Generalized Method of Moments with Tail Trimming

Jonathan B. Hill* and Eric Renault†
Dept. of Economics
University of North Carolina - Chapel Hill
Feb. 2010

...comments are welcome...

Abstract

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

1. INTRODUCTION

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

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

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

JEL subject classifications: C13, C22.

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

Let \( m_t(\theta) \) denote estimating equations, a stochastic mapping
\[
m_t : \Theta \to \mathbb{R}^q, \quad \text{compact } \Theta \subset \mathbb{R}, \quad q \geq r,
\]
induced from some moment condition. The strong global identification condition is
\[
E [ m_t(\theta) ] = 0 \text{ if and only if } \theta = \theta^0 \text{ for unique } \theta^0 \in \Theta.
\]
As an example consider a strong-ARCH(1) process \( \{y_t\} \),
\[
y_t = \sqrt{\epsilon_t}, \quad h_t^2 = \alpha^0 + \beta^0 y_{t-1}^2, \quad \theta^0 = [\alpha^0, \beta^0]' \quad \epsilon_t \overset{iid}{\sim} (0, 1) \quad \text{and } \Theta := \sigma(y_r : \tau \leq t)
\]
with equations
\[
m_t(\theta) = \{ y_t^2 - \alpha - \beta y_{t-1}^2 \} z_{t-1}, \quad z_{t-1} = [1, y_t^2, \ldots] \in \mathbb{R}^q, \quad q \geq 2.
\]
The GMM estimator solves
\[
\hat{\theta}_g = \arg\min_{\theta \in \Theta} \left\{ \left( \frac{1}{n} \sum_{t=1}^n m_t(\theta) \right)' \hat{\Sigma}_n \left( \frac{1}{n} \sum_{t=1}^n m_t(\theta) \right) \right\}
\]
for some stochastic positive semi-definite matrix \( \hat{\Sigma}_n \in \mathbb{R}^{q \times q} \), and \( n \geq 1 \) is the sample size. Under mild conditions \( \hat{\theta}_g \) is asymptotically linear (e.g. Newey and McFadden 1994)
\[
\sqrt{n} \left( \hat{\theta}_g - \theta^0 \right) = A_n \times \frac{1}{\sqrt{n}} \sum_{t=1}^n m_t(\theta^0) + o_p(1)
\]
for some \( A_n \in \mathbb{R}^{q \times q} \), so asymptotics are grounded on \( \sum_{t=1}^n m_t(\theta^0) \).

Finite variances \( E[m_t^2(\theta^0)] < \infty \), along with standard regulatory conditions, ensures Gaussian asymptotics, but this requires \( \epsilon_t \) 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 \( z_{4,t-1} = |y_{t-1}|^{2+\delta}/2 \) and \( \delta > 0 \) then \( y_t \) must have a finite \((8 + \delta)^{th}\) moment, a very tall order for financial time series. Models with heterogeneous estimating equations include the multifactor factor Capital Asset Pricing Model with high risk (e.g. oil futures), composite market returns (e.g. NYMEX), low risk asset returns (e.g. U.S. Treasury Bill), and factor premia (e.g. market capitalization and book-to-price ratio); VARX for causality modeling of financial and macroeconomic returns; and multivariate random volatility. See French and Fama (1996), Ding and Granger (1996), Mikosch and Stårică (2000), and Embrechts et al (2003).

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

Define tail specific observations of \( m_{i,t}(\theta) \) and sample order statistics:
\[
m_{i,t}^{(-)}(\theta) := m_{i,t}(\theta) \times I(m_{i,t}(\theta) < 0) \quad \text{and } \quad m_{i,1}^{(-)}(\theta) \leq \cdots \leq m_{i,m(n)}^{(-)}(\theta) \leq 0
\]
\[
m_{i,t}^{(+)}(\theta) := m_{i,t}(\theta) \times I(m_{i,t}(\theta) > 0) \quad \text{and } \quad m_{i,1}^{(+)}(\theta) \geq \cdots \geq m_{i,m(n)}^{(+)}(\theta) \geq 0
\]
\[
m_{i,t}^{(a)}(\theta) := |m_{i,t}(\theta)| \quad \text{and } \quad m_{i,1}^{(a)}(\theta) \geq \cdots \geq m_{i,m(n)}^{(a)}(\theta) \geq 0.
\]
Now trim any equation \( m_{i,t}(\theta^0) \) that may have an infinite variance between its lower \( k_{1,i,n}/n \) th and upper \( k_{2,i,n}/n \) th sample quantiles:

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

\[
= m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta)
\]

\[
\hat{m}_{n,t}(\theta) = \left[ m_{i,t}(\theta) \times \hat{I}_{i,n,t}(\theta) \right]_{i=1}^q \text{ where } \hat{I}_{j,n,t}(\theta) = 1 \text{ if equation } j \text{ is not trimmed, and } I(\mathbf{A}) = 1 \text{ is } A \text{ is true, and } 0 \text{ otherwise}\].

If the data generating process is symmetric and \( m_{i,t}(\theta^0) \) is heavy-tailed (e.g. IGARCH) then symmetric trimming is appropriate: for \( k_{1,n} \rightarrow \infty \) and \( k_{i,n}/n \rightarrow 0 \)

\[
\hat{m}_{i,n,t}^*(\theta) := m_{i,t}(\theta) \times I\left( |m_{i,t}(\theta)| \leq m_{i,(k_{1,i,n})}(\theta) \right).
\]

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

\[
\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \left\{ \left( \frac{1}{n} \sum_{t=1}^n \hat{m}_{n,t}^*(\theta) \right) \times \hat{Y}_n \times \left( \frac{1}{n} \sum_{t=1}^n \hat{m}_{n,t}^*(\theta) \right) \right\}.
\]

As long as \( m_{i}(\theta^0) \) is an integrable martingale difference and standard smoothness conditions apply,

\[
V_n^{1/2} \left( \hat{\theta}_n - \theta^0 \right) \overset{d}{\rightarrow} N(0, I_r)
\]

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

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

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

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

\footnote{Other criteria for trimming exist, including trimming according to the Euclidean norm \( m_{t}^{(N)}(\theta) := ||m_{t}(\theta)|| \). In this case \( \hat{m}_{n,t}(\theta) = m_{t}(\theta)I(||m_{t}(\theta)|| \leq m_{t}(\theta)) \) where \( k_{n} \rightarrow \infty \) and \( k_{n}/n \rightarrow 0 \). Simulation work reveals the latter is massively dominated by component-wise trimming when \( q > 1 \), irrespective of distribution symmetry.}
Method of Moments literatures where symmetry is imposed $\lambda_{1,i} = \lambda_{2,i}$ (see below). In this case without further information the data generating process must be symmetric to ensure identification of $\theta^0$. Since key asymptotic arguments in this paper exploit negligibility and degeneracy properties under tail trimming, a direct extension to fixed quantile trimming is not evident.

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

1.2 EXTANT METHODS

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

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

Although robust estimation has a substantial history (Huber 1964, Stigler 1973, Ju-recková 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, Čízek 2005, 2008, 2009); most concern M-estimator frameworks (e.g. Čízek 2008 and the citations therein); and when trimming, truncation or weighting are employed only non-tail data quantiles are considered. Almost all uses of tail trimming appear in the robust location and central limit theory literatures, and to our knowledge extremum estimation and tail-trimming have never been combined. See Hill (2009b) for a review, and see Horowitz (1998) for a tail-trimmed covariance matrix without supporting theory.

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

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

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

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

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

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

2.1 TAIL-TRIMMING

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

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

and assume throughout \( q' \geq 1 \) since otherwise the following reduces to known results.

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

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

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

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

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

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

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

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

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

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

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

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

Write compactly throughout

\[
c_{i,n}(\theta) := \max\{l_{i,n}(\theta), u_{i,n}(\theta)\} \quad \text{and} \quad c_{n}(\theta) = \max_{1 \leq i \leq q}\{c_{i,n}(\theta)\}
\]

\[
k_{i,n} = \max\{k_{1,i,n}, k_{2,i,n}\} \quad \text{and} \quad k_{n} = \max_{1 \leq i \leq q}\{k_{i,n}\}
\]

\[
\{l_{i,n}, u_{i,n}, c_{i,n}\} = \{l_{i,n}(\theta^{0}), u_{i,n}(\theta^{0}), c_{i,n}(\theta^{0})\}.
\]

### 2.2 ASSUMPTIONS

Let \(\{\Upsilon_{n}\}\) be a sequence of positive definite weight matrices \(\Upsilon_{n} \in \mathbb{R}^{q \times q}\). The sample criterion is

\[
\hat{Q}_{n}(\theta) := \hat{m}^{*}_{n}(\theta)^{t} \times \Upsilon_{n} \times \hat{m}^{*}_{n}(\theta), \text{ where } \hat{m}^{*}_{n}(\theta) := \frac{1}{n} \sum_{t=1}^{n} \hat{m}^{*}_{n,t}(\theta) \text{ and } \Upsilon_{n} \in \mathbb{R}^{q \times q},
\]

hence the GMTTME solves

\[
\hat{\theta}_{n} = \arg\min_{\theta \in \Theta}\{\hat{Q}_{n}(\theta)\}.
\]

Under the identification and smoothness conditions detailed below, \(\hat{\theta}_{n}\) exists and is unique.

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

\[
\Sigma_{n}(\theta) := E \left[[m^{*}_{n,t}(\theta)\right]^{2}, \Sigma_{n} := \Sigma_{n}(\theta^{0}), \text{ and } m_{n} = \sup_{\theta} E \left[[m^{*}_{n,t}(\theta)\right]^{2};
\]

the population and sample Jacobia

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

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

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

6
and the Hessian and scale

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

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

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

Let \( \{\mathcal{F}_t\} \) be any sequence of increasing \( \sigma \)-fields adapted to \( \{m_i(\theta)\}, \theta \in \Theta \), where \( \{\mathcal{F}_t\} \) itself does not depend on \( \theta \). The first set characterizes matrix norms, weight limits and covariance definiteness. Denote by \( [\lambda_n(\theta)]_{i=1}^q \) the minimum eigenvalue of \( \Sigma_n(\theta) \) for each \( n \) and \( \theta \).

**M1 (weight).** \( \Sigma_n \) is positive definite for each \( n \geq N \); and \( ||\hat{\Sigma}_n - \Sigma_n|| \xrightarrow{P} 0 \) and \( ||\hat{\Sigma}_n - \Sigma_0|| \rightarrow 0 \) for some positive definite \( \Sigma_0 \), \( 0 < ||\Sigma_0|| < \infty \).

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

**M3 (smoothness).** \( a. \sup_{\theta} ||||\Sigma_n^{-1}(\theta)|||| = o(n^{1/2}); \quad b. \lim \inf_{n \geq N} \inf_{\theta} \lambda_n(\theta) > 0 \).

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

The second set promotes local identification of \( \theta^0 \).

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

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

**I3 (smoothness).** \( \inf_{n \geq N} \inf_{||\theta - \theta^0|| < \delta} \{m_{i,n}(\theta^0) ||E[m_{i,n,t}(\theta^0)]|| > 0 \text{ for tiny } \delta > 0, \text{ and} \lim \inf_{n \geq N} \{m_{i,n}>0 \} \rightarrow 0 \text{ for some } N \geq 1 \).

**I4 (orthogonality).** \( \max_{\neq i} ||||E[m_{i,n,s}(\theta^0)m_{i,n,t}(\theta^0)]|||| = o(||\Sigma_n^{-1/2}||^{-2/n}) \).

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

Remark 2: Both \( E[m_{i,n,t}(\theta^0)] = 0 \) and \( E[m_{i,n,s}(\theta^0)m_{i,n,t}(\theta^0)] = 0 \) are assured by I1 and Lebesgue’s dominated convergence theorem. Conditions I2 and I4 merely ensure \( m_{i,n,t}(\theta^0) \) behaves like a martingale difference sufficiently fast as \( n \rightarrow \infty \). If the joint distributions of \( \{m_{i,n,s}(\theta^0), m_{i,n,t}(\theta^0)\} \) are symmetric or independent then \( E[m_{i,n,t}(\theta^0)] = 0 \) and \( E[m_{i,n,s}(\theta^0)m_{i,n,t}(\theta^0)] = 0 \) under I1 for any thresholds \( i_{n,t}(\theta^0) = u_{i,n}(\theta^0) \). Otherwise by I2 and I4 \( ||\Sigma_n^{-1/2}||^{-1/n^{1/2}} \leq K ||\Sigma_n^{1/2}||/n^{1/2} = o(1) \) under M3.a suffices. In turn \( ||\Sigma_n^{-1/2}||/n^{1/2} = o(1) \) holds trivially in thin-tailed cases, and for equations with Paretoian tails under intermediate order tail-trimming.
Remark 3: Versions of smoothness I3 are standard for consistency (Huber 1967, Pakes and Pollard 1989, Newey and McFadden 1994). The envelope scale $m_n$ is required since $m_n(\theta)$ need not be integrable on $\Theta$-a.e. in heavy-tailed cases, whereas $E[m_n^*(\theta)]/m_n$ is always well defined. Consider an AR(1) $y_t = \theta y_{t-1} + \epsilon_t$ with $|\theta|^2 < 1$, $\Sigma_1 = \sigma(y_r: \tau \leq t)$, martingale difference innovations $E[\epsilon_t^2 \mid \Sigma_{t-1}] = 0$ with infinite variance $E[\epsilon_t^2] = \infty$, and one equation $m_n(\theta) = (y_t - \theta y_{t-1})/y_{t-1}$ and $E[m_n(\theta)^2 \mid \Sigma_{t-1}] = 0$ a.s. hence $E[m_n(\theta)^2] = 0$, but in general $m_n(\theta) = - (\theta - \theta^0) y_{t-1}$ is non-integrable for any coefficient $\theta \neq \theta^0$. This matters for a proof of consistency $\theta_n$-a.e. since that requires a uniform law of large numbers for $m_n^*(\theta)$ by well known arguments (e.g. Pakes and Pollard (1989)).

