}$, where $\{l, u\}$ may not be members of $\mathcal{F}$; an ε-L_2-bracket $\{l, u\}$ satisfies $||l - u|| \leq \varepsilon$; the L_2-bracketing numbers $\mathcal{N}_{[\,]}(\varepsilon, \Theta, || \cdot ||_2)$ are the number of ε-L_2-brackets required to cover $\mathcal{F}$, and metric entropy with L_2-bracketing is $\mathcal{H}_{[\,]}(\varepsilon, \Theta, || \cdot ||_2) = \ln(\mathcal{N}_{[\,]}(\varepsilon, \Theta, || \cdot ||_2))$. See Giné and Zinn (1984), Pollard (1984), van der Vaart and Wellner (1996) and Dudley (1999). Since $\mathcal{H}_{[\,]}(\varepsilon, \Theta, || \cdot ||_2) = O(|\ln(\varepsilon)|)$ under D6 implies $\int_0^1 \mathcal{H}^{1/2}_{[\,]}(\varepsilon, \Theta, || \cdot ||_2) d\varepsilon < \infty$, a required stochastic equicontinuity condition for weak convergence of a partial sum of $I_{T,t}(\theta)$ applies (see Dudley's 1978 landmark paper).

and $\varpi(\xi) = (2\pi)^{-1}\int_{-\infty}^{\infty} k(x)e^{i\xi x}dx < \infty$. *Further* $\sum_{s,t=1}^{T}|k((s-t)/\gamma_T)| = o(T^2)$, $\max_{1\leq s\leq T}\sum_{t=1}^{T} k((s-t)/\gamma_T) = o(T)$ *and bandwidth* $\gamma_T = o(T)$.

Remark: Class K includes Bartlett, Parzen, Quadratic Spectral, Tukey-Hanning and other kernels. See Davidson and de Jong (2000) and the citations therein.

APPENDIX B: Proofs of Main Results

The following proofs exploit threshold and moment properties Lemmas C.1-C.3 and limit theory Lemmas D.1-D.8. See Appendices C and D respectively.

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

Now, the Lemma D.5 uniform criterion probability bound implies

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

hence $Q_T(\hat{\theta}_T)(1 - o_p(1)) \leq \hat{Q}_T(\hat{\theta}_T) + o_p(\mathfrak{m}_T^2)$. By construction $\hat{Q}_T(\hat{\theta}_T) \leq \hat{Q}_T(\theta^0)$, while weight bound M1, the Lemma D.2.a asymptotic approximation and covariance bound Lemma C.2.c dictate $\hat{Q}_T(\theta^0)$ is bounded:

$$\hat{Q}_T(\theta^0) \leq K\left\|\hat{m}_T^*(\theta^0)\right\|^2 \leq K\left(\left\|m_T^*(\theta^0)\right\| + o_p\left(\left\|S_T\right\|^{1/2}/T^{1/2}\right)\right)^2 = K\left(\left\|m_T^*(\theta^0)\right\| + o_p(1)\right)^2.$$

Finally, the Lemma D.3 law of large numbers states $||m_T^*(\theta^0) - E[m_{T,t}^*(\theta^0)]|| \xrightarrow{p} 0$, and $||E[m_{T,t}^*(\theta^0)]|| = o(||S_T||^{1/2}/T^{1/2}) = o(1)$ by identification I2 and covariance bound Lemma C.2.c. Therefore $||m_T^*(\theta^0)|| = o_p(1)$ by Minkowski's inequality which completes the proof. ■

Proof of Theorem 2.2. Asymptotic linearity Lemma D.6 states

$$V_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = -V_T^{1/2}\left(H_T^{-1}J_T'\Upsilon_T\right)\frac{1}{T}\sum_{t=1}^{T}\hat{m}_{T,t}^*(\theta^0) \times (1 + o_p(1)).$$

Invoke asymptotic approximation Lemma D.2.a coupled with the construction of V_T to deduce

$$V_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = -V_T^{1/2}\left(H_T^{-1}J_T'\Upsilon_T\right)\frac{1}{T}\sum_{t=1}^{T} m_{T,t}^*(\theta^0) \times (1 + o_p(1)) + o_p(1).$$

Now invoke central limit theorem Lemma D.7 and $V_T^{1/2}(T^{-1/2}H_T^{-1}J_T'\Upsilon_T S_T^{1/2}) \to I_r$ to conclude for some $\{A_T\}$, $A_T S_T A_T \to I_r$,

$$V_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = -A_T^{-1/2}T^{-1/2}\sum_{t=1}^{T} m_{T,t}^*(\theta^0) \times (1 + o_p(1)) + o_p(1) \xrightarrow{d} N(0, I_r).$$

■

Proof of Lemma 2.4. Define $\hat{x}^*_{T,t}(\lambda) := \lambda'\hat{m}^*_{T,t}(\tilde{\theta}_T)$ and $x^*_{T,t}(\lambda) := \lambda' m^*_{T,t}(\theta^0)$ for any conformable $\lambda'\lambda = 1$, and write $k_{T,s,t} = k((s - t)/\gamma_T)$. Since it suffices to show

$$\frac{\sum_{s,t=1}^T k_{T,s,t}\hat{x}^*_{T,s}(\lambda)\hat{x}^*_{T,t}(\lambda)}{\sum_{s,t=1}^T E\left[x^*_{T,s}(\lambda)x^*_{T,t}(\lambda)\right]} \xrightarrow{p} 1 \;\forall \lambda'\lambda = 1,$$

and $x^*_{T,t}(\lambda)$ is geometrically β-mixing under D3 with the same bounds as $||\hat{m}^*_{T,t}(\theta_0)||$, we need only consider the univariate case $m_{T,t}(\theta_0) \in \mathbb{R}$ (e.g. Newey and West 1987). Thus, by Minkowski's inequality we must show

$$\mathcal{A}_T := S_T^{-1}\frac{1}{T}\sum_{s,t=1}^T k_{T,s,t}\left\{\hat{m}^*_{T,s}(\tilde{\theta}_T)\hat{m}^*_{T,t}(\tilde{\theta}_T) - m^*_{T,s}(\theta^0)m^*_{T,t}(\theta^0)\right\} \xrightarrow{p} 0$$

$$\mathcal{B}_T := S_T^{-1}\frac{1}{T}\sum_{s,t=1}^T k_{T,s,t}m^*_{T,s}(\theta^0)m^*_{T,t}(\theta^0) \xrightarrow{p} 1.$$

Cross-product expansion Lemma D.2.d with $||\tilde{\theta}_T - \theta^0||_2 = O(T^{-1/2}||S_T||^{1/2}||J_T||^{-1})$ implies $\mathcal{A}_T \xrightarrow{p} 0$.

In order to show $\mathcal{B}_T \xrightarrow{p} 1$ we will apply Theorem 2.1 of Davidson and de Jong (2000) = DJ. It suffices to verify their Assumptions 1-3. DJ's Assumption 1 holds by K1.

Their Assumptions 2 and 3 concern a Near Epoch Dependence property and relate the NED property to the bandwidth γ_T. Both conditions are used solely to promote partial sum variance bounds. It helps to translate DJ's environment into ours. In their setting $m^*_{T,t}(\theta^0)$ is assumed L_r-bounded for $r > 2$, or uniformly square integrable, with a standardization $\mathcal{X}_{T,t} := T^{-1/2}\Sigma_T^{-1/2}m^*_{T,t}(\theta^0)$. Under geometric β-mixing D3 $\{m^*_{T,t}(\theta^0), \Im_t\}$ forms a geometric L_2-mixingale with constants $e_{T,t}$ (cf. McLeish 1975: Theorem 2.1). Therefore $\{\mathcal{X}_{T,t}, \Im_t\}$ forms a geometric L_2-mixingale with constants $\mathcal{E}_{T,t} := T^{-1/2}\Sigma_T^{-1/2}e_{T,t}$.

Their Assumption 2 is used only to ensure $E(\sum_{t=1}^T \mathcal{X}_{T,t})^2 \leq K\sum_{t=1}^T \mathcal{E}^2_{T,t} \leq K$ by invoking McLeish's (1975: Theorem 1.6) maximal inequality and by bounding $e_{T,t}$. But under Lemma C.2.a $||\Sigma_T^{-1}S_T|| \leq K$ hence $E(\sum_{t=1}^T \mathcal{X}_{T,t})^2 \leq K$ without any reference to mixingale constants. A careful inspection of DJ's proof of their Theorem 2.1 reveals $E(\sum_{t=1}^T \mathcal{X}_{T,t})^2 \leq K$ suffices in place of their Assumption 2.

Finally, Assumption 3 states $\gamma_T \times \max_{1\leq t\leq T}\{\mathcal{E}^2_{T,t}\} = o(1)$ and is used, like Assumption 2, only to ensure partial sum bounds for L_2-mixingale functions of $\mathcal{X}_{T,t}$. See especially the proofs of their Lemmas A.3 and A.4. Lemma C.2.a, however, implies we can always side-step the use of mixingale coefficients in partial sum variance bounds for geometrically β-mixing data, in particular we can always replace $e_{T,t}$ with $K||\Sigma_T||^{1/2}$, hence $\mathcal{E}_{T,t} = T^{-1/2}\Sigma_T^{-1/2}e_{T,t}$ with $T^{-1/2}$. Therefore $\gamma_T \times T^{-1} = o(T/T)$ under K1. This completes the proof. ■

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

Denote by $e_i \in \mathbb{R}^r$ the unit vector (e.g. $e_2 = [0, 1, 0, ..., 0]'$), define a sequence of bounded positive numbers $\{\varepsilon_T\}$ that satisfies $\liminf_{T\geq 1}\varepsilon_T||J_T|| > 0$ and $||\tilde{\theta}_T - \theta^0||/\varepsilon_T \xrightarrow{p} 0$. This is always possible in lieu of the plug-in rate and Lemma C.2.c: $||\tilde{\theta}_T - \theta^0||/\varepsilon_T$

$= O_p(T^{-1/2}||S_T||^{1/2}) = o_p(1)$. Define

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

Minkowski's inequality implies for arbitrary θ

$$\left\|\hat{J}^*_T(\tilde{\theta}_T) - J_T\right\| \leq \left\|\hat{J}^*_T(\tilde{\theta}_T) - \check{J}^*_T(\theta, \varepsilon_T)\right\| + \left\|\check{J}^*_T(\theta, \varepsilon_T) - J_T\right\|$$

Apply asymptotic expansion Lemma D.1.b to deduce for some $\tilde{\theta}_{T,*} \in \{\tilde{\theta}_T - e_i\varepsilon_T, \tilde{\theta}_T + e_i\varepsilon_T\}$

$$\hat{J}^*_T(\tilde{\theta}_T) = \check{J}^*_{i,j,T}(\tilde{\theta}_{T,*}, \varepsilon_T) + o_p\left(\|J_T\|\right), \quad \text{hence} \quad \left\|\hat{J}^*_T(\tilde{\theta}_T) - \check{J}^*_T(\theta, \varepsilon_T)\right\| = o_p\left(\|J_T\|\right).$$

Since $||\tilde{\theta}_{T,*} - \theta^0|| \leq ||\tilde{\theta}_T - \theta^0|| = o_p(1)$ it remains to show $||\check{J}^*_T(\tilde{\theta}_T, \varepsilon_T) - J_T|| = o_p(||J_T||)$ for any $||\tilde{\theta}_T - \theta^0|| \xrightarrow{p} 0$. Define

$$U^0\left(\delta_1, \delta_2\right) := \left\{\theta \in \Theta : \delta_1 \leq \left\|\theta - \theta^0\right\| \leq \delta_2\right\} \text{ for } 0 \leq \delta_1 \leq \delta_2$$

$$\mathcal{J}_T\left(\delta_1, \delta_2\right) := \sup_{\theta \in U^0(\delta_1, \delta_2)} \left\{\frac{\|J^*_T(\theta) - J_T\|}{\|J_T\|}\right\}$$

