Moment Condition Tests for Heavy-Tailed Time Series

Jonathan B. Hill[1] and Mike Aguilar[2]

Dept. of Economics, University of North Carolina

July 2011

ABSTRACT

We develop an asymptotically chi-squared statistic for testing moment conditions $E[m_t(\theta^0)] = 0$, where $m_t(\theta^0)$ may be weakly dependent, scalar components of $m_t(\theta^0)$ may have an infinite variance, and $E[m_t(\theta)]$ need not exist for any θ under the alternative. Score tests are a natural application, and in general a variety of tests can be heavy-tail robustified by our method, including white noise, GARCH affects, omitted variables, distribution, functional form, causation, volatility spillover and over-identification. The test statistic is derived from a tail-trimmed sample version of the moments evaluated at a consistent plug-in $\hat{\theta}_T$ for θ^0. Depending on the test in question and heaviness of tails, $\hat{\theta}_T$ may be *any* consistent estimator including sub-$T^{1/2}$-convergent and/or asymptotically non-Gaussian ones, since $\hat{\theta}_T$ can be assured not to affect the test statistic asymptotically. We adapt bootstrap, p-value occupation time, and covariance determinant methods for selecting the trimming fractile in any sample, and apply our statistic to tests of white noise, omitted variables and volatility spillover. We find it obtains sharp empirical size and strong power, while conventional tests exhibit size distortions.

[1] Corresponding author: Dept. of Economics, University of North Carolina-Chapel Hill, www.unc.edu/~jbhill, jbhill@email.unc.edu.

[2] Dept. of Economics, University of North Carolina-Chapel Hill, maguilar@email.unc.edu.

Key words and phrases: moment condition test; heavy tails; tail trimming; robust inference.

JEL subject classifications. C13, C20, C22.

AMS subject classifications. Primary 62F35; secondary 62F07.

The authors thank Eric Renault for many helpful conversations. We also thank two referees and the special issue editorial board for helpful suggestions.

1. INTRODUCTION We propose an asymptotically chi-squared statistic for testing moment conditions in the presence of heavy tails.

Casting inference within a moment condition framework covers a broad array of tests. Moment condition [MC] tests, as well as those with MC interpretations (like score tests), include tests of omitted or instrumental variables, functional form, distribution, conditional heteroscedasticity, over-identification, causation, volatility spillover, structural change, order selection and encompassing tests. Notable theory contributions for iid data include Hausman (1978), White (1981), Hansen (1982), Newey (1985), Bierens (1990), Wooldridge (1990), Imbens et al (1998) and Kitamura et al (2004) to name a very few. Moment equality and inequality tests for time series are developed in de Jong (1996), Florens et al (1998), Bai (2003), Ghysels and Guay (2003), Hill (2008) and Bontemps and Meddahi (2011), amongst others.

Applying moment based inference to heavy tailed data is particularly relevant since evidence for heavy tails across disciplines is substantial, ranging from financial, macroeconomic, auction, actuarial, meteorological to network telecommunication data. The literature is vast, but notable surveys include Embrechts et al (1997) and Finkenstädt and Rootzén (2003). See also Hill and Shneyerov (2010). In general, by their sample moment form, MC tests require the existence of higher moments to ensure standard asymptotics. As a consequence, this forces possibly very severe moment restrictions on an underlying process that may fail for even mildly volatile data. One example is estimating equations for GARCH models as the basis of a test of over-identifying restrictions or model mis-specification. A finite variance may require the GARCH process to have a finite eighth moment, or an error to have a finite fourth moment, depending on the equation form (Hill and Renault 2010). The problem of requiring "more than just the hypotheses of interest" was noted in Wooldridge (1990: p. 18) concerning a variety of contexts, although not heavy tails.

Let $m_t : \Theta \to \mathbb{R}^q$ be parametric test equations where Θ is a compact subset of $\mathbb{R}^r$, and $q, r \geq 1$, and for simplicity of exposition assume $m_t(\theta)$ is continuous and differentiable. The null hypothesis is

$$H_0 : E\left[m_t(\theta^0)\right] = 0 \text{ for unique } \theta^0 \in \Theta \tag{1}$$

with a general alternative

$$H_1 : \text{the null is false.} \tag{2}$$

We allow for heavy tails such that $E[m_{i,t}^2(\theta^0)] = \infty$ and do not require the moment $E[m_t(\theta)]$ to exist under H_1 for any θ. If the test equation $m_t(\theta)$ is integrable uniformly on Θ, then the alternative becomes H_1: $E[m_t(\theta^0)] \neq 0$. In general θ may represent a subset of parameters, such as when testing for the autoregression order in an AR-GARCH. Alternatively, the equations may be parameter free $m_t(\theta) = m_t$, such as in a test of white noise on an observable time series. See Section 3 for examples.

In order to conquer the challenge of heavy tails, and arrive at a test statistic that is easy to compute and interpret due to a standard limit distribution, we trim a negligible sample fraction of those $m_{i,t}(\theta)$ that may be heavy tailed. Let $\{k_{1,i,T}, k_{2,i,T}\}$ be integer *fractile* sequences representing the number of trimmed left-tailed and right-tailed observations from each sample $\{m_{i,t}(\theta)\}_{t=1}^T$ with sample size T. We enforce negligible trimming by assuming $\{k_{1,i,T}, k_{2,i,T}\}$ are intermediate order sequences: $k_{j,i,T} \to \infty$ and $k_{j,i,T}/T \to 0$ (Leadbetter et al 1983, Hahn et al 1991). Define tail specific observations of $m_{i,t}(\theta)$ and their sample order statistics:

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

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

If an equation $m_{i,t}(\theta^0)$ has an infinite variance, or its higher moments are unknown, we trim $m_{i,t}(\theta)$

such that it lies between its lower $k_{1,i,T}/T^{th}$ and upper $k_{2,i,T}/T^{th}$ sample quantiles:

$$\hat{m}^*_{T,i,t}(\theta) = m_{i,t}(\theta) \times I\left(m^{(-)}_{i,(k_{1,i,T})}(\theta) \leq m_{i,t}(\theta) \leq m^{(+)}_{i,(k_{2,i,T})}(\theta)\right) = m_{i,t}(\theta) \times \hat{I}_{i,T,t}(\theta) \tag{3}$$

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

Note $I(A) = 1$ is A is true, and 0 otherwise. If $m_{i,t}(\theta^0)$ is symmetric then use[3] $I(|m_{i,t}(\theta)| \leq m^{(a)}_{i,(k_{i,T})}(\theta))$ where $m^{(a)}_{i,t}(\theta) := |m_{i,t}(\theta)|$, $k_{i,T} \to \infty$ and $k_{i,T}/T \to 0$.

Now let $\hat{\theta}_T$ be any consistent estimator of θ^0. The proposed Tail-Trimmed Moment Condition [TTMC] test statistic has a quadratic form

$$\hat{W}_T = \left(\sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\theta}_T)\right)' \hat{S}^{-1}_T(\hat{\theta}_T)\left(\sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\theta}_T)\right) \tag{4}$$

where $\hat{S}_T(\theta)$ is a kernel covariance estimator

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

and $\hat{m}^*_T(\theta) := 1/T\sum_{t=1}^{T}\hat{m}^*_{T,t}(\theta)$, $k(\cdot)$ is a kernel function and $\gamma_T \to \infty$ is a bandwidth parameter. Trimming introduces spurious dependence unless $m_{i,t}(\theta^0)$ is iid, so a HAC is in general preferred. See Section 2.3.

Our framework is built on the principles of Generalized of Method of Tail-Trimmed Moments by Hill and Renault (2010), denoted HR (2010). Indeed, $\sum_{t=1}^{T}\hat{m}^*_{T,t}(\theta)' \times \hat{S}^{-1}_T(\hat{\theta}_T) \times \sum_{t=1}^{T}\hat{m}^*_{T,t}(\theta)$ is simply the efficiently weighted GMTTM criterion with consistent plug-in $\hat{\theta}_T$, provided there are at least as many equations as parameters $q \geq r$, and (1) holds. The primary contribution of this paper is to provide an accompanying theory of heavy tail robust inference for GMM, with an arbitrary plug-in $\hat{\theta}_T$ that may not have a Gaussian limit nor be $T^{1/2}$-convergent.

We prove $\hat{W}_T$ is asymptotically chi-squared under the null (1) and suitable regularity conditions. Further, under these conditions $\hat{W}_T$ has non-negligible power against a sequence of local alternatives, which implies $\hat{W}_T \to \infty$ under (2) with probability one. In both cases $\hat{W}_T$ has the same limiting properties whether tails are heavy or not due to self-standardization and trimming negligibility. This is a major advantage over other methods (see below) since other than $E[m_t(\theta^0)] = 0$ under H_0 we never need to know if $m_t(\theta^0)$ lacks higher moments. Small sample power, however, can in principle be affected by trimming when tails are thin. See Theorem 2.2 for details, and see Section 5 for simulation evidence that trimming need not reduce power.

In the presence of heavy tails $\sum_{t=1}^{T}\hat{m}^*_{T,t}(\theta^0)$ and $\hat{\theta}_T$ typically have different rates of convergence. In many cases the test component $\sum_{t=1}^{T}\hat{m}^*_{T,t}(\theta^0)$ can be slowed simply by trimming more (i.e. faster $k_{j,i,T} \to \infty$), implying many plug-ins $\hat{\theta}_T$ will not affect $\hat{W}_T$, hence $\hat{\theta}_T$ need not be $T^{1/2}$-convergent *nor*

[3] If $m_{i,t}(\theta^0)$ is symmetric, in *theory* it suffices to use $I(m^{(-)}_{i,(k_{1,i,T})}(\theta) \leq m_{i,t}(\theta) \leq m^{(+)}_{i,(k_{2,i,T})}(\theta))$ with $k_{1,i,T} = k_{2,i,T}$ $\forall T$. In *practice*, however, $m^{(-)}_{i,(k_{1,i,T})}(\theta)$ and $m^{(+)}_{i,(k_{2,i,T})}(\theta)$ may be quite different, even for large T, and our simulation experiments show $I(|m_{i,t}(\theta)| \leq m^{(a)}_{i,(k_{i,T})}(\theta))$ works exceptionally well because it forces the same left- and right-tail threshold.

possess a Gaussian limit. This is evidently the first study to allow sub-$T^{1/2}$ *and* super-$T^{1/2}$-convergent estimation, although Antoine and Renault (2010) tackle heterogeneous rates at or below $T^{1/2}$ for GMM. We further develop these ideas by example in Section 3 since deep results require a specification for $m_t(\theta)$. Depending on the form of $m_{i,t}(\theta^0)$ and tails, valid plug-ins include at least OLS, LAD, QML and GMM, and information theoretic estimators like Empirical Likelihood, each with non-Gaussian limits when tails are heavy. Similarly, we may use estimators robust to data contamination that are *not* robust to heavy tails, like Least Trimmed Squares and Quasi-Maximum Trimmed Likelihood (see Čižek 2008 and his references). Finally, heavy-tailed robust estimators with Gaussian limits are valid, like Peng and Yao's (2003) Log-LAD, Ling's (2005, 2007) Least Weighted Absolute Deviations [LWAD] and Quasi-Maximum Weighted Likelihood [QMWL], and HR's (2010) Generalized Method of Tail-Trimmed Moments [GMTTM].

Further transformations of the equations may lead to robustness of $\hat{W}_T$ to *any* $\hat{\theta}_T$ that converges no slower than GMTTM, including orthogonal projections (Wooldridge 1990, Bai 2003, Bontemps and Meddahi 2011). In the literature robustness is evidently only ensured for $T^{1/2}$-convergent plug-ins[4]. We only briefly discuss orthogonal transformations in Section 2 due to space constraints.

In Section 4 we discuss data-driven methods for choosing the number of trimmed $m_t(\theta)'s$, including p-value occupation time, wild bootstrap, and a covariance determinant technique. We then investigate tests of white noise and omitted variables in a simulation study in Section 5, and we study tests of volatility spillover in Hill and Aguilar (2011), the technical appendix to this paper. Our simulations serve two purposes. First, they demonstrate heavy tails may distort empirical size of non-robust tests, adding to existing evidence (e.g. de Lima 1997, Runde 1997). Second, trimming remarkably few large $m_t(\theta)$ leads to sharp empirical size while still retaining power in most cases, and substantial power in some cases. Third, data-driven methods for handling $k_{j,i,T}$ show great promise.

Under *fixed* quantile trimming, wherein $k_{j,i,T}/T \rightarrow (0,1)$, $\hat{m}^*_{T,t}(\theta^0)$ has a finite variance even asymptotically. However, there is no guarantee the null (1) is identified in the sense $1/T\sum_{t=1}^{T}\hat{m}^*_{T,t}(\theta^0) \xrightarrow{p} 0$ when the equations are asymmetric. Indeed, only *tail*-trimming robustifies against heavy tails *and* bias in general settings (HR 2010). Intermediate order trimming predominantly appears in the central limit theory literature for iid sequences, with few applications in the econometrics literature and none concerning robust inference for regression models. See the compendium Hahn et al (1991), and see Hill (2010a, 2011) and HR (2010) for detailed literature reviews.

The proposed TTMC statistic is obviously general, but other tail-trimming strategies may be optimal. A robust test of white noise, for example, can be couched in terms of (1) and tested by $\hat{W}_T$, while it is also conceivable to tail-trim sample covariances for a robust portmanteau statistic. Similarly, if $m_t(\theta) = \epsilon_t(\theta) \times x_t(\theta)$ for scalar $\epsilon_t(\theta)$ and $x_t(\theta) \in \mathbb{R}^q$ then we may trim by $\epsilon_t(\theta)$ and $x_t(\theta)$ separately, a great convenience when either component is asymmetric. We briefly discuss the latter in in the sequel, but leave the full scope of strategies for future research.

There are at least four major classes of inference in the presence of heavy tails. First, re-scaling a test statistic so it is non-degenerate, yet obtains a non-standard limit distribution, has been used for t-tests, and tests of white noise, covariance stationarity, unit roots, cointegration and GARCH (Davis and Resnick 1986, Phillips and Loretan 1991, Davis et al 1992, Runde 1997, Caner 1998, Hall and Yao 2003). Runde (1997), for example, exploits results in Davis and Resnick (1986) to show the Box-Pierce Q-statistic obtains a non-standard distribution under the null, but fails to characterize the statistic under the alternative. Further, the test statistic form requires knowledge of tail thickness. Our TTMC statistic is asymptotically chi-squared under the null *and* consistent against any global deviation from the null, and we do not need to know tail thickness to perform our test. Similarly, so-called "chi-

[4]GMTTM for heavy tailed GARCH is $o_p(T^{1/2})$ and is guaranteed *not* to affect $\hat{W}_T$ for a test of volatility spillover couched in a GARCH framework (see HR 2010 and Hill and Aguilar 2011). Sub-$T^{1/2}$-convergence also arises due to kernel smoothing in estimating equations, in-fill asymptotics, and nearly-weak GMM. See Antoine and Renault (2010) for examples and references.

squared" statistics of the form $\hat{W}_T$ have non-standard limits in simple environments, in particular when $m_t(\theta) = m_t$ is an iid nonparametric scalar in the domain of attraction of an infinite variance stable law. In this case the statistic is merely $(\sum_{t=1}^T m_t)^2 / \sum_{t=1}^T m_t^2$ which by self-standardization does not require knowledge of tail thickness (e.g. Davis and Resnick 1986, Mittnik et al 1998)[5]. This form is therefore well suited for a test of white noise on an observable process, while critical values have to be simulated as in Runde (1997).

Second, tests specialized to heavy tailed data, like tail dependence, obtain Gaussian limits and can in principle be used to test regression model mis-specification (e.g. Davis and Mikosch 2009). There are few attempts in the econometrics literature to exploit such methods for specification tests, although the tail behavior of regression estimators is used to model breakdown points (e.g. He et al 1990). See Hill (2011) for a review. The third class is distribution free tests and non-parametric inference, including rank-order tests of unit roots, correlation integral-based tests of dependence, bootstrapped confidence bands, and sub-sampling (e.g. Giné and Zinn 1990, Breitung and Gouriéroux 1997, Brock et al 1996, Mason and Shao 2001).

We use a version of Hansen's (1996) p-value method and Kline and Santos' (2010) score wild bootstrap and demonstrate their potential for by-passing fractile selection in TTMC. In general "m out of n" subsampling with $m = o(n)$ and block bootstraps can be used for conventional statistics when tails are heavy, offering a valid substitute for our approach[6]. See Arcones and Gine (1989), Lahiri (1995), Chernick (2007) and Cornea and Davidson (2010) for background theory and extensions to heavy tailed mixing processes, and see Horowitz et al (2006) for the bootstrap of the Box-Pierce Q-statistic[7].

The fourth class includes statistics derived from heavy tail robust methods which therefore have standard limits. Examples are Ling's (2005) Wald statistic for LWAD, and Hill and Renault's (2010a) kernel HAC for tail-trimmed GMM. By contrast, our test equations $\hat{m}^*_{T,t}(\theta)$ need not be based on an estimation problem, and if they are the plug-in $\hat{\theta}_T$ need not be based on the same method.

We use the following notational conventions. We drop θ^0 whenever it is understood (e.g. $\hat{m}^*_{T,t} = \hat{m}^*_{T,t}(\theta^0)$). $L(T) \to \infty$ is a slowly varying function whose value and rate may change with the context[8]. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the minimum and maximum eigenvalues of square matrix A. The L_p-norm is $||x||_p = (\sum_{i,j} E|x_{i,j}|^p)^{1/p}$, and the spectral (matrix) norm is $||A|| = (\lambda_{\max}(A'A))^{1/2}$. $(z)_+ := \max\{0, z\}$. K denotes a positive finite constant whose value may change from line to line; $\iota > 0$ is a tiny constant; N is a whole number. $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$ denote probability and distribution convergence. $[z]$ denotes the integer part of z.

2. ROBUST MOMENT CONDITION TESTS The potential problem with testing (1) is at least one $m_{i,t}$ may have an unbounded support and infinite variance. Equations that are known to have a bounded variance are logically left untrimmed for the sake of efficiency. Assume the first $\underline{q} \in$

---

[5] Evidently quadratic forms of an infinite variance mixing vector $m_t(\theta^0)$ with a general plug-in $\hat{\theta}_T$ have not been treated in the literature. This is non-trivial because the potentially non-Gaussian limit distribution of $\hat{\theta}_T$ may impact the quadratic form, and if $\hat{\theta}_T$ has a Gaussian limit (e.g. Log-LAD, GMTTM) the quadratic form's asymptotics may be even more complicated. This will certainly be an interesting topic once it is fully developed.

[6] Some bootstrap methods are not applicable for heavy tailed data. Athreya (1987) shows the naïve bootstrap for a standardized mean of infinite variance iid data fails to approximate the limit distribution.

[7] Horowitz et al (2006: Assumption A2) assume the data are bounded, but clearly this can be relaxed (cf. Lahiri 1995).

[8] Recall a slowly varying function satisfies $L(\lambda x)/L(x) \to 1$ as $x \to \infty$ for any scale $\lambda > 0$. The classic example is $\ln(x)$ and its powers. See Resnick (1987).

$\{1,...,q\}$ equations are trimmed:

$$\hat{m}^*_{T,t}(\theta) = \left[m_{i,t}(\theta) \times \hat{I}_{i,T,t}(\theta)\right]_{i=1}^{q} = \left[\left\{m_{i,t}(\theta) \times \hat{I}_{i,T,t}(\theta)\right\}_{i=1}^{\underline{q}}, \{m_{i,t}(\theta)\}_{i=\underline{q}+1}^{q}\right].$$

Throughout $\underline{q} \geq 1$ since otherwise the following reduces to known results. If the analyst does not know whether an equation has an infinite variance[9] than *all* equations are trimmed: $\underline{q} = q$. Unnecessary trimming does not affect the limiting null distribution nor asymptotic power (see Theorems 2.1 and 2.2 and subsequent remarks), and does not appear to impact the performance of $\hat{W}_T$ based on our simulation study (see Section 5). Nevertheless, trimming when tails are thin removes usable information, which may reduce small sample power. See Theorem 2.2.