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

The last set concerns properties of the equations $m_n(\theta)$ and the random Jacobian matrices $J_n$ and $\hat{J}_n$.

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

D2 (differentiability). $m_n(\theta)$ is continuous and differentiable on $\Theta$-a.e.

D3 (mixing). $\{m_n(\theta)\}$ is strictly stationary, geometrically $\beta$-mixing (absolutely regular):

$$\beta_1 := \sup_{A \subset \Omega_{i+1}} E[|P(A)S_{i-\infty}^t - P(A)|] = o(\rho)$$
for some $\rho \in (0,1)$.

D4 (envelope bounds). $\sup_\theta ||m_n(\theta)||$ and $\sup_\theta ||(\partial/\partial \theta)m_n(\theta)||$ are $L_\infty$-bounded.

D5 (thresholds and fractiles).

i. $\sup_\theta \{ |(n/k_{1,n})P(m_n(\theta) < -l_{i,n}(\theta)) - 1| = O(1/k_{1,n}^{1/2}) \} \text{ and } \sup_\theta \{ |(n/k_{2,n})P(m_n(\theta) > u_{i,n}(\theta)) - 1| = O(1/k_{2,n}^{1/2}) \}$.

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

D6 (Jacobian).

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

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

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

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

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

\[\text{sup_\theta [1/n \sum_{t=1}^{n} (\hat{m}_n^*(\theta) - E[m_n^*(\theta)])] = o_p(m_n) \text{ in Lemma C.3 in Appendix C, which suffices for consistency.} \]
GARCH models are covered since geometric ergodicity implies β-mixing (An and Huan 1996, Carrasco and Chen 2002, Meitz and Saikonnen 2008).

Remark 3: Since by construction $\langle n/k_{1,n} \rangle P(\langle m_{i,t}(\theta) - l_{i,n}(\theta) \rangle < 1)$ and $\langle n/k_{2,n} \rangle P(\langle m_{i,t}(\theta) - u_{i,n}(\theta) \rangle < 1)$ the D5.i probability orders merely sharpen the rates. This is required for uniform asymptotics concerning trimming indicators $I_{i,n,t}(\theta)$. The rates hold for a large array of probability tails that satisfy second order regular variation or slow variation with uniform asymptotics concerning trimming indicators $I_{i,n,t}(\theta)$. The result helps prove for any fractiles $\mathbf{f}$ of $\mathbf{F}$ for $\mathbf{F}$, the D5.ii bound is irrelevant if equation variances are finite, and holds the distribution limit of a sum of an iid array to be Gaussian (e.g. Kallenberg 2002: Lemma 2.6.18) which holds, for example, for finite dimensional functions (e.g. Pakes and Pollard 1989, and van der Vaart and Wellner 1994: Lemma 2.4), covering at least $m_i(\theta)$ polynomial in $\theta$, hence dynamic linear regressions and ARCH.

2.3 MAIN RESULTS

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

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

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

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

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

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

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

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

---

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

9
J_n has two components that enter V_n asymmetrically as n → ∞, so an optimal weight does not exist (Čížek 2009). Under tail trimming, however, each J_{i,j;n} also decomposes into two components E[(∂/∂θ_j)m_{i,t}(θ)]_y \times I_{i,n,t}(θ^0) + (∂/∂θ_j)E[m_{i,t}(θ^0) \times I_{i,n,t}(θ)]_y. The latter is asymptotically dominated by the former due to negligibility I_{i,n,t}(θ) → 1 a.s. so J_{i,j;n} = E[(∂/∂θ_j)m_{i,t}(θ)]_y \times I_{i,n,t}(θ^0) \times (1 + O(1)). See Lemma C.1 in Appendix C.

If the Jacobian and covariance are asymptotically bounded J_n → J and Σ_n → Σ then the rate is exactly √n since V_n^{1/2} = n^{1/2}V_1^{1/2} for some positive definite V ∈ R^r×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 θ_{i,n} and may be greater or less than √n. See Section 3.

**Lemma 2.3** If \( \limsup_{n \to \infty} ||E[m_{n,t}(θ^0)m_{n,t}(θ^0)']|| < \infty \) and \( \limsup_{n \to \infty} ||(∂/∂θ)E[m_{n,t}(θ)]_y|| < \infty \) then the rate of convergence of each \( \hat{θ}_{i,n} \) is √n.

### 2.4 Covariance and Jacobian Matrix Estimation

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

\[
\hat{Σ}_n(\hat{θ}_n) = \frac{1}{n} \sum_{t=1}^{n} \hat{m}_{n,t}(\hat{θ}_n)\hat{m}_{n,t}(\hat{θ}_n)'
\]

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

**Lemma 2.4** Under D1-D6.i, D7, I1, I2, I4 and M2 \( ||Σ_n^{-1}\hat{Σ}_n(\hat{θ}_n) - I_q|| = o_p(1) \) for any \( \hat{θ}_n = θ^0 + O_p(||V_n^{1/2}||^{-1}) \).

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

**Lemma 2.5** Under D1-D7 and M1-M3 \( J_n(\hat{θ}_n) = J_n \times (1 + o_p(1)) \) and \( \hat{J}_n(\hat{θ}_n) = J_n \times (1 + o_p(1)) \) for any \( ||\hat{θ}_n - θ^0|| \approx 0 \).

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

\[
\hat{V}_n^{-1}(θ) = n \times \hat{H}_n(θ) \left( \hat{J}_n(θ)\hat{Σ}_n(θ)\hat{Y}_n\hat{J}_n(θ) \right)^{-1} \hat{H}_n(θ)
\]

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

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

### 2.5 Robust M-Estimates

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
\[
y_t = \theta^\prime x_t + \epsilon_t \quad \text{with} \quad s_t(\theta) := (y_t - \theta^\prime x_t)^2.
\]
Assume \(\epsilon_t\) is zero-mean with distribution function \(F_\epsilon(\epsilon) := P(\epsilon_t \leq \epsilon)\), and two-tailed inverse \(F_\epsilon^{-1}(\lambda) := \inf\{\epsilon \geq 0 : P(|\epsilon_t| \leq \epsilon) \leq \lambda\}\). The fixed quantile LTSE is (Ruppert and Carrol 1980, Rousseeuw 1985, Čížek 2008)
\[
\hat{\theta}_n = \arg\min_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_{t=1}^{n} s_t(\theta) \times I(s_t(\theta) \leq s_{(n\lambda)}(\theta)) \right\}, \quad \lambda \in (0, 1).
\]
If the distribution governing \(s_t(\theta)\) is absolutely continuous on \(\Theta\)-a.e., \(\{x_t, \epsilon_t\}\) have finite variance marginal distributions, \(\{\epsilon_t, x_t\}\) are geometrically \(\beta\)-mixing, and \(J(\lambda) := -E[x_t x_t' I(|\epsilon_t| \leq F^{-1}_\epsilon(\lambda))]\) is non-singular, then for the given linear DGP
\[
\sqrt{n} \left( \hat{\theta}_n - \theta^0 \right) = J(\lambda)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t x_t I(\epsilon_t \leq F^{-1}_\epsilon(\lambda)) \right) + o_p(1) \overset{d}{\to} N \left(0, \tilde{V}^{-1}(\lambda)\right)
\]
where \(\tilde{V}(\lambda) = J(\lambda)^T \tilde{\Sigma}^{-1} J(\lambda)\) and \(\tilde{\Sigma}(\lambda) := E[\epsilon_t^2 x_t x_t' I(|\epsilon_t| \leq F^{-1}_\epsilon(\lambda))]\). See Čížek (2005, 2008). Clearly if the error \(\epsilon_t\) is independent of \(x_t\) and any stochastic element of \(x_t\) has an infinite variance then \(1/\sqrt{n} \sum_{t=1}^{n} \epsilon_t x_t I(\epsilon_t^2 \leq \epsilon_{(n\lambda)}^2)\) does not have a Gaussian limit even under fixed quantile trimming, and \(\tilde{\Sigma}(\lambda)\) and \(J(\lambda)\) do not exist. The same conclusion applies to Least Absolute Trimmed Deviations. Since the object that governs asymptotics is the gradient \(\partial(\theta)/\partial(\theta)_s(\theta)|_{\theta^0}\), its components \(\epsilon_t x_t, t\) must be trimmed to ensure asymptotic normality, not simply \(\epsilon_t\).

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

Since a standard question is whether a conditional heterscedastic effect exists, suppose not for simplicity: \(\beta^0 = 0\). If the distribution governing \(\epsilon_t\) is absolutely continuous then by Lemma 2.1 of Čížek (2008) \(g_{n,t}(\theta) := (\partial/\partial(\theta))_s_{n,t}(\theta) = (\partial/\partial(\theta)) s_t(\theta) \times I(s_t(\theta) \leq F^{-1}_s(\theta)\lambda)\) a.s. on \(\Theta\)-a.e. By direct computation it follows under \(\beta^0 = 0\)
\[
g_{n,t}(\theta^0) = -\left(\epsilon_t^2 - 1\right) [1, y_{t-1}^2]' \times I(\epsilon_t \leq F^{-1}_\epsilon(\lambda)).
\]
Now exploit independence to deduce
\[
E[ g_{2,n,t}(\theta^0) ] = E[ (\epsilon_t^2 - 1)^2 I(\epsilon_t \leq F^{-1}_\epsilon(\lambda)) ] \times E[ y_{t-1}^4 ].
\]
Since \(\beta^0 = 0\) we know \(y_t\) has an unbounded fourth moment \(E[ y_t^4 ] = \infty\ if and only if E[ \epsilon_t^4 ] = \infty\). The QMTL Jacobian is unbounded and by asymptotic linearity and independence between \(\epsilon_t\) and \(y_{t-1}\), the QMTLE is not asymptotically normal.

Adaptive M-Estimation: Ling’s (2005, 2007) smooth symmetrically weighed LAD and QML criteria work like smooth trimming. But theory is only delivered for symmetric DGP’s, and only fixed quantiles of the data \(y_t\) are considered for the weight function. Although heavy-tails are allowed super-\(\sqrt{n}\)-convergence cannot be achieved due to the weight structure.
3. CONVERGENCE RATE FOR HEAVY-TAILED DATA Consider the efficient weight \( \Upsilon_n = \Sigma^{-1}/||\Sigma^{-1}|| \) for brevity. By positive definiteness and the Cauchy-Schwartz inequality we can define diagonal matrices \( \Gamma_n \in \mathbb{R}^{q \times q} \) with

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

Now write

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

Assume \( \sigma_{i,j} \neq 0 \forall i,j \), to simplify exposition. The component-wise rates \( n_{\theta_i} \) are

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

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

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

3.1 DYNAMIC REGRESSION WITH IID ERRORS Consider a stationary dynamic linear regression with an intercept

\[
y_t = \theta^0 x_t + \epsilon_t, \quad x_{1,t} = 1, \quad x_t \in \mathbb{R}^r \quad \text{with} \quad m_t(\theta) = (y_t - \theta x'_t) x_t,
\]

where \( \epsilon_t \) and \( x_t \) are mutually independent. Assume stochastic \( x_{1,t} \) are measurable with \( \mathbb{R} \)-a.e. continuous, stationary, symmetric distributions. Independence rules out random volatility errors: see Sections 3.2-3.3 for this case.

Define the moment suprema of \( z_t \in \{\epsilon_t, x_{1,t}\} \),

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

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

\[
P(|z_t| > z) = d_z z^{-\kappa_z} (1 + o(1)) \quad \text{with} \quad \kappa_z \in \{\kappa_{\epsilon}, \kappa_{x}\} = (1,2].
\]

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

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

Heavy-tailed convolutions \( m_{I,t}(\theta^0) = \epsilon_t x_{1,t} \) also satisfy \( (7) \) with index \( \kappa_{\epsilon,j} := \min \{\kappa_{\epsilon}, \kappa_{x}\} \) (Cline 1986), so the thresholds \( c_{i,n} = K(n/k_n)^{1/\kappa_{\epsilon,i}} \) by the construction of \( c_{i,n} \) and \( k_n \).
Define
\[
a^*_e(i) := \min_{j \notin \{1,i\}} \left\{ \frac{1}{\kappa_{e,j}} + (1 - 1/\kappa_j) \kappa_i / \kappa_{e,i} \right\}
\]
\[
A_n = \left( n/k_n \right)^{1/2+1/\kappa_{e,i}+\kappa_i/\kappa_{e,i}-2a^*_e(i)}
\]
\[
B_n = \max_{j \neq i; \kappa_j \leq 2} \left\{ \left( n/k_n \right)^{2/\kappa_j-1} \right\}
\]
\[
C_n = \max_{j \neq i; \kappa_j \leq 2} \left\{ \left( n/k_n \right)^{2-2/\kappa_j-2/\kappa_i+2\kappa_i^{-1}(\kappa_i/\kappa_j+1-\kappa_i)} \right\}.
\]

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

**Lemma 3.1 (ARX with IID Error)**

a. Let \( \max\{\kappa_e, \kappa_2, ..., \kappa_r\} \leq 2 \). Each \( \Gamma_{i,i,n} = (n/k_n)^{1/\kappa_{e,i}+1/2} \) and \( J_{i,j,n} \sim -E[x_{i,t}x_{j,t}I(|\epsilon_i|+\epsilon_j)|x_{i,t},x_{j,t}] \leq c_{j,n} \). For stochastic \( \{x_{i,t},x_{j,t}\} \) in general \( J_{i,i,n} \sim K(n/k_n)^{\kappa_e(2-\kappa_i)} \), \( J_{i,j,n} = O((n/k_n)^{\kappa_e(\kappa_i/\kappa_j+1-\kappa_i)}) \) \( \forall i \neq j \), and \( J_{j,j,n} \sim K \) if \( x_{i,t} \) is independent of \( x_{j,t} \). Hence the slope rates are
\[
n_{\theta_i} \sim K n^{1/2} (n/k_n)^{1/2-\kappa_i/\kappa_e,i+1/\kappa_{e,i}} \left[ K + O(A_n) \right]^{1/2}, \ i = 2, ..., r.
\]

Further \( J_{1,1,n} = -1+o(1) \) and \( J_{1,i,n}, J_{1,1,n} = O(1) \times (1 + o(1)) \), hence the intercept rate is
\[
n_{\theta_i} = K n^{1/2} \times \left( K + O(A_n) \right)^{1/2} \times (1 + O(1)).
\]

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

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

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

In the following assume all random variables are heavy tailed \( \max\{\kappa_e, \kappa_2, ..., \kappa_r\} \leq 2 \). General examples are similar.