Stochastic differentiability Lemma D.8 and the fact that $U^0(\delta_1, \delta_2) \subseteq U^0\left(0, \delta_2\right)$, and consistency $\tilde{\theta}_T \xrightarrow{p} \theta^0$ imply for large K and any non-zero constant vector $a \in \mathbb{R}^r/0$

$$\left\|\left\{\hat{m}_T(\tilde{\theta}_T + a\varepsilon_T) - \hat{m}_T\left(\theta^0\right)\right\} - \left\{E\left[m^*_{T,t}\left(\tilde{\theta}_T + a\varepsilon_T\right)\right] - E\left[m^*_{T,t}\left(\theta^0\right)\right]\right\}\right\|$$

$$\leq K\left\{1 + \|J_T\| \times \left\|\tilde{\theta}_T + a\varepsilon_T - \theta^0\right\|\right\} \times o_p\left(1\right) \times \left(\mathcal{J}_T\left(\delta_1, \delta_2\right) + o_p\left(1\right)\right)$$

$$\leq K\left\{1 + \|J_T\| \times \left\|\tilde{\theta}_T - \theta^0\right\| + \|J_T\| \times \|a\varepsilon_T\|\right\} \times \left(\mathcal{J}_T\left(\delta_1, \delta_2\right) + o_p\left(1\right)\right)$$

$$= o_p\left(\varepsilon_T\left\|J_T\right\|\right) + O_p\left(\varepsilon_T\left\|J_T\right\| \times \mathcal{J}_T\left(\delta_1, \delta_2\right)\right).$$

Similarly, by differentiability of $E[m^*_{T,t}(\theta)]$,

$$\left\|\frac{E\left[m^*_{T,t}(\tilde{\theta}_T + a\varepsilon_T)\right] - E\left[m^*_{T,t}\left(\theta^0\right)\right]}{\varepsilon_T} - aJ_T\right\|$$

$$= \left\|J_T\varepsilon_T^{-1}\left(\tilde{\theta}_T + a\varepsilon_T - \theta^0\right) - aJ_T + o_p\left(\left\|J_T\right\|\varepsilon_T^{-1}\left(\tilde{\theta}_T + \varepsilon_T - \theta^0\right)\right)\right\|$$

$$= \left\|J_T\varepsilon_T^{-1}\left(\tilde{\theta}_T - \theta^0\right)\right\| + o_p\left(\|J_T\|\right) = o_p\left(\|J_T\|\right).$$

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

$$\left\|\check{J}^*_T(\tilde{\theta}_T, \varepsilon_T) - J_T\right\| = \left\|\frac{\hat{m}^*_T(\tilde{\theta}_T + \varepsilon_T) - \hat{m}^*_T(\tilde{\theta}_T - \varepsilon_T)}{2\varepsilon_T} - J_T\right\| = o_p\left(\|J_T\|\right) + O_p\left(\|J_T\| \times \mathcal{J}_T(\delta_1, \delta_2)\right)$$

hence we have shown $\hat{J}^*_T(\tilde{\theta}_T) = J_T(1 + o_p(1)) + O_p(||J_T|| \times \mathcal{J}_T(\delta_1, \delta_2))$.

Since $0 \leq \delta_1 < \delta_2$ are arbitrary, the proof is complete if we show for some sequence of positive numbers $\{\delta_{1,T}\}$, $\delta_{1,T} \to 0$ and $\delta_{2,T} = 2\delta_{1,T}$:

$$\mathcal{J}_T(\delta_{1,T}, \delta_{2,T}) \xrightarrow{p} 0.$$

Define

$$\mathfrak{m}_T(\delta_1, \delta_2) = \sup_{\theta \in U^0(\delta_1,\delta_2)} \left\| E\left[m^*_{T,t}(\theta)\right]\right\|.$$

The required limit follows from expansion Lemma D.1.b, and a generalization of ULLN Lemma D.3 to $U^0(\delta_1, \delta_2)$. For each $\theta \in U^0(\delta)$ we can always find a sequence $\{\theta_{T,\delta}\} \in U^0(\delta_1, \delta_2)$, $\theta_{T,\delta} \neq \theta^0$ for each finite $T \geq N$, such that

$$\begin{aligned}
\frac{E\left[m^*_{T,t}(\theta_{T,\delta})\right] - E\left[m^*_{T,t}\left(\theta^0\right)\right]}{\left\|\theta_{T,\delta} - \theta^0\right\|} &= \frac{m^*_T(\theta_{T,\delta}) - m^*_T\left(\theta^0\right)}{\left\|\theta_{T,\delta} - \theta^0\right\|} + o_p(1) \times \frac{\mathfrak{m}_T(\delta_1, \delta_2)}{\left\|\theta_{T,\delta} - \theta^0\right\|} \\
&= J^*_T(\theta) \times \frac{\left(\theta_{T,\delta} - \theta^0\right)}{\left\|\theta_{T,\delta} - \theta^0\right\|} \times (1 + o_p(1)) + o_p(1) \times \frac{\mathfrak{m}_T(\delta_1, \delta_2)}{\left\|\theta_{T,\delta} - \theta^0\right\|},
\end{aligned}$$

where each $o_p(1)$ term does not depend on θ. Moreover, by moment expansion Lemma C.3

$$\frac{E\left[m^*_{T,t}(\theta_{T,\delta})\right] - E\left[m^*_{T,t}\left(\theta^0\right)\right]}{\left\|\theta_{T,\delta} - \theta^0\right\|} = J_T \times \frac{\left(\theta_{T,\delta} - \theta^0\right)}{\left\|\theta_{T,\delta} - \theta^0\right\|} \times (1 + o(1)).$$

Further, by construction $||\theta_{T,\delta} - \theta^0|| \geq \delta_{2,T}/2$. Together it follows

$$\sup_{\theta \in U^0(\delta)} \left\{\frac{\|J^*_T(\theta) - J_T\|}{\|J_T\|}\right\} = o_p(1) + o_p\left(\frac{\mathfrak{m}_T(\delta_1, \delta_2)}{\delta_{2,T}\|J_T\|}\right).$$

Therefore $\mathcal{J}_T(\delta_{1,T}, \delta_{2,T}) \xrightarrow{p} 0$ if $\mathfrak{m}_T(\delta_{1,T}, \delta_{2,T})/[\delta_{2,T}||J_T||] = O(1)$. By the definition of a derivative, the construction $U^0(\delta_{1,T}, \delta_{2,T}) \subseteq U^0(0, \delta_{2,T}) = U^0(\delta_{2,T})$ and identification I2,

$$\mathfrak{m}_T(\delta_{1,T}, \delta_{2,T}) \leq K\delta_{2,T} \sup_{\theta \in U^0(\delta_{2,T})} \|J_T(\theta)\| \times (1 + o(1))$$

Now invoke Jacobian smoothness D5.ii to conclude

$$\frac{\mathfrak{m}_T(\delta_1, \delta_2)}{\delta_{2,T}\|J_T\|} \leq \frac{K\delta_{2,T}\|J_T\|(1 + o(1)) + o(1)}{\delta_{2,T}\|J_T\|} = O(1).$$

■

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

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

$$m_{i,t}(\theta^0) = \epsilon_t x_{i,t} \sim (7) \text{ with index } \kappa_{\epsilon,i} := \min\{\kappa_\epsilon, \kappa_i\}, \tag{9}$$

where $x_{1,t} = 1$ implies (10) with $\kappa_{\epsilon,i} = \kappa_\epsilon$. Therefore by the construction of $c_{i,T}$ and k_T, and tail (7),

$$c_{i,T} = K(T/k_T)^{1/\kappa_{\epsilon,i}}. \tag{10}$$

Finally, processes z_t with regularly varying tails (7) and index $\kappa_z \in (0, 2]$ satisfy by Karamata's Theorem (e.g. Feller 1971: §IX.8; Resnick 1987: Theorem 0.6):

$$\kappa_z < 2 : E\left[z_t^2 I(|z_t| \leq c)\right] \sim Kc^2 P(|z_t| > c) \text{ as } c \to \infty. \tag{11}$$

Use (6), and $\Gamma_{i,i,T} > 0$ and $J_{i,i,T} > 0$ for all $T \geq N$ and some $N \in \mathbb{N}$ by D5.i and M2, to express T_{θ_i} as

$$T_{\theta_i} = KT^{1/2}\frac{J_{i,i,T}}{\Gamma_{i,i,T}} \times \left[K + K\left(\frac{\max_{j\neq i}\left\{\Gamma_{j,j,T}^{-1}J_{j,i,T}\right\}}{\Gamma_{i,i,T}^{-1}J_{i,i,T}}\right)^2 \times (1 + O(1))\right]^{1/2}. \tag{12}$$

Step 1 (Σ_T, Γ_T): Properties (10) and (11) imply

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

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

Step 2 (J_T): By Lemma C.4 $J_{i,j,T} = -E[x_{i,t}x_{j,t}I(|\epsilon_t x_{j,t}| \leq c_{j,T})] \times (1 + o(1))$. Assume initially all regressors are stochastic. Since $\kappa_i, \kappa_\epsilon \leq 2$ it follows $\kappa_i < \kappa_\epsilon + 2$. Thus, by independence, (10) and (11)

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

We bound the cross-products $E[x_{i,t}x_{j,t}I(|\epsilon_t x_{j,t}| \leq c_{j,T})]$ by case. If $x_{i,t}$ and $x_{j,t}$ are independent then $E[x_{i,t}x_{j,t}I(|\epsilon_t x_{i,t}| \leq c_{i,T})] \sim K$ given $\sup_{t\in\mathbb{Z}} E|x_{i,t}| < \infty\ \forall i$. If they are perfectly positively dependent then since $x_{i,t}, x_{j,t} \sim$ (7) with indices $\kappa_i, \kappa_j \in (1, 2)$ it can only be the case that $x_{i,t} = sign(x_{j,t}) \times |x_{j,t}|^p$ where $p = \kappa_j/\kappa_i$. But this implies $\kappa_j - p - 1 < \kappa_\epsilon$ hence

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

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

Finally, use $x_{1,t} = 1$ to deduce $E[x_{1,t}^2 I(|\epsilon_t x_{i,t}| \leq c_{i,T})] \sim 1 - k_T/T$, $E[x_{i,t}x_{1,t}I(|\epsilon_t x_{1,t}| \leq c_{i,T})] = E[x_{j,t}I(|\epsilon_t| \leq c_{i,T})] = O(1)$ and $E[x_{1,t}x_{i,t}I(|\epsilon_t x_{i,t}| \leq c_{i,T})] = E(E[x_{i,t}I(|x_{i,t}| \leq c_{i,T}/|\epsilon_t|]|\epsilon_t) = O(1)$. Therefore $J_{1,1,T} = -1 + o(1)$ and $J_{i,1,T}, J_{1,i,T} = O(1) \times (1 + o(1))$.

Step 3 (T_{θ_i}): Consider the slope rates, the intercept rate being similar. The claim follows from (12) by noting

$$\Gamma_{i,i,T}^{-1}J_{i,i,T} = \left(T/k_T\right)^{1/2-1/\kappa_{\epsilon,i}}\left(T/k_T\right)^{(2-\kappa_i)/\kappa_{\epsilon,i}} = \left(T/k_T\right)^{1/2+(1-\kappa_i)/\kappa_{\epsilon,i}}$$

and

$$\max_{j \neq i} \left\{ \Gamma_{j,j,T}^{-1} J_{j,i,T} \right\} = O\left(\max_{j \notin \{1,i\}} \left\{ (T/k_T)^{1/2 - 1/\kappa_{\epsilon,j}} \times (T/k_T)^{(\kappa_i/\kappa_j + 1 - \kappa_i)/\kappa_{\epsilon,i}} \right\} \right)$$
$$= O\left((T/k_T)^{1/2 + 1/\kappa_{\epsilon,i} - \min_{j \notin \{1,i\}} \{1/\kappa_{\epsilon,j} + (1 - 1/\kappa_j)\kappa_i/\kappa_{\epsilon,i}\}} \right)$$

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

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

If $\kappa_i > 2 > \kappa_j$ then by Case 1 and $\kappa_{\epsilon,j} = \kappa_j$ it follows $J_{j,j,T} = K(T/k_T)^{2/\kappa_j - 1}$, $J_{i,j,T} = O((T/k_T)^{1/\kappa_i + 1/\kappa_j - 1})$, and $J_{j,i,T} = O((T/k_T)^{\kappa_{\epsilon,i}^{-1}(\kappa_j/\kappa_i + 1 - \kappa_i)})$. The rate can be deduced from (12) by noting if $\kappa_i > 2$ then

$$\Gamma_{i,i,T}^{-1} J_{i,i,T} = 1 \times K, \quad \left| \max_{j \notin \{1,i\}: \kappa_j > 2} \left\{ \Gamma_{j,j,T}^{-1} J_{j,i,T} \right\} \right| = K$$

$$\max_{j \notin \{1,i\}: \kappa_j \leq 2} \left\{ \Gamma_{j,j,T}^{-1} J_{j,i,T} \right\} = O\left((T/k_T)^{1/\kappa_j - 1/2} \right)$$

hence $T_{\theta_i} \sim T^{1/2} \times [K + O(\max_{j \notin \{1,i\}: \kappa_j \leq 2}\{(T/k_T)^{2/\kappa_j - 1}\})]^{1/2}$.