2.1 Threshold and Fractile Sequences

The TTMC statistic (4) involves $\hat{m}^*_{T,t}(\theta)$ in (3), but asymptotic theory and how $\hat{\theta}_T$ impacts $\hat{W}_T$ is grounded on non-random thresholds. The following borrows heavily from HR's (2010) treatment of the Generalized Method of Tail-Trimmed Moments. Consult that source for details. Let the trimming fractile sequences $\{k_{1,i,T}, k_{2,i,T} : 1 \leq i \leq \underline{q}\}$ and sequences of positive threshold functions $\{l_{i,T}(\theta), u_{i,T}(\theta) : 1 \leq i \leq \underline{q}\}$ satisfy

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

and (e.g. Leadbetter et al 1983: Theorem 1.7.13),

$$P\left(m_{i,t}(\theta) < -l_{i,T}(\theta)\right) = \frac{k_{1,i,T}}{T} \ \text{ and } \ P\left(m_{i,t}(\theta) > u_{i,T}(\theta)\right) = \frac{k_{2,i,T}}{T}. \tag{5}$$

Thus, $\{k_{j,i,T}\}$ are intermediate order sequences[10], and $l_{i,T}(\theta)$ and $u_{i,T}(\theta)$ are identically the equation specific lower $k_{1,i,T}/T \to 0$ and upper $k_{2,i,T}/T \to 0$ tail quantiles. We are guaranteed the existence of such quantiles $\{l_{i,T}(\theta), u_{i,T}(\theta)\}$ on Θ for any choice of fractile $\{k_{1,i,T}, k_{2,i,T}\}$ since we assume $m_t(\theta)$ has continuous marginal distributions. See Appendix A for all assumptions with discussion. The deterministically trimmed equations are

$$m^*_{T,i,t}(\theta) = m_{i,t}(\theta) \times I\left(-l_{i,T}(\theta) \leq m_{i,t}(\theta) \leq u_{i,T}(\theta)\right) = m_{i,t}(\theta) \times I_{i,T,t}(\theta) \ \ : \ \ 1 \leq i \leq \underline{q} \tag{6}$$

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

If $m_{i,t}(\theta)$ is symmetric then use $I(|m_{i,t}(\theta)| \leq c_{i,T}(\theta))$ where $P(|m_{i,t}(\theta)| > c_{i,T}(\theta)) = k_{i,T}/T$.

It is straightforward to show $\hat{m}^*_{T,t}$ is sufficiently close to $m^*_{T,t}$ in the sense

$$S_T^{-1/2} \sum_{t=1}^{T} \left\{\hat{m}^*_{T,t} - m^*_{T,t}\right\} = o_p(1),$$

where S_T is the covariance matrix for $\sum_{t=1}^{T} m^*_{T,t}$, defined below (cf. Hill 2010a, HR 2010). Thus, all asymptotic arguments are grounded on $m^*_{T,t}$, which is much simpler to work with for theory purposes.

Since trimming may affect inference we can now only say $m^*_{T,t}(\theta)$ *eventually* identifies θ^0 under the null:

$$H_0 : E\left[m^*_{T,t}(\theta^0)\right] \to 0.$$

[9] In many cases $m_t(\theta^0)$ reduces to a function of random variables whose tail decay rates are either known, or can be directly assessed by standard extreme value theory methods. See Section 3 for examples, and see Hill (2010b, 2011) for related references.

[10] Intermediate order sequences appear widely in the extreme value theory and tail-trimming literatures for sequestering "extreme values" (for analysis, or removal). See Leadbetter et al (1983) and Hahn et al (1991) for seminal contributions, and Hill (2011) for a detailed review.

The condition is easily guaranteed by Lebesgue's dominated convergence since tail trimming is negligible and m_t is integrable under H_0, while trivially $E[m^*_{T,t}] = 0$ if the equations are symmetric and symmetrically trimmed.

2.2 Plug-In Properties

In simple contexts $m_t(\theta) = m_t$ is non-parametric, as in a test of white noise or distribution form, on some observable time series. In parametric contexts we assign to $\hat{\theta}_T$ a sequence of positive definite scale matrices $\{\tilde{V}_T\}$, $\tilde{V}_T \in \mathbb{R}^{r \times r}$, with diagonal components $\tilde{V}_{i,i,T} \to \infty$, and assume under either hypothesis

$$\tilde{V}_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = O_p(1).$$

Under the alternative this translates to $\tilde{V}_T^{1/2}$-consistency for some point $\theta^0 \in \Theta$ (e.g. the minimizer of the Kullback-Leibler Information Criterion: see White 1982 and his references). As long as $\hat{\theta}_T \xrightarrow{p} \theta^0$ sufficiently fast we do not need to say anything else about $\hat{\theta}_T$. Stationarity and thin tails typically rule out this possibility since both $||\tilde{V}_T^{1/2}|| \sim KT^{1/2}$ and $\sum_{t=1}^T \{m^*_{T,t} - E[m^*_{T,t}]\} = O_p(T^{1/2})$[11]. But, as we detail below and by example in Section 3, in many cases heavy tails introduce a unique advantage for ensuring in principle *any* estimator $\hat{\theta}_T$ does not impact $\hat{W}_T$, depending on the test equations.

In order to gauge the impact $\hat{\theta}_T$ has on the limit distribution of $\hat{W}_T$, we exploit the fact that equation differentiability and negligibility of trimming ensure $\sum_{t=1}^T m^*_{T,t}(\hat{\theta}_T)$ can be asymptotically expanded around θ^0. We therefore need the following covariance, Jacobian, and scale matrices associated with the expansion:

$$S_T(\theta) := \sum_{s,t=1}^T E\left[\left\{m^*_{T,s}(\theta) - E[m^*_{T,s}(\theta)]\right\}\left\{m^*_{T,t}(\theta) - E[m^*_{T,t}(\theta)]\right\}'\right] \in \mathbb{R}^{q \times q} \text{ and } S_T := S_T\left(\theta^0\right)$$

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

$$V_T(\theta) := T^2\left[J'_T(\theta) S_T^{-1}(\theta) J_T(\theta)\right]^{-1} \in \mathbb{R}^{r \times r} \text{ and } V_T := V_T(\theta^0).$$

Under regularity conditions detailed in Appendix A, we obtain a fundamentally important asymptotic expansion:

$$\text{(7)} \quad \hat{S}_T^{-1/2}(\hat{\theta}_T)\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T) = \left\{S_T^{-1/2}\sum_{t=1}^T m^*_{T,t} + TS_T^{-1/2}J_T\left(\hat{\theta}_T - \theta^0\right)\right\} \times (1 + o_p(1)) + o_p(1),$$

where $TS_T^{-1/2}J_T$ satisfies

$$\{TS_T^{-1/2}J_T\} \times V_T^{-1} \times \{TS_T^{-1/2}J_T\}' \to I_q.$$

See Lemmas B.3-B.6 in Appendix B, and see HR (2010). In order to ensure $\hat{W}_T$ actually tests (1) we must assume the plug-in rates $\tilde{V}_{i,i,T}^{1/2} \to \infty$ sufficiently fast in the sense $||V_T\tilde{V}_T^{-1}|| = O(1)$, hence

$$V_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = O_p(1).$$

All conditions concerning $\{\hat{\theta}_T, \tilde{V}_T, V_T\}$ are detailed under P1 and P2 of Appendix A.

A test of over-identifying restrictions based on GMTTM provides the intuition. If $\hat{\theta}_T$ is the efficiently weighted GMTTM estimator based on $m_t(\theta)$ with trimming fractile $\{k_{1,i,T}, k_{2,i,T}\}$, then

[11] See, e.g., Newey and McFadden (1994) for broad details on M- and GMM-estimation theory under standard regularity conditions.

$V_T^{1/2}$ is exactly the GMTTM scale: under (1) and regularity conditions outlined in Appendix A, the GMTTM estimator is asymptotically normal $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{d} N(0, I_r)$ by Theorem 2.2 in HR (2010). If all $m_{i,t}^2$ and $(\partial/\partial\theta_j)m_{i,t}(\theta)|_{\theta^0}$ are integrable then each $V_{i,i,T}^{1/2} \sim KT^{1/2}$.

Roughly speaking, the general requirement here $V_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(1)$ forces $\hat{\theta}_T$ to have a compound rate $\tilde{V}_T^{1/2}$ at least as fast as efficient GMTTM in the sense $||V_T\tilde{V}_T^{-1}|| = O(1)$. This simply mimics the nearly universal thin tail case of $T^{1/2}$-convergence (see e.g. Wooldridge 1990).

Depending on the equation form, when a component $m_{i,t}$ has an infinite variance many estimators $\hat{\theta}_T$ inherently satisfy $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{p} 0$. Alternatively, for a chosen $\hat{\theta}_T$ we can force $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{p} 0$ by trimming $m_t(\theta)$ fast enough that $V_T^{1/2}$ is slow relative to $\hat{\theta}_T \xrightarrow{p} \theta^0$. Thus, $\hat{\theta}_T$ may not influence $\hat{W}_T$ depending on the test equations, their tails, and the fractiles. In such a case $\hat{\theta}_T$ need not have a Gaussian limit nor be $T^{1/2}$-convergent. The general context is obviously complicated by the fact that $m_t(\theta)$ may be very different from the estimating equations used to obtain $\hat{\theta}_T$.

If the compound rates are proportional $\tilde{V}_T \sim \mathcal{K}V_T$ for some positive definite $\mathcal{K} \in \mathbb{R}^{r\times r}$, then to assess the impact of $\hat{\theta}_T$ on $\hat{W}_T$ it is common to assume $\hat{\theta}_T$ is grounded on some array of parametric estimating equations

$$\{\tilde{m}_{T,t}(\theta,\zeta)\}, \text{ where } \tilde{m}_{T,t}(\theta,\zeta) \in \mathbb{R}^p,\ p \geq r, \text{ and } \zeta \in \mathbb{R}^s,$$

that may depend on additional parameters ζ. Typically $\tilde{m}_{T,t}(\theta,\zeta)$ will be first order or estimating equations from a minimum discrepancy criterion (e.g. QML, GMTTM). Since ζ does not play any role here, and treatment of it merely deviates from the central theme, without much loss of generality assume[12]

$$\tilde{m}_{T,t}(\theta,\zeta) = \tilde{m}_{T,t}(\theta).$$

In this case we require $\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0)$ to be asymptotically linear in $\sum_{t=1}^T\{\tilde{m}_{T,t} - E[\tilde{m}_{T,t}]\}$, which satisfies a Gaussian central limit theorem. This necessarily rules out non-smooth M-estimators like quantile regression, hence LAD. Although we assume $m_t(\theta)$ is continuous and differentiable, we make no such assumptions on the estimating equations $\tilde{m}_{T,t}(\theta)$.

This is important to note: if $\hat{\theta}_T \xrightarrow{p} \theta^0$ is faster than GMTTM such that $||V_T\tilde{V}_T^{-1}|| \to 0$, then $\hat{\theta}_T$ need not be $T^{1/2}$-convergent in many cases, nor asymptotically normal, nor linear; and otherwise we assume for simplicity that $\hat{\theta}_T$ is grounded on equations $\tilde{m}_{T,t}$ that belong to the normal domain of attraction. The latter implies either the data are sufficiently thin tailed so that a conventional linear plug-in like OLS, QML and GMM has a normal limit; or $\hat{\theta}_T$ is asymptotically linear *and* heavy tail robust, like QMWL and GMTTM. We show in Section 3 how OLS, LAD, LWAD, QML, QMWL, Log-LAD and GMTTM variously satisfy the required rates of convergence for different tests, and use many of those estimators in the simulation studies of Section 5 and Hill and Aguilar (2011), while other estimators are similarly valid (e.g. EL, Least Trimmed Squares).

We rule out the perverse case $||\tilde{V}_T V_T^{-1}|| \to 0$ since this implies some component $\hat{\theta}_{j,T}$ is slow enough that it dominates $\hat{W}_T$ in the limit, and therefore erodes the possibility of testing (1) with $m_t(\hat{\theta}_T)$.

Assuming asymptotic linearity when $\hat{\theta}_T$ converges at the GMTTM rate allows us to quantify its impact. An alternative approach in the literature exploits an orthogonal projection error $m_t^{\perp}(\theta)$ that satisfies $E[m_t^{\perp}] = 0$ *if and only if* $E[m_t] = 0$, and under standard regularity conditions a $T^{1/2}$-convergent plug-in $\hat{\theta}_T$ will not asymptotically impact an untrimmed quadratic statistic based on $m_t^{\perp}(\theta)$. See Wooldridge (1990), Bai (2003) and Bontemps and Meddahi (2011) to name a few[13]. Although we do not pursue this method here due to space constraints, it appears a test statistic based on an

[12] This is the same as assuming the "true" value ζ^0 is known. The theory that follows easily allows for a plug-in $\hat{\zeta}_T$ that is consistent $\hat{\zeta}_T \xrightarrow{p} \zeta^0$ sufficiently fast (e.g. no slower than $\hat{\theta}_T \xrightarrow{p} \theta^0$).

[13] Wooldridge (1990), for example, considers the projection operator $\mathcal{P}_t(\theta) = I_q - J_t(\theta)(E[J_t(\theta)'J_t(\theta)])^{-1}E[J_t(\theta)']$, where $J_t(\theta) := (\partial/\partial\theta)m_t(\theta)$, hence $m_t^{\perp}(\theta) := m_t(\theta)\mathcal{P}_t(\theta)$. Bontemps

orthogonalized and trimmed $\hat{m}^{\perp}_{T,t}(\theta)$ will be insensitive asymptotically to *any* $||V^{\perp}_T||^{1/2}$-convergent plug-in $\hat{\theta}_T$, where $V^{\perp}_T$ is identically V_T evaluated with $m^{\perp}_{T,t}$, the orthogonalized $m^*_{T,t}$. This will allow non-linear estimators, e.g. LAD and LWAD, and sub-$T^{1/2}$- convergence since $||V^{\perp}_T||^{1/2} = o(T^{1/2})$ is possible depending on the equation tails.

2.3 Main Results

The main results of the paper follow. First, the TTMC statistic is asymptotically chi-squared under the null. See Appendix A for all assumptions concerning distribution properties (D1-D6), identification and moment smoothness (I1-I4), the HAC kernel (K1), and the plug-in (P1-P2); and see Appendix B for all proofs.

THEOREM 2.1. *Let D1-D6, I1-I4, and K1 hold, and if a plug-in is required then assume P1 or P2 holds. Under the null (1)* $\hat{W}_T \xrightarrow{d} \chi^2(\xi)$ *where degrees of freedom* ξ *depend on the rate of convergence of* $\hat{\theta}_T$. *In particular, if P1 holds such that* $\hat{\theta}_T \xrightarrow{p} \theta^0$ *fast enough such that* $||V_T\tilde{V}_T^{-1}|| \to 0$ *then* $\xi = q$ *the number of estimating equations; and if P2 holds such that* $\tilde{V}_T \sim \mathcal{K}V_T$ *for some positive definite* $\mathcal{K} \in \mathbb{R}^{r\times r}$ *then* $\xi = s - r$ *the difference between the total number of unique equations* s *and the dimension* r *of* θ^0.

Remark 1: The null chi-squared limit holds irrespective of tail thickness due to trimming negligibility and self-standardization. The limit only requires the trimming fractiles $\{k_{j,i,T}\}$ to be intermediate order sequences. Whether $k_{j,i,T} \to \infty$ is fast or slow, or $k_{j,i,T}$ is near 1 or $T - 1$ *for a particular sample* is irrelevant for asymptotics as long as $k_{j,i,T} \to \infty$ and $k_{j,i,T} = o(T)$. Small sample performance, however, is a different matter. In general very few observations need to be trimmed to ensure $\hat{W}_T$ obtains reasonable empirical size and power, while large values of $k_{j,i,T}$ can result in size and power distortions[14]. See Sections 4 and 5.

Remark 2: If $m_t(\theta) = m_t$ is non-parametric then degrees of freedom are q.

Remark 3: Unless m_t is independent, a HAC estimator is preferred: even if m_t is a martingale difference, in general $m^*_{T,t}$ *may not be*, and $\hat{m}^*_{T,t}$ *is not*, a martingale difference for each T. Trimming introduces spurious dependence by eradicating the mds property, and as $T \to \infty$ there are opposing forces at play: there are an increasing number of cross products $E[m^*_{T,s}m^{*\prime}_{T,t}]$ in S_T, yet $m^*_{T,t} \to m_t$ *a.s.* so $m^*_{T,t}$ becomes a martingale difference. Although any one $E[m^*_{T,s}m^{*\prime}_{T,t}] \to 0$ by dominated convergence, in sum $S_T(TE[m^*_{T,t}m^{*\prime}_{T,t}])^{-1} \nrightarrow I_q$ is possible. See Hill (2010a) and HR (2010).

Remark 4: Degrees of freedom fall out of expansion (7) based on plug-in rate of convergence, and the number of estimating equations $\tilde{m}_{T,t}$. Although the latter follows from classic arguments (e.g. Newey and McFadden 1994) we present details in Appendix A.

Expansion (7) requires centering for Gaussian asymptotics, in particular

$$S_T^{-1/2}\sum_{t=1}^{T} m^*_{T,t} = S_T^{-1/2}\sum_{t=1}^{T}\left\{m^*_{T,t} - E\left[m^*_{T,t}\right]\right\} + TS_T^{-1/2}E\left[m^*_{T,t}\right].$$

This forces us to assume identification of (1) occurs sufficiently fast $||TS_T^{-1/2}E[m^*_{T,t}]|| \to 0$. See condition I2 in Appendix A. A logical class of local alternatives with so-called Pitman drift is therefore

$$H_{1,L} : TS_T^{-1/2}E\left[m_t\left(\theta\right)\right] \to v \in \mathbb{R}^q,\ v'v \in [0,\infty),\ \textit{if and only if}\ \theta = \theta^0.$$

---

and Meddahi (2011) treat the case $m_t(\theta) = m(z_t, \theta)$ for random z_t, and exploit the score based on the density of z_t for their operator.

[14] It is important to stress that "large k_T" and "fast k_T" are different properties. Consider $k_T \sim 20\ln(T)$ is slow as $T \to \infty$, but is large relative to $T = 100$ (about 92) and very small relative to $T = 10,000$ (about 184). We only require $k_T \to \infty$ and $k_T = o(T)$, so even $k_T \sim 20\ln(T)$ is valid in theory as an asymptotic policy.

We assume m_t is geometrically β-mixing in Appendix A, so if all $E|m_{i,t}|^{2+\iota} < \infty$ for tiny $\iota > 0$ then $S_T \sim T \times S$ where $S \in \mathbb{R}^{q \times q}$ is a positive definite covariance matrix. Thus $H_{1,L}$ represents a sequence of $T^{1/2}$-local alternatives. Otherwise, if some $E[m_{i,t}^2] = \infty$ then $H_{1,L}$ depicts $||TS_T^{-1/2}|| \to \infty$ convergent yet $o(T^{1/2})$-alternatives. This follows from the facts that $||S_T||/T \to \infty$ if some $m_{i,t}$ has an infinite variance (Hill 2010a), and under intermediate order trimming and geometric β-mixing $||S_T||$ is $o(T^2)$. See Lemma B.1 in Appendix B.

THEOREM 2.2. *Let D1-D6, I3-I4, K1, P1 or P2 if a plug-in is required, and $H_{1,L}$ hold. Then $\hat{W}_T \xrightarrow{d} \chi^2_\xi(v'v)$ a noncentral chi-squared law with noncentrality parameter $v'v \in [0, \infty)$, and degrees of freedom ξ characterized in Theorem 2.1.*

Remark 1: $\hat{W}_T$ has non-negligible local power irrespective of tail thickness, hence asymptotic power of one under global alternative H_1, where $E[m_t(\theta)]$ may not exist for any θ. Further, asymptotic power is one irrespective of the plug-in as long as P1 or P2 hold.