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

**Case 1 (\( \kappa_i \leq \kappa_e \)):** Here \( n_{\theta_i} \geq K n^{1/2} (n/k_n)^{1/2} \times \left[ K + O(C_n) \right]^{1/2} \) only reflects the tails of \( x_{i,t} \). Simply choose a light fractile \( k_n = [\lambda \ln n] \) for \( \lambda > 0 \) to obtain
\[
n_{\theta_i} \geq K n^{1/2}/\ln(n).
\]

The GMTTME rate is therefore arbitrarily close to the highest achievable rate \( n^{1/\kappa_e} \) for stationary data with Paretian tails by any estimation method, including untrimmed OLS.
The AR process since under tail trimming, so the GMTTME is sub-
minimal trimming is optimal (slow \( k \)). If \( \kappa_i = 1.5 \), for example, then any \( \kappa_i \geq 1 \) applies.

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

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

\[
n_{\theta_i} \sim Kn^{1/2} (n/k_n)^{1/\kappa-1/2}.
\]

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

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

**EXAMPLE 4 (Location):** Consider estimating location

\[
y_t = \theta^0 + \epsilon_t, \, \epsilon_t \, i.i.d. \, (7), \, \kappa \in (1, 2],
\]

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

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

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

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

\[
y_t = \sum_{i=1}^{r} \theta_i y_{t-i} + \epsilon_t, \, \epsilon_t \, i.i.d. \, (7), \, \kappa \in (1, 2].
\]

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

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

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

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

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

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

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

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

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

**3.2 ARCH**

Consider a Strong-ARCH(\( p \)) model

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

\[
m_t(\theta) = (y_t^2 - \theta x_t) x_t, \quad x_t = [1, y_{t-1}^2, \ldots, y_{t-p}^2]^T,
\]

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

Integrability of \( m_t(\theta^0) \) requires \( E[|\epsilon_t^2|] < \infty \) and \( E[h_t^2 y_t^2] < \infty \), so assume

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

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

Since \((\partial / \partial \theta)m_t(\theta) = -x_t x_t' \) have indices in \( \{\kappa_y/4, \kappa_y/2, \infty\} \) and \( \kappa_y > 4 \) each component is integrable, thus \( J_n \sim J \). The standard deviations \( \Gamma_{i,n} \) are almost as simple to compute since \( \kappa_y > 4 \) implies the intercept term \( \Gamma_{1,1,n} \leq \left(E[(\epsilon_t^2 - 1)^2 k_t^4]\right)^{1/2} = K \), and the remaining \( \Gamma_{i,m} = \left(E[m_{i,t}(\theta^0)^2 I_{i,n,m}(\theta^0)]\right)^{1/2} \sim K c_{i,n}(k_n/n)^{1/2} = K(\alpha/\kappa_n)^{1/2} - 1/2 \) by Feller property (FE) and \( c_{i,n} = K(n/k_n)^{1/2(\kappa_y/4)} \) for tail (7). Therefore \( J_{1,1,n}/\Gamma_{1,1,n} \sim K \) and all other \( J_{i,m}/\Gamma_{i,m} \sim K(n/k_n)^{-1/2(\kappa_y-1/2)} = o(1) \). Now use (6) to deduce \( n_{\alpha, i}, n_{\beta_i} \sim Kn^{1/2} \). This proves the next claim.
**Lemma 3.2 (Strong-ARCH)** Any stationary strong-ARCH process with negative Lyapunov exponent and $\kappa_y > 4$ has GMTTME rates $n_\alpha, n_\beta \sim Kn^{1/2}$.

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

**Remark 2:** Strong-ARCH are AR in squares $y_t^2 = \theta^0 x_t + v_t$, where $E[v_t|\mathcal{F}_{t-1}] = 0$. But stationary AR equations all have the same tail index $\kappa$ when $\epsilon_t$ is iid with tail (7). In the ARCH case the "error" $y_t^2 - h_t^2 = (\epsilon_t^2 - 1)\theta^0 x_t$ depends on $x_t$, thus $m_{1,t}(\theta^0) = (\epsilon_t^2 - 1)h_t^2$ has tail index $\kappa/2$ and all other $m_{i,t}(\theta^0) = (\epsilon_t^2 - 1)h_t^2 y_{t-i+1}$ for $i \geq 2$ have index $\kappa/4$ due to feedback. The intuition from Section 3.1 suffices to explain the rate: models with disproportionately heavy-tailed "errors" render less sharp estimates (in this case, less than super-$\sqrt{n}$-consistent).

### 3.3 Autoregressions with ARCH-Errors

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

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

$$h_t^2 = \alpha^0 + \beta^0 u_{t-1}^2, \quad \alpha^0 > 0, \beta^0 > 0, \quad \theta = [\rho, \alpha, \beta]^t$$

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

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

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

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

Integrability again requires $\kappa_y > 4$, hence $J_n \sim J$. If $\kappa_y > 8$ then $\Gamma_{1,n} = n^{1/2}/I_3$, otherwise use $E[y_t^2] < \infty$ to deduce $E[u_t^2] < \infty$, $E[h_t^2] < \infty$ and $E[\epsilon_t^2] < \infty$ so $\Sigma_{1,1,n} \sim K$, $\Sigma_{2,2,n} \sim K$, and $\Sigma_{3,3,n} = E[(\epsilon_t^2 - 1)^2 h_t^4 y_{t-1}^4 \leq c_3, n] \sim K(n/k)^{8/\kappa_y - 1}$. This gives trimmed standard deviations $\Gamma_{1,1,n} = \Gamma_{2,2,n} = 1$ and $\Gamma_{3,3,n} = (n/k)^{4/\kappa_y - 1/2}$. The same conclusion as the Strong-ARCH case is therefore reached by working through formula (6): all rates are $\sqrt{n}$ since feedback between $u_t$ and $y_{t-1}$ dulls the rate below super-$\sqrt{n}$-convergence.

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

**Example 7 (AR with ARCH Error):** Estimate only the AR slope $\rho^0$ 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 $\kappa_y^2/2$ due to feedback, half that in the iid case Example 5. But this implies $m_t(\rho^0)$ is integrable only if $\kappa_y > 2$. Arguments in the proof of Lemmas 3.1 can be used to deduce $n_\rho \sim K^{1/2}$ if $\kappa_y \geq 4$, and $n_\rho \sim K^{1/2}$ if $\kappa_y > 4$.
 optional. Feedback between error \( u_t \) and regressor \( y_{t-1} \) substantially elevates estimating equation tail thickness relative to the Jacobian, hence the convergence rate \( n_{\nu} \) falls below \( n^{1/2} \).

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

### 3.4 TAIL-TRIMMED QML FOR GARCH

Consider an ARCH(1) with QML equations

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

\[
m_t(\theta) = (y_t^2 - \theta' x_t)(\theta' x_t)^{-2} x_t,
\]

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

If \( E[\epsilon_t^4] = \infty \) assume \( \epsilon_t \) 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 \( \epsilon_t \) matter, and ensures \( J_n \sim J \). This will imply a diminished rate of convergence when \( E[\epsilon_t^4] = \infty \) for the reasons given in Section 3.1.

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

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

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

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

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

### 4. AUTOREGRESSIONS AND GARCH

We now verify the major assumptions for heavy-tailed stationary AR and GARCH under symmetric trimming \( m_{i,t}(\theta) = m_{i,t}(\theta)I(|m_{i,t}(\theta)| \leq c_{i,n}(\theta)) \) where \( (n/k_{\nu})P(|m_{i,t}(\theta)| > c_{i,n}(\theta)) \rightarrow 1 \), the same fractile \( k_n \) for each equation, and weight \( \Upsilon_n = \Sigma_n^{1/2}/||\Sigma_n^{1/2}|| \).

#### 4.1 Autoregression

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

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

\[
m_t(\theta) = (y_t - \theta' x_t)x_t.
\]
Assume $\epsilon_t$ has an absolutely continuous marginal distribution symmetric at zero and positive $\mathbb{R}$-a.e. Then $y_t$ is uniformly $L_{1+i}$-bounded geometrically $\beta$-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 $\kappa$ (Cline 1986, 1989).

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

Further $I2$ and $I4$ are trivial since $E[m_{n,t}^*(\theta^0)] = 0$ and $E[m_{n,t}^*(\theta^0)m_{n,t}^*(\theta^0)^\top] = 0$ under symmetry, integrability and orthogonality; $M1$ holds under $I1$ and $D1$-$D7$ given $\mathcal{T}_n = \Sigma_n^{-1}/||\Sigma_n^{-1}||$ and Lemma 2.4. Finally, $D5.i$ merely defines the threshold sequence. What remains is envelope moment bounds $D4$, threshold bound $D5.ii$, Jacobia property $D6$, indicator property $D7$, smoothness $I3$, and covariance bound $M3$.

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

**D6 (Jacobia).**

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

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

**I3 (smoothness).** Define $Q_n(\theta) := E[m_{n,t}^*(\theta)]^\top \mathcal{T}_n \times E[m_{n,t}^*(\theta)]$. Identification and the definition of a derivative imply $E[m_{n,t}^*(\theta)] = J_n(\theta - \theta^0) + o(||J_n|| \times ||\theta - \theta^0||)$, hence $m_{n,t} = \sup_{\theta} ||E[m_{n,t}^*(\theta)]|| \leq K||J_n|| \times (1 + o(1))$ given compactness of $\Theta$. Therefore since $\mathcal{T}_n$ is bounded

$$\inf_{||\theta - \theta^0|| > \delta} \left\{ m_{n,t}^2 Q_n(\theta) \right\} \geq \inf_{||\theta - \theta^0|| > \delta} \left\{ (\theta - \theta^0)^\top J_n^\top \mathcal{T}_n J_n \frac{J_n}{||J_n||} (\theta - \theta^0) \right\} \times (1 + o(1)) + o(1)$$

But boundedness and positive definiteness of $\mathcal{T}_n$ imply $J_n^\top \mathcal{T}_n J_n / ||J_n||^2$ is positive definite for sufficiently large $n$, so I3 follows.

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

Asymptotic normality is therefore a primitive property of the GMTTME for stationary AR data. The above verification, Theorems 2.2 and 2.6 and Lemma 3.1 suffice to prove the following claim.
COROLLARY 4.1 Let $y_t$ satisfy (8), let $\Sigma = \Sigma_n^{-1} / ||\Sigma_n^{-1}||$, and assume the thresholds satisfy D5.i. Then $V_n^{1/2}(\hat{\theta}_n - \theta^0) \overset{d}{\rightarrow} N(0, I)$ and $\hat{V}_n = V_n(1 + o_p(1))$. In particular, $n^{1/2}(n/k_n)^{1/\alpha - 1/2}(\hat{\theta}_{i,n} - \theta^0_i) \overset{d}{\rightarrow} N(0, V_i)$ for any $k_n \to \infty$ and $k_n = o(n)$ each $i$ and some $V_i < \infty$, and $(n^{1/\alpha} / \ln(n))(\hat{\theta}_{i,n} - \theta^0_i) \overset{d}{\rightarrow} N(0, V_i)$.

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

4.2 Nonlinear AR-Nonlinear GARCH

Recall the ARCH model of Sections 3.2 and 3.5, assume at least one slope $\beta_i > 0$ and the distribution of $\epsilon_i \overset{iid}{\sim} (0, 1)$ does not have an atom at zero. Then $\{y_t\}$ is stationary, $L_p$-bounded, geometrically $\beta$-mixing (Meitz and Saikkonen 2008) with tail (7) (Basrak et al 2002: Theorem 3.1). Least squares and QML type equations $m_1(\theta)$ are differentiable in $\theta$ with absolutely continuous marginal distributions, and integrable at $\theta^0$ for all $\kappa_y > 2$. Similar to the AR model all conditions are either trivial (e.g. I2), regulatory (e.g. D5.i), or can be verified using arguments from Section 4.1 (e.g. M3).