The rate under $\kappa_i < 2$ follows similarly since $\Gamma_{i,i,T}^{-1} J_{i,i,T} = (T/k_T)^{1/\kappa_i - 1/2}$ and $\max_{j \notin \{1,i\}: \kappa_j \leq 2}\{\Gamma_{j,j,T}^{-1} J_{j,i,T}\} = O(\max_{j \notin \{1,i\}: \kappa_j \leq 2}\{(T/k_T)^{1/2 - 1/\kappa_j + \kappa_{\epsilon,i}^{-1}(\kappa_i/\kappa_j + 1 - \kappa_i)}\})$. ■

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

APPENDIX C : Threshold and Moment Properties

LEMMA C.1 (threshold bound) *Under power-law tail decay D1* $\sup_\theta\{c_T(\theta)/||\Sigma_T(\theta)||^{1/2}\} = o(T^{1/2})$.

LEMMA C.2 (covariance bounds) *Let D1, D3, and M2 hold.*

a. $\limsup_{T \geq N} \sup_\theta ||\Sigma_T^{-1}(\theta) S_T(\theta)|| \leq K$ *and* $\limsup_{T \geq N} \sup_\theta ||S_T^{-1}(\theta) \Sigma_T(\theta)|| \leq K$.

b. $\Sigma_T = o(T)$, $\Sigma_T(\theta) = o(T||E[m_{T,t}^*(\theta)]||^2)$ *and* $\sup_\theta ||\Sigma_T(\theta)|| = o(T \sup_\theta ||\, E[m_{T,t}^*(\theta)]||^2)$.

c. $S_T = o(T)$.

LEMMA C.3 (moment expansion) *Under D5.i* $E[m_{T,t}^*(\theta)] - E[m_{T,t}^*(\tilde{\theta})] = J_T(\tilde{\theta})(\theta - \tilde{\theta}) + o(||J_T(\tilde{\theta})|| \times ||\theta - \tilde{\theta}||)$ *for any* $\theta, \tilde{\theta} \in \Theta$.

LEMMA C.4 (Jacobian approximation) *Under D1-D6* $J_T = E[J_{T,t}^*] \times (1 + o(1))$.

Proof of Lemma C.1. Use the arguments from the proof of Lemma C.2 under power-law decay D1.ii to deduce

$$\sup_\theta \left\{ \frac{\max_{1 \leq i \leq q}\{c_{i,T}(\theta)\}}{\|\Sigma_T(\theta)\|^{1/2}} \right\} \leq K \times \sup_\theta \left\{ \frac{\max_{1 \leq i \leq q}\{c_{i,T}(\theta)\}}{\left(\sum_{i=1}^q c_{i,T}^2(\theta)(k_{i,T}/T)\right)^{1/2}} \right\}$$
$$= O\left(\frac{T^{1/2}}{\min_{1 \leq i \leq q}\{k_{i,T}\}^{1/2}} \right) = o\left(T^{1/2}\right).$$

■

Proof of Lemma C.2. The covariance relationships follow from a partial sum variance bound for tail-trimmed mixing processes. Write $z_{T,t}(\theta, r) := r'T^{-1/2}\Sigma_T^{-1/2}(m^*_{T,t}(\theta) - E[m^*_{T,t}(\theta)])$ for any conformable $r'r = 1$, where $\Sigma_T^{-1/2}$ exists by M2 for sufficiently large T.

Claim (a): Apply partial sum bound Theorem 2.1 in Hill (2010b) to deduce $E[(\sum_{t=1}^T z^2_{T,t}(\theta, r)]$ $\leq K \sum_{t=1}^T E[z^2_{T,t}(\theta, r)] = K$. An identical argument reveals $\sup_\theta E[(\sum_{t=1}^T z^2_{T,t}(\theta, r)] \leq K \sup_\theta \sum_{t=1}^T E[z^2_{T,t}(\theta, r)] = K$, hence $\sup_\theta ||\Sigma_T^{-1}(\theta) S_T(\theta)|| \leq K$. The remaining claim $\sup_\theta ||S_T^{-1}(\theta)\Sigma_T(\theta)|| \leq K$ follows from the construction of $S_T(\theta)$ and non-degeneracy M2.

Claim (b): If $||\Sigma_T|| < \infty$ the claim is trivial, so assume at least one $E[m^2_{i,t}(\theta^0)] = \infty$, and assume without loss of generality $m_{i,t}(\theta)$ is symmetrically trimmed with two-tailed thresholds $c_{i,T}(\theta)$ and fractiles $k_{i,T}$: $(T/k_{i,T})P(|m_{i,t}(\theta)| > c_{i,T}(\theta)) = 1$. Power-law tail D1.ii implies $c_{i,T} = d(\theta^0)^{1/\kappa_i(\theta^0)}(T/k_{i,T})^{1/\kappa_i(\theta^0)}$ for some $\kappa_i(\theta^0) \in (1, 2]$. Coupled with properties of trimmed variances for regularly varying tails if $\kappa_i(\theta^0) \in (1, 2)$ then

$$E\left[\left(m^*_{i,T,t}(\theta^0)\right)^2\right] \sim K c^2_{i,T} P\left(\left|m_{i,t}(\theta^0)\right| > c_{i,T}\right) \sim K c^2_{i,T}(k_{i,T}/T) = K(T/k_{i,T})^{2/\kappa_i(\theta^0)-1}$$

It is easy to show $(T/k_{i,T})^{2/\kappa_i(\theta^0)-1} = o(T)$ for all $\kappa_i(\theta^0) \geq 1$. Similarly if $\kappa_i(\theta^0) = 2$ then $E[(m^*_{i,T,t}(\theta^0))^2] \sim L(T) \to \infty$ a slowly varying function which is trivially $o(T)$. Now invoke the Cauchy-Schwarz inequality to deduce $\Sigma_T = o(T)$.

The uniform case is identical. If $\sup_\theta ||\Sigma_T(\theta)|| < \infty$ the claim is trivial. Otherwise under D1.ii at least one equation tail index $\inf_\theta \kappa_i(\theta) = \inf_{\theta \in \Theta_{2,i}} \kappa_i(\theta) < 2$ by the definition of subset $\Theta_{2,i} \subseteq \Theta$. Therefore $\sup_\theta E[(m^*_{i,T,t}(\theta))^2] \sim K \sup_{\theta \in \Theta_{2,i}} c^2_{i,T}(\theta)(k_{i,T}/T)$ $= K(T/k_{i,T})^{2/\inf_{\theta \in \Theta_{2,i}} \kappa_i(\theta)-1} = o(T)$ if $\inf_{\theta \in \Theta_{2,i}} \kappa_i(\theta) \geq 1$. However, if $\inf_{\theta \in \Theta_{2,i}} \kappa_i(\theta) < 1$ then $\sup_{\theta \in \Theta_{2,i}} |E[m^*_{i,T,t}(\theta)]| \sim \sup_{\theta \in \Theta_{2,i}} c_{i,T}(\theta)(k_{i,T}/T) = K(T/k_{i,T})^{1/\inf_{\theta \in \Theta_{2,i}} \kappa_i(\theta)-1}$. Therefore

$$\frac{\sup_{\theta \in \Theta_{2,i}} E\left[\left(m^*_{i,T,t}(\theta)\right)^2\right]}{\sup_{\theta \in \Theta_{2,i}} \left|E\left[m^*_{i,T,t}(\theta)\right]\right|^2} \sim K(T/k_{i,T}) = o(T),$$

which proves (b).

Claim (c): The final claim follows from (a) and (b). ■

Proof of Lemma C.3. Apply Jacobian existence D5.i and the definition of a derivative.

■

Proof of Lemma C.4. Lemma D.1.b implies for $r_T \to 0$ arbitrarily fast, some $||\theta_* - \theta^0|| \leq ||\theta - \theta^0||$ and tiny $\iota > 0$

$$\frac{E\left[m^*_{T,t}(\theta)\right] - E\left[m^*_{T,t}(\theta^0)\right]}{\left\|\theta - \theta^0\right\|} = E\left[J^*_{T,t}(\theta_*)\right] \times \frac{(\theta - \theta^0)}{\left\|\theta - \theta^0\right\|} + o\left(\frac{r_T \times \left\|\theta - \theta^0\right\|^{1/\iota}}{\left\|\theta - \theta^0\right\|}\right).$$

Further, moment expansion Lemma C.3 asserts

$$\frac{E\left[m^*_{T,t}(\theta)\right] - E\left[m^*_{T,t}(\theta^0)\right]}{\left\|\theta - \theta^0\right\|} = J_T \times \frac{(\theta - \theta^0)}{\left\|\theta - \theta^0\right\|} \times (1 + o(1)).$$

Equate the right-hand-side of each equation and take $||\theta_* - \theta^0|| \leq ||\theta - \theta^0|| \to 0$ to prove the claim. ■

APPENDIX D: Limit Theory for Tail-Trimmed Sums

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

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

$$\hat{m}_T^*(\theta) := \frac{1}{T}\sum_{t=1}^{T}\hat{m}_{T,t}^*(\theta) \text{ and } m_T^*(\theta) := \frac{1}{T}\sum_{t=1}^{T} m_{T,t}^*(\theta).$$

LEMMA D.1 (expansions) *Assume D1-D6 hold. Let* $\check{m}_T^*(\theta) \in \{m_T^*(\theta), \hat{m}_T^*(\theta)\}$, $\check{J}_T^*(\theta)$ $\in \{J_T^*(\theta), \hat{J}_T^*(\theta)\}$ *and* $\check{I}_{T,t}(\theta) \in \{I_{T,t}(\theta), \hat{I}_{T,t}(\theta)\}$, *choose* $\theta, \tilde{\theta} \in \Theta$ *and let* $\{r_T\}$ *be a sequence of strictly positive numbers where* $r_T \to 0$ *arbitrarily fast. In the following* $o(1)$ *and* $o_p(1)$ *are not functions of* $t \in \mathbb{Z}$ *or* $\theta \in \Theta$. *For some sequence* $\{\theta_{T,*}\}$ *satisfying* $||\theta_{T,*} - \tilde{\theta}|| \leq ||\theta - \tilde{\theta}||$ *that may be different in different places, and for infinitessimal* $\iota > 0$*:*

a. $1/T\sum_{t=1}^{T} m_t(\theta)\left\{\check{I}_{T,t}(\theta) - \check{I}_{T,t}(\tilde{\theta})\right\} = o_p(r_T) \times ||\theta - \tilde{\theta}||^{1/\iota}$

$1/T\sum_{t=1}^{T} J_t(\theta)\left\{\check{I}_{T,t}(\theta) - \check{I}_{T,t}(\tilde{\theta})\right\} = o_p(r_T) \times ||\theta - \tilde{\theta}||^{1/\iota}$

b. $\check{m}_T^*(\theta) = \check{m}_T^*(\tilde{\theta}) + \check{J}_T^*(\theta_{T,*})(\theta - \tilde{\theta}) + o_p(r_T) \times ||\theta - \tilde{\theta}||^{1/\iota}$.

LEMMA D.2 (approximation) *Under D1-D4, and D6:*

a. $\left\|\sum_{t=1}^{T}\left\{\hat{m}_{T,t}^*(\theta) - m_{T,t}^*(\theta)\right\}\right\| = o_p\left(T^{1/2}\left\|S_T(\theta)\right\|^{1/2}\right)$ *for any* $\theta \in \Theta$

b. $\sup_\theta\left\{\left\|1/T\sum_{t=1}^{T}\{\hat{m}_{T,t}^*(\theta) - m_{T,t}^*(\theta)\}\right\|\right\} = o_p\left(\sup_\theta\left\|E[m_{T,t}^*(\theta)]\right\|\right)$.