Remark 2: Test consistency is not without a price. We must have a consistent plug-in $\hat{\theta}_T \xrightarrow{p} \theta^0$ under the null *and* global alternative. In robust tests of omitted variables or functional form, for example, this implies consistency when variables are omitted or when the functional form is misspecified. In the former case we must implicitly correctly specify an encompassing model as in Example 1 of Section 3. In the latter case $\hat{\theta}_T \to \theta^0$ even if the regression model error $\epsilon_t = y_t - f(x_t, \theta^0)$ satisfies $E[\epsilon_t|x_t] \neq 0$ with positive probability. Evidently the literature on regression model estimation for heavy tailed data predominantly imposes independence or $E[\epsilon_t|x_t] = 0$ (e.g. Davis et al 1992, Ling 2005, 2007, Hall and Yao 2003, Linton et al 2010). GMTTM with dependent equations is a notable exception.

Remark 3: Heavy trimming and plug-in sampling error due to $\hat{\theta}_T$ may diminish small sample power. By the proof of Theorem 2.2 $\hat{S}_T^{-1/2}(\hat{\theta}_T) \sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T) \sim S_T^{-1/2} \sum_{t=1}^T \{m^*_{T,t} - E[m^*_{T,t}]\} + TS_T^{-1/2} E[m^*_{T,t}] + V_T^{1/2}(\hat{\theta}_T - \theta^0)$. The summation approaches $N(0,1)$,while $TS_T^{-1/2} E\left[m^*_{T,t}\right] \to v$ under $H_{1,L}$, and under fast plug-in P1 $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{p} 0$, hence consistency against global H_L. But for any small sample $V_T^{1/2}(\hat{\theta}_T - \theta^0)$ may be very different from zero. Conversely, with heavy trimming $E[m^*_{T,t}]$ may be very small hence $TS_T^{-1/2} E[m^*_{T,t}] \approx 0$ is possible even under the global alternative. Thus $TS_T^{-1/2} E\left[m^*_{T,t}\right] + V_T^{1/2}(\hat{\theta}_T - \theta^0)$ may reveal too much information about the plug-in and too little information about $E[m_t] \neq 0$, effectively eroding small sample test power. The slow plug-in case P2 is similar. Plug-in and fractile choices are tackled below.

3. EXAMPLES In order to appreciate more fully the implications of Theorem 2.1, a specification of $m_{i,t}(\theta)$ is necessary. In the following we explore three examples: testing for omitted variables by for testing white noise, volatility spillover, and distribution form. Our examples involve both *test equation* trimming with fractiles $\{k_{j,i,T}\}$, and efficiently weighted GMTTM with *estimating equation* trimming and associated fractiles $\{\tilde{k}_T\}$. As usual $k_{j,i,T}, \tilde{k}_T \to \infty$ and $k_{j,i,T}, \tilde{k}_T = o(T)$.

EXAMPLE 1 (Omitted Variables/White Noise): Consider testing for omitted variables in a stationary AR(r) model by testing for white noise. The model is

$$y_t = \sum_{i=1}^r \beta_i^0 y_{t-i} + \epsilon_t = \beta^{0\prime} x_t + \epsilon_t, \ \epsilon_t \sim iid, \, E[\epsilon_t] = 0.$$

Assume ϵ_t has a symmetric absolutely continuous distribution with bounded density on $\mathbb{R}$. If $E[\epsilon_t^2] = \infty$ assume ϵ_t exhibits power-law tail decay:

$$P(|\epsilon_t| > \epsilon) = d\epsilon^{-\kappa}(1 + o(1)) \text{ where } d > 0 \text{ and } \kappa \in (1, 2]. \tag{8}$$

We want to test whether a subset of $1 \leq q \leq r$ parameters $\beta^{(q)} := \{\beta_{i_1}, ..., \beta_{i_q}\} = 0$ by removing the associated regressors $x_t^{(q)} := \{y_{t-i_1}, ..., y_{t-i_q}\}$ and testing the resulting residuals for white noise. A test of AR order $r-1$ against r is a test of $\beta_r = 0$, hence $\beta^{(1)} = \{\beta_r\}$; and a test of white noise in y_t tests all slopes $\beta^{(r)} = \{\beta_1, ..., \beta_r\} = 0$.

Define the remaining parameters $\theta := \beta/\beta^{(q)} \in \mathbb{R}^{r-q}$, regressors $w_t := x_t/x_t^{(q)}$, and associated error $u_t(\theta) := y_t - \theta' w_t$ and write $u_t = u_t(\theta^0)$. By convention $u_t = y_t$ if $q = r$. The null hypothesis is

$$H_0 : E\left[u_t u_{t-i}\right] = 0 \text{ for } i = 1, 2, ...$$

The test equations and trimmed version are

$$m_t(\theta) = \left[u_t(\theta)u_{t-i}(\theta)\right]_{i=1}^q \text{ and } \hat{m}_{T,t}(\theta) = \left[u_t(\theta)u_{t-i}(\theta)\hat{I}_{i,T,t}(\theta)\right]_{i=1}^q.$$

Notice a finite mean $\kappa > 1$ is required to ensure $E[m_{i,t}] = 0$ given independence of the errors ϵ_t, while $E[m_{i,t}^2] < \infty$ *if and only if* $E[\epsilon_t^2] < \infty$. Now impose common symmetric trimming across equations $k_{1,i,T} = k_{2,i,T} = k_T$ since $m_{i,t}$ are identically and symmetrically distributed under the null. The indicators are simply $\hat{I}_{i,T,t}(\theta) = I(|m_{i,t}(\theta)| \leq m_{i,(k_T)}^{(a)}(\theta))$ where $m_{i,t}^{(a)}(\theta) := |m_{i,t}(\theta)|$. Let $\hat{\beta}_T$ be the OLS, LAD, LWAD or GMTTM estimator, and define the subset $\hat{\theta}_T = \hat{\beta}_T/\hat{\beta}_T^{(q)}$.

All assumptions detailed in Appendix A hold, where GMTTM, OLS and LAD plug-ins $\hat{\beta}_T$ are valid depending on the trimming policy and tails. Recall $L(T) \to \infty$ is slowly varying and may change from line to line. There is little loss in generality if the reader simply assumes $L(T) = a(\ln(T))^b$ for $a, b > 0$.

LEMMA 3.1 (Omitted Variables in AR). *The above AR data generating process satisfies I1-I4 and D1-D6. Further* $||V_T|| \sim KT$ *if the variance is finite* $\kappa > 2$, $||V_T|| \sim KT/L(T)$ *in the hairline infinite variance case* $\kappa = 2$, *and* $||V_T|| \sim KT^{2/\kappa}/k_T^{2/\kappa - 1}$ *if the variance is infinite* $\kappa < 2$. *Let* $\{k_T\}$ *be any intermediate order sequence:*

i. LAD satisfies P1 if $\kappa \leq 2$, *and if* $\kappa > 2$ *then neither P1 or P2 hold. LWAD never satisfies P1 or P2.*

ii. OLS satisfies P1 if $\kappa \leq 2$, *and if* $\kappa > 2$ *then P2 holds.*

iii. Impose heavy equation trimming $k_T \sim T/L(T)$ *or* $k_T \sim T^\lambda$ *for any* $\lambda \in (1/2, 1)$. *If* $\kappa \leq 2$ *then GMTTM with minimal estimating equation trimming* $\tilde{k}_T \sim L(T)$ *satisfies P1, and if* $\kappa > 2$ *then GMTTM for any* $\tilde{k}_T$.

Remark 1: If variance is infinite $\kappa < 2$, then $||V_T^{1/2}|| \sim KT^{1/\kappa}/k_T^{1/\kappa - 1/2}$ is trumped by $T^{1/\kappa}$-convergent OLS and LAD, and in general by a large array of smooth M-estimators (Davis et al 1992). Similarly, $T^{1/\kappa}/L(T)$-convergent GMTTM trumps $||V_T^{1/2}||$ as long as we accelerate test equation trimming (e.g. $k_T \sim T/L(T)$) and diminish estimating equation trimming $\tilde{k}_T \sim L(T)$. See HR (2010) for GMTTM rate characterizations for AR-GARCH models.

Remark 2: LWAD is $T^{1/2}$-convergent, so it is too slow for P1 if $\kappa \leq 2$ no matter how fast $k_T \to \infty$, since $||V_T^{1/2}||/T^{1/2} \to \infty$ as long as $k_T = o(T)$. Otherwise in thin tailed cases P2 requires an asymptotically linear estimator so LAD and LWAD are not applicable.

EXAMPLE 2: Volatility Spillover: A rich literature has emerged on testing for market associations and contagion, and stock price/volume relationships during volatile periods (e.g. Forbes and Rigobon 2002, Ito and Hashimoto 2005). Let $\{y_{1,t}, y_{2,t}\}$ be a joint process of interest with GARCH(1,1) coordinates: each $y_{i,t}$ satisfies

$$y_{i,t} = h_{i,t}(\theta_i^0)\epsilon_{i,t} \text{ and } h_{i,t}^2(\theta) = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1}^2(\theta), \ \omega_i^0 > 0, \ \alpha_i^0, \beta_i^0 \geq 0,$$

where $\epsilon_{i,t}$ are serially independent with zero mean and unit variance. Hong (2001) argues volatility spillover reduces to testing whether $y_{1,t}^2/h_{1,t}^2 - 1$ and $y_{2,t-j}^2/h_{2,t-j}^2 - 1$ are correlated at some lag $j \geq 1$. He proposes a standardized portmanteau statistic and requires $E[\epsilon_{i,t}^8] < \infty$, although $y_{i,t}$ may be IGARCH or mildly explosive GARCH, as long as $y_{i,t}$ is stationary.

Define $\theta = [\theta_1', \theta_2']'$ and test equations

$$m_{j,t}(\theta) = \left(\frac{y_{1,t}^2}{h_{1,t}^2(\theta_1)} - 1\right) \times \left(\frac{y_{2,t-j}^2}{h_{2,t-j}^2(\theta_2)} - 1\right).$$

Under the compound null of correct marginal strong-GARCH(1,1) and no spillover from $y_{2,t}$ to $y_{1,t}$ it follows that $E[m_{j,t}] = E[(\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-j}^2 - 1)] = 0$, where $E[m_{j,t}^2] < \infty$ requires at least $E[\epsilon_{i,t}^4] < \infty$. Under tail-trimming we only need m_t to be integrable under the null, hence $E[\epsilon_{i,t}^2] < \infty$, which is a substantial improvement over Hong (2001).

Hong (2001) uses QML to estimate univariate GARCH(1,1) models $y_{i,t} = \epsilon_{i,t}h_{i,t}(\theta_i^0)$ where $h_{i,t}^2(\theta_i) = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1}^2(\theta)$. Assuming a correct strong-GARCH(1,1) specification, under the null of no volatility spillover the errors $\epsilon_{i,t} \overset{iid}{\sim} (0,1)$ are mutually independent.

Define the moment supremum $\kappa_i := \sup\{\alpha > 0 : E|\epsilon_{i,t}|^\alpha < \infty\}$, impose $E[\ln(\alpha_i^0\epsilon_{i,t}^2 + \beta_i^0)] < 0$ to ensure stationarity under the null, and assume the distributions of $\epsilon_{i,t}$ are sufficiently smooth to ensure $y_{i,t}$ have Paretian tails (8) with indices $\kappa_{y_i} > 0$ (Basrak et al 2002). Further, if $E[\epsilon_{i,t}^4] = \infty$ assume $\epsilon_{i,t}$ has Paretian tail (8) with index $\kappa_i \in (2, 4]$, cf. Hall and Yao (2003) and HR (2010). Under the stated conditions $\{y_{1,t}, y_{2,t}\}$ are geometrically β-mixing (see Francq and Zakoïan 2006).

Define test equations

$$m_t(\theta) = \left[\left(\frac{y_{1,t}^2}{h_{1,t}^2(\theta)} - 1\right) \times \left(\frac{y_{2,t-j}^2}{h_{2,t-j}^2(\theta)} - 1\right)\right]_{j=1}^q. \tag{9}$$

If $E[\epsilon_{i,t}^4] = \infty$ then tail-trimming is appropriate with trimmed equations

$$\hat{m}_{T,t}^*(\theta) = \left[\left(\frac{y_{1,t}^2}{h_{1,t}^2(\theta)} - 1\right) \times \left(\frac{y_{2,t-j}^2}{h_{2,t-j}^2(\theta)} - 1\right) \times \hat{I}_{j,T,t}(\theta)\right]_{j=1}^q.$$

In general, $m_{j,t}$ is skewed right and has the same tail thickness for each j, so the same fractiles $\{k_{1,T}, k_{2,T}\}$ are used for each j. Let $\hat{\theta}_T$ be the QML, QMWL, GMTTM or Log-LAD estimator.

LEMMA 3.2 (Volatility Spillover). *Define* $\kappa := \min\{\kappa_1, \kappa_2\}$. *The above GARCH data generating process satisfies I1-I4 and D1-D6. Further, if* $\kappa > 4$ *then* $||V_T|| \sim KT$, if $\kappa = 4$ *then* $||V_T|| \sim KT/L(T) = o(T)$, *and if* $\kappa \in (2, 4)$ *then* $||V_T|| \sim KT(k_T/T)^{4/\kappa - 1} = o(T)$. *In particular for any intermediate order* $\{k_{1,T}, k_{2,T}\}$*:*

i. GMTTM with maximal estimating equation trimming $\tilde{k}_T = T/L(T)$ *satisfies P1 if either* $E[\epsilon_{i,t}^4] = \infty$, *and satisfies P2 if both* $E[\epsilon_{i,t}^4] < \infty$.

ii. Log-LAD satisfies P1 if either $E[\epsilon_{i,t}^4] = \infty$, *and satisfies neither P1 nor P2 if both* $E[\epsilon_{i,t}^4] < \infty$.

iii. QML and QMWL do not satisfy P1 or P2 if either $E[\epsilon_{i,t}^4] = \infty$, *and satisfy P2 if both* $E[\epsilon_{i,t}^4] < \infty$.

Remark: The QML estimator is $T^{1/2}$-convergent if $\kappa_i > 4$, and if $\kappa_i \leq 4$ then under our regularity conditions it converges at a rate $T^{1-2/\kappa_i} = o(T^{1/2})$. See Hall and Yao (2003). In the latter case the rate is too slow since $||V_T^{1/2}|| = T^{1-2/\kappa}k_T^{2/\kappa-1/2} > T^{1-2/\kappa_i}$ for any $k_T \to \infty$ and that index

κ_i that is identically the minimum $\kappa = \min\{\kappa_1, \kappa_2\}$. Similarly, since QMWL only weights by lagged y_{t-j} and not ϵ_t (Ling 2007), its rate can be shown to be the same as QML when $E[\epsilon_{i,t}^4] = \infty$ by using arguments in Hall and Yao (2003).

EXAMPLE 3: Distribution Test: Let y_t have a differentiable density on $\mathbb{R}$. Define a mapping $m(y_t) := \psi'(y_t) + \psi(y_t)s(y_t)$ for any differentiable function ψ with derivative ψ', and a score function $s(y) := (\partial/\partial y) \ln f(y)$ of any chosen differentiable density $f(y)$. Assume $m(y_t)$ is integrable with respect to $(\partial/\partial y)P(y_t \leq y)$. It is known $E[m(y_t)] = 0$ provided $f(y) = (\partial/\partial y)P(y_t \leq y)$, hence an MC test of distribution form is possible. See Bontemps and Meddahi (2011 and their citations.

Since ψ is otherwise arbitrary an MC test may require substantially higher moments on y_t, even under the unspecified alternative[15], and there does not exist a theory for selecting an optimal ψ. Thus, the TTMC test is well suited to permit a heavy tail robust test of distribution. Further, since we do not require $m(y_t)$ to be integrable under the alternative, the alternative may include any distribution lacking higher moments (e.g. Stable-Paretian and Student's t).

4. OPTIMAL FRACTILE SELECTION

Any *intermediate order sequences* $\{k_{j,i,T}\}$ in principle are valid for trimming. But even if $m_{i,t}$ is symmetrically distributed under the null where $E[m_{T,t}^*] = 0$ for any $k_{1,i,T} = k_{2,i,T}$, the question of how *fast* we should trim $k_{i,T} \to \infty$ and precisely *which* number of large equations $k_{i,T} \in \{1, ..., T-1\}$ to trim still remains. In this section we give details on appropriate classes of $k_{j,i,T}$ based on test properties in theory, and we discuss several data-driven strategies for choosing $k_{j,i,T}$ in practice.

Throughout Λ is a compact subset of $(0, 1]$.

4.1 Fractile Selection - Theory (Asymmetric Equations)

Expansion (7) reveals that the accuracy and efficiency of $\hat{W}_T$ depends on the rate of convergence of $\hat{\theta}_T$, the rate of identification $E[m_{T,t}^*] \to 0$ for asymmetric equations, and the rate of convergence of $m_T^* := 1/T \sum_{t=1}^T m_{T,t}^*$ under the null. Since we treat convergence rates in Section 3, consider identification.

Suppose m_t is a scalar for simplicity, and asymmetrically distributed with an exact Pareto tail to simplify exposition: for all $m \geq M$ and some $M \geq 1$

$$P(m_t < -m) = d_1 m^{-\kappa_1} \text{ and } P(m_t > m) = d_2 m^{-\kappa_2}, \tag{10}$$

where $d_i > 0$ and $\min\{\kappa_i\} > 1$. Notice $\min\{\kappa_i\} > 1$ ensures (1) is valid. Then from (5) it is easy to show $E[m_{T,t}^*] \approx 0$ for each T and any policy $\{k_{1,i,T}, k_{2,i,T}\}$ that satisfies (HR 2010: Section 4)

$$\frac{k_{2,T}^{1-1/\kappa_2}}{k_{1,T}^{1-1/\kappa_1}} = T^{1/\kappa_1 - 1/\kappa_2} \times \frac{d_1^{1/\kappa_1}}{d_2^{1/\kappa_2}} \frac{(1-1/\kappa_2)}{(1-1/\kappa_1)}. \tag{11}$$

If the tails decay identically $\kappa_1 = \kappa_2 = \kappa$ then (11) reduces to $k_{2,T} = k_{1,T}(d_1/d_2)^{1/(\kappa-1)}$. If m_t is a product of random variables then (11) may be by-passed entirely. See "focused trimming" below.

Relation (11) implies the *heavier* tail is *trimmed less.* For example, if the right tail is heavier $d_2 > d_1$ and/or $\kappa_1 > \kappa_2$ and we use symmetric trimming $k_{1,T} = k_{2,T}$ then a disproportionate number of large positive values will occur on average and therefore be trimmed, resulting in $E[m_{T,t}^*] < 0$. By (11) $k_{2,T} < k_{1,T}$ ensures $E[m_{T,t}^*] \approx 0$. In practice plug-ins for κ_j and d_j can be used to enforce (11), while estimators due to Hill (1975) and Hall (1982) are consistent for a vast array of dependent and heterogeneous processes (Hill 2010b, 2011).

4.2 Fractile Selection - Practice

---

[15] Bontemps and Medahhi's (2011) test of $E[m(y_t)] = 0$ cannot be consistent since the alternative must always exclude distributions where $E[m(y_t)]$ does not exist.

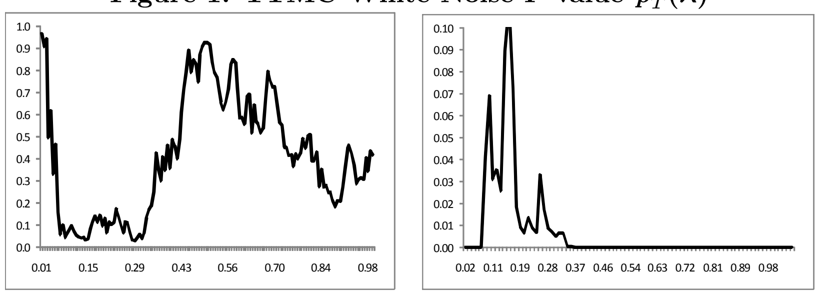
In practice the analyst must commit to particular integers $k_{1,i,T}$ and $k_{2,i,T}$ which clearly must lead to approximate identification $E[m^*_{T,t}] \approx 0$ and $\hat{W}_T \approx \chi^2(\xi)$ for large T under the null. The examples in Section 3 result in homogeneous equations, and any $k_{j,T} \to \infty$ or policy type $k_{j,T} = [\lambda_j T/\ln(T)]$ for any $\lambda_j \in (0,1]$ so assume $k_{1,T} \sim \lambda_1 T/\ln(T)$, λ_2 satisfies (11), and both[16] $\lambda_i \in \Lambda \subset (0,1]$. Since the problem now concerns selecting just λ_1, we simply write λ.