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

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

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

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Functional Form: $f(\theta^*, \epsilon_t, x_t)$</th>
<th>$\theta^*_t$</th>
<th>iid errors $\epsilon_t$</th>
<th>$\kappa_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCATION</td>
<td>$1 + \epsilon_t$</td>
<td></td>
<td>$P_{1.5}, P_{2.5}$</td>
<td>1.50, 2.50</td>
</tr>
<tr>
<td>AR(1)</td>
<td>$0.9 \times y_{t-1} + \epsilon_t$</td>
<td>$0.9$</td>
<td>$P_{1.5}, P_{2.5}$</td>
<td>1.50, 2.50</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>${3 + 5y_{t-1}^2}^{1/2} \epsilon_t$</td>
<td>$0.5$</td>
<td>$N_{0.1}$</td>
<td>4.65$^5$</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>${3 + 6y_{t-1}^2}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$N_{0.1}$</td>
<td>3.80$^7$</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>${3 + 6y_{t-1}^2}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$P_{2.5}$</td>
<td>1.80</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>${3 + 6y_{t-1}^2}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$P_{2.1}$</td>
<td>1.40</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>${3 + 3y_{t-1}^2 + 6h_{t-1}^2}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$N_{0.1}$</td>
<td>4.10$^3$</td>
</tr>
<tr>
<td>IGARCH(1,1)</td>
<td>${3 + 4y_{t-1}^2 + 6h_{t-1}^2}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>${3 + 3y_{t-1}^2 + 6h_{t-1}^2}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
</tr>
<tr>
<td>TAR(1)</td>
<td>${3 + 6y_{t-1}^2 \times I(y_{t-1} &lt; 0)}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$N_{0.1}$</td>
<td>5.25$^5$</td>
</tr>
<tr>
<td>TIA(1)</td>
<td>${3 + 6y_{t-1}^2 \times I(y_{t-1} &lt; 0)}^{1/2} \epsilon_t$</td>
<td>$0.6$</td>
<td>$P_{2.5}$</td>
<td>2.60</td>
</tr>
<tr>
<td>QARCH(1)</td>
<td>${3 + 8y_{t-1}} \epsilon_t$</td>
<td>$0.8$</td>
<td>$N_{0.1}$</td>
<td>3.50$^6$</td>
</tr>
<tr>
<td>QIARCH(1)</td>
<td>${3 + y_{t-1}} \epsilon_t$</td>
<td>$1$</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Model</th>
<th>iid errors $\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>$u_t(\theta) \in \mathbb{R}$</th>
<th>$z_t \in \mathbb{R}^q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>$P_{1.5}, P_{2.5}$</td>
<td>1.5, 2.5</td>
<td>LS</td>
<td>$[1, y_{t-1}]$</td>
</tr>
<tr>
<td>AR(1)</td>
<td>$P_{1.5}, P_{2.5}$</td>
<td>1.5, 2.5</td>
<td>LS</td>
<td>$[y_{t-1}]_{i=1}^2$</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>$N_{0.1}$</td>
<td>4.65</td>
<td>LS</td>
<td>$1, {y_{t-1}^2}_{i=1}^2$</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>$N_{0.1}, P_{2.5}, P_{2.1}$</td>
<td>3.80, 1.80, 1.40</td>
<td>QML</td>
<td>$1, {y_{t-1}^2}_{i=1}^2$</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>LS</td>
<td>$1, {y_{t-1}^2, h_{t-1}^2}_{i=1}^2$</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>$N_{0.1}, P_{2.5}$</td>
<td>2.00, 1.50</td>
<td>QML</td>
<td>$1, {y_{t-1}^2, h_{t-1}^2}_{i=1}^2$</td>
</tr>
<tr>
<td>TAR(1)</td>
<td>$N_{0.1}$</td>
<td>5.25</td>
<td>LS</td>
<td>$1, y_{t-1}^2, y_{t-2}^2$</td>
</tr>
<tr>
<td>TAR(1)</td>
<td>$N_{0.1}, P_{2.5}$</td>
<td>3.40, 2.60</td>
<td>QML</td>
<td>$1, y_{t-1}^2, y_{t-2}^2$</td>
</tr>
<tr>
<td>QARCH(1)</td>
<td>$N_{0.1}, N_{0.1}$</td>
<td>3.50, 2.00</td>
<td>QML</td>
<td>$1, {y_{t-1}^2}_{i=1}^2$</td>
</tr>
</tbody>
</table>

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

$^4$Basrak et al (2002: eq. 2.10) show $E[|\beta_t^2 + \gamma|^{\kappa/2}] = 1$ for GARCH(1,1) $y_t = h_t\epsilon_t$ with iid $\epsilon_t$ and $h_t^2 = \alpha + \rho y_{t-1}^2 + \gamma h_{t-1}^2$, provided the Lyapunov index is negative. The index $\kappa$ is computed as $\hat{\kappa} = \arg\min_{\kappa \in K} \{|1/N \sum_{i=1}^N (\beta_t^2 + \gamma)^{\kappa/2} - 1\}$ over $K \in \{0.1, 0.2, \ldots, 1.0\}$ based on $N = 100,000$ iid random draws $\epsilon_t$ from $N_{0.1}, P_{2.5}$ or $P_{2.1}$. The 1% bands are less than .001 in all cases.

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

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

20
The collective GARCH group have heavy tails due to the innovations $\epsilon_t \sim P_{2.5}$ and/or the parametric structure. The kurtosis of $y_t$ is infinite in most cases, and variance is infinite for IGARCH, ARCH and GARCH with Pareto errors and QARCH. Thus, in all random volatility models the GMME is not asymptotically normal, and the QMLE has not been shown to be asymptotically normal when $E[\epsilon_t^4] = \infty$. Nevertheless, the QMLE is consistent in all cases\(^\ast\).

5.1 Tail Fractile

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

$$k_n = \lceil n^{\lambda} \rceil$$

where $\lambda \in \{.01, .02, ..., .99\}$.

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

5.2 Evaluation

We analyze $r^{th}$ parameter estimates $\hat{\theta}_r$ for brevity. Consult the third column of Table 1 for the true $\theta_r$. Estimator performance is gauged by simulation means, mean-squared errors, and Kolmogorov-Smirnov tests of standard normality. Let $\{\hat{\theta}_{j,r}\}_{j=1}^{1000}$ be the independently drawn sequence of estimates of $\theta_r$. In a first experiment we fix the sample size $n = 1000$ and use the simulation $\text{mse} = (1/1000) \sum_{j=1}^{1000} (\hat{\theta}_{j,r} - \theta_r)^2$ to generate an iid sequence of ratios $\{\tilde{T}_{j,r}\}_{j=1}^{1000}$, $\tilde{T}_{j,r} = (\hat{\theta}_{j,r} - \theta_r)/\hat{s}_{n,r}^2$. We report $(1/1000) \sum_{j=1}^{1000} \hat{\theta}_{j,r}, \hat{s}_{n,r}$, and the KS test based on $\{\tilde{T}_{j,r}\}_{j=1}^{1000}$. In the case of GMTTM only the KS minimizing $\lambda$ or pair $(\lambda_1, \lambda_2)$ is used. See Tables 3 and 4.

5.3 GMTTM Weight

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

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

5.4 Rate of Convergence

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

\[^\ast\]Let $Q_0(\theta) \equiv \text{QML criterion}$. It can be easily verified for all models above $(\partial/\partial \theta)Q_0(\theta) = \sum_{t=1}^n g_t(\theta)$ for some $L_1$-bounded martingale difference sequence $(g_t(\theta), \mathcal{F}_t)$. Since an $L_{1+a}$-bounded mds trivially forms a uniformly integrable $L_1$-mixingale (McLeish 1975), Andrews’ (1988: Theorem 1) law of large numbers applies: $1/n \sum_{t=1}^n g_t(\theta_0) \xrightarrow{a.s.} 0$. Further, each $g_t(\theta)$ satisfies Andrew’s (1992: W-LIP) Lipschitz condition given differentiability and the QML criterion form, so $\sup_{\theta} \max_{1 \leq n \leq N} g_t(\theta) = E[|g_t(\theta)|] \leq 0$ by Theorem 3 of Andrews (1992). Consistency of the QMLE is now a standard exercise (e.g. Pakes and Pollard 1989: Corollary 3.4). See also Hall and Yao (2003) for the GARCH case with infinite kurtosis errors.
by exactly identified GMTTM based on the Table 2 equation forms over sample sizes $n \in N = \{1000, \ldots, 10000\}$ with increments of 20. In the ARCH case we focus on the slope $\beta^0$. This produces 450 estimates of the simulation mse’s $\hat{s}_{n,r}^2$.

Since the rate satisfies $n \theta_1 \sim K \hat{s}_{n,r}^{-1}$, according to location and AR Examples 4 and 5 if $y_t$ has an infinite variance $\kappa_y < 2$ then

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

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

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

- **ARCH**: $\hat{s}_{n,r}^{-1}/n^{1/2} \sim Kn^{-(1-\lambda)(2/\kappa_e-1/2)}$

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

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

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

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

5.5 **Summary of Results**

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

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

ii. Asymmetric trimming for asymmetrically distributed equations is always optimal, where more observations are trimmed from the heavier tail. Consider the TIARCH model (see Table 6): left-tailed $y_t$ have an infinite variance and right-tailed $y_t$ are Gaussian. The optimal trimming pair $\{\lambda_1, \lambda_2\} = \{0.40, 0.25\}$ translates to trimming $k_{1,n} = 16$ left-tailed and $k_{2,n} = 6$ right-tailed observations when $n = 1000$.

iii. Typically only a few tail observations need to be trimmed to ensure approximate normality. Examples include symmetric GARCH with Pareto errors: $k_n = [1000^{0.35}] = 11$; and TIARCH with a heavier left-tail: $k_{1,n} + k_{2,n} = 16 + 6 = 22$.

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

5.6 **Rate of Convergence**
See Table 4 for the simulation 95% confidence bands of the rate log-trend slope $b$. In most cases simulation results closely match theory, and in all cases the correct sign is deduced. The finite variance location and AR the slope is $b = 0$ and the respective bands are $0.31 \pm 0.051$ and $0.27 \pm 0.039$.

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

\begin{align*}
\text{location : } b &= -(1 - \lambda)(1/\kappa_y - 1/2) = -(1 - .22)(1/1.5 - 1/2) = -0.112 \\
\text{AR : } b &= (1 - \lambda)(1/\kappa_y - 1/2) = (1 - .32)(1/1.5 - 1/2) = 0.113.
\end{align*}

The 95% band for location is $-0.130 \pm 0.024$, and for AR is $0.116 \pm 0.012$, both quite sharp and containing the true $b$. In the AR case the maximum rate of convergence amongst M-estimators is $n^{1/\kappa} \equiv n^{67} > n^5$ while the GMTTME achieved on average $n^{1/2+b} = n^{5 + .116} = n^{616} \in (n^{5}, n^{67})$, exactly as theory dictates.

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

Use Lemma 3.5 and Table 4 to deduce

\begin{align*}
\text{ARCH (}\kappa_e = 2.5): b &= -(1 - \lambda)(2/\kappa_e - 1/2) = -(1 - .4)(2/2.5 - 1/2) = -0.18 \\
\text{ARCH (}\kappa_e = 2.1): b &= -(1 - \lambda)(2/\kappa_e - 1/2) = -(1 - .38)(2/2.1 - 1/2) = -0.28.
\end{align*}

The 95% simulation bands are $-0.387 \pm 0.306$ and respectively $-0.412 \pm 0.393$. Feedback erodes efficiency substantially.

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

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

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

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

Future work should involve details on the convergence rate for broader classes of linear and nonlinear processes, the trade-off between small sample distribution and efficiency,
APPENDIX A: Proofs of Main Results

We repeatedly use the following properties under M3:

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

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

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

Proof of Theorem 2.1. Define \( \hat{m}_n^*(\theta) := 1/n \sum_{t=1}^n \tilde{m}_{n,t}(\theta), m_n^*(\theta) := 1/n \sum_{t=1}^n m_{n,t}(\theta), \) and \( Q_n(\theta) := E[m_{n,t}^*(\theta)]' \times \mathcal{T}_n \times E[m_{n,t}^*(\theta)] \). The following is similar to Pakes and Pollard’s (1989: p. 1039) argument. Use smoothness I3 and weight boundedness M1 to define \( \epsilon(\delta) := \inf_{n \geq N} \inf_{\| \theta - \theta_0 \| > \delta} \{ m_{n}^2 \times Q_n(\theta) \} > 0 \) for arbitrarily large \( N \) and tiny \( \delta > 0 \). Since \( P(\| \hat{\theta}_n - \theta_0 \| > \delta) \leq P(m_{n}^2 Q_n(\hat{\theta}_n) > \epsilon(\delta)) \) it suffices to show \( Q_n(\hat{\theta}_n) = o_p(m_{n}^2) \) to prove \( \| \hat{\theta}_n - \theta_0 \| = o_p(1) \).

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

\[
Q_n(\hat{\theta}_n) \leq \hat{Q}_n(\hat{\theta}_n) + \left| \hat{Q}_n(\hat{\theta}_n) - Q_n(\hat{\theta}_n) \right| \leq \hat{Q}_n(\hat{\theta}_n) + \left( m_{n}^2 + Q_n(\hat{\theta}_n) \right) \times o_p(1),
\]

hence \( Q_n(\hat{\theta}_n)(1 - o_p(1)) \leq \hat{Q}_n(\hat{\theta}_n) + o_p(m_{n}^2) \). By construction \( \hat{Q}_n(\hat{\theta}_n) \leq Q_n(\theta_0) \), and \( \hat{Q}_n(\theta_0) \leq K ||\hat{m}_n^*(\theta_0)||^2 \leq K ||m_n^*(\theta_0)||^2 + o_p(\|\Sigma_n^{-1/2}\|^2/n^{1/2}) = K ||m_n^*(\theta_0)||^2 + o_p(1) \).