Recall the kernel function $k_{T,s,t}$ *under K1. If additionally I2 and K1 hold then*

c. $\sup_{\theta\in U^0(\delta)}\left\{\|\hat{m}_T^*(\theta) - m_T^*(\theta)\| / \left[1 + \|J_T\| \times \left\|\theta - \theta^0\right\|\right]\right\} = o_p(1)$ *for any* $\delta > 0$.

d. $\left\|S_T^{-1}T^{-1}\sum_{s,t=1}^{T} k_{T,s,t}\left\{\hat{m}_{T,s}^*(\tilde{\theta}_T)\hat{m}_{T,t}^*(\tilde{\theta}_T)' - m_{T,s}^*(\theta^0)m_{T,t}^*(\theta^0)'\right\}\right\| = o_p(1)$ *for any* $\tilde{\theta}_T = \theta^0 + O_p\left(T^{-1/2}\left\|S_T\right\|^{1/2}\left\|J_T\right\|^{-1}\right)$

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

LEMMA D.3 (LLN and ULLN) *Under D1-D4 and I2* $1/T\sum_{t=1}^{T} m_{T,t}^*(\theta^0) = o_p(1)$. *If additionally I3 holds* $\sup_\theta\{||1/T\sum_{t=1}^{T}(m_{T,t}^*(\theta) - E\left[m_{T,t}^*(\theta)\right])||\} = o_p(\sup_\theta E[||m_{T,t}^*(\theta)||])$.

LEMMA D.4 (uniform indicator laws) *Let D1-D4 and D6 hold. Define* $\mathrm{I}_{T,t}^*(\theta) :=$ $((T/k_T)^{1/2})\{I_{i,T,t}(\theta) - E[I_{i,T,t}(\theta)]\}$ *for any* i, *and let* $\{\mathcal{I}(\theta) : \theta \in \Theta\}$ *be a Gaussian process with a* version[17] *that has uniformly bounded and uniformly continuous sample paths with respect to the* L_2*-norm. Then* $\{T^{-1/2}\sum_{t=1}^{T}\mathcal{I}_{T,t}^*(\theta) : \theta \in \Theta\} \Longrightarrow^*$ $\{\mathcal{I}(\theta) : \theta \in \Theta\}$ *and* $E[(\sup_\theta\{|T^{-1/2}\sum_{t=1}^{T}\mathcal{I}_{T,t}^*(\theta)|\})^2] = O(1)$, *and* $\Longrightarrow^*$ *denotes weak convergence in the sense of Hoffmann-Jørgensen (1984).*

[17] Two random variables are *versions* of each other if they have the same finite dimensional distributions.

Remark: Hoffmann-Jørgensen (1984) defines weak convergence for some index function space $\mathcal{Z}$ as

$$\{\mathcal{I}_T(\xi) : \xi \in \mathcal{Z}\} \Longrightarrow^* \{\mathcal{I}(\xi) : \xi \in \mathcal{Z}\}$$

if and only if $\lim_{T\to\infty} E^*[g(\mathcal{I}_T(\xi))] = E^*[g(\mathcal{I}(\xi))]$ for all uniformly bounded g. See Dudley (1978), Pollard (1984), and van der Vaart and Wellner (1994)[18].

LEMMA D.5 (uniform criterion bound) *Under D1-D4 and D6*

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

Asymptotic linearity of $\hat{\theta}_T$ and asymptotic normality of the tail-trimmed equations follow.

LEMMA D.6 (asymptotic linearity) *Under D1-D6, I1-I3 and M1-M2*

$$V_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = A_T \sum_{t=1}^{T} \hat{m}_{T,t}^*(\theta^0) \times (1 + o_p(1)) + o_p(1) \ \ a.s.$$

where $A_T = -V_T^{1/2}(H_T^{-1} J_T' \Upsilon_T)T^{-1} \in \mathbb{R}^{r\times q}$.

LEMMA D.7 (CLT) *Under D1, D3 and I2* $T^{-1/2} S_T^{-1/2} \sum_{t=1}^T m_{T,t}^*(\theta^0) \xrightarrow{d} N(0,1)$.

Stochastic differentiability aids proving Jacobian estimator consistency.

LEMMA D.8 (stochastic differentiability) *Under D1-D6 and M2 for any* $\delta \geq 0$

$$\sup_{\theta\in U^0(\delta)} \left\{ \frac{\left\| \left\{\hat{m}_T^*(\theta) - \hat{m}_T^*(\theta^0)\right\} - \left\{E\left[m_{T,t}^*(\theta)\right] - E\left[m_{T,t}^*(\theta^0)\right]\right\}\right\|}{1 + \|J_T\| \times \left\|\theta - \theta^0\right\|} \right\}$$
$$= \sup_{\theta\in U^0(\delta)} \left\{ \frac{\|J_T^*(\theta) - J_T\|}{\|J_T\|} \right\} + o_p(1).$$

Proof of Lemma D.1. Assume θ and $m_t(\theta)$ are scalars and $m_t(\theta)$ is symmetrically trimmed to simplify notation.

We only expand $m_T^*(\theta)$ since $\hat{m}_T^*(\theta)$ is similar. Write $m_{T,t}^*(\theta) = m_t(\theta) \times I_{T,t}(\theta)$ where $I_{T,t}(\theta) = I(|m_t(\theta)| \leq c_T(\theta))$, and choose $||\theta - \tilde{\theta}|| \leq \delta$ for any $\delta > 0$. Use differentiability D2 to deduce by Taylor's theorem

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

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

$$\begin{aligned} m_T^*(\theta) - m_T^*(\tilde{\theta}) \quad = \quad & J_T^*(\theta_{T,\delta}) \times (\theta - \tilde{\theta}) + \frac{1}{T}\sum_{t=1}^T m_t(\theta) \times \left\{I_{T,t}(\theta) - I_{T,t}(\tilde{\theta})\right\} \qquad (13) \\ & + \frac{1}{T}\sum_{t=1}^T J_t(\theta_{T,\delta}) \times \{I_{T,t}(\theta) - I_{T,t}(\theta_{T,\delta})\} \times (\theta - \tilde{\theta}). \end{aligned}$$

[18] In a landmark paper Dudley (1978) shows it suffices to prove convergence in finite dimensional distributions and the metric entropy with bracketing bound $\int_0^1 \mathcal{H}_{[\,]}^{1/2}(\varepsilon, \mathcal{Z}, \rho)d\varepsilon < \infty$, where ρ is the metric under which the brackets are defined. This justifies D6. See also Ossiander (1987) and Doukhan et al (1995).

We will show the second and third terms are $o_p(r_T) \times ||\theta - \tilde{\theta}||^{1/\iota}$, proving both claims (a) and (b).

Consider the second term in (13) and use $I_{T,t}(\theta) - I_{T,t}(\tilde{\theta}) \in \{-1, 0, 1\}$ to bound

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

The threshold construction (4), $I_{T,t}(\theta) \in \{0, 1\}$ and triangle inequality imply for any $p > 0$

$$\sup_{\theta,\tilde{\theta}\in\Theta} E\left[\left|I_{T,t}(\theta) - I_{T,t}(\tilde{\theta})\right|^p\right] = O(k_T/T)$$

where $O(\cdot)$ is not a function of θ. Combined with D1.i continuity and boundedness of the finite dimensional distributions of $m_t(\theta)$ and the mean-value-theorem, it follows $E|I_{T,t}(\theta) - I_{T,t}(\tilde{\theta})|^p = O((k_T/T)) \times ||\theta - \tilde{\theta}||$. Now invoke stationarity D1.i, envelope bound D4 and the Cauchy-Schwarz inequality to deduce for tiny $\iota > 0$

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

Since $\iota > 0$ can be chosen arbitrarily small and $k_T/T \to 0$ by tail trimming, invoke Markov's inequality to conclude for some $r_T \to 0$ arbitrarily fast and $o_p(\cdot)$ not a function of θ

$$A_T(\theta,\tilde{\theta}) = o_p\left(T^{1/2}\left(k_T^{1/2}/T\right)^{1/\iota}\left\|\theta - \tilde{\theta}\right\|^{1/\iota}\right) = o_p(r_T) \times \left\|\theta - \tilde{\theta}\right\|^{1/\iota}.$$

Since $E|B_T(\theta,\tilde{\theta})| \leq T^{1/2}$ follows trivially from $|I_{T,t}(\theta) - I_{T,t}(\tilde{\theta})| \in \{0, 1\}$ we have shown for some $r_T \to 0$ arbitrarily fast

$$\left|\frac{1}{T}\sum_{t=1}^{T} m_t(\theta)\left\{I_{T,t}(\theta) - I_{T,t}(\tilde{\theta})\right\}\right| \leq A_T(\theta,\tilde{\theta}) \times B_T(\theta,\tilde{\theta}) = o_p(r_T) \times \left\|\theta - \tilde{\theta}\right\|^{1/\iota}.$$

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

■

The proof of approximation Lemma D.2 requires consistency of the intermediate order statistics $m^{(\cdot)}_{i,(k_{j,i,T})}(\theta)$. Simplify notation by considering the two-tailed equations $m^{(a)}_{i,t}(\theta) := |m_{i,t}(\theta)|$, and two-tailed fractiles and thresholds that satisfy $(T/k_{i,T})P(|m_{i,t}(\theta)| > c_{i,T}(\theta)) = 1$. The order statistic $m^{(a)}_{i,(k_{i,T})}(\theta)$ therefore estimates $c_{i,T}(\theta)$.

LEMMA D.2.1 (uniform order statistic law) *Under D1-D4 and D6* $\sup_\theta |m^{(a)}_{i,(k_T)}(\theta)/c_{i,T}(\theta) - 1| = O_p(k_{i,T}^{-1/2})$.

Proof. See Hill and Renault (2010). ■

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

Claim (a): Let $\theta \in \Theta$ be arbitrary, and write $m_t = m_t(\theta)$, $c_T = c_T(\theta)$, $\hat{m}^*_{T,t} = \hat{m}^*_{T,t}(\theta)$, $m^*_{T,t} = m^*_{T,t}(\theta)$, $\bar{I}_{T,t} = 1 - I_{T,t}(\theta)$, $\hat{I}_{T,t} = \hat{I}_{T,t}(\theta)$, and $S_T := S_T(\theta)$. First bound

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

By construction $||m_t\{\hat{I}_{T,t} - I_{T,t}\}|| \leq 2||m^{(a)}_{(k_T)} - c_T||$, where $m^{(a)}_{(k_T)}/c_T = 1 + O_p(k_T^{-1/2})$ given Lemma D.2.1. Now use threshold bound Lemma C.1 and covariance relation Lemma C.2.a to deduce

$$\max_{1 \leq t \leq T} \left\{ \left\| m_t \left\{ \hat{I}_{T,t} - I_{T,t} \right\} \right\| \right\} \leq 2 \left\| m^{(a)}_{(k_T)} - c_T \right\| = 2c_T \left\| m^{(a)}_{(k_T)}/c_T - 1 \right\| = o_p \left(\|S_T\|^{1/2} (T/k_T)^{1/2} \right).$$

Next, by construction and the triangle inequality

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

which is $O_p(k_T^{1/2})$ by the threshold construction (4) and an application of Lemma D.4. Therefore $\sum_{t=1}^{T}\{\hat{m}^*_{T,t} - m^*_{T,t}\} = o_p(||S_T||^{1/2}(T/k_T)^{1/2}k_T^{1/2}) = o_p(||S_T||^{1/2}T^{1/2})$.

Claim (b): Define

$$M^*_T := \max_{1 \leq t \leq T} \left\{ \sup_{\theta} \left\| m_t(\theta) \{ \hat{I}_{T,t}(\theta) - I_{T,t}(\theta) \} \right\| \right\}.$$

and repeat the above argument to reach

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

Uniform indicator law Lemma D.4 and threshold construction (4) imply the right-hand-side is $O_p(M^*_T k_T^{1/2}/T)$.