Write $\hat{m}^*_{T,t}(\hat{\theta}_T,\lambda)$, $\hat{S}_T(\hat{\theta}_T,\lambda)$ and $\hat{W}_T(\lambda)$ to denote dependence on λ. We discuss using occupation time, the covariance determinant and bootstrap methods for handling the trimming parameter. In each case due to space constraints we either leave asymptotic arguments for future consideration or give a brief sketch[17]. We divide methods according to whether all or any $\lambda \in \Lambda$ is used, or one λ is optimally selected.

4.2.1 Arbitrary $\boldsymbol{\lambda \in \Lambda}$

Occupation Time: Let $p_T(\lambda)$ denote the p-value for $W_T(\lambda)$ evaluated under the asymptotic null chi-squared distribution, and compute the occupation time $\tau_T(\alpha) := \int_{\lambda\in\Lambda} I(p_T(\lambda) \leq \alpha)d\lambda$ for some level $\alpha \in (0,1)$. Under the null and regularity conditions for weak convergence $\sup_{\alpha\in[0,1]} |\tau_T(\alpha) - \alpha| \xrightarrow{p} 0$, and under the alternative $\tau_T(\alpha) \xrightarrow{p} 1$. Figure 1 shows one sample path $\{p_T(\lambda) : \lambda \in \Lambda\}$ for the white noise TTMC test for iid and AR(2) Pareto data with index $\kappa = 1.5$, level $\alpha = .05$ and sample size $T = 500$. We compute a discretized version $\hat{\tau}_T(\alpha) := 1/T \sum_{i=1}^{T} I(p_T(i/T) \leq \alpha)$. A p-value occupation time test is performed by rejecting the null at α-level if $\tau_T(\alpha) > \alpha$. See Section 5 for simulation details.

Figure 1: TTMC White Noise P-value $\boldsymbol{p_T(\lambda)}$



iid, $\kappa = 1.5$, $\hat{\tau}_T(.05) = .0599$. AR(2), $\kappa = 1.5$, $\hat{\tau}_T(.05) = .9496$.

Wild Bootstrap: Another method is to draw observations from the null distribution of $\hat{W}_T(\lambda)$ by wild bootstrap (Liu 1988, Hansen 1996) for $\lambda \in \Lambda$. Assume $m_{i,t}$ are constructed from random variables $z_t \in \mathbb{R}^k$ (i.e. $m_t(\theta) = m(z_t,\theta)$), and let $\{\xi_t^{(r)}\}_{t=1}^T$ be a sample of iid $N(0,1)$ random variables where $r = 1,...,R$ samples are generated. Since $E[(\xi_t^{(r)})^2] = 1$, as long as $m^*_{T,t}$ are sufficiently orthogonal that $S_T \sim T \times E[m^*_{T,t} m^{*\prime}_{T,t}]$ then

$$\hat{W}_T^{(r)}(\lambda) = \left(\sum_{t=1}^T \xi_t^{(r)} \hat{m}^*_{T,t}(\hat{\theta}_T,\lambda)\right)' \hat{S}_T^{-1}(\hat{\theta}_T,\lambda) \left(\sum_{t=1}^T \xi_t^{(r)} \hat{m}^*_{T,t}(\hat{\theta}_T,\lambda)\right)$$

[16] Since $\lambda T/\ln(T) < n$ for some $\lambda > 1$ when $T \geq 100$ clearly a space with greater measure than $\Lambda \subset (0,1]$ is allowed. Our simulation study points overwhelming to small $\lambda \in (0,1)$ in order to achieve sharp empirical size.

[17] If tails are thin then each method discussed here is valid by known results. However, we are not aware of weak limit theory for tail-trimmed heavy tailed data, nor theory results that explicitly handle bootstrap methods under tail-trimming, including bootstrapping for the sake of selecting the trimming fractile. Arguments in Pakes and Pollard (1989), Hansen (1996), Gonçalves and White (2002), Hill (2008, 2009) and Doukhan et al (1996), amongst others point the way for characterizing processes like $\{W_T(\lambda) : \lambda \in \Lambda\}$ or $\{\tau_T(\alpha) : \alpha \in [0,1]\}$, but explicit results are apparently not yet available.

is a draw from the limit null distribution of $\hat{W}_T(\lambda)$, conditional on the sample $\{z_t\}_{t=1}^T$. This simply generalizes the score wild-bootstrap to TTCM (e.g. Kline and Santos 2010).

Note $\hat{S}_T(\theta,\lambda)$ is constructed from $\hat{m}^*_{T,t}(\theta,\lambda)$, and the asymptotic p-value approximation is simply $\hat{p}_{R,T}(\lambda) := 1/R_T\sum_{r=1}^{R_T} I(\hat{W}_T(\lambda) < \hat{W}_T^{(r)}(\lambda))$ for $R \in \mathbb{N}$, cf. Hansen (1996). It is not difficult to show $\hat{p}_{R,T}(\lambda)$ is consistent for the asymptotic p-value if we take $R \rightarrow \infty$. A sketch of the argument follows (see also de Jong 1996 and Hill 2008). By the Glivenko-Cantelli theorem it can be shown $\text{plim}_{R\rightarrow\infty}\,\hat{p}_{R,T}(\lambda)$ $= P(\hat{W}_T(\lambda) < \hat{W}_T^{(r)}(\lambda))$, and conditional on the sample $\{z_t\}_{t=1}^T$ the bootstrap statistic $\hat{W}_T^{(r)}(\lambda)$ is distributed $\chi^2(\xi)$, hence $P(\hat{W}_T(\lambda) < \hat{W}_T^{(r)}(\lambda))$ is the asymptotic p-value. We can perform a test of level $\alpha \in (0,1)$, for example, by rejecting the null if $\hat{p}_{R,T}(\lambda) < \alpha$. Under the alternative by Theorem 2.2 it can be shown $\lim_{T\rightarrow\infty} P(\hat{W}_T(\lambda) < \hat{W}_T^{(r)}(\lambda)) = 0$ hence $\hat{p}_{R,T}(\lambda) \xrightarrow{p} 0$ as $T \rightarrow \infty$.

A disadvantage is $\xi_t^{(r)}\hat{m}^*_{T,t}(\hat{\theta}_T,\lambda)$ is orthogonal hence $m^*_{T,t}$ must be asymptotically orthogonal for $\text{plim}_{R\rightarrow\infty}\,\hat{p}_{R,T}(\lambda)$ to be the asymptotic p-value, cf. Theorem 2.1. The advantage is we may use any $\lambda \in \Lambda \subset (0,1]$.

Block Bootstrap: We may bootstrap the distribution of $\hat{W}_T(\lambda)$ under either hypothesis. Background theory is presented in Lahiri (1995) for heavy-tailed self-standardized sums of mixing data, and Gonçalves and White (2002) for dependent processes. See also Hall and Horowitz (1996), Mason and Shao (2001), and Chernick (2007). Assume $\{\hat{W}_T^{(r)}(\lambda)\}_{r=1}^R$ denotes a bootstrapped draw. Under our regularity conditions the two block, moving block and stationary block bootstrap $\widehat{\mathcal{W}}_T(\lambda) := 1/R\sum_{r=1}^R \hat{W}_T^{(r)}(\lambda) \xrightarrow{p} E[\hat{W}_T(\lambda)]$ as $R \rightarrow \infty$ (Gonçalves and White 2002: Theorem 2.2), a bootstrap confidence band is easily obtained, and $\lim_{T\rightarrow\infty} E[\hat{W}_T(\lambda)] = \xi$ (cf. Theorem 2.1). In the case of asymmetry either (11) must be enforced for asymmetric trimming, or focused symmetric trimming with re-centering may be used. Alternatively, a bootstrap estimate of the covariance of the moment estimator $1/T\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T)$ may itself be used for fractile selection (e.g. Léger and Romano 1990).

Similarly "m out of n" bootstrap and related sub-sampling methods evidently apply since they apply to untrimmed thin or heavy tailed processes (cf. Arcones and Giné 1989, Lahiri 1995, Cornea and Davidson 2010).

Focused Trimming with Re-Centering: If $m_{i,t}(\theta) = \epsilon_t(\theta)x_{i,t}(\theta)$ then we may focus trimming symmetrically on $\epsilon_t(\theta)$ and $x_{i,t}(\theta)$ separately, and re-center, regardless of symmetry in the underlying distributions. Denote fractiles $\{k_T(\lambda_\epsilon), k_T(\lambda_{x_i})\}$, and for any $z_t(\theta) \in \{\epsilon_t(\theta), x_{i,t}(\theta)\}$ define $\hat{z}^*_{T,t}(\theta,\lambda) := z_t(\theta)I(|z_t(\theta)| \leq z^{(a)}_{(k_T(\lambda_z))}(\theta))$, $\bar{\hat{z}}^*_T(\theta,\lambda) := 1/T\sum_{t=1}^T \hat{z}^*_{T,t}(\theta,\lambda)$ and

$$\hat{m}_{T,t}(\theta,\lambda) = \left[\left(\hat{\epsilon}^*_{T,t}(\theta,\lambda) - \bar{\hat{\epsilon}}^*_T(\theta,\lambda)\right)\left(\hat{x}^*_{i,T,t}(\theta,\lambda) - \bar{\hat{x}}^*_{i,T}(\theta,\lambda)\right)\right]_{i=1}^q.$$

Similarly $z^*_{T,t}(\theta,\lambda) := z_t(\theta)I(|z_t(\theta)| \leq c_{z,T}(\theta,\lambda_z))$ where $P(|z_t(\theta)| \geq c_{z,T}(\theta,\lambda_z)) = k_T(\lambda_z)/T$, and

$$m^*_{i,T,t}(\theta,\lambda) = \left(\epsilon^*_{T,t}(\theta,\lambda_\epsilon) - E[\epsilon^*_{T,t}(\theta,\lambda_\epsilon)]\right) \times \left(x^*_{i,T,t}(\theta,\lambda_{x_i}) - E[x^*_{i,T,t}(\theta,\lambda_{x_i})]\right).$$

Tail trimming implies $E[m^*_{i,T,t}(\lambda)] \rightarrow 0$ under the null, and if ϵ_t and $x_{i,t}$ are independent then centering ensures $E[m^*_{i,T,t}(\lambda)] = 0$. Further, under the alternative $\limsup_{n\rightarrow\infty} |E[m^*_{i,T,t}(\lambda)]| > 0$ since trimming is negligible. This allows for arbitrary choices of λ_ϵ and λ_{x_i} from $\Lambda \subset (0,1]$ irrespective of asymmetry. It is straightforward to prove Theorems 2.1 and 2.2 are valid under this type of trimming, omitted here for the sake of brevity.

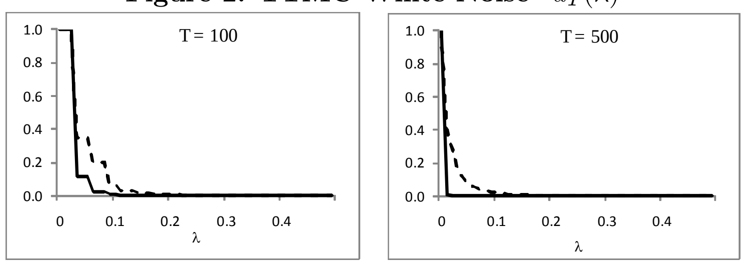
4.2.2 Optimal Selection of $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$

Pseudo-Minimum Covariance Determinant: In the literature on robust estimation, high breakdown points can be improved when "outliers" are detected by minimizing the estimator covariance determinant (see Rousseeuw et al 2004, and Agulló et al 2008 for reviews on related methods). The same mechanism can be exploited here to select λ since our "estimator" is $1/T\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T)$. Compute the normalized determinant $d_T(\lambda) := \det(\hat{S}_T(\hat{\theta}_T, \lambda))/\max_{\lambda \in \Lambda}\{\det(\hat{S}_T(\hat{\theta}_T, \lambda))\} \in [0,1]$ for each λ in a predetermined set $\Lambda_R = \{\lambda_1, \lambda_2, ..., \lambda_R\}$ where R is large. If a unique minimum exists the MCD solution is $\lambda := \arg\min_{\lambda \in \Lambda}\{d_T(\lambda)\}$.

Since $d_T(\lambda)$ measures the direct impact of trimming on $1/T \sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\theta}_T)$, by construction $d_T(\lambda)$ is monotonically decreasing in λ[18]. In simulations discussed below it falls precipitously for small λ and then flattens sharply. See Figure 2 for plots of $d_T(\lambda)$ based on white noise test equations for iid data with $\kappa \in \{1.5, 4.5\}$, and $T \in \{100, 500\}$. An identical pattern occurs for each test in Section 5 and either hypothesis. Minimizing $d_T(\lambda)$ is therefore not an option, while increasing λ beyond a point is irrelevant and will only damage the test statistic's ability to detect (1). We therefore recommend choosing the smallest $\lambda \in \Lambda$ such that the drop $|d_T(\lambda_j) - d_T(\lambda_{j+1})| < \delta$ for some tiny tolerance $\delta > 0$. We call the chosen λ the Pseudo-MCD solution. Other metrics are clearly viable (e.g. $||\hat{S}_T(\hat{\theta}_T, \lambda)||$), as well as other covariance methods (e.g. Ollila et al 2002) including bootstrap covariance (e.g. Léger and Romano 1990).

A drawback is the lack of criterion for selecting the tolerance δ, although inspection of $|d_T(\lambda_j) - d_T(\lambda_{j+1})|$ in our simulation study universally points to a small value that is clearly sensitive to sample size T and affected by tail thickness (see Figure 2).

Figure 2: TTMC White Noise[a] $d_T(\lambda)$



a. The data are iid with tail index $\kappa = 1.5$ (solid) or $\kappa = 4.5$ (dotted).

5. SIMULATION STUDY We now use the TTMC test statistic, its untrimmed version, and conventional statistics to perform tests of white-noise and omitted variables.

The data generating processes are IID and AR(2) for both white noise and omitted variables, and GARCH for white noise. We simulate 10,000 samples of each process for sample sizes $T \in \{100, 500, 1000\}$. Let P_κ denote a symmetric Pareto distribution: if $\epsilon_t \sim P_{\kappa_\epsilon}$ then $P(|\epsilon_t| > \epsilon) = .5(1+\epsilon)^{-\kappa_\epsilon}$, with index $\kappa_\epsilon \in \{1.5, 4.5\}$. The IID and AR models have $\epsilon_t \overset{iid}{\sim} P_{\kappa_\epsilon}$ with index $\kappa_\epsilon \in \{1.5, 4.5\}$ hence $y_t \sim P_{\kappa_\epsilon}$ (Cline and Brockwell 1985).The GARCH model is $y_t = h_t \epsilon_t$ where $\epsilon_t \overset{iid}{\sim} P_2$ or $\epsilon_t \overset{iid}{\sim} N(0,1)$, hence $\kappa_\epsilon \in \{2, \infty\}$, and GARCH y_t has a power law tail with index $\kappa_y < \kappa_\epsilon$ (Basrak et al 2001). See Table 1 for all model specifications and tail indices, and see Section 5.4 for robustness checks based on modified designs.

We use two possible tail indices for iid or GARCH errors in the baseline simulations. One is large enough to permit standard limit distributions for the conventional statistics used here. The other is small enough that conventional statistics have non-standard limit distributions under the null, and noise is substantial enough that even the TTMC statistic's performance under various alternatives may be challenged. Each process is stationary geometrically ergodic (Pham and Tran 1985, Francq and Zakoïan 2006) and therefore geometrically β-mixing.

[18] The proofs of Lemmas 3.1 and 3.2 show $\det(S_T(\lambda))$ is monotonic in λ for tests of white noise and volatility spillover. This carries to over a wide array of tests not treated here, including tests of functional form.

TABLE 1 - Data Generating Processes

Model Type	Test[a]	Process Specification for y_t	κ_ϵ^b	κ_y
IID	WN, OV	$y_t = \epsilon_t$	1.5, 4.5	1.5, 4.5
AR(2)	WN, OV	$y_t = .6 \times y_{t-1} + .3 \times y_{t-2} + \epsilon_t$	1.5, 4.5	1.5, 4.5
GARCH[c]	WN	$y_t = h_t \epsilon_t$, $h_t^2 = .3 + \alpha y_{t-1}^2 + \beta h_{t-1}^2$	2.0, ∞	1.7, 4.2

a. The test we apply to the particular model. WN = white noise; OV = omitted variables.
b. κ_ϵ is the moment supremum for ϵ_t. If ϵ_t is Paretian this is the tail index. κ_y is the moment supremum for y_t.
c. We use iid P_2 errors with $\{\alpha, \beta\} = \{.3,.2\}$ so $\kappa_y = 1.66$[19]. Or we use iid N(0,1) errors with $\{\alpha, \beta\} = \{.4,.4\}$ so $\kappa_y = 4.16$.

5.1 TTMC Test Equations and Control Tests

White Noise: We test y_t for serial correlation with equations

$$m_t = [y_t y_{t-i}]_{i=1}^q \, , \; q \in \{1, 5, 10\} \, .$$

The equation $m_{i,t} = y_t y_{t-i}$ has a power-law tail $\kappa_m = \kappa_y$ if y_t is IID (Embrechts and Goldie 1980: Theorem 3), hence $m_{i,t}$ is integrable for IID data in each case, but not for AR or GARCH data when $\kappa_y < 2$.

The GARCH $m_{i,t}$ has a tail index $\kappa_m \leq \kappa_y$, and only in the case $\kappa_y > 2$ can we claim $E|m_{i,t}| < \infty$. Thus, the GARCH case $\kappa_y < 2$ does not necessarily satisfy the fundamental assumption of integrability to ensure (1) is valid. We include this case to check how our test works.

We use a Bartlett kernel with a small HAC bandwidth since under the null m_t is a product of iid symmetrically distributed random variables, hence $m_{T,t}^*$ is a martingale difference with respect to $\Im_t = \sigma(y_\tau : \tau \leq t)$: $E[m_{i,T,t}^* | \Im_{t-1}] = y_{t-i} E[y_t I(|y_t y_{t-i}| \leq c_T) | \Im_{t-1}] = 0$. Simulation experiments not reported here suggest $\gamma_T = [T^{.25}]$ is optimal. Clearly there is a challenge in pinpointing *both* an optimal bandwidth γ_T *and* trimming fractile k_T. Simulation experiments uniformly suggest a small bandwidth γ_T is optimal irrespective of k_T since $\{m_{T,t}^*, \Im_t\}$ is a martingale difference under the null.

As control tests we compute an untrimmed version of the TTMC statistic $\hat{W}_T$ with the same bandwidth γ_T, and the Ljung-Box Q-statistic $T(T+2)\sum_{i=1}^5 \hat{\rho}(i)^2/(T-i)$. Note $\hat{\rho}(i)$ is the sample serial correlation coefficient of y_t at lag i.

Finally, we compute Runde's (1997) re-scaled version of the Ljung-Box Q-statistic $Q_{T,\kappa} := (T/\ln(T))^{2/\kappa_y} \times (T+2)\sum_{i=1}^5 \hat{\rho}(i)^2/(T-i)$ when $\kappa_y < 2$. This version both requires the tail index κ_y and has a non-standard limit. In order to give this statistic a competitive edge we simply use the true κ_y, and simulate critical values under Runde' (1997) iid null[20].