In lieu of weight bound M1, the Lemma C.2 asymptotic approximation and covariance bound (9.2). Finally \( ||m_n^*(\theta_0)|| \leq K ||m_n^*(\theta_0) - E[m_{n,t}^*(\theta_0)]|| + o(||\Sigma_n^{-1/2}\|^2/n^{1/2}) = o_p(1) \) by the triangular equality, identification I2, the Lemma C.3 law of large numbers, and (9.2). □

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

\[
V_n^{1/2} \left( \hat{\theta}_n - \theta_0 \right) = -V_n^{1/2} \left( H_n^{-1} J_n \mathcal{T}_n \right) \frac{1}{n} \sum_{t=1}^n \tilde{m}_{n,t}(\theta_0) \times (1 + o_p(1)).
\]

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

\[
V_n^{1/2} \left( \hat{\theta}_n - \theta_0 \right) = -V_n^{1/2} \left( H_n^{-1} J_n \mathcal{T}_n \right) \frac{1}{n} \sum_{t=1}^n m_{n,t}(\theta_0) \times (1 + o_p(1)) + o_p(1).
\]

Now invoke identification I2, central limit theorem Lemma C.6 and \( V_n^{1/2}(n^{-1/2} H_n^{-1} J_n \mathcal{T}_n \Sigma_n^{1/2}) \to I \) to conclude

\[
V_n^{1/2} \left( \hat{\theta}_n - \theta_0 \right) = -\Sigma_n^{-1/2} n^{-1/2} \sum_{t=1}^n \{ m_{n,t}(\theta_0) - E [m_{n,t}(\theta_0)] \} \times (1 + o_p(1)) + o_p(1)
\]

\[
\Rightarrow \frac{d}{d} N(0, I). \]

24
Proof of Lemma 2.4. The triangular inequality and the definitions of \( \hat{\Sigma}_n(\tilde{\theta}_n) \) and \( \Sigma_n \) imply \( ||\Sigma_n^{-1}\Sigma_n(\theta_n) - I_q|| \leq ||\Sigma_n^{-1}|| \times ||\Sigma_n(\theta_n) - \Sigma_n|| \) is bounded by

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

The first term is \( o_p(1) \) by an argument identical to the proof of uniform asymptotic approximation Lemma 2.2, and the third term is \( o_p(1) \) by Lemma C.8.

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

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

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

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

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

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

Minkowski’s inequality implies for arbitrary \( \theta \)

\[
\| \hat{J}^{*, i}_{i,n,t}(\theta) - J_n \| \leq \| \hat{J}^{*, i}_{i,n,t}(\theta) - J^*_{n}(\theta, \varepsilon_n) \| + \| J^*_{n}(\theta, \varepsilon_n) - J_n \|
\]

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

\[
\hat{J}^{*, i}_{i,n,t}(\theta) = \hat{J}^{*, i}_{i,n,t}(\tilde{\theta}_{n,*}, \varepsilon_n) + o_p(n^{-1/2}) + o_p(\|J_n\|) = \hat{J}^{*, i}_{i,n,t}(\tilde{\theta}_{n,*}, \varepsilon_n) + o_p(\|J_n\|),
\]

25
where $n^{-1/2} = o(||J_n||)$ under Jacobian non-degeneracy D6.i.

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

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

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

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

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

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

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

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

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

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

\[\blacksquare\]

**Proof of Lemma 3.1.** We treat the cases $\kappa = 2$ and $\kappa > 2$ separately.

**Case 1** $(\max\{\kappa, \kappa_2, \ldots, \kappa_r\} \leq 2)$: We require some properties of regularly varying tails. Since $\varepsilon_i$ and stochastic $x_{i,t}$ are mutually independent with tails $(7)$ and indices $\kappa_i$ and $\kappa$, the convolutions $\varepsilon_t x_{i,t}$ satisfy (Cline 1986)

$$m_{i,t}(\theta^0) = \varepsilon_t x_{i,t} \sim (7) \text{ with index } \kappa_{e,i} : = \min \{ \kappa, \kappa_i \}. \quad (10)$$

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

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

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

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

Express $n_{\theta_i}$ as

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

26
Step 1 (\(\Sigma_n, \Gamma_n\)): Properties (10)-(12) imply
\[
\Sigma_{i,i,n} = E \left[ m_{i,n}^2 (\theta^0) \right] \sim c_{i,n}^2 P \left( |m_{i,t}(\theta^0)| > c_{i,n} \right) \sim c_{i,n}^2 \frac{k_n}{n} = K \left( \frac{n}{k_n} \right)^{2/\kappa_{i,i} - 1}.
\]
Hence \(\Gamma_{i,i,n} = (n/k_n)^{1/\kappa_{i,i} - 1/2}\).

Step 2 (\(J_n\)): The Lemma C.1.c Jacobian approximation implies by the linear equation form \(J_{i,j,n} = -E[x_{i,t} x_{j,t} I(\epsilon_t x_{j,t} \leq c_{j,n})] \times (1 + o(1))\). Assume initially all regressors are stochastic. Since \(\kappa_i, \kappa_e \leq 2\) it follows \(\kappa_i < \kappa_e + 2\). Thus, by independence, (10) and (12)
\[
E \left[ x_{i,t}^2 I \left( |\epsilon_t x_{i,t} | \leq c_{i,n} \right) \right] = K \int \left[ \frac{c_{i,n}^2}{\epsilon^2} P \left( |x_{i,t}| > \frac{c_{i,n}}{\epsilon} \right) \right] f_{\epsilon} \left( d\epsilon \right)
\sim K \int E \left[ \frac{c_{i,n}^2}{\epsilon^2} \left( \frac{c_{i,n}}{\epsilon} \right)^{-\kappa_i} \right] f_{\epsilon} \left( d\epsilon \right)
= K c_{i,n}^{2-\kappa_i} E \left[ \epsilon_t^{\kappa_i/2} \right] = K c_{i,n}^{2-\kappa_i} \left( \frac{n}{k_n} \right)^{(2-\kappa_i)/\kappa_{i,i}}.
\]
The cross-products \(E[x_{i,t} x_{j,t} I(\epsilon_t x_{j,t} \leq c_{j,n})]\) are bounded by case. If \(x_{i,t}\) and \(x_{j,t}\) are independent then \(E[x_{i,t} x_{j,t} I(\epsilon_t x_{j,t} \leq c_{j,n})] \sim K \sup_{\epsilon \in \mathbb{E}} E|\epsilon_t x_{i,t}| < \infty \forall i\). If they are perfectly positively dependent with marginal tail (7) and \(\kappa_i, \kappa_j \leq 2\) it can only be the case that \(x_{i,t} = \text{sign}(x_{j,t}) \times |x_{j,t}|^p\) where \(p = \kappa_j/\kappa_i\). Therefore, since \(\kappa_j - p - 1 < \kappa_e\) is easy to verify,
\[
E \left[ x_{i,t} x_{j,t} I \left( |\epsilon_t x_{j,t} | \leq c_{j,n} \right) \right] = \int E \left[ (x_{j,t})^{p+1} I \left( |x_{j,t}|^{(p+1)/2} \leq \left( \frac{c_{j,n}}{\epsilon} \right)^{(p+1)/2} \right) \right] f_{\epsilon} \left( d\epsilon \right)
= K \int \left[ \left( \frac{c_{j,n}}{\epsilon} \right)^{p+1} P \left( |x_{j,t}| > \frac{c_{j,n}}{\epsilon} \right) \right] f_{\epsilon} \left( d\epsilon \right)
\sim K \int E \left[ \left( \frac{c_{j,n}}{\epsilon} \right)^{p+1} \left( \frac{c_{j,n}}{\epsilon} \right)^{-\kappa_j} \right] f_{\epsilon} \left( d\epsilon \right)
= K c_{j,n}^{p+1-\kappa_j} \int E \left[ \epsilon_t^{\kappa_j-1} \right] f_{\epsilon} \left( d\epsilon \right)
= K c_{j,n}^{p+1-\kappa_j} = K \left( \frac{n}{k_n} \right)^{(\kappa_j/\kappa_{i,i} + 1 - \kappa_j)/\kappa_{j,j}}.
\]
The perfect negative dependence case is similar. Hence \(J_{i,j,n} = O((n/k_n)^{\kappa_{i,i}+1}(\kappa_j/\kappa_{i,i}+1-\kappa_j))\).

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

Step 3 (\(n_0\)): Consider the slope rates, the intercept rate being similar. The claim follows from (13) by noting
\[
\Gamma_{i,i,n}^{-1} J_{i,i,n} = (n/k_n)^{1/2-\kappa_{i,i}} (n/k_n)^{(2-\kappa_i)/\kappa_{i,i}} = (n/k_n)^{1/2+(1-\kappa_i)/\kappa_{i,i}}.
\]
and
\[
\max_{j \neq i} \{ \Gamma_{j,i,n}^{-1} J_{j,i,n} \} = \frac{O \left( \max_{j \neq i} \left\{ (n/k_n)^{1/2-1/(\kappa_{i,j}+1)} \right\} \right)}{(n/k_n)^{1/2+1/(\kappa_{i,j}-\min_{j \neq i} (1/(\kappa_{i,j}+1/(\kappa_{i,j}-1))} \right) = \frac{O \left( (n/k_n)^{1/2+1/(\kappa_{i,j}-1)} \right)}{(n/k_n)^{1/2+1/(\kappa_{i,j}-1)}} \times (n/k_n)^{1/2-1/(\kappa_{i,j}+1)} \}.
\]

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

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

If $\kappa_i > 2$ then $J_{i,j,n} = K(n/k_n)^{1/\kappa_i-1}$, $J_{i,j,n} = O((n/k_n)^{1/\kappa_i+1/\kappa_i-1})$, and $J_{i,j,n} = O(\left( (n/k_n)^{\kappa_j^{-1} \left( \kappa_j+1-\kappa_i \right)} \right)$. The rate can be deduced from (13) by noting if $\kappa_i > 2$ then
\[
\Gamma_{i,i,n}^{-1} J_{i,i,n} = 1 \times K, \quad \max_{j \neq i, \kappa_j \geq 2} \left\{ \Gamma_{j,j,n}^{-1} J_{j,j,n} \right\} = K, \quad \max_{j \neq i, \kappa_j \leq 2} \left\{ \Gamma_{j,j,n}^{-1} J_{j,j,n} \right\} = O \left( (n/k_n)^{1/\kappa_j-1/2} \right)
\]

hence $n_{\theta_1} \sim n^{1/2} \times \left[ K + O(\max_{j \neq i, \kappa_j \leq 2} \{(n/k_n)^{2/(\kappa_j-1)}\}) \right]^{1/2}.
\]

The rate under $\kappa_i \leq 2$ follows similarly since $\Gamma_{i,i,n} J_{i,i,n} = (n/k_n)^{1/\kappa_i-1}$ and max $\max_{j \neq i, \kappa_j \leq 2} \left\{ \Gamma_{j,j,n}^{-1} J_{j,j,n} \right\} = O(\max_{j \neq i, \kappa_j \leq 2} \{(n/k_n)^{1/2-1/(\kappa_j+1)} \kappa_j^{-1} (\kappa_j+1-\kappa_i) \}).$


**APPENDIX B : GMTTM Criterion Properties**

**LEMMA B.1 (uniform criterion law)** Under $D1-D5$ and $D7$
\[
\sup_\theta \left\{ \frac{m_n^{-2} \times \bar{Q}_n(\theta) - Q_n(\theta)}{1 + m_n^{-2} \times Q_n(\theta)} \right\} = o_p(1) \quad \text{where} \quad m_n := \sup_\theta E \left[ ||m_n^*(\theta)|| \right].
\]

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

Proof of Lemma B.1. Write $\hat{m}_n^*(\theta) := 1/n \sum_{i=1}^n \hat{m}_{n,t}^*(\theta)$ and $m_n^*(\theta) := 1/n \sum_{i=1}^n m_{n,t}^*(\theta)$. By weight property M1 and the triangular inequality
\[
m_n^{-2} \left| \bar{Q}_n(\theta) - Q_n(\theta) \right| \leq m_n^{-2} \left| \bar{m}_n^*(\theta) \right|^2 \times ||\hat{\Upsilon}_n - \Upsilon_n|| + m_n^{-2} \left| \hat{m}_n^*(\theta) \hat{\Upsilon}_n \hat{m}_n^*(\theta) - Q_n(\theta) \right|
\]
\[
\leq \left| \hat{m}_n^*(\theta) \right|^2 + K \left| \hat{m}_n^*(\theta) - m_n^*(\theta) \right|^2
\]
\[
+ K m_n^{-2} \left| m_n^*(\theta) \right| \times \left| \hat{m}_n^*(\theta) - m_n^*(\theta) \right| + m_n^{-2} \left| m_n^*(\theta) \hat{\Upsilon}_n \hat{m}_n^*(\theta) - Q_n(\theta) \right|
\]
\[
= A_{1,n}(\theta) + A_{2,n}(\theta) + A_{3,n}(\theta) + A_{4,n}(\theta).
\]
Uniform approximation Lemma C.2 and \( \lim \inf_{n \geq N} m_n > 0 \) under smoothness I3 imply

\[
\sup_{\varrho} \{ A_{1,n}(\theta) \}^{1/2} = \sup_{\varrho} \left\{ \frac{\| \hat{m}_n^*(\theta) \|}{\sup_{\varrho} \| E[ m_n^*(\theta) ] \|} \right\} \times o_p(1)
\]

\[
\leq \sup_{\varrho} \frac{\| m_n^*(\theta) \|}{\sup_{\varrho} \| E[ m_n^*(\theta) ] \|} \times o_p(1) + o_p(1)
\]

\[
\leq \sup_{\varrho} \frac{\| m_n^*(\theta) \| - E[ m_n^*(\theta) ] \|}{\sup_{\varrho} \| E[ m_n^*(\theta) ] \|} \times o_p(1) + o_p(1).
\]