We need only prove $M^*_T = o_p(\sup_{\theta} ||E[m^*_{T,t}(\theta)]||T/k_T^{1/2})$ to complete the proof. Since

$$\left| m_t(\theta) \left\{ \hat{I}_{T,t}(\theta) - I_{T,t}(\theta) \right\} \right| \leq 2c_T(\theta) \left| m^{(a)}_{(k_T)}(\theta)/c_T(\theta) - 1 \right|,$$

use uniform law Lemma D.2.1, threshold bound Lemma C.1, and covariance bound Lemma C.2.b to deduce

$$\begin{aligned} M^*_T \quad &\leq \quad K \sup_{\theta} c_T(\theta) \sup_{\theta} \left| m^{(a)}_{(k_T)}(\theta)/c_T(\theta) - 1 \right| \leq o \left(\sup_{\theta} \|\Sigma_T(\theta)\|^{1/2} T^{1/2}/k_T^{1/2} \right) \\ &= \quad o \left(\sup_{\theta} \left\| E\left[m^*_{T,t}(\theta)\right] \right\| T/k_T^{1/2} \right). \end{aligned}$$

Claim (c): The claim follows from (b) and Jacobian smoothness $\sup_{\theta \in U^0(\delta)} ||J_T(\theta)||/||J_T||$ $= O(1)$ under D5.ii, since by the definition of a derivative and identification I2

$$\sup_{\theta \in U^0(\delta)} \left\{ \frac{\left\| E\left[m^*_{T,t}(\theta)\right] \right\|}{1 + \|J_T\| \times \left\| \theta - \theta^0 \right\|} \right\} \leq \sup_{\theta \in U^0(\delta)} \left\{ \frac{\|J_T(\theta)\| \times \left\| \theta - \theta^0 \right\|}{1 + \|J_T\| \times \left\| \theta - \theta^0 \right\|} \right\} \leq K.$$

Claim (d): Write $m_t = m_t(\theta^0)$, $\hat{I}_{T,t} = \hat{I}_{T,t}(\theta^0)$, $I_{T,t} = I_{T,t}(\theta^0)$, $\bar{I}_{T,t} := 1 - I_{T,t}$, $\hat{m}^*_{T,t} = m_t\hat{I}_{T,t}$, and $m^*_{T,t} = m_t I_{T,t}$. We prove $||T^{-1}S_T^{-1}\sum_{s,t=1}^T k_{T,s,t}\{\hat{m}^*_{T,s}\hat{m}^*_{T,t} - m^*_{T,s}m^*_{T,t}\}|| = o_p(1)$ and $||T^{-1}S_T^{-1}\sum_{s,t=1}^T k_{T,s,t}\{\hat{m}^*_{T,s}(\tilde{\theta}_T)\hat{m}^*_{T,t}(\tilde{\theta}_T) - \hat{m}^*_{T,s}\hat{m}^*_{T,t}\}|| = o_p(1)$ in two steps. The claim then follows by the triangle inequality.

Step 1: Observe

$$\begin{aligned}
&\left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^T k_{T,s,t}\left\{\hat{m}^*_{T,s}\hat{m}^*_{T,t} - m^*_{T,s}m^*_{T,t}\right\}\right\| \\
&\qquad \leq 2\left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^T k_{T,s,t}m_s\left(\hat{I}_{T,s} - I_{T,s}\right)m^*_{T,t}\right\| \\
&\qquad\qquad + \left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^T k_{T,s,t}m_s\left(\hat{I}_{T,s} - I_{T,s}\right)m_t\left(\hat{I}_{T,t} - I_{T,t}\right)\right\| \\
&\qquad = \mathcal{A}_{1,T} + \mathcal{A}_{2,T}.
\end{aligned}$$

We only bound $\mathcal{A}_{1,T}$ since $\mathcal{A}_{2,T}$ is similar. Define for any $\delta > 0$

$$\eta_\delta(x) := \frac{1}{\left(2\delta^2\pi\right)^{1/2}}\exp\left\{-x^2\delta^{-2}/2\right\} \quad \text{and} \quad \eta_{\delta,T,j} := \eta_\delta(j/\gamma_T)$$

$$\begin{aligned}
\mathcal{A}_{1,T,\delta} := \sum_{t=-T+1}^{2T}&\left(\frac{1}{\gamma_T^{1/2}}\sum_{l=1-t}^{T-t}k(l/\gamma_T)\frac{1}{T^{1/2}}S_T^{-1/2}m_{t+l}\left(\hat{I}^*_{T,t+l} - I^*_{T,t+l}\right)I\left(0 \leq l \leq [\gamma_T/\delta]\right)\right) \\
&\times\left(\frac{1}{\gamma_T^{1/2}}\sum_{j=1-t}^{T-t}\eta_{\delta,T,j}\frac{1}{T^{1/2}}S_T^{-1/2}m^*_{T,t+j}I\left(0 \leq j \leq [\gamma_T/\delta]\right)\right) \times (1 + o_p(1)).
\end{aligned}$$

By CLT Lemma D.7

$$\left\|\frac{1}{T^{1/2}}S_T^{-1/2}\sum_{t=1}^T m^*_{T,t}\right\|_2 = O(1). \tag{14}$$

Similarly, approximation Lemma D.2.a coupled with CLT Lemma D.7 and the Helly-Bray theorem imply

$$\left\|\frac{1}{T^{1/2}}S_T^{-1/2}\sum_{t=1}^T m_t\left(\hat{I}^*_{T,t} - I^*_{T,t}\right)\right\|_2 = o(1). \tag{15}$$

Now imitate Davidson and de Jong's (2000: Lemmas A.2-A.3) arguments to deduce[19]

$$\lim_{\delta\to 0}\limsup_{T\to\infty}\left\|\mathcal{A}_{1,T} - \mathcal{A}_{1,T,\delta} \times (1 + o_p(1))\right\|_1 = 0. \tag{16}$$

[19] Define $X_{T,t} := T^{-1/2}S_T^{-1/2}m^*_{T,t}$. Davidson and de Jong (2000: p. 414) invoke $E(\sum_{t=1}^T X_{T,t})^2 = O(1)$ under their Lemma A.1, which holds by a mixingale property and McLeish's (1975: Theorem 1.6) maximal inequality. But $E(\sum_{t=1}^T X_{T,t})^2 \leq K\sum_{t=1}^T E[X^2_{T,t}] \leq K(E[m^{*2}_{T,t}]/S_T)\sum_{t=1}^T(1/T) \leq K\sum_{t=1}^T(1/T) = K$ by partial sum variance bound Lemma E.1 in Hill and Renault (2010) and variance bound Lemma C.2.a, both without reference to McLeish (1975). The same argument applies to $X^*_{T,t} := T^{-1/2}S_T^{-1/2}m_t\{\hat{I}^*_{T,t} - I^*_{T,t}\}$ since $E(\sum_{t=1}^T X^*_{T,t})^2 = o(1)$ is trivially bounded by $K\sum_{t=1}^T E[X^2_{T,t}] \leq K(E[m^{*2}_{T,t}]/S_T)\sum_{t=1}^T(1/T) \leq K\sum_{t=1}^T(1/T) = K$. Close inspection of Davidson and de Jong's (2000: Lemmas A.2-A.3) proofs reveals $E(\sum_{t=1}^T X_{T,t})^2 \leq K\sum_{t=1}^T(1/T^{1/2})^2$ and $E(\sum_{t=1}^T X^*_{T,t})^2 \leq K\sum_{t=1}^T(1/T^{1/2})^2$ suffice.

Next, consider the components of $\mathcal{A}_{1,T,\delta}$. It is straightforward to generalize approximation Lemma D.2.a to a weighted version with $k(t/\gamma_T)$ under K1. Specifically, define $N_T(\delta) := \min\{T, [\gamma_T/\delta] + 1\}$ use stationarity to deduce for any δ

$$\begin{aligned}
T^{1/2} &\max_{-T+1 \leq t \leq 2T} \left\| \frac{1}{\gamma_T^{1/2}} \frac{1}{T^{1/2}} S_T^{-1/2} \sum_{l=1-t}^{T-t} k\left(l/\gamma_T\right) \left\{\hat{m}_{T,t+l}^* - m_{T,t+l}^*\right\} I\left(0 \leq l \leq [\gamma_T/\delta]\right) \right\|_2 \\
&\leq \left\| \frac{N_T^{1/2}(\delta)}{\gamma_T^{1/2}} S_{N_T(\delta)} S_T^{-1} \right\|^{1/2} \left\| \frac{1}{N_T^{1/2}(\delta)} S_{N_T(\delta)}^{-1/2} \sum_{t=1}^{N_T(\delta)} k\left(t/\gamma_T\right) \left\{\hat{m}_{T,t}^* - m_{T,t}^*\right\} \right\|_2 \\
&\to 0 \text{ as } T \to \infty,
\end{aligned}$$

since $||S_{N_T(\delta)} S_T^{-1}|| = O(1)$ and $N_T(\delta)/\gamma_T = O(1)$ by construction and covariance non-degeneracy M2. Similarly, by a straightforward generalization of Lemma D.7 for any δ

$$\begin{aligned}
T^{1/2} &\max_{-T+1 \leq t \leq 2T} \left\| \frac{1}{\gamma_T^{1/2}} \sum_{j=1-t}^{T-t} \eta_{\delta,T,j} \frac{1}{T^{1/2}} S_T^{-1/2} m_{T,t+j}^* I\left(0 \leq j \leq [\gamma_T/\delta]\right) \right\|_2 \\
&\leq \left\| \frac{N_T^{1/2}(\delta)}{\gamma_T^{1/2}} S_{N_T(\delta)} S_T^{-1} \right\|^{1/2} \left\| \frac{1}{N_T^{1/2}(\delta)} S_{[\gamma_T/\delta]}^{-1/2} \sum_{t=1}^{N_T(\delta)} \eta_{\delta,T,j} m_{T,t}^* \right\|_2 + o\left(1\right) \\
&\to 0 \text{ as } T \to \infty.
\end{aligned}$$

Therefore

$$\lim_{\delta \to 0} \limsup_{T \to \infty} \left\| \mathcal{A}_{1,T,\delta} \right\|_1 = 0. \tag{17}$$

Combine (17) and (18) to conclude $\mathcal{A}_{1,T} = o_p\left(1\right)$.

Step 2: Note

$$\begin{aligned}
&\left\| \frac{1}{T} S_T^{-1} \sum_{s,t=1}^{T} k_{T,s,t} \left\{ \hat{m}_{T,s}^*(\tilde{\theta}_T) \hat{m}_{T,t}^*(\tilde{\theta}_T) - \hat{m}_{T,s}^* \hat{m}_{T,t}^* \right\} \right\| \\
&\qquad \leq 2 \left\| \frac{1}{T} S_T^{-1} \sum_{s,t=1}^{T} k_{T,s,t} \left\{ \hat{m}_{T,s}^*(\tilde{\theta}_T) - \hat{m}_{T,s}^* \right\} \hat{m}_{T,t}^* \right\| \\
&\qquad\qquad + \left\| \frac{1}{T} S_T^{-1} \sum_{s,t=1}^{T} k_{T,s,t} \left\{ \hat{m}_{T,s}^*(\tilde{\theta}_T) - \hat{m}_{T,s}^* \right\} \left\{ \hat{m}_{T,t}^*(\tilde{\theta}_T) - \hat{m}_{T,t}^* \right\} \right\|.
\end{aligned}$$

We will bound the first term, the second is similar. Use the Taylor expansion argument in the proof of expansion Lemma D.1 to deduce for some $||\theta_{T,*} - \theta^0|| \leq ||\tilde{\theta}_T - \theta^0||$

$$\begin{aligned}
&\left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^{T}k_{T,s,t}\left\{\hat{m}_{T,s}^{*}(\tilde{\theta}_T)-\hat{m}_{T,s}^{*}\right\}\hat{m}_{T,t}^{*}\right\| \\
&\quad\leq\left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^{T}k_{T,s,t}\hat{J}_{T,s}(\theta_{T,*})\hat{m}_{T,t}^{*}\right\|\times\left\|\tilde{\theta}_T-\theta^0\right\| \\
&\qquad+\left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^{T}k_{T,s,t}J_s(\theta_{T,*})\left\{\hat{I}_{T,s}\left(\theta_{T,*}\right)-\hat{I}_{T,s}\left(\theta^0\right)\right\}\hat{m}_{T,t}^{*}\right\|\times\left\|\tilde{\theta}_T-\theta^0\right\| \\
&\qquad+\left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^{T}k_{T,s,t}J_s(\theta_{T,*})\left\{\hat{I}_{T,s}\left(\tilde{\theta}_T\right)-\hat{I}_{T,s}\left(\theta^0\right)\right\}\hat{m}_{T,t}^{*}\right\|\times\left\|\tilde{\theta}_T-\theta^0\right\| \\
&\qquad+\left\|\frac{1}{T}S_T^{-1}\sum_{s,t=1}^{T}k_{T,s,t}m_s(\theta^0)\left\{\hat{I}_{T,s}\left(\tilde{\theta}_T\right)-\hat{I}_{T,s}\left(\theta^0\right)\right\}\hat{m}_{T,t}^{*}\right\| \\
&\quad=\sum_{i=1}^{4}\mathcal{B}_{i,T}.
\end{aligned}$$