Omitted Variables: Define $\theta = \beta_1$ and generate errors $u_t(\theta) = y_t - \theta y_{t-1}$ by dropping y_{t-2}. We test $u_t(\theta)$ for white noise as a test of omitted y_{t-2}:

$$m_t(\theta) = [u_t(\theta) \, u_{t-i}(\theta)]_{i=1}^q \, , \; q \in \{1, 5, 10\} \, .$$

[19] Write $h_t^2 = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2$. Basrak et al (2002: eq. 2.10) show $E[(\alpha\epsilon_t^2 + \beta)^{\kappa/2}] = 1$ under regulatory conditions satisfied by our GARCH process. The index κ is computed as $\hat{\kappa} = \arg\min_{\kappa \in K}\{|1/N \sum_{t=1}^N (\alpha\epsilon_t^2 + \beta)^{\kappa/2} - 1|\}$ over $K \in \{.01, .02, ..., 10\}$ based on $N = 100,000$ iid random draws ϵ_t from $N(0,1)$. The 1% band is less than .001.

[20] We simulate 100,000 samples $\{y_t\}_{t=1}^T$ of iid stable random variables with index $\kappa_y < 2$, with zero skew and location, and unit scale. We then compute $Q_{T,\kappa}$ to obtain the critical values under Runde's null hypothesis. We obtain qualitatively similar results if we simply simulate the null limit distribution (Runde 1997: eq. (12)), although *neither* method results in critical values very similar to Runde's (1997: Table 1). For example, we obtain substantial under-rejection of the null when y_t is IID Pareto with index $\kappa = 1.5$ if we use Runde's values.

A finite mean and independence of the errors $u_t = \epsilon_t$ under the null ensures (1) is valid. Under the null $m_{i,t} = \epsilon_t \epsilon_{t-i}$ is Paretian with the same index $\kappa_m = \kappa$ as ϵ_t (Embrechts and Goldie 1980). We use an OLS plug-in here, and discuss robustness checks based on LAD and GMTTM below. Simulation experiments again support the use of a small HAC bandwidth $\gamma_T = [T^{.25}]$ since under the null $m^*_{T,t}$ is a martingale difference.

The control tests are an untrimmed version of $\hat{W}_T$, and a Wald statistic

$$\mathcal{W}_T = (R\hat{\theta})'[R\hat{\tilde{V}}_T^{-1} R']^{-1}(R\hat{\theta}) \text{ where } R = [0, 1],$$

where $\hat{\tilde{V}}_T$ estimates the plug-in scale $\tilde{V}_T$ (e.g. for OLS $\hat{\tilde{V}}_T = (x'x)\hat{\sigma}_T^{-2}$ with $\hat{\sigma}_T^2 = 1/T \sum_{t=1}^T (y_t - \hat{\theta}'_{LS} x_t)^2$).

5.2 Fractile Selection

All test equations defined above are homogenous and symmetric under the null, and the fractile class $\lambda T/\ln(T)$ minimizes plug-in impact (see Section 3). We symmetrically trim each $m_{i,t}(\theta)$ by $m_{i,t}(\theta)$, and as a benchmark we fix $\lambda = .05$ for each test in all cases, roughly the median Pseudo-MCD value across all simulations (see Table 2). We then perform p-value tests by occupation time $\hat{\tau}_T(\alpha) := 1/T \sum_{i=1}^T I(p_T(i/T) \leq \alpha)$ for significance levels $\alpha \in \{.01, .05, .10\}$.

We also select λ from $\Lambda = \{.01, .01 + 1/T, .01 + 2/T, ..., 1.01\}$ by the Pseudo-MCD method. In the case of white noise the tolerance is $\delta \in \{.5, .001, .001\}$ for $T \in \{100, 500, 1000\}$, and for omitted variables we use $\delta = .1$ in all cases. Table 2 displays the average Pseudo-MCD λ over all simulated samples for each model, test, and sample size T.

In our experiments fewer observations are trimmed by Pseudo-MCD when tails are heavy (i.e. small λ). Although in theory $\{k_T\}$ may be any intermediate order sequence by Theorems 2.1 and 2.2, and the class $k_T \sim \lambda T/\ln(T)$ diminishes plug-in impact, *for a particular sample* small λ and therefore small k_T arises because with heavy tails trimming has a much "faster" impact on $\hat{S}_T(\theta, \lambda)$: an increase from $\lambda = 0$ to tiny $\lambda > 0$ causes $\det(\hat{S}_T(\theta, \lambda))$ to plunge because a few very large values are removed. Once those massive values are removed $\det(\hat{S}_T(\theta, \lambda))$ tapers-off[21]. If tails are thin trimming has a less drastic impact on $\hat{S}_T(\theta, \lambda)$. See Figure 2. Overall using a data-driven selection of λ leads to a small but notable range of λ, thus fixing λ need not be optimal. As a robustness check we use focused trimming with fixed λ in Section 5.4.

Table 2: Pseudo-MCD λ

	Thin Tails[a]							Heavy Tails[b]						
	White Noise				Omit. Var.			White Noise				Omit. Var.		
T	IID	GH[c]	AR	δ^d	H_0	H_1	δ	IID	GH	AR	δ	H_0	H_1	δ
100	.04	.04	.11	.500	.06	.06	.10	.04	.04	.10	.500	.03	.05	.10
500	.16	.23	.53	.001	.03	.04	.10	.05	.04	.18	.001	.02	.03	.10
1000	.18	.24	.55	.001	.03	.04	.10	.06	.04	.11	.001	.02	.02	.10

a. The equation tail index $\kappa_m > 4$ in all cases. b. The equation tail index $\kappa_m < 2$ in all cases.
c. GH = GARCH. d. The Pseudo-MCD tolerance.

5.3 Simulation Results

[21] It is important to recognize that optimality of $k_T \sim \lambda T/\ln(T)$ for the moment condition tests treated here is only deduced by *asymptotic* arguments: trimming faster a la $\lambda T/\ln(T)$ is optimal only as $T \to \infty$, hence the particular $\lambda \in (0, 1)$ is irrelevant. That we find by experiment a small λ is optimal for a particular sample merely implies that nearly a *small constant percent* of T should be trimmed (i.e. $\lambda T/\ln(T)$). Notice trimming by exactly a constant percent λT can lead to bias: $E[m^*_{T,t}] \nrightarrow E[m_t]$.

Rejection frequencies are presented in Tables 3 and 4 for lag $q = 5$, where empirical powers are adjusted for size deviations based on the iid null. If equations have an infinite variance the untrimmed MC test displays size distortions. In all cases the null is strongly under-rejected demonstrating a deviation from the chi-squared limit distribution. We summarize all findings below.

i. The Ljung-Box Q-test and Wald test demonstrate size distortions in the presence of heavy tails. Notice the Q-test for strong-GARCH(1,1) data with thin or heavy tails substantially over-rejects the null of white noise. Thus an important advantage of the TTMC sandwich form with kernel covariance estimator: robustness to conditional heteroscedasticity.

ii. Runde's (1997) re-scaled Q-statistic with the true tail index and simulated critical values by construction works well for iid data, but logically suffers size distortions for GARCH data[22].

iii. The TTMC statistic exhibits relatively sharp size for each test under the null. This alone provides compelling evidence that removing a negligible number of large equation observations can sharpen a variety of moment condition tests in the presence of heavy tails.

iv. The Pseudo-MCD method for selecting the trimming fractile works exceptionally well for this design. P-value occupation is profoundly sharp for the test of white noise, but leads to slight over-rejection for the test of omitted variables, likely due to the plug-in. This suggests an orthogonal plug-in robust equation transformation is in principle useful.

v. In all cases the TTMC test obtains ample power that is lower for comparatively heavier tailed errors. This simply supports our prediction that even the TTMC test may perform less well when the signal is low relative to noise.

vi. Empirical power is not only highest for thin-tailed data, but the TTMC test performed by p-value occupation time or Pseudo-MCD matches untrimmed statistics: in this study unnecessary trimming does not affect empirical size or power.

5.4 Robustness Checks

We now discuss a battery of additional experiments used as robustness checks against the above design.

Check 1 (Plug-Ins) We perform the omitted variables test with GMTTM and LAD plug-ins. Recall our assumptions can only allow LAD when variance is infinite (Lemma 3.1). In general GMTTM and LAD perform roughly the same as OLS. See Hill and Aguilar (2011).

Check 2 (Test Lags) We repeat the white noise simulation with symmetric iid Pareto error for IID and AR cases, and use lags $q = 1$ and $q = 10$, with sample size $T = 1000$. See Table 5 for all TTCM results. Each test performs roughly the same as the baseline $q = 5$. Of particular note the TTCM p-value occupation time test is essentially flawless for any $q \in \{1, 5, 10\}$: nominal and empirical size are a near perfect match, and power is close to 1.0.

Check 3 (Runde's Test) We compute Runde's Q-statistic $Q_{T,\kappa} := (T/\ln(T))^{2/\kappa_y} \times (T + 2)\sum_{i=1}^{5} \hat{\rho}(i)^2/(T - i)$ for white noise with $q = 5$ lags using the Hill (1975) estimator $\hat{\kappa}_{\tilde{k}_T} = (1/\tilde{k}_T \sum_{i=1}^{\tilde{k}_T} \ln(y^a_{(i)}/y^a_{(\tilde{k}_T+1)}))^{-1}$ with integer fractile $\tilde{k}_T$. As long as $\tilde{k}_T \to \infty$ and $\tilde{k}_T = o(T)$ it is known $\hat{\kappa}_{\tilde{k}_T} \xrightarrow{p} \kappa_y$ for a massive array of dependent and heterogenous time series, including nonlinear AR-GARCH (Hill 2010b, 2011). We simply use the average $\hat{\kappa}_{\tilde{k}_T}$ over $\tilde{k}_T \in \{[aT/\ln(T)], ..., [bT/\ln(T)]\}$,

[22] It should be pointed out that Runde (1997) only considers iid data for asymptotic theory, and our empirical critical values are computed under Runde's iid null.

where $\{a, b\} = \{.1, .25\}$, $\{.1, .3\}$ and $\{1, 2\}$ respectively for IID, GARCH and AR[23]. The results presented in Table 6 show serious size distortions for either IID Pareto or GARCH data, despite the average $\hat{\kappa}_{\tilde{k}_T} \approx \kappa$.

Check 4 (Wild Bootstrap) We use the bootstrap to approximate the null distribution of each white noise test statistic. See Horowitz et al (2006) for a block bootstrap for the Box-Pierce test under standard regularity conditions. Since we must commit to a trimming policy we use the benchmark fixed value $\lambda = .05$. The TTMC and MC statistics are computed by the wild bootstrap detailed in Section 4.2.1 with $R = 10,000$ simulated samples $\{\xi_t^{(r)}\}_{t=1}^T$ of $\xi_t^{(r)} \overset{iid}{\sim} N(0,1)$. The Ljung-Box statistic $T(T+2)\sum_{i=1}^5 \hat{\rho}(i)^2/(T-i)$ is bootstrapped by computing $\hat{\rho}_r(i) := \hat{\gamma}_r(i)/\hat{\gamma}(0)$ where $\hat{\gamma}_r(i) := 1/T\sum_{t=i+1}^T y_t\xi_t^{(r)}y_{t-i}\xi_{t-i}^{(r)}$ and $\hat{\gamma}(0) := 1/T\sum_{t=i+1}^T y_t y_{t-i}$. Consult Table 7. If $E[\epsilon_t^2] < \infty$ then TTMC and the Ljung-Box test work roughly equally (unnecessary trimming does not impact power), and if $E[\epsilon_t^2] = \infty$ TTMC dominates likely because the Q-statistic under a standard scale has a degenerate limit distribution.

Check 5 (Asymmetry/Focused Trimming) Next we use asymmetric iid Pareto errors $P(\epsilon_t < -\epsilon) = .5(1+\epsilon)^{-\kappa_{1,\epsilon}}$ and $P(\epsilon_t > \epsilon) = .5(1+\epsilon)^{-\kappa_{2,\epsilon}}$ with indices $(\kappa_{1,\epsilon}, \kappa_{2,\epsilon}) = (1.5, 4.5)$ or $(4.25, 8.75)$. Nearly identical results arise for a large variety of pairs $(\kappa_{1,\epsilon}, \kappa_{2,\epsilon})$. Consider white noise equations $m_{i,t} = y_t y_{t-i}$. We symmetrically trim y_t and re-center: $\hat{m}^*_{T,i,t} = (\hat{y}^*_{T,t} - \overline{\hat{y}}^*_T)(\hat{y}^*_{T,t-i} - \overline{\hat{y}}^*_T)$ where $\hat{y}^*_{T,t} := y_t I(|y_t| \leq y^{(a)}_{(k_T)})$ and $\overline{\hat{y}}^*_T = 1/T\sum_{t=1}^T \hat{y}^*_{T,t}$. The same strategy is used for omitted variables test, and the HAC is computed as above. See Table 8 for results based on $T = 1000$. The TTMC test works roughly the same as with symmetric errors when we trim $m_{i,t}$ by $m_{i,t}$, although power is now higher in many cases.

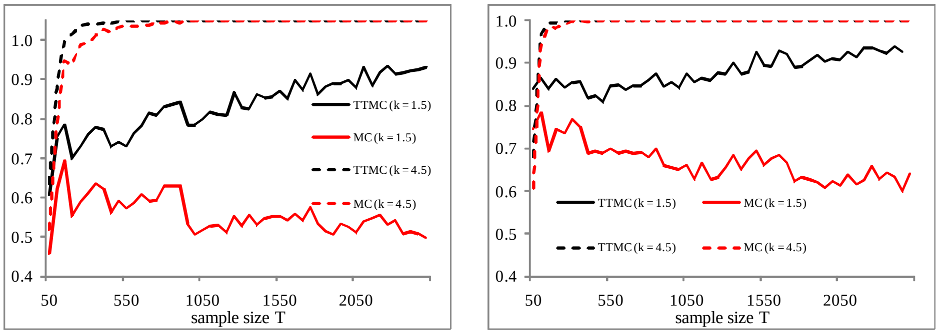
As a control we also include the symmetric case $\kappa_{1,\epsilon} = \kappa_{2,\epsilon} = 1.5$ for comparison with our baseline simulation. If we trim $m_{i,t}$ by $m_{i,t}$ then frequencies at the 5% level are .07 and .86 for IID and AR models (see Table 3), but with focused trimming we obtain .04 and .91. In conjunction with Check #6, below, in general re-centering appears to provide a non-negligible improvement in test performance.

If we asymmetrically trim $m_{i,t}$ by $m_{i,t}$ for asymmetric $m_{i,t}$, then in simulations not reported here we find at best empirical size is 2-5% too large, unless $T \geq 4000$. Despite formula (11), trimming is inherently non-smooth, so refinements in asymmetric cases either require a large sample, or focused trimming.

Check 6 (Power over T) Finally, we compute empirical power for white noise test statistics over $T \in \{50, 100, 150, ..., 2500\}$ for lag $q = 5$. We use iid symmetric Pareto errors with index $\kappa \in \{1.5, 4.5\}$, the benchmark trimming parameter $\lambda = .05$, and we either symmetrically trim $m_{i,t}$ by $m_{i,t}$, or we focus trimming with re-centering as in Check 5. As a direct comparison we also compute the MC statistic with re-centering: $m_{i,t} = (y_t - \bar{y})(y_{t-i} - \bar{y})$. See Figure 3. We omit the Q-statistic because it spikes to 1.00 at $T = 200$ if $\kappa = 1.5$ and at just $T = 100$ if $\kappa = 4.5$. The TTMC statistic exhibits linear power over nearly all T when $\kappa = 1.5$, focused trimming with re-centering leads to more power for each T, and power spikes when $\kappa = 4.5$, nearly matching the untrimmed MC statistic. If $q = 1$ then power spikes to 1.00 at $T \geq 100$ for all statistics (not shown).

FIGURE 3: Empirical Power ($q = 5$)

[23] There does not exist a theory for selecting tail estimator fractiles $\tilde{k}_T$ for non-iid data. See Hill (2010b) for references. Hill (2010b) establishes general conditions under which Hill's (1975) $\hat{\kappa}_{\tilde{k}_T(\xi)}$ is uniformly consistent over $\tilde{k}_T(\xi) := [\xi g(T)]$, where $g(T) \to \infty$ and $g(T) = o(T)$, including all processes treated here. Our choice of fractile range is due to Hill-plot dynamics under iid, GARCH and AR dependence. Note that Runde's test performs even less well if we use one $\hat{\kappa}_{\tilde{k}_T}$ as he proscribes, for any choice of $\tilde{k}_T$.



Left panel: trim m by m. Right panel: focused trimming m by y with re-recentering.

6. CONCLUSION We develop a moment condition test statistic that is robust to heavy tails by tail-trimming a sample version of the tested moment $E[m_t(\theta^0)]$. Under fairly general conditions the statistic is asymptotically chi-squared, and obtains non-negligible power against a sequence of local alternatives. Hypotheses covered encompass essentially any testable moment condition. The statistic can use as a plug-in $\hat{\theta}_T$ a large array of potential estimators that need not be $T^{1/2}$-convergent nor asymptotically normal, including conventional M- and GMM estimators, and robust versions based on weighting and trimming. This follows since in the presence of heavy tails $\hat{\theta}_T$ and $\sum_{t=1}^{T} \hat{m}^*_{T,t}$ may have different rates of convergence in favor of $\hat{\theta}_T$, and the difference can be manipulated by choice of fractile policy $\{k_{1,i,T}, k_{2,i,T}\}$. Both possibilities fail to exist for thin-tailed, stationary data. We explore data-driven methods for choosing the sample trimming fractile, and find p-value occupation time, covariance determinant, bootstrap and focused trimming methods offer compelling solutions.

We only scratch the surface of possibilities for heavy tail robust inference. Evidently tail trimming is applicable to specific statistics like portmanteau, and works in conjunction with plug-in robust orthogonal transformations. Finally, a complete theory for data-driven fractile selection requires a theory for determinant and bootstrap methods under tail trimming, and weak limit theory under tail-trimming for heavy tailed data. All such theory must be relegated to future studies.

APPENDIX A: Assumptions and Degrees of Freedom

We now present all assumptions, and provide additional details on the degrees of freedom for the TTMC statistic. In order to simplify notation assume all equations are trimmed. Write compactly throughout

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

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

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

In practice test and estimating equations $m_{T,t}(\theta)$ and $\tilde{m}_{T,t}(\theta)$ may have shared elements, so we define the total set of *unique* equations $\mathcal{M}^*_{T,t}(\theta)$:

$$\mathcal{M}^*_{T,t}(\theta) \in \mathbb{R}^s \text{ where } m^*_{T,t}(\theta), \tilde{m}_{T,t}(\theta) \in \mathcal{M}^*_{T,t}(\theta),\ s \geq \max\{p,q\}.$$

An extreme example is $m^*_{T,t}(\theta) = \tilde{m}_{T,t}(\theta)$ for a test of over-identification in GMTTM (HR 2010), hence $s = p = q$.