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

Finally, under M1

\[
A_{4,n}(\theta) = m_n^{-2}(\theta) \gamma_n m_n^*(\theta) - E[ m_n^*(\theta) ] \gamma_n E[ m_n^*(\theta) ]
\]

\[
\leq K m_n^{-2} \| m_n^*(\theta) - E[ m_n^*(\theta) ] \|^2 + K m_n^{-2} \| E[ m_n^*(\theta) ] \| \times \| m_n^*(\theta) - E[ m_n^*(\theta) ] \|
\]

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

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

### APPENDIX C: Limit Theory for Trimmed Sums

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

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

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

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

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

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

---

\(^8\)In this expansions, approximation and indicator bounds Lemmas C.1, C.2 and C.4 are trivial; the Lemma C.3 uniform law follows from Andrews (1992) in lieu of mixing D3; differentiability implies asymptotic linearity Lemma C.5; McLeish’s (1974) central limit theory for martingale differences suffices for Lemma C.6; and so on.
LEMMA C.2 (approximation) Under D1-D5 and D7 \( \| \sum_{t=1}^{n} (\tilde{m}_{n,t}^{*}(\theta) - m_{n,t}^{*}(\theta)) \| = o_p(n^{1/2}\|\Sigma_n^{-1/2}\|^{-1}) \) for any \( \theta \in \Theta \), and \( \sup_{\theta} \{ \| 1/n \sum_{t=1}^{n} (\tilde{m}_{n,t}^{*}(\theta) - m_{n,t}^{*}(\theta)) \| \} = o_p(1) \).

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

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

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

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

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

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

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

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

\[
V_{n}^{1/2} (\hat{\theta}_n - \theta^0) = A_{n} \sum_{t=1}^{n} \tilde{m}_{n,t}^{*}(\theta^0) \times (1 + o_p(1)) + o_p(1) \text{ a.s.}
\]

where \( A_{n} = -V_{n}^{1/2} (H_{n}^{-1} J_{n}^{r} \gamma_{n}) n^{-1} \in \mathbb{R}^{r \times q} \).

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

\( \Rightarrow \) \( N(0, 1) \).

Stochastic differentiability aids proving Jacobian estimator consistency.

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

\[
\sup_{\theta \in U(\delta_n)} \left\{ \frac{\| V_{n}^{1/2} \|}{\| J_{n} \|} \| \tilde{m}_{n}^{*}(\theta) - \tilde{m}_{n}^{*}(\theta^0) \| - \{ E [m_{n,t}^{*}(\theta^0)] - E [m_{n,t}^{*}(\theta^0)] \} \right\} \rightarrow 0.
\]

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

Proof of Lemma C.1.

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

\[
m_{n,t}^{*}(\theta) = \left\{ m_{t}(\hat{\theta}) + J_{t}(\theta, \delta)(\theta - \hat{\theta}) \right\} \times I_{n,t}(\theta)
\]
where \( || \theta_n, \delta - \hat{\theta} || \leq || \theta - \hat{\theta} || \), and \( J_t(\theta) := (\partial / \partial \theta)m_t(\theta) \). Therefore

\[
m^*_{m}(\theta) - m^*_{n}(\hat{\theta}) = J^*_{m}(\theta_n, \delta) \times (\theta - \hat{\theta}) + \frac{1}{n} \sum_{t=1}^{n} m_t(\theta) \times \left\{ I_{n,t}(\theta) - I_{n,t}(\hat{\theta}) \right\} \quad (14)
\]

+ \frac{1}{n} \sum_{t=1}^{n} J_t(\theta_n, \delta) \times \left\{ I_{n,t}(\theta) - I_{n,t}(\theta_n, \delta) \right\} \times (\theta - \hat{\theta}) .

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

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

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

\[
\left( E \left[ A_n(\theta, \hat{\theta})^\epsilon \right] \right)^{1/\epsilon} \leq n^{1/2} \left[ E \left| m_1(\theta) \left\{ I_{n,t}(\theta) - I_{n,t}(\hat{\theta}) \right\} \right|^{1/\epsilon} \right] = O \left( n^{1/2} (k_n/n)^{1/\epsilon} \right) .
\]

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

\[
A_n(\theta, \hat{\theta}) = o_p \left( n^{1/2} \left( k_n^{1/2}/n \right)^{1/\epsilon} \right) = r_n \times o_p(1) .
\]

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

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

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

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

Further, moment expansion Lemma B.2 asserts

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

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

The proof of approximation Lemma C.2 requires consistency of the intermediate order statistics \( m_{i,(k_1,i,n)}(\theta) \) and \( m_{i,(k_2,i,n)}(\theta) \). Simplify notation by considering the two-tailed \( m_{i,(k,i,n)}^\epsilon(\theta) \) which estimates \( c_{i,n}(\theta) \).
Lemma C.2.1 Under D1-D5 and D7 \( m_{(a)}(\theta)/c_{i,n}(\theta) = 1 + O_p(k_n^{-1/2}) \) for each \( \theta \in \Theta \), and \( \sup_{\theta} |m_{(a)}^{(a)}(\theta)/c_{i,n}(\theta) - 1| = O_p(k_n^{-1/2} n') \) for infinitessimal \( \iota > 0 \).


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

Claim 1: Let \( \theta \in \Theta \) be arbitrary, and write \( m_t = m_t(\theta), c_n = c_n(\theta), \hat{m}_n(t) = \hat{m}_n(t, \theta), m_n = m_n(\theta), \bar{I}_{n,t} = 1 - I_{n,t}(\theta), \bar{I}_{n,t} = \bar{I}_{n,t}(\theta), \text{ and } \bar{\sigma}_n := \|\Sigma_n^{-1/2}(\theta)\|^{-1} \). First bound
\[
\left| \sum_{t=1}^{n} \{ \hat{m}_{n,t} - m_{n,t} \} \right| \leq \max_{1 \leq t \leq n} \left\{ \left\| m_t \left( I_{n,t} - I_{n,t} \right) \right\| \right\} \times \sum_{t=1}^{n} \left| \bar{I}_{n,t} - I_{n,t} \right|.
\]
By construction \( \|m_t(I_{n,t} - I_{n,t})\| \leq 2\|m_{n}(I) - c_n\| \). The intermediate order statistic is consistent \( m_{n}(I)/c_n = 1 + O_p(k_n^{-1/2}) \) by Lemma C.2.1. Now use threshold bound D5 to deduce
\[
\max_{1 \leq t \leq n} \left\{ \left\| m_t \left( I_{n,t} - I_{n,t} \right) \right\| \right\} \leq 2 \left\| m_{n}(I) - c_n \right\| = 2c_n \left\| m_{n}(I) - c_n - 1 \right\| = o_p \left( \bar{\sigma}_n (n/k_n)^{1/2} \right).
\]
Next, by construction and the triangular inequality
\[
\sum_{t=1}^{n} \left| \bar{I}_{n,t} - I_{n,t} \right| \leq k_n^{1/2} \left\| \frac{1}{k_n^{1/2}} \sum_{t=1}^{n} \{ \bar{I}_{n,t} - E [\bar{I}_{n,t}] \} \right\| + k_n^{1/2} \left\| \frac{1}{k_n^{1/2}} \sum_{t=1}^{n} \{ \bar{I}_{n,t} - E [\bar{I}_{n,t}] \} \right\|.
\]
which is \( O_p(k_n^{1/2}) \) by D5 and an application of Lemma C.4.b. Therefore \( \sum_{t=1}^{n} \{ \hat{m}_{n,t} - m_{n,t} \} = o_p(\bar{\sigma}_n (n/k_n)^{1/2} k_n^{1/2}) = o_p(\bar{\sigma}_n n^{1/2}) \).

Claim 2: Define \( M_n := \max_{1 \leq t \leq n} \{ \sup_{\theta} \| m_\theta (\theta) I_{n,t}(\theta) - I_{n,t}(\theta) \| \} \) and \( \bar{\sigma}_n := \sup_{\theta} \|\Sigma_n^{-1/2}(\theta)\|^{-1} \) and repeat the above argument to reach
\[
\sup_{\theta} \left\| \frac{1}{n} \sum_{t=1}^{n} \{ \hat{m}_{n,t}(\theta) - m_{n,t}(\theta) \} \right\| \leq M_n \times \frac{k_n^{1/2}}{n} \sup_{\theta} \left\| \frac{1}{k_n^{1/2}} \sum_{t=1}^{n} \{ I_{n,t}(\theta) - E [\bar{I}_{n,t}(\theta)] \} \right\| + M_n \times \frac{k_n^{1/2}}{n} \sup_{\theta} \left\| \frac{1}{k_n^{1/2}} \left( \frac{n}{k_n} E [\bar{I}_{n,t}(\theta)] - 1 \right) \right\|.
\]
Uniform indicator law Lemma C.4.b and uniform probability orders D5 imply the right-hand-side is \( O_p(M_n k_n^{1/2}/n) \).

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

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

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

\[
P\left( \sup_\theta \sup_{|\theta - \hat{\theta}| \leq \delta} \left| m_{i,t}(\theta) \times \left( I_{i,n,t}(\theta) - I_{i,n,t}(\hat{\theta}) \right) \right| > \kappa \right) \leq \epsilon, 
\]

(15)

\[
P\left( \sup_\theta \sup_{|\theta - \hat{\theta}| \leq \delta} \left| m_{i,t}(\theta) - m_{i,t}(\hat{\theta}) \times I_{i,n,t}(\hat{\theta}) \right| > \kappa \right) \leq \epsilon. \tag{16}
\]

Since

\[
\sup_\theta \sup_{|\theta - \hat{\theta}| \leq \delta} \left| m_{i,t}(\theta) \left\{ I_{i,n,t}(\theta) - I_{i,n,t}(\hat{\theta}) \right\} \right|
\]

\[
\leq \sup_\theta \left| m_{i,t}(\theta) I_{i,n,t}(\theta) \right| \times \sup_\theta \sup_{|\theta - \hat{\theta}| \leq \delta} \left| I_{i,n,t}(\theta) - I_{i,n,t}(\hat{\theta}) \right|
\]

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

\[
P\left( \sup_\theta \sup_{|\theta - \hat{\theta}| \leq \delta} \left| m_{i,t}(\theta) - m_{i,t}(\hat{\theta}) \right| \times \left| I_{i,n,t}(\hat{\theta}) \right| > \kappa \right) \leq \frac{1}{\kappa^2} E \left[ \sup_\theta \left| \frac{\partial}{\partial \theta} m_{i,t}(\theta) \right|^{\epsilon} \right] \times \left| \theta - \hat{\theta} \right|^{\epsilon} \leq K \times \delta^{\epsilon} \to 0
\]

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

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

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

and

\[
\lim_{\delta \to 0} \lim_{n \to \infty} \left\{ \zeta_n \times P \left( \sup_\theta \sup_{|\theta - \hat{\theta}| \leq \delta} \left| M^*_{n,t}(\theta) - M^*_{n,t}(\hat{\theta}) \right| > \epsilon \right) \right\} = 0 \tag{18}
\]

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

**Proof.** The proof is similar to Andrews (1992: Lemma 3) except we bypass uniform integrability with the weaker (17). Define \( s^*_{n,t} := \sup_\theta |M^*_{n,t}(\theta)| \) and \( M^*_{n,t}(\delta) := \sup_{|\theta - \hat{\theta}| \leq \delta} |M^*_{n,t}(\theta)| - |M^*_{n,t}(\hat{\theta})| \). For given \( \epsilon > 0 \) by supposition there exists \( \delta > 0 \) such that, uniformly in \( n \),

\[
E[2s^*_{n,t} I(2s^*_{n,t} > \zeta_n)] \leq \epsilon^2/6 \quad \text{and} \quad P(M^*_{n,t}(\delta) > \epsilon^2/6) \leq \epsilon^2/[6\zeta_n].
\]

Then by stationarity, Marvok’s inequality and term-wise stochastic equicontinuity (18), for sufficiently large \( n \)
Proof of Lemma C.3. (e.g. Andrews 1992: p. 250)

The first term on the right-hand-side is

This proves the claim. ■

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

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

\[
E(X_n^*(\theta, a_n))^2 \leq \frac{1}{n} E \left[ (X_{n,t}^*(\theta, a_n))^2 \right] + \frac{1}{n} \sum_{i=2}^\infty E \left[ X_{n,1}(\theta, a_n) X_{n,t}^*(\theta, a_n) \right]
\]

\[
\leq K \left\{ \frac{1}{n} E \left[ (X_{n,t}^*(\theta, a_n))^2 \right] + \frac{1}{n} \sum_{i=2}^\infty \rho^{(i-1)\times i} \left( E \left| X_{n,1}(\theta, a_n) \right|^{2+i} \right)^{1-i} \right\}
\]

\[
\leq K \left\{ \frac{1}{n} E \left[ (m_{i,n,t}^*(\theta))^2 \right] + \frac{1}{n} \left( E \left| m_{i,n,t}^*(\theta) \right|^{2+i} \right)^{1-i} \right\}.
\]

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

Thus by Chebyshev’s inequality \( X_n^*(\theta, 1) \overset{P}{\to} 0 \), and since \( \liminf_{n \to \infty} \{m_n\} > 0 \) under I3 also \( X_n^*(\theta, m_n) \overset{P}{\to} 0 \).