The gist of Davidson and de Jong's (2000: p. 419-420) Fourier inversion argument applies. Extend their equation (A.51) to our environment to obtain

$$\begin{aligned}
\mathcal{B}_{1,T} &\leq K\int_{-\infty}^{\infty}\left(\|J_T\|^{-1}\left\|\frac{1}{T}\sum_{s=1}^{T}e^{-i\xi s/\gamma_T}\hat{J}_{T,s}(\theta_{T,*})\right\|\times\left\|T^{-1/2}S_T^{-1/2}\sum_{t=1}^{T}e^{i\xi t/\gamma_T}\hat{m}_{T,t}^{*}\right\|\right)\left|\varpi\left(\xi\right)\right|d\xi \\
&= K\int_{-\infty}^{\infty}\mathcal{C}_T\left(\xi\right)\mathcal{D}_T\left(\xi\right)\left|\varpi\left(\xi\right)\right|d\xi,
\end{aligned}$$

where $\varpi(\xi)$ is defined under K1. Approximation Lemma D.2.a and CLT Lemma D.7 render $\mathcal{D}_T\left(\xi\right) = O_p(1)$. Further, Jacobian consistency Lemma 2.5 with $||\theta_{T,*} - \theta^0|| \leq ||\tilde{\theta}_T - \theta^0|| = O_p(T^{-1/2}||S_T||^{1/2} \times ||J_T||^{-1})$, and K1 properties $\sum_{s,t=1}^{T}|k_{T,s,t}| = o(T^2)$, $\max_{1\leq s\leq T}\sum_{t=1}^{T}|k_{T,s,t}| = o(T)$ and $\gamma_T = o(T)$ imply $\mathcal{C}_T\left(\xi\right) = o_p(1)$. Therefore $\int_{-\infty}^{\infty}\mathcal{C}_T\left(\xi\right)\mathcal{D}_T\left(\xi\right)\left|\varpi\left(\xi\right)\right|d\xi = o_p\left(1\right)$ by dominated convergence and K1. Similar arguments extend to the remaining terms by exploiting expansion Lemma D.1.a. ■

Proof of Lemma D.3. The pointwise LLN

$$\frac{1}{T}\sum_{t=1}^{T}m_{T,t}^{*}\left(\theta^0\right)=\frac{1}{T}\sum_{t=1}^{T}\left\{m_{T,t}^{*}\left(\theta^0\right)-E[m_{T,t}^{*}\left(\theta^0\right)]\right\}+E[m_{T,t}^{*}\left(\theta^0\right)]\ =\ o_p(1)$$

follows from identification I2, covariance bound Lemma C.2.c and Chebyshev's inequality.

The ULLN follows from a classic bracketing argument (e.g. Blum 1955, DeHardt 1971, cf. Dudley 1999). Define for any $i \in \{1,...,q\}$

$$h_{T,t}^{*}(\theta):=\frac{m_{i,T,t}^{*}\left(\theta\right)-E\left[m_{i,T,t}^{*}(\theta)\right]}{\sup\limits_{\theta\in\Theta}\left\|E\left[m_{T,t}^{*}(\theta)\right]\right\|}.$$

Use the Lemma C.2.a,b to deduce the covariance bound $S_T\left(\theta\right) = o(T\sup_{\theta\in\Theta}||E\left[m_{T,t}^{*}(\theta)\right]||^2)$, hence $1/T\sum_{t=1}^{T}\{h_{T,t}^{*}(\theta) - E\left[h_{T,t}^{*}(\theta)\right]\} = o_p(1)$ by Chebyshev's inequality. Further,

$h^*_{T,t}(\theta)$ is uniformly L_1-bounded so it belongs to a separable Banach space (Royden 1988). Therefore the L_1-bracketing numbers satisfy $N_{[\,]}(\varepsilon, \Theta, ||\cdot||_1) < \infty$ (Dudley 1999: Proposition 7.1.7). Together, the pointwise LLN and bracketing numbers $N_{[\,]}(\varepsilon, \Theta, ||\cdot||_1) < \infty$ deliver $\sup_\theta |1/T \sum_{t=1}^T (h^*_{T,t}(\theta) - E[h^*_{T,t}(\theta)])| = o_p(1)$. See Theorem 7.1.5 of Dudley (1999), cf. Blum (1955) and DeHardt (1971). ■

Proof of Lemma D.4. Threshold construction (4) and D3 imply $\mathcal{I}^*_{T,t}(\theta)$ is L_2-bounded uniformly on $1 \le t \le T$, $T \ge 1$, and Θ, and geometrically β-mixing. Coupled with metric entropy with L_2-bracketing D6 we may extend Doukhan et al's (1995: Theorem 1; eq. (2.17)) uniform central limit theorem to triangular arrays $\{\mathcal{I}^*_{T,t}(\theta)\}$. See especially Application 4 in Doukhan et al (1995). Therefore $\{1/T^{1/2} \sum_{t=1}^T \mathcal{I}^*_{T,t}(\theta) : \theta \in \Theta\} \Longrightarrow \{\mathcal{I}(\theta) : \theta \in \Theta\}$, a Gaussian process with a version that has uniformly bounded and uniformly continuous sample paths with respect to $||\cdot||_2$.

Further, the conditions for Doukhan et al's (1995: Theorem 2) uniform maximal inequality are satisfied since their required bound (2.10) holds under their (2.17), which D6 ensures. Therefore $E[(\sup_\theta\{|T^{-1/2}\sum_{t=1}^T \mathcal{I}^*_{T,t}(\theta)|\})^2] = O(1)$ where $O(1)$ which completes the proof. ■

Proof of Lemma D.5. Write $\hat{m}^*_T(\theta) := 1/T \sum_{t=1}^T \hat{m}^*_{T,t}(\theta)$ and $m^*_T(\theta) := 1/T \sum_{t=1}^T m^*_{T,t}(\theta)$. By weight property M1 and the triangle inequality

$$\begin{aligned}
\mathfrak{m}_T^{-2}\left|\hat{Q}_T(\theta) - Q_T(\theta)\right| &\le \mathfrak{m}_T^{-2}\|\hat{m}^*_T(\theta)\|^2 \times \left\|\hat{\Upsilon}_T - \Upsilon_T\right\| + \mathfrak{m}_T^{-2}\left|\hat{m}^*_T(\theta)'\Upsilon_T\hat{m}^*_T(\theta) - Q_T(\theta)\right| \\
&\le \left\{\|\hat{m}^*_T(\theta)\|^2 \times o_p\left(\mathfrak{m}_T^{-2}\right)\right\} + \left\{K\|\hat{m}^*_T(\theta) - m^*_T(\theta)\|^2\right\} \\
&\quad + \left\{K\mathfrak{m}_T^{-2}\|m^*_T(\theta)\| \times \|\hat{m}^*_T(\theta) - m^*_T(\theta)\|\right\} + \left\{\mathfrak{m}_T^{-2}\left|m^*_T(\theta)'\Upsilon_T m^*_T(\theta) - Q_T(\theta)\right|\right\} \\
&= A_{1,T}(\theta) + A_{2,T}(\theta) + A_{3,T}(\theta) + A_{4,T}(\theta).
\end{aligned}$$

Uniform approximation Lemma D.2.b and $\liminf_{T\ge N} \mathfrak{m}_T > 0$ under smoothness I3 imply

$$\begin{aligned}
\sup_\theta \{A_{1,T}(\theta)\}^{1/2} &= \sup_\theta \left\{\frac{\|\hat{m}^*_T(\theta)\|}{\sup_{\theta\in\Theta}\left\|E\left[m^*_{T,t}(\theta)\right]\right\|}\right\} \times o_p(1) \\
&\le \frac{\sup_\theta \|m^*_T(\theta)\|}{\sup_{\theta\in\Theta}\left\|E\left[m^*_{T,t}(\theta)\right]\right\|} \times o_p(1) + o_p(1) \\
&\le \frac{\sup_\theta \|m^*_T(\theta) - E[m^*_t(\theta)]\|}{\sup_{\theta\in\Theta}\left\|E\left[m^*_{T,t}(\theta)\right]\right\|} \times o_p(1) + o_p(1).
\end{aligned}$$

Now apply the ULLN Lemma D.3 to deduce $\sup_\theta\{A_{1,T}(\theta)\} = o_p(1)$. Similar arguments based on approximation Lemma D.2.b reveal $\sup_\theta\{A_{2,T}(\theta)\}$ and $\sup_\theta\{A_{3,T}(\theta)\}$ are $o_p(1)$.

Finally, under M1

$$\begin{aligned}
A_{4,T}(\theta) &= \mathfrak{m}_T^{-2}\left|m^*_T(\theta)'\Upsilon_T m^*_T(\theta) - E[m^*_t(\theta)]'\,\Upsilon_T E[m^*_t(\theta)']\right| \\
&\le K\mathfrak{m}_T^{-2}\|m^*_T(\theta) - E[m^*_t(\theta)]\|^2 + K\mathfrak{m}_T^{-2}\|E[m^*_t(\theta)]\| \times \|m^*_T(\theta) - E[m^*_t(\theta)]\|.
\end{aligned}$$

Lemma D.3 implies each term is $o_p(1)$ uniformly on Θ. ■

Proof of Lemma D.6. Apply Čižek's (2008: Lemma 2.1) argument to deduce absolute continuity of the equation distributions D1.i and equation differentiability D2 ensures $\hat{Q}_T(\theta)$ is continuous and differentiable at $\hat{\theta}_T$ *a.s.* Now use $\hat{Q}_T(\hat{\theta}_T) \leq \hat{Q}_T(\theta)$ $\forall \theta \in \Theta$ to deduce by Čižek's (2008: Lemma 2.1) argument

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

The Lemma D.1.b asymptotic expansion for $1/T\sum_{t=1}^{T}\hat{m}_{T,t}^*(\hat{\theta}_T)$ implies we may write

$$\hat{J}_T^*(\hat{\theta}_T)'\hat{\Upsilon}_T\left\{\hat{J}_T^*(\theta_{T,*})'\left(\hat{\theta}_T-\theta^0\right)+\frac{1}{T}\sum_{t=1}^{T}\hat{m}_{T,t}^*(\theta^0)\right\}+\hat{J}_T^*(\hat{\theta}_T)'\hat{\Upsilon}_T\times o_p\left(r_T\times\left\|\hat{\theta}_T-\theta^0\right\|^{1/\iota}\right)=0 \ a.s.$$

for some $||\theta_{T,*}-\theta^0|| \leq ||\hat{\theta}_T-\theta^0||$, $r_T \rightarrow 0$ arbitrarily fast and tiny $\iota > 0$.