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

$$\Sigma_T(\theta) = E\left[\left\{m^*_{T,t}(\theta) - E[m^*_{T,t}(\theta)]\right\}\left\{m^*_{T,t}(\theta) - E[m^*_{T,t}(\theta)]\right\}'\right] \text{ and } \Sigma_T = \Sigma_T\left(\theta^0\right) \in \mathbb{R}^{q\times q}$$

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

$$\tilde{\Sigma}_T(\theta) = E\left[\left\{\tilde{m}^*_{T,t}(\theta) - E[\tilde{m}^*_{T,t}(\theta)]\right\}\left\{\tilde{m}^*_{T,t}(\theta) - E[\tilde{m}^*_{T,t}(\theta)]\right\}'\right] \text{ and } \tilde{\Sigma}_T = \tilde{\Sigma}_T\left(\theta^0\right)$$

$$\tilde{S}_T(\theta) := \sum_{s,t=1}^{T} E\left[\left\{\tilde{m}^*_{T,s}(\theta) - E[\tilde{m}^*_{T,s}(\theta)]\right\}\left\{\tilde{m}^*_{T,t}(\theta) - E[\tilde{m}^*_{T,t}(\theta)]\right\}'\right] \text{ and } \tilde{S}_T = \tilde{S}_T\left(\theta^0\right),$$

and

$$\mathfrak{S}^*_T(\theta) := \sum_{s,t=1}^{T} E\left[\left\{\mathcal{M}^*_{T,s}(\theta) - E[\mathcal{M}^*_{T,s}(\theta)]\right\}\left\{\mathcal{M}^*_{T,t}(\theta) - E[\mathcal{M}^*_{T,t}(\theta)]\right\}'\right],$$

where $m^*_{T,t}(\theta), \tilde{m}^*_{T,t}(\theta) \in \mathcal{M}^*_{T,t}(\theta)$. We abuse notation since $\tilde{\Sigma}_T(\theta)$, $\tilde{S}_T(\theta)$ and $\mathfrak{S}^*_T(\theta)$, which depict covariances in $\tilde{m}^*_{T,t}(\theta)$, may not exist for any θ. See conditions P1-P2 below. Population and sample Jacobia are

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

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

and a scale matrix is

$$V_T(\theta) := T^2 J'_T(\theta) S_T^{-1}(\theta) J_T(\theta) \in \mathbb{R}^{r\times r} \text{ and } V_T := V_T(\theta^0).$$

A.1 Assumptions

Four sets of assumptions ensure $\hat{\theta}_T$ estimates θ^0; $\sum_{t=1}^T \hat{m}^*_{T,t}(\theta)$ is sufficiently close to $\sum_{t=1}^T m^*_{T,t}(\theta)$ uniformly on Θ; $S_T^{-1/2}\sum_{t=1}^T\{m^*_{T,t} - E[m^*_{T,t}]\}$ is asymptotically normal; and $\hat{J}^*_T(\hat{\theta}_T)$ and $\hat{S}_T(\hat{\theta}_T)$ are consistent.

The first portrays the plug-in $\hat{\theta}_T$. Define a sequence of positive definite matrices $\{\tilde{V}_T\}$ on $\mathbb{R}^{r\times r}$, with divergent diagonal components $\tilde{V}_{i,i,T} \to \infty$.

P1 (*fast plug-in convergence*). *$\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(1)$ and $||V_T\tilde{V}_T^{-1}|| \to 0$ where $\tilde{\Sigma}_T^2$ and $\tilde{S}_T$ may not exist.*

P2 (*slow plug-in convergence*).

a. $\tilde{V}_T \sim \mathcal{K}V_T$ for some positive definite $\mathcal{K} \in \mathbb{R}^{r\times r}$;

b. $\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = \tilde{A}_T\sum_{t=1}^T\{\tilde{m}_{T,t} - E[\tilde{m}_{T,t}]\} \times (1 + o_p(1)) + o_p(1)$ for unique $\theta^0 \in \Theta$ where non-stochastic $\tilde{A}_T \in \mathbb{R}^{r\times p}$ has full column rank $\tilde{A}_T\tilde{S}_T^{-1}\tilde{A}'_T \to I_p$;

c. The limiting finite dimensional distributions for $\mathfrak{S}^{*\,-1/2}_T\{\mathcal{M}^*_{T,t} - E[\mathcal{M}^*_{T,t}]\}$ belong to the same class as those for $S_T^{-1}\{m^*_{T,t} - E[m^*_{T,t}]\}$.

Remark 1: P1 states $\hat{\theta}_T$ is consistent with a compound rate of convergence $\tilde{V}_T^{1/2}$ faster than the GMTTM-type $V_T^{1/2}$ in the sense $||V_T \tilde{V}_T^{-1}|| \to 0$ for given fractile sequences $\{k_{1,i,T}, k_{2,i,T}\}$. Under stationarity and the D3 mixing property, below, $||V_T|| \sim KT$ if $m_{i,t}^2$ and $(\partial/\partial\theta)m_t(\theta)|_{\theta^0}$ are integrable. Similarly, a wide range of minimum distance estimators under standard regularity conditions satisfy $||\tilde{V}_T|| \sim KT$ for sufficiently thin-tailed data. Thus, for stationary data P1 evidently can only occur when $m_{i,t}$ has an infinite variance. In heavy tailed cases, however, we can often choose a rapidly converging plug-in $\hat{\theta}_T$ that need not be linear, and/or choose $\{k_{1,i,T}, k_{2,i,T}\}$ to slow down $\tilde{V}_T$. In either case $||V_T|| = o(||\tilde{V}_T||)$ covering smooth and non-smooth M-, Method of Moments and Empirical Likelihood estimators. See Section 3 for examples.

Remark 2: P2 imposes proportionality $\tilde{V}_T \sim \mathcal{K}V_T$. In this case $\hat{\theta}_T \xrightarrow{p} \theta^0$ slowly enough that $\hat{\theta}_T$ effects the test statistic, hence we assume $\hat{\theta}_T$ is asymptotically linear in equations $\tilde{m}_{T,t}$ in P2.b. Since the test equations m_t are geometrically β-mixing by D3, the tail-trimmed equations $m_{T,t}^*$ satisfy a Gaussian central limit theorem (Hill 2010a). Property P2.c therefore ensures $\tilde{m}_{T,t}$ has the same central limit property, and together P2.b and P2.c imply $\tilde{m}_{T,t}$ may be estimating equations from conventional and outlier robust M- and Method of Moment estimators under thin tails, and asymptotically linear heavy-tailed robust estimators like GMTTM (HR 2010).

Remark 3: The omitted case $||\tilde{V}_T||/||V_T|| \to 0$ is unsatisfactory since the plug-in equations $\tilde{m}_{T,t}$ dominate, so a test of (1) cannot be performed.

The second set promotes local identification of θ^0.

I1 (*integrability*). *m_t is integrable under the null (1).*

I2 (*identification*). *Under the null (1) the thresholds $\{l_{i,T}, u_{i,T}\}$ satisfy a sequence of fixed point bounds: $||TS_T^{-1/2}E[m_{T,t}^*]|| = o(1)$.*

I3 (*covariance*). $\sup_\theta ||A_T(\theta)|| < \infty$ *and* $\liminf_{T\geq N} \inf_\theta \{\lambda_{\min}(A_T(\theta))\} > 0$ *for each* $A_T(\theta) \in \{\Sigma_T(\theta), \tilde{\Sigma}_T(\theta), \tilde{S}_T(\theta), \mathfrak{S}_T^*(\theta)\}$ *provided* $\tilde{\Sigma}_T(\theta)$, $\tilde{S}_T(\theta)$ *and* $\mathfrak{S}_T^*(\theta)$ *exist.*

I4 (*moment smoothness*). $\liminf_{T\geq N} \sup_{||\theta-\theta^0||\leq\delta}\{||E[m_{T,t}^*(\theta)]||\} > ||E[m_{T,t}^*]||$ *for some* $N \geq 1$ *and any* $\delta > 0$.

Remark 1: I2 is required due to the quadratic test statistic form and expansion (7), where $||S_T||^{1/2}/T = o(1)$ under the null by Lemma B.2 in Appendix B. In general $E[m_{T,t}^*] \to 0$ by Lebesgue's dominated convergence under the null; and $E[m_{i,T,t}^*] = 0$ if $m_{i,t}$ is symmetrically distributed and trimmed.

Remark 2: Positive definiteness I3 is standard, although we must assume it for sufficiently large T to overcome the small sample impact of trimming.

The third set concerns properties of $m_t(\theta)$, $J_{T,t}^*(\theta)$ and $I_{i,T,t}(\theta)$.

D1 (*distribution*).

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

ii. If $\sup_\theta E[m_{i,t}^2(\theta)] = \infty$ *then* $m_{i,t}(\theta)$ *have for each* t *a common power-law tail* $P(|m_{i,t}(\theta)| > m) = d_i(\theta)m^{-\kappa_i(\theta)}(1 + o(1))$ *where* $\inf_\theta \kappa_i(\theta) > 0$, $\kappa_i = \kappa_i(\theta^0) > 1$ *under the null (1),* $\inf_\theta d_i(\theta) > 0$ *where* $d_i(\theta)$ *is for each* θ *a constant, and* $\sup_\theta\{d_i^{-1}(\theta)m^{\kappa_i(\theta)}P(|m_{i,t}(\theta)| > m)\} \to 1$.

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

D3 (*mixing*). *$m_t(\theta)$ is for each T strictly stationary over $1 \leq t \leq T$, and geometrically β-mixing:* $\beta_l := \sup_{\mathcal{A}\subset\Im_{t+l}^{+\infty}} E|P(\mathcal{A}|\Im_{-\infty}^t) - P(\mathcal{A})| = o(\rho^l)$ *for* $\rho \in (0,1)$, *where* $\{\Im_t\}$ *is some sequence of σ-fields adapted to* $\{m_t(\theta)\}$, *and* $\Im_t$ *does not depend on T or θ.*

D4 (*moment envelopes*). $\sup_\theta |m_{i,t}(\theta)|$ *and* $\sup_\theta |(\partial/\partial\theta_j)m_{i,t}(\theta)|$ *are L_ι-bounded* $\forall i, j$.

D5 (*Jacobia rank and smoothness*).

i. $\sup_{\theta} ||A_T(\theta)|| < \infty$ *and* $A_T(\theta)$ *has full column rank for each* $A_T(\theta) \in \{J_T(\theta),\ J_T^*(\theta),\ E[J_{T,t}^*(\theta)]\}$.

ii. For all $\{\delta_T\}$, $\delta_T \to 0$, $\sup_{||\theta-\theta^0||\leq\delta_T}\{||J_T(\theta)||/||J_T||\} = 1 + o(1)$.

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

Remark 1: D1-D6 are essentially identical to conditions imposed in HR (2010: D1-D6) for GMTTM. See that source for details. Distribution continuity D1 greatly simplifies asymptotics in lieu of the trimming indicators $I_{i,T,t}(\theta)$, cf. Čižek (2008).

Remark 2: Mixing D3 and indicator metric entropy property D6 ensure partial sums of $I_{i,T,t}(\theta)$ satisfy a uniform central limit theorem (Doukhan et al 1995)[24]. This is used to prove $\sum_{t=1}^{T} \hat{m}_{T,t}^*(\theta)$ uniformly approximates $\sum_{t=1}^{T} m_{T,t}^*(\theta)$ sufficiently fast, a key step toward deriving expansion (7), cf. HR (2010).

Remark 3: Jacobian D5 ensures $||J_T(\theta)||$ has the same rate as $||J_T||$ for θ "close to" θ^0 with a distance vanishing in T. In thin tail cases $\sup_{\theta\in\Theta} ||J_T(\theta)|| \to K < \infty$ the property is trivial.

The last concerns kernel properties for the HAC kernel estimator.

K1 (*kernel*). $k(\cdot)$ *is integrable, and a member of the class* $\{k : \mathbb{R} \to [-1, 1] \mid k(0) = 1,\ k(x) = k(-x)\ \forall x \in \mathbb{R},\ \int_{-\infty}^{\infty} |k(x)|dx < \infty,\ \int_{-\infty}^{\infty} |\varpi(\xi)|d\xi < \infty,\ k(\cdot)$ *is continuous at* 0 *and all but a finite number of points*$\}$, *where* $\varpi(\xi) := (2\pi)^{-1} \int_{-\infty}^{\infty} k(x)e^{i\xi x}dx < \infty$. *Further* $\sum_{s,t=1}^{T} |k((s-t)/\gamma_T)| = o(T^2)$, $\max_{1\leq s\leq T} \sum_{t=1}^{T} k((s-t)/\gamma_T) = o(T)$ *and bandwidth* $\gamma_T = o(T)$.

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

A.2 Degrees of Freedom

In Theorem 2.1 degrees of freedom ξ depend on whether $m_t(\theta)$ is parametric, how fast $\hat{\theta}_T \xrightarrow{p} \theta^0$ in parametric cases, how many unique test and estimating equations there are when $\hat{\theta}_T \xrightarrow{p} \theta^0$ is relatively slow, and whether over-identifying restrictions are used to estimate θ^0. Under fast plug-in convergence P1 we have $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{p} 0$ hence by expansion (7) the test statistic is

$$\hat{W}_T = \left(\sum_{t=1}^{T} m_{T,t}^*\right)' S_T^{-1} \left(\sum_{t=1}^{T} m_{T,t}^*\right) (1 + o_p(1)) + o_p(1).$$

A mixing property then ensures $\hat{W}_T \xrightarrow{d} \chi^2(q)$ by a tail-trimmed central limit theorem (cf. Hill 2010a).

Under slow plug-in convergence P2 where $\tilde{V}_T \sim \mathcal{K}V_T$, the degrees of freedom are exactly $\xi = s - r = q$ when estimating and test equations are unique (i.e. $s = q + p$) and θ^0 is exactly identified ($r = p$). This case applies to many tests and estimators, including a test of white noise on regression errors estimated by QMWL, or a test of functional form with exactly identified GMTTM, and so on, and applies to all cases in our simulation study in Section 5.

[24] The brackets $\{l, u\}$ of an index function class $\mathcal{F}$ satisfies $l \leq f \leq u$ for every member $f \in \mathcal{F}$, where $\{l, u\}$ may not be members of $\mathcal{F}$; an ε-L_2-bracket $\{l, u\}$ satisfies $||l - u|| \leq \varepsilon$; the L_2-bracketing numbers $\mathcal{N}_{[\,]}(\varepsilon, \mathcal{B}, ||\cdot||_2)$ are the number of ε-L_2-brackets required to cover $\mathcal{F}$, and metric entropy with L_2-bracketing is $\mathcal{H}_{[\,]}(\varepsilon, \mathcal{B}, ||\cdot||_2) = \ln(\mathcal{N}_{[\,]}(\varepsilon, \mathcal{B}, ||\cdot||_2))$. See Pollard (1984). Since $\mathcal{H}_{[\,]}(\varepsilon, \Theta, ||\cdot||_2) = O(|\ln(\varepsilon)|)$ under D6 clearly $\int_0^1 \mathcal{H}_{[\,]}^{1/2}(\varepsilon, \Theta, ||\cdot||_2)d\varepsilon < \infty$, hence a required stochastic equiconuity condition for weak convergence of a partial sum of $I_{T,t}(\theta)$ applies (Dudley 1978, Doukhan et al 1995).

The more challenging case is P2 with over-identifying conditions used to estimate θ^0. We assume $\hat{\theta}_T$ is asymptotically a linear function in $\tilde{m}_{T,t}$:

$$\tilde{V}_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = \tilde{A}_T \sum_{t=1}^{T} \{\tilde{m}_{T,t} - E[\tilde{m}_{T,t}]\} (1 + o_p(1)) + o_p(1),$$

where $\tilde{A}_T \in \mathbb{R}^{r \times p}$ has full column rank and $\tilde{A}_T \tilde{S}_T \tilde{A}_T' \rightarrow I_r$ and $\tilde{S}_T$ is the covariance matrix of $\sum_{t=1}^{T} \tilde{m}_{T,t}$. In this case $\hat{S}_T^{-1/2} \sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\theta}_T)$ reduces to

$$\hat{S}_T^{-1/2}(\hat{\theta}_T) \sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\theta}_T) = S_T^{-1/2} \sum_{t=1}^{T} \left\{m^*_{T,t} - E\left[m^*_{T,t}\right]\right\}$$
$$+ T S_T^{-1/2} J_T \tilde{V}_T^{-1} \tilde{A}_T \sum_{t=1}^{T} \{\tilde{m}_{T,t} - E[\tilde{m}_{T,t}]\} (1 + o_p(1)) + o_p(1).$$

Degrees of freedom are therefore governed by the overlap of $m^*_{T,t}$ and $\tilde{m}_{T,t}$, and the dimensions of J_T and $\tilde{A}_T$ which depend upon over-identifying conditions.

One extreme is a test of over-identifying restrictions in GMTTM: $\tilde{m}_{T,t}(\theta) = m^*_{T,t}(\theta)$ hence $s = p + q$, $\tilde{V}_T = V_T$, $\tilde{A}_T S_T \tilde{A}_T' \rightarrow I_r$ and

$$\begin{aligned}\hat{S}_T^{-1/2}(\hat{\theta}_T) \sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\theta}_T) &= \left\{S_T^{-1/2} + T S_T^{-1/2} J_T V_T^{-1} \tilde{A}_T\right\} \sum_{t=1}^{T} \left\{m^*_{T,t} - E\left[m^*_{T,t}\right]\right\} (1 + o_p(1)) + o_p(1) \\ &= \mathcal{B}_T \sum_{t=1}^{T} \left\{m^*_{T,t} - E\left[m^*_{T,t}\right]\right\} (1 + o_p(1)) + o_p(1),\end{aligned}$$

say. Since $J_T \in \mathbb{R}^{q \times r}$, $V_T \in \mathbb{R}^{r \times r}$, $\tilde{A}_T \in \mathbb{R}^{r \times p}$ and $q = p = s$, there are only $q - r = s - r$ linearly independent columns in $\mathcal{B}_T$, hence $\hat{W}_T \xrightarrow{d} \chi^2(s-r)$. See also Hansen (1982) and Newey and McFadden (1994: Section 9).

The other extreme is no shared elements: $s = p + q$. Then $[I_q, T J_T \tilde{V}_T^{-1/2} \tilde{A}_T]$ contains $q + p - r = s - r$ linearly independent columns, hence $\hat{W}_T \xrightarrow{d} \chi^2(s - r)$. Finally, apply Corollary 3.3 to deduce $s - r = q$ when there are no over-identifying conditions ($p = r$).

APPENDIX B: Proofs of Main Results

The proofs of Theorems 2.1 and 2.2 require limit theory for the tail-trimmed arrays $\{\hat{m}^*_{T,t}(\theta), m^*_{T,t}(\theta)\}$ and the Jacobian and HAC estimators $\hat{J}_T(\theta)$ and $\hat{S}_T(\theta)$. The following results come directly from, or after slight alterations are consequences of, theory developed in Hill (2010a) and HR (2010)[25]. We present all proofs in the technical appendix Hill and Aguilar (2011) for ease of reference.

Throughout r_T, $o_p(1)$, $O_p(1)$, $o(1)$ and $O(1)$ do not depend on θ and t, where $r_T \rightarrow 0$ arbitrarily fast, and matrix inverses exist under the positive definiteness and rank properties I3 and D5 for large T.

First, the long run covariance S_T is bounded.

LEMMA B.1 (*covariance properties*). *Under D1 and D3* $||S_T|| = o(T^2)$.

[25] See HR (2010) for Lemmas B.1-B.5 and Hill (2010a) for Lemma B.6.

Next, the stochastically trimmed $\hat{m}^*_{T,t}(\theta)$ is sufficiently close to the deterministically trimmed $m^*_{T,t}(\theta)$.

LEMMA B.2 (*approximations*). *Under D1-D4, D6, and P1 or P2* $||\sum_{t=1}^T\{\hat{m}^*_{T,t}(\theta) - m^*_{T,t}(\theta)\}|| = o_p(||S_T(\theta)||^{1/2})$ *for any* $\theta \in \Theta$.

Further, $m_{T,t}(\theta)$ can be expanded around θ essentially as a first-order asymptotic Taylor expansion.

LEMMA B.3 (*expansions*). *Under D1-D6* $m^*_T(\theta) = m^*_T(\tilde{\theta}) + J^*_T(\theta_*)(\theta - \tilde{\theta}) + r_T \times o_p(1)$ *and* $\hat{m}^*_T(\theta) = \hat{m}^*_T(\tilde{\theta}) + \hat{J}^*_T(\theta_*)(\theta - \tilde{\theta}) + r_T \times ||\theta - \tilde{\theta}||^{1/\iota} \times o_p(1)$ *for* $||\theta_* - \theta|| \leq ||\theta - \tilde{\theta}||$ *that may be different in different in each case, and tiny* $\iota > 0$.

The sample Jacobian of the trimmed equations is consistent.

LEMMA B.4 (*Jacobian*). *Under D1-D6, and P1 or P2* $\hat{J}^*_T(\hat{\theta}_T) = J_T(1 + o_p(1))$.