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

Consider two possibly over-lapping cases. First, if \( \lim_{n \to \infty} E[\sup_{\theta} |M_{n,t}^*(\theta)|] < \infty \) then \( \lim_{n \to \infty} E[\sup_{\theta} |M_{n,t}^*(\theta)| \times I(\sup_{\theta} |M_{n,t}^*(\theta)| > \zeta)] \leq K(\zeta) \) for all \( \zeta > 0 \) and some mapping \( K(\zeta) \to 0 \) as \( \zeta \to \infty \). Further, Andrews’ (1992) Assumption TSE holds by Lemma C.3.1. Therefore stochastic equicontinuity follows from Lemma 3 of Andrews (1992).
Second, if \( m_n := \sup_{\theta} E[|m^*_{n,t}(\theta)|] \to \infty \) then we need only verify conditions (17) and (18) of Lemma C.3.2. Define for arbitrary \( \epsilon > 0 \)

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

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

\[
P \left( \sup_{\bar{\theta}} \sup_{|\theta - \bar{\theta}| \leq \delta} \left| M^*_{n,t}(\theta) - M^*_{n,t}(\bar{\theta}) \right| > \epsilon \right) \leq K \frac{E \left( \sup_{\theta} |m^*_{n,t}(\theta)| \right)}{m_n} \leq K m_n^{-\iota} = o(1).
\]

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

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

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

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

\[
P \left( \sup_{\bar{\theta}} \sup_{|\theta - \bar{\theta}| \leq \delta} \left| m_t(\theta) - m_t(\bar{\theta}) \right| > \epsilon \right) \leq K \times E \left[ \sup_{\theta} \left| \frac{\partial}{\partial \theta} m_t(\theta) \right| \right] \times \delta^2 \leq \kappa.
\]

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

\[
P_n(\delta) := P \left( \sup_{\bar{\theta}} \sup_{|\theta - \bar{\theta}| \leq \delta} \left| I(|m_t(\theta)| \leq c_n(\theta)) - I \left( |m_t(\bar{\theta})| \leq c_n(\bar{\theta}) \right) \right| = 1 \right)
\]

\[
\leq P \left( \exists \theta, \bar{\theta} \in \Theta, |\theta - \bar{\theta}| \leq \delta : |m_t(\theta)| \leq c_n(\theta) \text{ and } c_n(\bar{\theta}) < |m_t(\bar{\theta})| \right)
\]

\[
+ P \left( \exists \theta, \bar{\theta} \in \Theta, |\theta - \bar{\theta}| \leq \delta : |m_t(\bar{\theta})| \leq c_n(\bar{\theta}) \text{ and } c_n(\theta) < |m_t(\theta)| \right)
\]

\[
\leq 2 \times P \left( \exists \theta \in \Theta : |m_t(\theta)| > c_n(\theta) \right) \left\{ [1 - \epsilon] - \frac{K\delta}{\sup_{\theta} \{c_n(\theta)\}} \right\}
\]

\[
- 2 \times P \left( \exists \theta \in \Theta : |m_t(\theta)| \geq c_n(\theta) \right) \left\{ [1 + \epsilon] + \frac{K\delta}{\inf_{\theta} \{c_n(\theta)\}} \right\} .
\]

But this implies \( P_n(\delta) = O(k_n/n) = o(1) \leq \epsilon \) for each \( \epsilon \) and sufficiently large \( n \) by D5, non-uniqueness of the thresholds \( \{c_n(\theta)\} \) and \( k_n/n \to 0 \). Since \( \epsilon \) is arbitrary the proof is complete.
Claim (b): Recall $X^*_n, \theta(t) := ((n/k_n)^{1/2+\epsilon}) \{ I_{n,t,\theta(t)} - E[I_{n,t,\theta(t)}] \}$ for tiny $\epsilon > 0$. 
$\{X^*_n, \theta(t) : \theta \in \Theta \}$ forms a VC class under $D_7$, while mixing $D_3$ and probability orders $D_5$ ensure by measurability $\{X^*_n, \theta(t)\}$ is geometrically $\beta$-mixing and $L_{2+\epsilon}$-bounded uniformly on $1 \leq t \leq n$, $n \geq 1$ and $\theta \in \Theta$. It is now straightforward to extend Arcones and Yu’s (1994: Theorem 2.1) uniform central limit theorem for stationary $\beta$-mixing sequences to geometrically $\beta$-mixing triangular arrays $\{X^*_n, \theta(t) : 1 \leq t \leq n\}$ stationary over $t$: 
\[
\{ n^{-1/2} \sum_{t=1}^n X^*_n, \theta(t) \} \Rightarrow \{ X(t) : \theta \in \Theta \} \text{ a Gaussian process with uniformly bounded and uniformly continuous sample paths with respect to } L_2\text{-norm.}
\]
Therefore $E[\{ \sup_{\theta \in \Theta} \{ n^{-1/2} \sum_{t=1}^n X^*_n, \theta(t) \} \}] = O(1)$ by a straightforward generalization of Doukhan et al’s (1995: Theorem 2) maximal inequality in lieu of mixing $D_2$ and VC class $D_7$.$^6$

Proof of Lemma C.5. Apply Čížek’s (2008: Lemma 2.1) argument to deduce absolute continuity of the equation distributions $D_1$ and equation differentiability $D_2$ ensures $\bar{Q}_n(\theta)$ is continuous and differentiable at $\theta_n$. Therefore $\bar{Q}_n(\theta_n) \leq \bar{Q}_n(\theta) \forall \theta \in \Theta$ implies
\[
\hat{J}_n^*(\theta_n)^T \hat{Y}_n \left\{ \hat{J}_n^*(\theta_n, \theta) ^T (\hat{\theta}_n - \theta) + \frac{1}{n} \sum_{t=1}^n \hat{m}_{n,t}^*(\theta(t)) \right\} + \hat{J}_n^*(\theta) \hat{Y}_n \times o_p(r_n) = 0 \ a.s.
\]
for some $||\theta_n, \theta - \theta^0|| \leq ||\hat{\theta}_n - \theta^0||$ and $r_n \to 0$ arbitrarily fast.

Consistency $||\hat{\theta}_n - \theta^0|| \overset{P}{\to} 0$ under Theorem 2.1 and Jacobian consistency Lemma 2.5 ensure both $\hat{J}_n^*(\theta_n) = J_n(1 + o_p(1))$ and $\hat{J}_n^*(\theta_n,s) = J_n(1 + o_p(1))$. Further $H_n^{-1} := (J_n^* \hat{Y}_n J_n)^{-1}$ exists given weight and Jacobian properties $M_1$ and $D_6.1$. Re-arrange terms and exploit the construction of $V_n$ and $r_n \to 0$ arbitrarily fast to deduce
\[
V_n^{1/2} (\hat{\theta}_n - \theta^0) = - \left\{ V_n^{1/2} H_n^{-1} J_n^* \hat{Y}_n \right\} \frac{1}{n} \sum_{t=1}^n \hat{m}_{n,t}^*(\theta(t)) \times (1 + o_p(1)) + o_p(1)
\]
\[
= A_n \sum_{t=1}^n \hat{m}_{n,t}^*(\theta(t)) \times (1 + o_p(1)) + o_p(1).
\]

In order to prove central limit theorem Lemma C.6 we need to define a mixingale property for tail-trimmed arrays. Define for arbitrary $\xi \in \mathbb{R}^q$, $\xi^t \xi = 1$:
\[
z_{n,t}^*(\xi, \theta) = \xi^t \{ n^{-1/2} \sum_{t=1}^n m_{n,t}^*(\theta(t)) \}
\]
Orthogonality $I_4$ and $||\Sigma_n^{-1/2}||^{-2} \leq K||\Sigma_n||$ ensure
\[
E \left[ \sum_{t=1}^n m_{n,t}(\theta(t)) \sum_{t=1}^n m_{n,t}(\theta(t))' \right] = n \Sigma_n + \sum_{s \neq t} E \left[ m_{n,s}(\theta(t)) m_{n,t}(\theta(t))' \right] = n \Sigma_n (1 + o(1))
\]
hence $E[(z_{n,t}^*(\xi, \theta(t))]^2) = n^{-1/2} an E[\sum_{t=1}^n z_{n,t}(\xi, \theta(t)) \sum_{t=1}^n z_{n,t}(\xi, \theta(t))'] = 1 + o(1)$.
Although \( \{z_{n,t}^*(\xi, \theta^0)\} \) is \( L_2 \)-bounded geometrically \( \beta \)-mixing due to trimming and mixing D3, it is not necessarily uniformly square integrable since that requires uniformity over \( n \). We therefore exploit central limit theory for \( L_2 \)-mixingale arrays \( \{x_{n,t}, \mathcal{F}_n\} \) with size \( \lambda \); there exists an array of positive constants \( \{e_{n,t}\} \) and a sequence of positive coefficients \( \{\zeta_q\} \), \( \zeta_q = o(q^{-\lambda}) \) such that

\[
\begin{align*}
\left(E \left( x_{n,t} - E \left[ x_{n,t} \mid \mathcal{F}_n \right] \right)^2 \right)^{1/2} &\leq e_{n,t} \times \zeta_{q+1} = e_{n,t} \times o(q^{-\lambda}) \quad \text{(MA)} \\
\left(E \left( E \left[ x_{n,t} \right] - E \left[ x_{n,t} \mid \mathcal{F}_n \right] \right)^2 \right)^{1/2} &\leq e_{n,t} \times \zeta_q = e_{n,t} \times o(q^{-\lambda}).
\end{align*}
\]

Since \( \beta \)-mixing implies \( \alpha \)-mixing, invoke D3 and Lemma 2.1 of McLeish 1975) to deduce \( \{z_{n,t}^*(\xi, \theta^0), \mathcal{F}_n\} \) satisfies (MA) with constants \( e_{n,t} = Kn^{-1/2}||\Sigma_n^{-1/2}|| \times (E||m_{n,t}^*||^{2+\delta})^{1/(2+\delta)} \) for tiny \( \delta > 0 \) and arbitrary size \( \lambda > 0 \).

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

\[
h_n = o(e_i^d) \quad \text{for tiny } i > 0,
\]

then \( d_n = [n/h_n] \), and \( j_n = o(h_n) \), and define \( \Gamma_{n,i} := \sigma(\cup_{r \leq i} \mathcal{F}_r) \) and blocks

\[
Z_{n,i}(\theta) = \sum_{t=(i-1)h_n+j_n}^{ih_n} z_{n,t}^*(\xi, \theta) \quad \text{and} \quad W_{n,i}(\theta) = E[Z_{n,i}(\theta) \mid \mathcal{F}_{n,i}] - E[Z_{n,i}(\theta) \mid \mathcal{F}_{n,i-1}].
\]

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

**Lemma C.6.1** If \( \{x_{n,t}, \mathcal{F}_n\} \) forms an \( L_2 \)-mixing array with size \( \lambda = 1/2 \) then

\[
E[\max_{1 \leq j \leq n}(\sum_{i=1}^{j} \{y_{n,t} - E[y_{n,t}]\}]^2 = O(\sum_{i=1}^{n} E[y_{n,t}^2]).
\]

**Remark:** Notice the upper bound is based on \( E[y_{n,t}^2] \) rather the constants \( c_{n,t}^2 \) as in McLeish (1975).

Define

\[
\sigma_n^2 := n \times \left\| \Sigma_n^{-1/2} \right\|^{-2}, \quad u_{n,i}^2 := \sigma_n^2 W_{n,i}, \quad \text{and} \quad K_n \sim K_{n+1},
\]

and a truncation function

\[
\tilde{u}_{n,i}^2(K) := u_{n,i}^2 I(u_{n,i}^2 \leq K) = \sigma_n^2 W_{n,i} I(\sigma_n^2 W_{n,i} \leq K) \quad \text{and} \quad \tilde{u}_{n,i}^2 = \tilde{u}_{n,i}^2(K_n).
\]

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

**Lemma C.6.2** \( \sigma_n^{-2} \sum_{i=1}^{d_n} E[\tilde{u}_{n,i}^2 - u_{n,i}^2] = o(1) \) and \( \tilde{\sigma}_n^{-2} \sum_{i=1}^{d_n} (\tilde{u}_{n,i}^2 - u_{n,i}^2) = o_p(1) \).

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

**Lemma C.6.4** \( \|\sum_{i=1}^{d_n}(W_{n,i}^2 - Z_{n,i}^2)\|_2 = o(1) \).

**Lemma C.6.5** \( \sum_{i=1}^{d_n} E[Z_{n,i}^2] \overset{p}{=} 1 \).
Proof of Lemma C.6. We will show \( \sum_{i=1}^{n} z_{n,t}^* (\xi, \theta^0) = \sum_{i=1}^{d_n} W_{n,i} (\theta^0) + o_p(1) \) and \( \sum_{i=1}^{d_n} W_{n,i} (\theta^0) \xrightarrow{d} N(0, 1) \). The claim then follows by the Cramér-Wold theorem. Drop the arguments \( \xi \) and \( \theta^0 \).