Consistency $||\hat{\theta}_T - \theta^0|| \xrightarrow{p} 0$ under Theorem 2.1 and Jacobian consistency Lemma 2.5 ensure both $\hat{J}_T^*(\hat{\theta}_T) = J_T(1 + o_p(1))$ and $\hat{J}_T^*(\theta_{T,*}) = J_T(1 + o_p(1))$. Further $H_T^{-1} := (J_T'\Upsilon_T J_T)^{-1}$ exists given weight and Jacobian properties M1 and D5.i. Re-arrange terms and exploit the construction of V_T and $r_T \rightarrow 0$ arbitrarily fast to deduce

$$\begin{aligned}
V_T^{1/2}\left(\hat{\theta}_T-\theta^0\right) &= -\left\{V_T^{1/2}H_T^{-1}J_T'\Upsilon_T\right\}\frac{1}{T}\sum_{t=1}^{T}\hat{m}_{T,t}^*(\theta^0)\times\left(1+o_p\left(1\right)\right)+o_p\left(1\right) \\
&= A_T\sum_{t=1}^{T}\hat{m}_{T,t}^*(\theta^0)\times\left(1+o_p\left(1\right)\right)+o_p\left(1\right).
\end{aligned}$$

■

Proof of Lemma D.7. Define $z_{T,t}^*(r) = r'T^{-1/2}S_T^{-1/2}(\theta^0)m_{T,t}^*(\theta^0)$ for arbitrary $r \in \mathbb{R}^s$, $r'r = 1$, and observe identification I2 implies both $E[z_{T,t}^*(r)] = o(1)$ and $E[z_{T,t}^{2*}(r)]$ $= 1 + o(1)$. The proof that $\sum_{t=1}^{T} z_{T,t}^*(r) \xrightarrow{d} N(0, 1)$ follows from a standard martingale difference decomposition under geometric β-mixing D3. See Hill (2010c: Lemma 3.3). ■

Proof of Lemma D.8. Apply Minkowski's inequality and the Lemma D.2.c uniform approximation to obtain

$$\begin{aligned}
&\sup_{\theta\in U^0(\delta)}\left\{\frac{\left\|\left\{\hat{m}_T^*(\theta)-\hat{m}_T^*(\theta^0)\right\}-\left\{E\left[m_{T,t}^*(\theta)\right]-E\left[m_{T,t}^*(\theta^0)\right]\right\}\right\|}{1+\|J_T\|\times\left\|\theta-\theta^0\right\|}\right\} \\
&\quad\leq \sup_{\theta\in U^0(\delta)}\left\{\frac{\left\|\left\{m_T^*(\theta)-m_T^*(\theta^0)\right\}-\left\{E\left[m_{T,t}^*(\theta)\right]-E\left[m_{T,t}^*(\theta^0)\right]\right\}\right\|}{1+\|J_T\|\times\left\|\theta-\theta^0\right\|}\right\} \\
&\qquad+2\sup_{\theta\in U^0(\delta)}\left\{\frac{\left\|\hat{m}_T^*(\theta)-m_T^*(\theta)\right\|}{1+\|J_T\|\times\left\|\theta-\theta^0\right\|}\right\} \\
&\quad= \sup_{\theta\in U^0(\delta)}\left\{\frac{\left\|\left\{m_T^*(\theta)-m_T^*(\theta^0)\right\}-\left\{E\left[m_{T,t}^*(\theta)\right]-E\left[m_{T,t}^*(\theta^0)\right]\right\}\right\|}{1+\|J_T\|\times\left\|\theta-\theta^0\right\|}\right\}+o_p\left(1\right)
\end{aligned}$$

Moment expansion Lemma C.3 and equation expansion Lemma D.1.b imply the last line is bounded by $\sup_{\theta\in U^0(\delta)}\{||J_T^*(\theta) - J_T||/||J_T||\} + o_p(1)$. ■

REFERENCES

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

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

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

Andrews, B. (2008). Rank-Based Estimation for Autoregressive Moving Average Time Series Models, Journal of Time Series Analysis 29, 51-73.

Antoine, B. and E. Renault (2009). Efficient GMM with Nearly-Weak Instruments, Econometrics Journal (Tenth Anniversary Issue) 12, S135-S171.

Antoine, B. and E. Renault (2010). Efficient Minimum Distance Estimation with Multiple Rates of Convergence, Journal of Econometrics: forthcoming.

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

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

Berkes, I., L. Horváth and J. Schauer (2011). Asymptotics of Trimmed CUSUM Statistics, Bernoulli: in press.

Blum, J.R. (1955). On Convergence of Empirical Distribution Functions, Annals of Mathematical Statistics 26, 527-529.

Bonnal, H. and E. Renault (2004). On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood, Working Paper, CIREQ, Université de Montréal.

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.

Bougerol, P. and N. Picard (1992). Stationarity of GARCH Processes and of Some Nonnegative Time Series, Journal of Econometics 52, 115–127.

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.

Campbell, J. and L. Hentschel (1992). No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns, Journal of Fi6ancial Economics 31, 281-313.

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

Chernozhukov, V. (2010). Inference for Extremal Conditional Quantile Models, with an Application to Birthweights, Review of Economic Studies, forthcoming.

Čižek, P. (2005). Least Trimmed Squares in Nonlinear Regression Under Dependence, Journal of Statistical Planning and Inference 136, 3967–3988.

Čižek, P. (2008). General Trimmed Estimation: Robust Approach to Nonlinear and Limited Dependent Variable Models, Econometric Theory 24 , 1500-1529.

Čižek, P. (2009). Generalized Method of Trimmed Moments, Working Paper, CentER, University of Tilburg.

Cline, D.B.H. (1986). Convolution Tails, Product Tails and Domains of Attraction, Probability Theory and Related Fields 72, 529–557.

Cline, D.B.H. (1989). Consistency for Least Squares Regression Estimators with Infinite Variance Data, Journal of Statistical Planning and Inference 23, 163-179.

Cline, D.B.H. (2007). Regular Variation of Order 1 Nonlinear AR-ARCH models, Stochastic Processes and Theiry Applications 117, 840-861.

Davidson, J. (1994). Stochastic Limit Theory. Oxford University Press: Oxford.

Davidson, J. and R. de Jong (2000). Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices, Econometrica 68, 407-423.

Davis, R.A., K. Knight and J. Liu (1992). M-Estimation for Autoregressions with Infinite Variance, Stochastic Processes and Their Applications 40, 145-180.

Davis, R.A. and T. Mikosch (2008). Extreme Value Theory for GARCH Models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (eds.): Handbook of Financial Time Series, 186–200. Springer: New York.

Davis, R.A. and S.I. Resnick (1986). Limit Theory for Sample for the Sample Covariance and Correlation Functions of Moving Averages, Annals of Statistics 14, 533-558.

Davis, R., and S. Resnick (1996). Limit Theory for Bilinear Processes with Heavy-Tailed Noise, Annals of Applied Probability 6, 1191–1210.

de Haan, L., S.I. Resnick, H. Rootzén, and C.G. de Vries (1989). Extremal Behaviour of Solutions to a Stochastic Difference Equation with Applications to ARCH Processes, Stochastic Processes and their Applications 32, 213–224.

Dell'Aquila, R., E. Ronchetti, F. Trojani (2003). Robust GMM Analysis of Models for the Short Tate Process, Journal of Empirical Finance 10, 373-397.

DeHardt, J. (1971). Generalizations of the Glivenko-Cantelli Theorem Generalizations of the Glivenko-Cantelli Theorem, Annals of Mathematical Statistics 42, 2050-2055.

Ding, Z. and C.W.J. Granger (1996). Modeling Volatility Persistence of Speculative Returns: A New Approach, Journal of Econometrics 73, 185-215.

Doukhan, P. (1994). Mixing: Properties and Examples, Lecture Notes in Statistics 85. Springer: Berlin.

Doukhan, P., P. Massart, and E. Rio (1995). Invariance Principles for Absolutely Regular Empirical Processes, Annales de l'Institut Henri Poincaré 31, 393-427.

Dudley, R. M. (1978). Central Limit Theorem for Empirical Processes. Annals of Probaility 6, 899-929.

Dudley, R. M. (1999). Continuous Central Limit Theorems. Cambridge University Press: Cambridge .

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag: Frankfurt. Engle, R. amd V. Ng (1993). Measuring and Testing the Impact of News On Volatility, Journal of Finance 48, 1749-1778.

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.

Feller, W. (1971). An Introduction to Probability Theory and its Applications, 2 ed., Vol. 2. Wiley: New York.

Finkenstadt, B., and H. Rootzén (2003). Extreme Values in Finance, Telecommunications and the Environment. Chapman and Hall: New York.

Francq, C. and J.M. Zakoïan (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes, Bernoulli 10, 605–637.

Franc, C. And J-M. Zakoïan (2010). GARCH Models: Structure, Statistical Inference and Financial Applications, Wiley: New York.

Giné, E. and J. Zinn (1984). Some Limit Theorems for Empirical Processes, Annals of Probability 12, 929-989.

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.

Hall, A. (2005). Generalized Method of Moments. Oxford University Press: Oxford.

Hall, P. (1982). On Some Estimates of an Exponent of Regular Variation, Journal of the Royal Statistical Society Series B 44, 37-42.

Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH Models with Heavy-Tailed Errors, Econometrica 71, 285-317.

Hannan, E.J. and M. Kanter (1977). Autoregressive Processes with Infinite Variance, Journal of Applied Probability 14, 411-415.

Hansen, L.P. (1982). Large Sample Properties of Method of Moments Estimators, Econometrica 50, 1029-1054.

Hansen, B. and S.W. Lee (1994). Asymptotic Theory for the Garch (1,1) Quasi-Maximum Likelihood Estimator, Econometric Theory 10, 29-52.

He, X., J. Jurecková, R. Koenker and S. Portnoy (1990). Tail Behavior of Regression Estimators and their Breakdown Points, Econometrica 58, 1195-1214.

Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution, Annals of Mathematical Statistics 3, 1163-1174.

Hill, J.B. (2010a). On Tail Index Estimation for Dependent Heterogeneous Data, Economeric Theory 26, 1398-1436.

Hill, J.B. (2010b). A New Moment Bound and Weak Laws for Mixingale Arrays without Memory or Heterogeneity Restrictions, with Applications to Tail Trimmed Arrays, University of North Carolina - Chapel Hill.

Hill, J.B. (2010c). Central Limit Theory for Kernel-Self Normalized Tail-Trimmed Sums of Dependent, Heterogeneous Data with Applications, University of North Carolina - Chapel Hill.

Hill, J.B. and E. Renault (2010). Generalized Method of Moments with Tail Trimming : Appendix E : Omitted Proofs and Simulations, University of North Carolina - Chapel Hill; www.unc.edu/∼jbhill/gmttm_tech_append.pdf.

Hill, J.B. and A. Shneyerov (2009). Are There Common Values on BC Timber Sales? A Tail-Index Nonparametric Test, under revision for Journal of Econometrics.

Hoffmann-Jørgensen, J. (1984). Convergence of Stochastic Processes on Polish Spaces. Unpublished manuscript.

Honoré, B. (1992). Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects, Econometrica 60, 533-565.

Horowitz, J. L. (1998). Bootstrap Methods for Covariance Structures, Journal of Human Resources 33, 39-61.

Huber, P.J. (1964). Robust Estimation of Location Parameter, Annals of Mathematical Statistics 35 73-101.

Huisman, R., K. Koedijk, and C. Kool (2001). Tail-Index Estimates in Small Samples, Journal of Business and Economic Statistics 19, 208-216.

Ibragimov, I. A. and Y. V. Linnik (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhof: Berlin.

Jensen, S.T. and A. Rahbek (2004). Asymptotic Normality of the QML Estimator of ARCH in the Nonstationary Case, Econometrica 72, 641-646.

Jurecková, J. and P.K. Sen (1996). Robust Statistical Procedures: Asymptotics and Interrelations. Wiley: New York.

Khan, S., and A. Lewbel (2007). Weighted and Two-Stage Least Squares Estimations of Semiparamertic Truncated Regression Models, Econometric Theory 23, 309-347.

Kitamura, Y, T. Gautam, and A. Hyungtaik (2004). Empirical Likelihood-Based Inference in Conditional Moment Restriction Models, Econometrica 72, 1667-1714.

Knight, K. (1987). Rate of Convergence of Centred Estimates of Autoregressive Parameters for Infinite Variance Autoregressions, Journal of Time Series Analysis 8, 51-60.

Koul, H.L., and A.K. Md. E. Saleh (1995). Autoregression Quantiles and Related Rank-Scores Processes, Annals of Statistics 23, 670-689.

Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag: New York.

Ling, S. (2005). Self-Weighted LAD Estimation for Infinite Variance Autoregressive Models, Journal of the Royal Statistical Society Series B 67, 381–393.

Ling, S. (2007). Self-Weighted and Local Quasi-Maximum Likelihood Estimators for ARMA-GARCH/IGARCH Models, Journal of Econometrics 150, 849-873.

Linton, O., J. Pan and H. Wang (2010). Estimation for a Non-stationary Semi-Strong GARCH(1,1) Model with Heavy-Tailed Errors, Econometric Theory 26, 1-28.

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.

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.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices, Journal of Business 36, 394-419.

McFadden D.L. (1989). A Method of Simulated Moments for Estimation of Multinomial Discrete Response Models Without Numerical Integration, Econometrica 57, 995-1026.

McLeish, D.L. (1975). Maximal Inequality and Dependent Strong Laws, Annals of Probability 3, 829-839.