The HAC estimator is consistent for S_T and $\check{S}_T(\theta)$.

LEMMA B.5 (*HAC estimator*). *Under D1-D6, K1, I3, and P1 or P2* $\hat{S}_T(\hat{\theta}_T) = S_T(1 + o_p(1))$ *and* $\hat{S}_T(\hat{\theta}_T) = \check{S}_T(\hat{\theta}_T)(1 + o_p(1))$.

Finally, the test equations satisfy a Gaussian central limit theorem.

LEMMA B.6 (*clt*). *Under D1, D3 and I3* $r'S_T^{-1/2}\sum_{t=1}^T\{m^*_{T,t} - E[m^*_{T,t}]\} \xrightarrow{d} N(0,1)$ *for any conformable* $r'r = 1$. *If P2 also holds then* $r'\mathfrak{S}_T^{-1/2}\sum_{t=1}^T\{\mathcal{M}^*_{T,t} - E[\mathcal{M}^*_{T,t}]\} \xrightarrow{d} N(0,1)$.

We are now ready to prove the main results.

PROOF OF THEOREM 2.1. Let H_0 hold. We prove the claim by case according to plug-in property P1 or P2. Define

$$M^*_{T,t} := m^*_{T,t} - E\left[m^*_{T,t}\right] \quad \text{and} \quad \tilde{M}^*_{T,t} := \tilde{m}^*_{T,t} - E\left[\tilde{m}^*_{T,t}\right]$$

$$\check{S}_T(\theta) = \sum_{s,t=1}^T E\left[\left\{\hat{m}^*_{T,s}(\theta) - E[\hat{m}^*_{T,s}(\theta)]\right\}\left\{\hat{m}^*_{T,t}(\theta) - E[\hat{m}^*_{T,t}(\theta)]\right\}\right]$$

We require the following properties under either case. The plug-in is consistent:

$$\hat{\theta}_T - \theta^0 = O_p\left(||\tilde{V}_T||^{-1/2}\right) = O_p\left(||V_T^{-1/2}||\right) = o_p\left(1\right).$$

Identification I2 states under the null $S_T^{-1/2}E[m^*_{T,t}] = o(T^{-1})$, hence under the null

$$S_T^{-1}\sum_{t=1}^T m^*_{T,t} = S_T^{-1}\sum_{t=1}^T M^*_{T,t} + o\left(1\right). \tag{12}$$

Further, asymptotic expansion Lemma B.3 coupled with Jacobian consistency Lemma B.4 and $\hat{\theta}_T \xrightarrow{p} \theta^0$ imply for some non-stochastic $r_T \to 0$ arbitrarily fast

$$\frac{1}{T}\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T) = \frac{1}{T}\sum_{t=1}^T \hat{m}^*_{T,t} + J_T\left(\hat{\theta}_T - \theta^0\right)\left(1 + o_p\left(1\right)\right) + o_p\left(r_T\right). \tag{13}$$

Finally, by approximation Lemma B.2

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

Case 1 (P1): In this case $||\tilde{V}_T||/||V_T|| \to \infty$, and by construction $\{TS_T^{-1/2}J_T\}V_T^{-1}\{J_T'S_T^{-1/2}T\} \to I_q$, hence the plug-in satisfies

$$TS_T^{-1/2}J_T\left(\hat{\theta}_T - \theta^0\right) = \left\{TS_T^{-1/2}J_T\tilde{V}_T^{-1/2}\right\}\tilde{V}_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = o_p(1) \tag{15}$$

since $||TS_T^{-1/2}J_T\tilde{V}_T^{-1/2}|| \leq ||V_T||^{1/2}/||\tilde{V}_T||^{1/2} \to 0$. Further $\hat{S}_T(\hat{\theta}_T) = S_T(1 + o_p(1))$ by HAC consistency Lemma B.5. Since $r_T \to 0$ arbitrarily fast in (13), combine (12)-(15) to obtain

$$\begin{aligned}
\hat{W}_T &= T^2\left(\frac{1}{T}\sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T)\right)'\hat{S}_T^{-1}(\hat{\theta}_T)\left(\frac{1}{T}\sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T)\right) \\
&= T^2\left(\frac{1}{T}\sum_{t=1}^T \hat{m}_{T,t}^* + J_T\left(\hat{\theta}_T - \theta^0\right)(1 + o_p(1)) + o_p(r_T)\right)' \\
&\quad \times S_T^{-1}\left(\frac{1}{T}\sum_{t=1}^T \hat{m}_{T,t}^* + J_T\left(\hat{\theta}_T - \theta^0\right)(1 + o_p(1)) + o_p(r_T)\right) \times (1 + o_p(1)) \\
&= \left(S_T^{-1/2}\sum_{t=1}^T M_{T,t}^* + TS_T^{-1/2}J_T\left(\hat{\theta}_T - \theta^0\right)(1 + o_p(1)) + o_p(1)\right)' \\
&\quad \times \left(S_T^{-1/2}\sum_{t=1}^T M_{T,t}^* + TS_T^{-1/2}J_T\left(\hat{\theta}_T - \theta^0\right)(1 + o_p(1)) + o_p(1)\right) \times (1 + o_p(1)) + o_p(1) \\
&= \left(S_T^{-1/2}\sum_{t=1}^T M_{T,t}^* + o_p(1)\right)' \times \left(S_T^{-1/2}\sum_{t=1}^T M_{T,t}^* + o_p(1)\right) \times (1 + o_p(1)) + o_p(1) \\
&= \mathcal{Z}_T'\mathcal{Z}_T \times (1 + o_p(1)) + o_p(1),
\end{aligned}$$

say. Invoke central limit theorem Lemma B.6 to deduce

$$\mathcal{Z}_T = S_T^{-1/2}\sum_{t=1}^T \left\{m_{T,t}^* - E\left[m_{T,t}^*\right]\right\} + o_p(1) \xrightarrow{d} N(0, I_q)$$

hence $\hat{W}_T = \mathcal{Z}_T'\mathcal{Z}_T \times (1 + o_p(1)) + o_p(1) \xrightarrow{q} \chi^2(q)$ by the mapping theorem.

Case 2 (P2): In this case $\tilde{V}_T \sim \mathcal{K}V_T$ for positive definite $\mathcal{K} \in \mathbb{R}^{r \times r}$, and some non-stochastic sequence $\{\tilde{A}_T\}$, $\tilde{A}_T \in \mathbb{R}^{r \times p}$, with full column rank, $\tilde{A}_T\tilde{S}_T\tilde{A}_T' \to I_p$ and

$$\tilde{V}_T^{1/2}\left(\hat{\theta}_T - \theta^0\right) = \tilde{A}_T\sum_{t=1}^T \tilde{M}_{T,t}^* \times (1 + o_p(1)) \; + \; o_p(1). \tag{16}$$

Further, HAC consistency Lemma B.5 states

$$\hat{S}_T(\hat{\theta}_T) \; = \check{S}_T(\hat{\theta}_T) \times (1 + o_p(1)). \tag{17}$$

Substitute for $\hat{\theta}_T - \theta^0$ in (13), and invoke properties (12) and (14) to obtain

$$\begin{aligned}
\sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\theta}_T) &= \sum_{t=1}^{T} \hat{m}^*_{T,t} + TJ_T\left(\hat{\theta}_T - \theta^0\right)(1 + o_p(1)) + o_p(r_T) \\
&= \sum_{t=1}^{T} m^*_{T,t} + TJ_T\tilde{V}_T^{-1/2}\tilde{A}_T\sum_{t=1}^{T}\tilde{M}^*_{T,t} \times (1 + o_p(1)) + o_p\left(\|S_T\|^{1/2}\right) \\
&= \sum_{t=1}^{T} M^*_{T,t} + S_T^{1/2}\left\{TS_T^{-1/2}J_T\right\}\tilde{V}_T^{-1/2} \times \tilde{A}_T\sum_{t=1}^{T}\tilde{M}^*_{T,t} \times (1 + o_p(1)) + o_p\left(\|S_T\|^{1/2}\right) \\
&= \sum_{t=1}^{T} M^*_{T,t} + \tilde{\mathcal{B}}_T\sum_{t=1}^{T}\tilde{M}^*_{T,t} \times (1 + o_p(1)) + o_p\left(\|S_T\|^{1/2}\right),
\end{aligned}$$

say, where $\tilde{\mathcal{B}}_T \in \mathbb{R}^{q\times p}$. The second equality substitutes for $\hat{\theta}_T - \theta^0$, and uses the facts that $r_T \to 0$ arbitrarily fast and $\liminf_{T\geq N} ||S_T|| > 0$ under I3 ensure $o_p(r_T) = o_p(||S_T||^{1/2})$. Observe the scale for $\tilde{\mathcal{B}}_T\sum_{t=1}^{T}\tilde{M}^*_{T,t}$ is proportional to $S_T^{-1/2}$: by construction of $V_T^{1/2}$, and $\tilde{V}_T \sim \mathcal{K}V_T$ under P2, there exists a positive definite $\mathcal{C} \in \mathbb{R}^{q\times q}$ that satisfies

$$\begin{aligned}
E&\left[\left(S_T^{-1/2}\tilde{\mathcal{B}}_T\sum_{t=1}^{T}\tilde{M}^*_{T,t}\right)\left(S_T^{-1/2}\tilde{\mathcal{B}}_T\sum_{t=1}^{T}\tilde{M}^*_{T,t}\right)'\right] \\
&= S_T^{-1/2}\tilde{\mathcal{B}}_T\tilde{S}_T\tilde{\mathcal{B}}_T'S_T^{-1/2} \\
&= \left\{TS_T^{-1/2}J_T\right\}\tilde{V}_T^{-1/2}\left\{\tilde{A}_T\tilde{S}_T\tilde{A}_T'\right\}\tilde{V}_T^{-1/2}\left\{TJ_T'S_T^{-1/2}\right\} \\
&\sim \left\{TS_T^{-1/2}J_T\right\}V_T^{-1/2}\left\{V_T^{1/2}\tilde{V}_T^{-1}V_T^{1/2}\right\}V_T^{-1/2}\left\{TJ_T'S_T^{-1/2}\right\} \to \mathcal{C},
\end{aligned}$$

hence the scale for $\tilde{\mathcal{B}}_T\sum_{t=1}^{T}\tilde{M}^*_{T,t}$ is identically $\mathcal{C}^{-1/2}S_T^{-1/2}$.

Now recall $\mathcal{M}^*_{T,t} \in \mathbb{R}^s$ contains all unique equations in $m^*_{T,t} \in \mathbb{R}^q$ and $\tilde{m}_{T,t} \in \mathbb{R}^p$, $s \geq \max\{p, q\}$. Define the selection matrix $\mathcal{R}_T = [I_q, \tilde{\mathcal{B}}_T] \in \mathbb{R}^{q\times s}$ with column rank $s - r$, hence

$$\mathcal{R}_T\left\{\mathcal{M}^*_{T,t} - E\left[\mathcal{M}^*_{T,t}\right]\right\} = M^*_{T,t} + \tilde{\mathcal{B}}_T\tilde{M}^*_{T,t}$$

and

$$\sum_{t=1}^{T}\hat{m}^*_{T,t}(\hat{\theta}_T) = \sum_{t=1}^{T}\mathcal{R}_T\left\{\mathcal{M}^*_{T,t} - E\left[\mathcal{M}^*_{T,t}\right]\right\}(1 + o_p(1)) + o_p\left(\|S_T\|^{1/2}\right). \tag{18}$$

Define $\mathbb{S}_T := \mathcal{R}_T\mathfrak{S}^*_T\mathcal{R}_T' \in \mathbb{R}^{q\times q}$ where $\mathfrak{S}^*_T \in \mathbb{R}^{s\times s}$ is the covariance matrix for $\sum_{t=1}^{T}\mathcal{M}^*_{T,t}$. By construction $||\mathbb{S}_T^{-1}S_T|| = O(1)$, thus by central limit theorem Lemma B.6

$$\mathbb{S}_T^{-1/2}\sum_{t=1}^{T}\hat{m}^*_{T,t}(\hat{\theta}_T) = \mathbb{S}_T^{-1/2}\sum_{t=1}^{T}\mathcal{R}_T\left\{\mathcal{M}^*_{T,t} - E\left[\mathcal{M}^*_{T,t}\right]\right\}(1 + o_p(1)) + o_p(1) \xrightarrow{d} N(0, I_q), \tag{19}$$

where $\mathcal{R}_T\mathbb{S}_T^{-1}\mathcal{R}_T'$ has rank $s - r$.

Limit (19) implies $E(\mathbb{S}_T^{-1/2}\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T)) \to 0$ by the Helly-Bray theorem, hence

$$\mathbb{S}_T^{-1/2}\sum_{t=1}^T \left\{\hat{m}^*_{T,t}(\hat{\theta}_T) - E\left[\hat{m}^*_{T,t}(\hat{\theta}_T)\right]\right\} = \mathbb{S}_T^{-1/2}\sum_{t=1}^T \mathcal{R}_T\left\{\mathcal{M}^*_{T,t} - E\left[\mathcal{M}^*_{T,t}\right]\right\}(1+o_p(1))+o_p(1) \xrightarrow{d} N(0, I_q).$$

But this ensures by HAC consistency (17) and the definition of $\check{S}_T(\hat{\theta}_T)$,

$$\mathbb{S}_T = \check{S}_T(\hat{\theta}_T) \times (1+o_p(1)) = \hat{S}_T(\hat{\theta}_T) \times (1+o_p(1)), \tag{20}$$

therefore by (17)-(20)

$$\hat{S}_T^{-1/2}(\hat{\theta}_T)\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T) = \mathbb{S}_T^{-1/2}\sum_{t=1}^T \mathcal{R}_T\left\{\mathcal{M}^*_{T,t} - E\left[\mathcal{M}^*_{T,t}\right]\right\}(1+o_p(1)) + o_p(1) \xrightarrow{d} N(0, I_q). \tag{21}$$

Now combine (21) with rank $s-r$ of $\mathcal{R}_T\mathbb{S}_T^{-1}\mathcal{R}_T'$ and invoke the mapping theorem to prove the claim:

$$\hat{W}_T = \left(\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T)\right)' \hat{S}_T^{-1}(\hat{\theta}_T)\left(\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T)\right) \xrightarrow{d} \chi^2(s-r). \quad \blacksquare$$

PROOF OF THEOREM 2.2. Notice $H_{1,L}$ does not affect the supporting Lemmas B.1-B.6 since none require null identification I2. The proof of Theorem 2.1 therefore carries over with only minor changes.

Under $H_{1,L}$ the equations satisfy $TS_T^{-1/2}E[m_t] \to v$ where $v'v \in [0,\infty)$. The trimmed equations have the same limit by Lebesgue's dominated convergence:

$$TS_T^{-1/2}E\left[m^*_{T,t}\right] \to v. \tag{22}$$

Under P1 and $H_{1,L}$, apply (13)-(15), (22) and HAC Lemma B.5 to deduce

$$\begin{aligned}\hat{S}_T^{-1/2}(\hat{\theta}_T)\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T) &= S_T^{-1/2}\sum_{t=1}^T m^*_{T,t} \times (1+o_p(1)) + TS_T^{-1/2}J_T\left(\hat{\theta}_T - \theta^0\right) \times (1+o_p(1)) \\ &= \left(S_T^{-1/2}\sum_{t=1}^T\left\{m^*_{T,t} - E\left[m^*_{T,t}\right]\right\} + TS_T^{-1/2}E\left[m^*_{T,t}\right]\right) \times (1+o_p(1)) + o_p(1) \\ &= \left(\mathcal{Z}_{1,T} \; + TS_T^{-1/2}E\left[m^*_{T,t}\right]\right) \times (1+o_p(1)) + o_p(1),\end{aligned}$$

where $\mathcal{Z}_{1,T} \xrightarrow{d} N(0,1)$ by CLT Lemma B.6, and so on. Similarly, under P2 use (16), (21) and (22) to arrive at

$$\begin{aligned}\hat{S}_T^{-1/2}(\hat{\theta}_T)\sum_{t=1}^T \hat{m}^*_{T,t}(\hat{\theta}_T) &= \mathbb{S}_T^{-1/2}\sum_{t=1}^T \mathcal{R}_T\left\{\mathcal{M}^*_{T,t} - E\left[\mathcal{M}^*_{T,t}\right]\right\} \times (1+o_p(1)) + TS_T^{-1/2}E\left[m^*_{T,t}\right] \times (1+o_p(1)) \\ &= \left(\mathcal{Z}_{2,T} + TS_T^{-1/2}E\left[m^*_{T,t}\right]\right) \times (1+o_p(1)),\end{aligned}$$

where $\mathcal{Z}_{2,T} \xrightarrow{d} N(0,1)$ $\blacksquare$

The proofs of Lemmas 3.1 and 3.2 require the following properties. In the special case $m_t = u_t x_{t-1}$ for iid zero mean $u_t \in \mathbb{R}$ with symmetric distribution, and stochastic measurable $x_t \in \mathbb{R}^q$, then symmetrically trimmed $m^*_{T,t}$ forms a martingale difference array with respect to $\Im_t = \sigma(\{u_\tau, x_\tau\} : \tau \leq t)$ since

$$\text{(MDA)} \qquad E\left[m^*_{i,T,t}|\Im_{t-1}\right] = x_{i,t-1}E\left[u_t I(|u_t x_{i,t-1}| \leq c_{T,i})|x_{t-1}\right] = 0.$$

Second, under D1-D6 J_T is proportional to $E[(\partial/\partial\theta)m_t(\theta)|_{\theta^0} I_{T,t}]$ (HR 2010: Lemma C.4):

$$\text{(JAC)} \qquad J_T = E\left[\frac{\partial}{\partial\theta}m_t(\theta)|_{\theta^0} I_{T,t}\right] \times (1 + o(1)).$$

PROOF OF LEMMA 3.1. The verification of conditions D1-D6 and I1-I4 is similar to arguments in HR (2010: Section 5.1).

Consider plug-in properties P1 or P2, and assume for simplicity $m_t(\theta)$ and θ are scalars. Let $L(T)$ be a slowly varying function, $L(T) \to \infty$, that may change from place to place. We will show $V_T \sim KT$ if $\kappa > 2$, $V_T \sim KT/L(T)$ if $\kappa = 2$ and $V_T \sim KT(T/k_T)^{2/\kappa - 1} > T$ for any $\kappa \in (1,2)$. Apply properties (MDA) and (JAC) defined above to deduce under the null $S_T = T \times E[\epsilon_t^2 \epsilon_{t-1}^2 I_{T,t}]$ and

$$J_T = E\left[\{\epsilon_t y_{t-2} + \epsilon_{t-1} y_{t-1}\} I_{T,t}\right] \times (1 + o(1)) = E\left[\epsilon_{t-1}^2 I(|\epsilon_t \epsilon_{t-1}| \leq c_T)\right] \times (1 + o(1)).$$

If $\kappa > 2$ then by independence of the errors $S_T = KT$ and $J_T = K$, hence $V_T \sim KT$.

Otherwise consider $\kappa < 2$. Tail (8) implies $c_T = K(T/k_T)^{1/\kappa}$, and by independence $\epsilon_t \epsilon_{t-1}$ has tail (8) with the same index κ (Embrechts and Goldie 1980). An application of Karamata's Theorem therefore implies (Resnick 1987: Theorem 0.6; cf. Problem 4.2.8)

$$E\left[\epsilon_t^2 \epsilon_{t-1}^2 I(|\epsilon_t \epsilon_{t-1}| \leq c_T)\right] \sim K c_T^2 P(|\epsilon_t \epsilon_{t-1}| > c_T) = K(T/k_T)^{2/\kappa - 1}.$$

Similarly, by independence, distribution continuity and dominated convergence

$$\begin{aligned} J_T &\sim E\left[\epsilon_{t-1}^2 I(|\epsilon_t \epsilon_{t-1}| \leq c_T)\right] = E\left(E\left[\epsilon_{t-1}^2 I(|\epsilon_t \epsilon_{t-1}| \leq c_T)|\epsilon_{t-1}\right]\right) \\ &\sim K c_T^2 \int_{-\infty}^{\infty} \epsilon^{-2} P(|\epsilon_t \epsilon| > c_T) f(d\epsilon) = K c_T^{2-\kappa} \int_{-\infty}^{\infty} \epsilon^{\kappa-2} f(d\epsilon) = K(T/k_T)^{2/\kappa - 1}. \end{aligned}$$

Therefore as claimed

$$V_T = T^2 J_T' S_T^{-1} J_T \sim KT \frac{\left((T/k_T)^{2/\kappa - 1}\right)^2}{(T/k_T)^{2/\kappa - 1}} = KT(T/k_T)^{2/\kappa - 1}.$$

If $\kappa = 2$ then by similar arguments the claim follows from Karamata's Theorem.