Step 1 (decomposition): Decompose

\[
\sum_{t=1}^{n} z_{n,t}^* = \sum_{i=1}^{d_n} W_{n,i} + \sum_{i=1}^{d_n} (Z_{n,i} - E[Z_{n,i}|F_{n,i}]) + \sum_{i=1}^{d_n} E[Z_{n,i}|F_{n,i-1}]
\]

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

\( \{z_{n,t}, \mathcal{F}_t\} \) satisfies (MA) with constants \( c_{n,t} = Kn^{-1/2}||\Sigma^{-1/2}|| \times (E||m_{n,t}^*||^2)^{1/(2+\delta)} \) and coefficients \( \psi_q = o(q^{-\lambda}) \) for any \( \lambda > 0 \). Define the index set \( B_{n,t} = \{t : t \in (i-1)h_n + j_n + 1, \ldots, ih_n]\) It can similarly be shown \( \{z_{n,t}^* - E[z_{n,t}^*|F_{n,i}], \mathcal{F}_t\}_{t \in B_{n,t}} \) satisfy (MA) with constants \( c_{n,t} \psi_j \) and coefficients \( \psi_j^{1-\delta} \) of arbitrary size. (de Jong 1997: A.7-A.12)

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

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

since 1 - \( n/h_n \) tends to 0 arbitrarily fast. Further, \( E[(\sum_{i=1}^{d_n} \sum_{t=(i-1)h_n + j_n}^{ih_n} z_{n,t}^*)^2] = O(\sum_{i=1}^{d_n} \sum_{t=(i-1)h_n + j_n}^{ih_n} E[|z_{n,t}^*|^2]) = O(d_n j_n/h_n) = o(1) \). Similarly, \( E[(\sum_{i=1}^{d_n} \sum_{t=(i-1)h_n + j_n}^{ih_n} z_{n,t}^* - E[Z_{n,i}|F_{n,i}])^2] = O(d_n \sum_{i=1}^{d_n} \sum_{t=(i-1)h_n + j_n}^{ih_n} c_{n,t}^2 \psi_j) = o(1) \), and \( E[|\sum_{i=1}^{d_n} E[Z_{n,i}|F_{n,i-1}]|]^2 = o(1) \). Therefore \( E_n = o_p(1) \) by Chebyshev’s inequality.

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

Step 2.1 (LLN \( \sum_{i=1}^{d_n} W_{n,i} \overset{P}{\rightarrow} 1 \)): By the triangular inequality \( |\sum_{i=1}^{d_n} W_{n,i} - 1| \leq \sum_{i=1}^{d_n} E_{n,i} \) where

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

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

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

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

Step 2.2 (Lindeberg \( \sum_{i=1}^{d_n} E[|W_{n,i}^2 I(|W_{n,i} > \epsilon)|] \rightarrow 0 \)): We first show \( E[|W_{n,i}^2 \hat{\sigma}_n^2 I(W_{n,i} > \epsilon)|^2] \leq K\epsilon_n h_n + K\epsilon_n^2 u^{-2/3} du \) for any \( \epsilon > 0 \), tiny \( \epsilon > 0 \) and some \( \kappa \in (0, 2] \). Use sub-
additivity and the triangular inequality to deduce

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

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

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

Next, trimming implies \( |W_{n,i} - Z_{n,i}| \hat{\sigma}_n \leq Kh_n c_n \). By Markov’s inequality

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

Since \( h_n = o(\epsilon_n) \) from (19) the right-hand-side is 0 for large \( n \).

Now use stationarity, \( d_n h_n \sim n \), and covariance property (9.1) to deduce

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

Since (9.1) and D5 together imply \( \hat{\sigma}_n^2/c_n^2 \sim \infty \) for sufficiently tiny \( \epsilon > 0 \), the right-hand-side is 0 for large \( n \). \( \blacksquare \)

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

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

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

\[
\sup_{\theta \in \mathcal{U}_0(\delta_n)} \left\{ \left( \frac{\| V_n^{1/2} \|}{\| J_n \|} \right) \left\{ \| m^*_n(\theta) - m^*_n(\theta^0) \| - \left\{ E \left[ m^*_{n,t}(\theta) \right] - E \left[ m^*_{n,t}(\theta^0) \right] \right\} \right\} \right\} \\
\leq \sup_{\theta \in \mathcal{U}_0(\delta_n)} \left\{ \frac{\| J^*_n(\theta) - J_n \|}{\| J_n \|} \right\} + o_p(1)
\]

\[
\leq \sup_{\theta \in \mathcal{U}_0(\delta_n)} \left\{ \frac{\| J^*_n(\theta) - J_n \|}{\| J_n \|} \right\} + \left\{ \frac{\| J^*_n - J_n \|}{\| J_n \|} \right\} + o_p(1).
\]

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

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

**REFERENCES**


### Table 3: Location, AR

<table>
<thead>
<tr>
<th>Location</th>
<th>AR $\epsilon_t \sim P_{2,5}, \kappa_y = 2.5^a, \theta_r^a = 1^b$</th>
<th>Location</th>
<th>AR $\epsilon_t \sim P_{2,5}, \kappa_y = 1.5, \theta_r^a = .9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>Mean$^c$</td>
<td>$\lambda^d$</td>
<td>KS$^e$</td>
</tr>
<tr>
<td>GMTTM$^f$</td>
<td>.994 (.03)</td>
<td>.47</td>
<td>.101</td>
</tr>
<tr>
<td>GMM</td>
<td>1.00 (.08)</td>
<td>-</td>
<td>.076</td>
</tr>
<tr>
<td>QML</td>
<td>.999 (.04)</td>
<td>-</td>
<td>.121</td>
</tr>
</tbody>
</table>

### Table 4: ARCH

| $\epsilon_t \sim P_{1,5}, \kappa_y = 1.5, \theta_r^a = 1$ | $\epsilon_t \sim P_{1,5}, \kappa_y = 1.5, \theta_r^a = .9$ |
|----------|------------------------------------------------|----------|------------------------------------------------|---|
| $\theta$ | Mean $^c$ | $\lambda$ | KS | $\theta$ | Mean | $\lambda$ | KS |
| GMTTM | .994 (.13) | .22 | .051 | GMTTM | .900 (.01) | .32 | .069 |
| GMM | .939 (.21) | - | .172 | GMM | .900 (.01) | - | .187 |
| QML | 1.01 (.15) | - | .272 | QML | .889 (.01) | - | .253 |

---

a. True parameter value.
b. Error distribution ($P_{1,5}$ or $N_{0,1}$), and moment supremum or tail index of $y_t$:

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

### Table 4: ARCH

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

| $\epsilon_t \sim P_{2,5}, \kappa_y = 1.8, \theta_r^a = .6$ | $\epsilon_t \sim P_{2,1}, \kappa_y = 1.4, \theta_r^a = .6$ |
|----------|------------------------------------------------|----------|------------------------------------------------|---|
| $\theta$ | Mean $^c$ | $\lambda$ | KS | $\theta$ | Mean | $\lambda$ | KS |
| GMTTM | .602 (.38) | .50 | .093 | GMTTM | .582 (.34) | .38 | .105 |
| GMM | .547 (.22) | - | .289 | GMM | .542 (.26) | - | .213 |
| QML | .620 (.27) | - | .110 | QML | .683 (.18) | - | .414 |
TABLE 5 : GMMT Convergence Rate

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\theta^e$</th>
<th>$\kappa_s$</th>
<th>$\kappa_y$</th>
<th>$m_e(\theta)$</th>
<th>$b$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCATION</td>
<td>$P_{2.5}$</td>
<td>1</td>
<td>2.50</td>
<td>2.50</td>
<td>LS</td>
<td>.000</td>
<td>.031±.051</td>
</tr>
<tr>
<td>LOCATION</td>
<td>$P_{1.5}$</td>
<td>1</td>
<td>1.50</td>
<td>1.50</td>
<td>LS</td>
<td>-.129</td>
<td>-.130±.024</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>.9</td>
<td>2.50</td>
<td>2.50</td>
<td>LS</td>
<td>.000</td>
<td>.027±.039</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>.9</td>
<td>1.50</td>
<td>1.50</td>
<td>LS</td>
<td>.112</td>
<td>.116±.012</td>
</tr>
<tr>
<td>ARCH</td>
<td>$N_{0,1}$</td>
<td>.5</td>
<td>$\infty$</td>
<td>4.65</td>
<td>LS</td>
<td>.000</td>
<td>-.094±.162</td>
</tr>
<tr>
<td>ARCH</td>
<td>$N_{0,1}$</td>
<td>.6</td>
<td>$\infty$</td>
<td>3.80</td>
<td>QML</td>
<td>.000</td>
<td>.195±.422</td>
</tr>
<tr>
<td>ARCH</td>
<td>$P_{2.1}$</td>
<td>.6</td>
<td>2.5</td>
<td>1.80</td>
<td>QML</td>
<td>-.180</td>
<td>-.387±.306</td>
</tr>
<tr>
<td>ARCH</td>
<td>$P_{2.1}$</td>
<td>.6</td>
<td>2.5</td>
<td>1.40</td>
<td>QML</td>
<td>-.280</td>
<td>-.412±.393</td>
</tr>
</tbody>
</table>

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

TABLE 6 : GARCH, TARCH, QARCH

<table>
<thead>
<tr>
<th>Model</th>
<th>$\theta$</th>
<th>Mean</th>
<th>$\theta$</th>
<th>Mean</th>
<th>$\theta$</th>
<th>Mean</th>
<th>$\theta$</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH, $N_{0,1}$, $\kappa_y = 4.1$, $\theta_y^e = .6$</td>
<td>GMTTM</td>
<td>.595 (.17)</td>
<td>GMM</td>
<td>.577 (.20)</td>
<td>QML</td>
<td>.597 (.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.064</td>
<td>$\lambda^G$</td>
<td>.26</td>
<td>KS</td>
<td>.177</td>
<td>$\lambda^G$</td>
<td>.177</td>
</tr>
<tr>
<td>TARCH, $N_{0,1}$, $\kappa_y = 5.25$, $\theta_y^e = .6$</td>
<td>GMTTM</td>
<td>.617 (.16)</td>
<td>GMM</td>
<td>.536 (.15)</td>
<td>QML</td>
<td>.595 (.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.083</td>
<td>$\lambda^G$</td>
<td>.29,16</td>
<td>KS</td>
<td>.269</td>
<td>$\lambda^G$</td>
<td>.269</td>
</tr>
<tr>
<td>IGARCH, $N_{0,1}$, $\kappa_y = 2.0$, $\theta_y^e = .6$</td>
<td>GMTTM</td>
<td>.599 (.17)</td>
<td>GMM</td>
<td>.628 (.23)</td>
<td>QML</td>
<td>.586 (.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.078</td>
<td>$\lambda^G$</td>
<td>.31</td>
<td>KS</td>
<td>.273</td>
<td>$\lambda^G$</td>
<td>.273</td>
</tr>
<tr>
<td>TIARCH, $N_{0,1}$, $\kappa_y = 3.4$, $\theta_y^e = 1$</td>
<td>GMTTM</td>
<td>.998 (.24)</td>
<td>GMM</td>
<td>.839 (.39)</td>
<td>QML</td>
<td>1.00 (.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.064</td>
<td>$\lambda^G$</td>
<td>.41,24</td>
<td>KS</td>
<td>.284</td>
<td>$\lambda^G$</td>
<td>.284</td>
</tr>
<tr>
<td>GARCH, $P_{2.5}$, $\kappa_y = 1.5$, $\theta_y^e = .6$</td>
<td>GMTTM</td>
<td>.594 (.23)</td>
<td>GMM</td>
<td>.264 (.35)</td>
<td>QML</td>
<td>.605 (.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.094</td>
<td>$\lambda^G$</td>
<td>.35</td>
<td>KS</td>
<td>.448</td>
<td>$\lambda^G$</td>
<td>.448</td>
</tr>
<tr>
<td>TARCH, $P_{2.5}$, $\kappa_y = 2.6$, $\theta_y^e = .5$</td>
<td>GMTTM</td>
<td>.619 (.29)</td>
<td>GMM</td>
<td>.270 (.20)</td>
<td>QML</td>
<td>.516 (.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.104</td>
<td>$\lambda^G$</td>
<td>.67,09</td>
<td>KS</td>
<td>.532</td>
<td>$\lambda^G$</td>
<td>.532</td>
</tr>
<tr>
<td>QARCH, $P_{2.5}$, $\kappa_y = 3.5$, $\theta_y^e = .8$</td>
<td>GMTTM</td>
<td>.798 (.09)</td>
<td>GMM</td>
<td>.673 (.09)</td>
<td>QML</td>
<td>.896 (.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.066</td>
<td>$\lambda^G$</td>
<td>.32,10</td>
<td>KS</td>
<td>.454</td>
<td>$\lambda^G$</td>
<td>.454</td>
</tr>
<tr>
<td>QAARCH, $P_{2.5}$, $\kappa_y = 2.0$, $\theta_y^e = 1$</td>
<td>GMTTM</td>
<td>1.01 (.09)</td>
<td>GMM</td>
<td>.721 (.09)</td>
<td>QML</td>
<td>.997 (.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KS</td>
<td>.050</td>
<td>$\lambda^G$</td>
<td>.57,37</td>
<td>KS</td>
<td>.643</td>
<td>$\lambda^G$</td>
<td>.643</td>
</tr>
</tbody>
</table>

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