Meddahi, N. and E. Renault (2004). Temporal Aggregation of Volatility Models, Journal of Econometrics 119, 355-379.

Meitz M. and P. Saikkonen (2008). Stability of Nonlinear AR-GARCH Models, Journal of Time Series Analysis 29, 453-475.

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.

Mikosch, T and D. Straumann (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach, Annals of Statistics 34, 2449-2495.

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.

Ossiander, M. (1987). A Central Limit Theorem Under Metric Entropy with L2 Bracketing, Annals of Probability 15, 897-919.

Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators, Econometrica 57, 1027-1057.

Pan, J., H. Wang, and Q. Yao (2007). Weighted Least Absolute Deviations Estimation for ARMA Models with Infinite Variance, Econometric Theory 23, 852-879.

Peng, L. and Q. Yao (2003). Least Absolute Deviations Estimation for ARCH and GARCH Models, Biometrika 90, 967-975.

Pollard, D. (1984). Convergence of Stochastic Processes. Springer-Verlang New York.

Pollard, D. (1989). Bracketing Methods in Statistics and Econometrics, in Proceedings of the Conference on Nonparametric and Semiparametric Methods in Econometrics and Statistics by W.A. Barnett, J. Powell and G. Tauchen (eds.), Cambridge University Press: Cambridge.

Pollard, D. (2002). Maximal Inequalities via Bracketing with Adaptive Truncation, Annales de l'Institut Henri Poincaré 38, 1039-1052.

Powell, J.L. (1986). Symmetrically Trimmed Least Squares Estimation for Tobit Models, Econometrica 54, 1435-1460.

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.

Resnick, S.I. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag: New York.

Resnick, S.I. (1997). Heavy Tail Modeling and Traffic Data, Annals of Statistics 25, 1805-1849.

Ronchetti, E. and F. Trojani (2001a). Robust Inference with GMM Estimators, Journal of Econometrics 101, 37-69.

Ronchetti, E. and F. Trojani (2001b). Robust Inference for Generalized Linear Estimators, Journal of the American Statistical Association 96, 1022-1030.

Royden, H.L. (1988). Real Analysis, Prentice Hall: New Jersey.

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.

Ruppert, D. and J. Carroll (1980). Trimmed Least Squares Estimation in the Linear Model, Journal of the American Statistical Association 75, 828-838.

Stock, J.H. and J.H. Wright (2000). GMM with Weak Identification, Econometrica 68, 1055-1096.

Tableman, M. (1994). The Asymptotics of the Least Trimmed Absolute Deviations (LTAD) Estimator, Statistics and Probability Letters 19, 329-337.

van der Vaart, A.W., and J.A. Wellner (1994). Weak Convergence and Empirical Processes, Springer-Verlang: New York.

TABLE 4 : LOCATION, AR, $k_T \sim T^{\lambda}$

LOCATION						**AR**				
$\epsilon_t \sim P_{2.5}$, LS[a], $\kappa_y = 2.5$[b], $\theta_r^0 = 1$[c]						$\epsilon_t \sim P_{2.5}$, LS, $\kappa_y = 2.5$, $\theta_r^0 = .9$				
$\hat{\theta}$	Mean[d]	KS[e]	λ[f]	k_T		$\hat{\theta}$	Mean	KS	λ	k_T
GMTTM[g]	1.01 (.33)	.083	.44	21		GMTTM	.900 (.01)	.042	.48	28
GMM	1.00 (.41)	.111	-	-		GMM	.899 (.01)	.097	-	-
OLS	.999 (.04)	.121	-	-		OLS	.899 (.01)	.132	-	-
LAWD[h]	.996 (.01)	.064	-	-		LAWD	.900 (.01)	.046	-	-
$\epsilon_t \sim P_{1.5}$, LS, $\kappa_y = 1.5$, $\theta_r^0 = 1$						$\epsilon_t \sim P_{1.5}$, LS, $\kappa_y = 1.5$, $\theta_r^0 = .9$				
$\hat{\theta}$	Mean	KS	λ	k_T		$\hat{\theta}$	Mean	KS	λ	k_T
GMTTM	1.01 (.31)	.072	.32	9		GMTTM	.900 (.00)	.066	.40	16
GMM	1.07 (.46)	.145	-	-		GMM	.898 (.02)	.162	-	-
OLS	1.01 (.15)	.272	-	-		OLS	.889 (.01)	.253	-	-
LAWD	1.00 (.02)	.052	-	-		LAWD	.900 (.00)	.061	-	-

a. LS = least squares type estimating equations for GMM and GMTTM.

b. True parameter value for the r$^{\text{th}}$ element θ_r^0.

c. Error distribution ($P_{1.5}$ or $N_{0,1}$), and moment supremum or tail index of y_t.

d. Simulation mean of parameter estimation (square root of mse in parentheses).

e. Kolmogorov-Smirnov test p-value. In the case of GMTTM, KS p-values are evaluated at that $k_T = [T^{\lambda}]$ which minimizes KS. 1%, 5%, 10% critical values: .136, .122, .107.

f. KS minimizing λ in the trimming fractile $k_T = [T^{\lambda}]$, $\lambda \in \{.01, .02, ..., .99\}$.

g. GMTTM and GMM are computed in two steps with efficient weight and QMLE plug-in, under *exact* identification. See Hill and Renault (2010) for GMTTM and GMM simulations allowing over-identifcation.

h. LAWD = Least Absolute Weighted Deviations.

TABLE 5 : ARCH, GARCH, $k_{j,T} \sim T^{\lambda_j}$

ARCH					GARCH				
$\epsilon_t \sim N_{0,1}$, LS, $\kappa_y = 4.65$, $\theta_r^0 = .5$					$\epsilon_t \sim N_{0,1}$, LS, $\kappa_y = 4.1$, $\theta_r^0 = .6$				
$\hat{\theta}$	Mean	KS	λ_1, λ_2[a]	k_1, k_2[b]	$\hat{\theta}$	Mean	KS	λ_1, λ_2	k_1, k_2
GMTTM[c]	.505 (.10)	.073	.36,.07	12,2	GMTTM	.594 (.19)	.070	.07,.05	2,1
GMM	.421 (.14)	.327	-	-	GMM	.615 (.24)	.145	-	-
QML[c]	.492 (.07)	.051	-	-	QML	.597 (.08)	.098	-	-
QMWL[d]	.496 (.07)	.095	-	-	QMWL	.596 (.08)	.096	-	-
$\epsilon_t \sim N_{0,1}$, QML[d], $\kappa_y = 3.8$, $\theta_r^0 = .6$					$\epsilon_t \sim N_{0,1}$, QML, $\kappa_y = 2.0$, $\theta_r^0 = .6$				
$\hat{\theta}$	Mean	KS	λ_1, λ_2	k_1, k_2	$\hat{\theta}$	Mean	KS	λ_1, λ_2	k_1, k_2
GMTTM	.602 (.14)	.083	.41,.17	17,4	GMTTM	.608 (.18)	.064	.47,.26	26,6
GMM	.506 (.19)	.248	-	-	GMM	.523 (.18)	.246	-	-
QML	.599 (.07)	.082			QML	.586 (.20)	.262	-	-
QMWL	.600 (.06)	.079	-	-	QMWL	.602 (.09)	.095	-	-
$\epsilon_t \sim P_{2.5}$, QML, $\kappa_y = 1.8$, $\theta_r^0 = .6$					$\epsilon_t \sim P_{2.5}$, QML, $\kappa_y = 1.5$, $\theta_r^0 = .6$				
$\hat{\theta}$	Mean	KS	λ_1, λ_2	k_1, k_2	$\hat{\theta}$	Mean	KS	λ_1, λ_2	k_1, k_2
GMTTM	.604 (.24)	.087	.48,.16	28,4	GMTTM	.598 (.20)	.063	.49,.16	30,3
GMM	.571(.38)	.175	-	-	GMM	.543 (.22)	.195	-	-
QML	.620 (.27)	.110	-	-	QML	.605 (.18)	.569	-	-
QMWL	.598 (.22)	.108	-	-	QMWL	.569 (.23)	.117	-	-

a. KS minimizing pair $\{\lambda_1, \lambda_2\}$. b. $k_j = k_{j,T} = [T^{\lambda_j}]$.
c. GMTTM and GMM are computed under *exact* identification.
d. QMWL = Quasi-Maximum Weighted Likelihood.
e. QML-type estimating equations for GMM and GMTTM.

TABLE 6 : TARCH, QARCH, $k_T \sim T^{\lambda}$

TARCH					QARCH				
$\epsilon_t \sim N_{0,1}$, LS, $\kappa_y = 5.25$, $\theta_r^0 = .6$					$\epsilon_t \sim N_{0,1}$, QML, $\kappa_y = 3.5$, $\theta_r^0 = .8$				
$\hat{\theta}$	Mean	KS	λ_1, λ_2	k_1, k_2	$\hat{\theta}$	Mean	KS	$\{\lambda_1, \lambda_2\}, \lambda$[b]	k_1, k_2
GMTTM[a]	.61 (.16)	.046	.19,.04	4,1	GMTTM	.792 (.44)	.064	{.11,.05},.05	{2,1},1
GMM	.557 (.15)	.203	-	-	GMM	.863 (.53)	.092	-	-
QML	.595 (.09)	.073	-	-	QML	.896 (.66)	.389	-	-
QMWL	.606 (.10)	.063	-	-	QMWL	.894 (.53)	.144	-	-
$\epsilon_t \sim P_{2.5}$, QML, $\kappa_y = 2.6$, $\theta_r^0 = .6$					$\epsilon_t \sim N_{0,1}$, QML, $\kappa_y = 2.0$, $\theta_r^0 = 1$				
$\hat{\theta}$	Mean	KS	λ_1, λ_2	k_1, k_2	$\hat{\theta}$	Mean	KS	$\{\lambda_1, \lambda_2\}, \lambda$	k_1, k_2
GMTTM	.602 (.22)	.083	.26,.11	6,2	GMTTM	1.08 (.46)	.079	{.17,.13},.05	{3,2},1
GMM	.526 (.24)	.224	-	-	GMM	1.16 (.54)	.177	-	-
QML	.516 (.38)	.323	-	-	QML	.997 (.64)	.287	-	-
QMWL	.676 (.27)	.239	-	-	QMWL	1.11 (.61)	.186	-	-

a. GMTTM and GMM are computed under *exact* identification.
b. The first QARCH equation is asymmetrically trimmed with $k_{1,j,T} = [T^{\lambda_j}]$, and the second equation is symmetrically trimmed with $k_T = [T^{\lambda}]$.

TABLE 7 : GMTTM, $k_T \sim \delta \ln(T)$, $\delta T/\ln(T)$

Model	ϵ_t	κ_y	θ_r^0	Mean	KS	δ	k_T
LOCATION	$P_{2.5}$	2.5	1	.996 (.32)	.088	.15[a]	22
LOCATION	$P_{1.5}$	1.5	1	.995 (.13)	.994	.06	9
AR	$P_{2.5}$	2.5	.9	.898 (.01)	.053	.20[b]	29
AR	$P_{1.5}$	1.5	.9	.901 (.01)	.089	.12	17
ARCH	$N_{0,1}$	3.8	.6	.593 (.09)	.105	.12, .02[c]	17,3
ARCH	$P_{2.5}$	1.8	.6	.606 (.25)	.116	.19, .02	28,3
IGARCH	$N_{0,1}$	2.0	.6	.597 (.21)	.076	.19, .04	28,6
GARCH	$P_{2.5}$	1.5	.6	.595 (.26)	.104	.23, .02	33,3
TARCH	$P_{2.5}$	2.6	.6	.618 (.29)	.105	.04, .02	6,3
QARCH	$N_{0,1}$	3.5	.8	.812 (.48)	.069	{.02, .01},.01[d]	{3,1},1
QIARCH	$N_{0,1}$	2.0	1	.110 (.45)	.074	{.03, .01},.01	{4,1},1

a. KS minimizing δ for LOCATION where $k_T = [\delta T/\ln(T)]$.

b. KS minimizing δ for AR where $k_T = [\delta \ln(T)]$.

c. KS minimizing $\{\delta_1, \delta_2\}$ for all GARCH except QARCH where $k_{j,T} = [\delta_j T/\ln(T)]$.

d. KS minimizing $\{\delta_1, \delta_2\}$ for the first QARCH equation and two-tailed δ for the all remaning QARCH equations.