OLS and LAD in general, and GMTTM with slowly varying trimming fractiles $\tilde{k}_T \sim \tilde{L}(T)$, all have scale elements $\tilde{V}_T^{1/2} \sim KT^{1/\kappa}/L(T)$ or $KT^{1/2}$ respectively if $\kappa < 2$ and $\kappa > 2$ (Davis et al 1992, HR 2010). Throughout the case $\kappa = 2$ is similar.

LWAD is $T^{1/2}$-convergent for any $\kappa > 1$ (Ling 2005), and neither LAD or LWAD are asymptotically linear. Thus, LWAD never satisfies P1 for any fractile $\{k_T\}$, and LAD and LWAD do not satisfy P2 due to non-asymptotic linearity.

Now consider maximal equation trimming $k_T \sim T/L(T)$ since $V_T \sim T \times L(T)$ is at its lowest rate (the following arguments also extend to $k_T \sim T^\lambda$ for any $\lambda \in (1/2, 1)$). Then OLS, LAD, and GMTTM satisfy P1 if $\kappa < 2$ since $V_T^{1/2}/\tilde{V}_T^{1/2} \sim KT^{-1/\kappa + 1/2}L(T) \to 0$. Finally, OLS, and GMTTM satisfy P2 when $E[\epsilon_t^2] < \infty$ since $V_T^{1/2}/\tilde{V}_T^{1/2} \to (0, \infty)$. ■

PROOF OF LEMMA 3.2. See HR (2010: Section 5.2) for arguments demonstrating D1-D6 and

I1-I4. Consider P1 or P2. If either $E[\epsilon_{i,t}^4] = \infty$ then under the null $m_{j,t} = (\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-j}^2 - 1)$ has tail (8) with index $\kappa/2 := \min\{\kappa_1, k_2\}/2 \leq 2$. Assume $\kappa < 4$, the case $\kappa = 4$ being similar. Define $c_{i,T}$ $= \max\{l_{i,T}, u_{i,T}\}$ and $k_T = \min\{k_{1,T}, k_{2,T}\}$, and apply Karamata's Theorem to deduce

$$E\left[m_{T,i,t}^{*2}\right] \sim Kc_{i,T}^2 P\left(\left|\left(\epsilon_{1,t}^2 - 1\right)\left(\epsilon_{2,t-i}^2 - 1\right)\right| > c_{i,T}\right) = K(T/k_T)^{4/\kappa - 1}.$$

Further, by the martingale difference property and Lebesgue's dominated convergence $E[m_{T,s}^* m_{T,t}^{*\prime}] \rightarrow 0$ $\forall s \neq t$. In particular we can always choose the trimming fractiles $\{k_{j,i,T}\}$ to satisfy $S_T \sim TE[m_{T,t}^* m_{T,t}^{*\prime}]$ (i.e. $E\left[m_{T,s}^* m_{T,t}^{*\prime}\right] = o(||E[m_{T,t}^* m_{T,t}^{*\prime}]||/T))$ hence $S_{i,i,T} \sim KT(T/k_T)^{4/\kappa - 1}$. Cf. HR (2010).

The Jacobian is analyzed as in Example 1. Define $x_{1,t} := [1, y_{1,t-1}^2, h_{1,t-1}^2] + \beta^0(\partial/\partial\theta)h_{1,t-1}^2|_{\theta^0}$, and observe under the null

$$J_{j,t} := \frac{\partial}{\partial\theta} m_{j,t}(\theta)|_{\theta^0} = -\epsilon_{1,t}^2 \frac{x_{1,t}}{h_{1,t}^2} \times \left(\epsilon_{2,t-j}^2 - 1\right) - \epsilon_{2,t}^2 \frac{x_{2,t-j}}{h_{2,t-j}^2}\left(\epsilon_{1,t}^2 - 1\right).$$

If there are GARCH effects then $J_{j,t}$ is integrable since $\epsilon_{i,t}^2$ is iid and integrable and $||x_{i,t}||/h_{i,t}^2$ is integrable (Francq and Zakoïan 2006). If there are no GARCH effects then $h_{i,t}^2 = K$, and $y_{i,t-1}^2 = \epsilon_{i,t-1}^2$ is independent of $\epsilon_{i,t}^2$, hence again $J_{j,t}$ is integrable. Therefore $J_T \sim E[J_t] \times (1 + o(1))$ by (JAC).

Together $||V_T|| \sim KT$ if $\kappa > 4$ and $||V_T|| \sim KT(k_T/T)^{4/\kappa - 1} = o(T)$ if if $\kappa < 4$. GMTTM with QML equations and fractiles $\tilde{k}_T \sim T/L(T)$ has a rate $||\tilde{V}_T|| \sim T/L(T)$ if $E[\epsilon_{i,t}^4] = \infty$ hence P1 applies, and $||\tilde{V}_T|| \sim KT$ otherwise hence P2. Log-LAD is $T^{1/2}$-convergent and therefore satisfies P1 if $E[\epsilon_{i,t}^4]$ $= \infty$, and otherwise P2 (Peng and Yao 2003).

QML is $T^{1/2}$-convergent if $E[\epsilon_{i,t}^4] < \infty$, and $T^{1-2/\kappa_i}/L(T)$-convergent if $\kappa_i \in (2, 4)$ (Hall and Yao 2003: Theorem 2.1). Neither P1 or P2 holds in heavy tailed cases since that requires $||V_T|| \sim T^{2-4/\kappa} k_T^{4/\kappa - 1} = O(T^{2-4/\kappa_i}/L(T))$ for *both* $i = 1, 2$. Since one $\kappa_i = \kappa := \min\{\kappa_1, k_2\}$ by construction, that one satisfies $T(k_T/T)^{4/\kappa - 1} = T^{2-4/\kappa} k_T^{4/\kappa - 1} > T^{2-4/\kappa}/L(T)$ due to $4/\kappa > 1$ by assumption and $k_T \rightarrow \infty$ by construction, Therefore $||V_T||/(T^{1-2/\kappa_i}/L(T)) \rightarrow \infty$ for at least one i. QWML only weighs by lagged y_{t-i} and not ϵ_t, hence its rate of convergence will be the same as QML when $E[\epsilon_{i,t}^4] = \infty$.■

REFERENCES

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

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

[3] Arcones, M. and E. Giné (1989). The Bootstrap of the Mean with Arbitrary Bootstrap Sample Size, Annales de l'Institut Henri Poincaré 25, 457-481.

[4] Athreya, K.B. (1987). Bootstrap of the Mean in the Infinite Variance Case, Annals of Statistics 15, 724-731.

[5] Bai, J. (2003). Testing Parametric Conditional Distributions of Dynamic Models, Review of Economics and Statistics 85, 531-549.

[6] Basrak, B., R.A. Davis, and T. Mikosch (2002). Regular Variation of GARCH Processes, Stochastic Processes and Their Applications 99, 95-115.

[7] Bierens, H. J. (1990). A Consistent Conditional Moment Test of Functional Form, Econometrica 58, 1443-1458.

[8] Bontemps, C. and N. Meddahi (2011). Testing Distributional Assumptions: A GMM Approach, Journal of Applied Econometrics: forthcoming.

[9] Breitung, J. and C. Gouriéroux (1997). Rank Tests for Unit Roots, Journal of Econometrics 81, 7-27.

[10] Brock, W.A., J.A. Scheinkman, W.D. Dechert, B. LeBaron (1996). A Test for Independence Based on the Correlation Dimension, Econometric Reviews 15, 197-235.

[12] Caner, M. (1998). Tests for Cointegration with Infinite Variance Errors, Journal of Econometrics 86, 155-175.

[13] Chernick, M.R. (2007). Bootstrap Methods: A Guide for Practitioners and Researchers, Wiley: New York.
[14] Čižek, P. (2008). General Trimmed Estimation: Robust Approach to Nonlinear and Limited Dependent Variable Models, Econometric Theory 24, 1500-1529.
[15] Cline, D.B.H. and P.J. Brockwell (1985). Linear Prediction of ARMA Processes with Infinite Variance, Stochastic Processes and their Applications 19, 281-296.
[17] Cornea, A. and R. Davidson (2010). A Parametric Bootstrap for Heavy-Tailed Distributions, Working Paper, McGill University.
[18] Davis, R.A. and T. Mikosch (2009). The Extremogram: A Correlogram for Extreme Events, Bernoulli 15, 977-1009.
[19] Davis, R. and S. Resnick (1986). Limit Theory for the Sample Covariance and Correlation Function of Moving Averages, Annals of Statistiscs 14, 533-558.
[20] Davis, R.A., K. Knight and J. Liu (1992). M-Estimation for Autoregressions with Infinite Variance, Stochastic Processes and Their Applications 40, 145-180.
[21] de Jong, R. (1996). The Bierens Test under Data Dependence, Journal of Econometrics 72, 1-32.
[22] de Jong, R. and J. Davidson (2000). Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices, Econometrica 68, 407-423.
[23] de Lima, P. (1997). On the Robustness of Nonlinearity Tests to Moment Condition Failure, Journal of Econometrics 76, 251-280.
[24] Doukhan, P., P. Massart, and E. Rio (1995). Invariance Principles for Absolutely Regular Empirical Processes, Annales de l'Institut Henri Poincaré 31, 393-427.
[25] Dudley, R. M. (1978). Central Limit Theorem for Empirical Measures. Annals of Probability 6, 899-929.
[26] Embrechts, P., and C.M. Goldie (1980). On Closure and Factorization Properties of Subexponential Distributions, Journal of the Australian Mathematical Society Series A, 29, 243-256.
[27] Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag: Berlin.
[28] Finkenstädt, B., and H. Rootzén (2003). Extreme Values in Finance, Telecommunications and the Environment. Chapman and Hall: New York.
[29] Florens, J-P., E. Renault and N. Touzi (1998). Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes, Econometric Theory 14, 744-769.
[30] Forbes, K. and R. Rigobon (2002). No Contagion, Only Interdependence: Measuring Stock Market Comovements, Journal of Finance 57, 2223-2261.
[31] Francq, C. and J-M. Zakoïan (2006). Mixing Properties of a General Class of GARCH(1,1) Models without Moment Assumptions on the Observed Process, Econometric Theory 22, 815-834.
[32] Ghysels, E. and A. Guay (2003). Structural Change Tests for Simulated Method of Moments, Journal of Econometrics 115, 91-123.
[33] Giné, E. and J. Zinn (1990). Bootstrapping General Empirical Measures, Annals of Probability 18, 851-869.
[34] Gonçalves, S. and H. White (2002). The Bootstrap of the Mean for Dependent Heterogeneous Arrays, Econometric Theory 18, 1267-1384.
[35] Hahn, M., D.M. Mason and D. C. Weiner (1991). Sums, Trimmed Sums, and Extremes (Progress in Probability), Birkhäuser.
[36] Hall, P. (1982). On Some Estimates of an Exponent of Regular Variation, Journal of the Royal Statistical Society Series B 44, 37-42.
[37] Hall, P. and J.L. Horowitz (1996). Bootstrap Critical Values for Tests based on Generalized Method of Moments Estimators, Econometrica 64, 891-916.
[38] Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH Models with Heavy-Tailed Errors, Econometrica 71, 285-317.
[39] Hansen, B.E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis, Econometrica 64, 413-430.
[40] Hansen, L.P. (1982). Large Sample Properties of Method of Moments Estimators, Econometrica 50, 1029-1054.
[41] Hausman, J.A. (1978). Specification Testing in Econometrics, Econometrica 46, 1251-1271.

[42] He, X., J. Jurečková, R. Koenker and S. Portnoy (1990). Tail Behavior of Regression Estimators and their Breakdown Points, Econometrica 58, 1195-1214.
[43] Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution, Annals of Mathematical Statistics 3, 1163-1174.
[44] Hill, J.B. (2008). Consistent and Non-Degenerate Model Specification Tests Against Smooth Transition and Neural Network Alternatives, Annales D'Economie et de Statistique 90, 145-179.
[45] Hill, J.B. (2009). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation, Journal of Statistical Planning and Inference 139, 2091-2110.
[46] Hill, J.B. (2010a). Central Limit Theory for Tail-Trimmed Sums of Heavy-Tailed Dependent, Heterogeneous Data: under revision for Stochastic Processes and Their Applications.
[47] Hill, J.B. (2010b). On Tail Index Estimation for Dependent, Heterogeneous Data, Econometric Theory 26, 1398-1436.
[48] Hill, J.B. (2011). Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics: Econometric Theory 27, 844–884.
[49] Hill, J.B. and M. Aguilar (2011). Moment Condition Tests for Heavy-Tailed Time Series - Appendix C : Omitted Simulations and Proofs, www.unc.edu/∼jbhill/tech_append_cm_test_tail_trim.pdf.
[50] Hill, J.B. and E. Renault (2010). Generalized Method of Moments with Tail Trimming, Dept. of Economics, University of North Carolina - Chapel Hill.
[51] Hill, J.B. and A. Shneyerov (2010). Are There Common Values in First-Price Auctions? A Tail-Index Nonparametric Test, Journal of Econometrics: revised and resubmitted.
[52] Hong, Y. (2001). A Test for Volatility Spillover with Application to Exchange Rates, Journal of Econometrics 103, 183-224.
[53] Horowitz, J.L., I.N. Lobato, J.C. Nankervis, and N.E. Savin (2006). Bootstrapping the Box-Pierce Q Test: A Robust Test of Uncorrelatedness, Journal of Econometrics 113, 841-862.
[54] Imbens, G., R.H. Spady, and P. Johnson (1998). Information Theoretic Approaches to Inference in Moment Condition Models, Econometrica 66, 333-357.
[55] Ito, T. and Y. Hashimoto (2005). High Frequency Contagion Between Exchange Rates and Stock Prices in the Asian Currency Crisis, in M. Dungey and D. Tambakis (eds.), Identifying International Financial Contagion: Progress and Challenges, Oxford University Press, 111-149.
[56] Kitamura, Y, T. Gautam, and A. Hyungtaik (2004). Empirical Likelihood-Based Inference in Conditional Moment Restriction Models, Econometrica 72, 1667-1714.
[57] Kline, P. and A. Santos (2010). A Score Based Approach to Wild Bootstrap Inference, NBER Working Paper No. 16127.
[58] Lahiri, S.N. (1995). On the Asymptotic Beahiour of the Moving Block Bootstrap for Normalized Sums of Heavy-Tailed Random Variables, Annals of Statistics 23, 1331-1349.
[59] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag: New York.
[60] Léger, C. and J.P. Romano (1990). Bootstrap Adaptive Estimation: The Trimmed-Mean Example, Canadian Journal of Statistics 18, 297-314.
[61] Ling, S. (2005). Self-Weighted Least Absolute Deviation Estimation for Infinite Variance Autoregressive Models, Journal of the Royal Statistical Society Series B 67, 381-393.
[62] Ling, S. (2007). Self-Weighted and Local Quasi-Maximum Likelihood Estimators for ARMA-GARCH/IGARCH Models, Journal of Econometrics 140, 849-873.
[63] Linton, O., J. Pan and H. Wang (2010). Estimation for a Non Stationary Semi-Strong GARCH(1,1) Model with Heavy-Tailed Errors, Econometric Theory 26, 1-28.
[64] Liu, R.Y. (1988). Bootstrap Procedures Under some Non-I.I.D. Models, Annals of Statistics 16, 1696-1708.
[65] Mason, D.M. and Q-M. Shao (2001). Bootstrapping the Student t-Statistic, Annals of Probability 29, 1435-1450.
[66] Mittnik, S., S.T. Rachev, J-R. Kim (1998). Chi-Square-Type Distributions for Heavy-Tailed Variates, Econometric Theory 14, 339-354.

[67] Newey, W.K. (1985). Maximum Likelihood Specification Testing and Conditional Moment Tests, Econometrica 53, 1047-1070.

[68] Newey, W.K., and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing, in: R. Engle and D. McFadden, eds., Handbook of Econometrics, Vol. 4. Amsterdam: North-Holland.

[69] Ollila, E., H.O. Thomas, and P. Hettmansperger (2002). Estimates of Regression Coefficients Based on the Sign Covariance Matrix, Journal of the Royal Statistical Society Series B 64, 447-466.

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

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

[72] Pham, T. and L. Tran (1985). Some Mixing Properties of Time Series Models. Stochastic Processes and Their Applications 19, 297-303.

[73] Phillips, P.C.B. and M. Loretan (1991). The Durbin Watson Ratio under Infinite Variance Errors, Journal of Econometrics 47, 85-114.

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

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

[77] Rousseeuw, P.J., S. Van Aelst, K. Van Driessen, J. Agulló (2004). Robust multivariate regression, Technometrics 46, 293-305.

[78] Runde, R. (1997). The Asymptotic Null Distribution of the Box-Pierce Q-Statistic for Random Variables with Infinite Variance: An Application to German Stock Returns, Journal of Econometrics 78, 205-216.

[79] White H. (1981). Consequences and Detection of Misspecified Nonlinear Regression Models, Journal of the American Statistical Association 76, 419-433.

[80] White H. (1982). Maximum Likelihood Estimation of Misspecified Models, Econometrica 50, pp. 1-25.

[81] Wooldridge, J.M. (1990). A Unified Approach to Robust, Regression-Based Specification Tests, Econometric Theory 6, 17-43.

Table 3 - Rejection Frequencies : White Noise

Heavy Tails				Thin Tails			
	H_0 : IID	H_0 : GH[a]	H_1 : AR		H_0 : IID	H_0 : GH	H_1 : AR
κ_m under H_0	1.5[b]	1.7	1.5		4.5	4.2	4.5
T = 100							
TT-FIX[c]	.04, .07, .11[d]	.00, .02, .05	.58, .81, .91	TT-Fix	.01, .04, .09	.00, .03, .07	.63, .84, .91
TT-MCD	.03, .07, .12	.01, .02, .05	.66, .86, .93	TT-MCD	.01, .05, .12	.01, .04, .07	.88, .97, .99
TT-OT	.01, .05, .10	.02, .07, .12	.94, .95, .97	TT-OT	.01, .06, .11	.02, .06, .11	.93, .94, .96
MC[e]	.01, .02, .06	.00, .03, .06	.51, .73, .84	MC	.01, .03, .06	.01, .04, .09	.92, .95, .98
Q-LB[f]	.03, .05, .07	.17, .30, .39	1.0, 1.0, 1.0	Q-LB	.01, .04,. 08	.10, .19, .25	.96, .97, .99
Q-Runde	.01, .05, .09	.14, .30, .40	1.0, 1.0, 1.0	Q-Runde	.01, .04,. 08	.10, .19, .25	.96, .97, .99
T = 500							
TT-Fix	.04, .08, .13	.01, .02, .06	.68, .80, .85	TT-Fix	.01, .06, .11	.01, .04, .09	.69, .82, .89
TT-MCD	.02, .06, .11	.00, .02, .05	.81, .87, .91	TT-MCD	.01, .04, .10	.01, .04, .07	1.0, 1.0, 1.0
TT-OT	.01, .06, .11	.01, .06, .10	.97, .98, .99	TT-OT	.01, .05, .11	.01, .05, .10	1.0, 1.0, 1.0
MC	.00, .01, .03	.00, .03, .07	.56, .69, .77	MC	.00, .03, .08	.01, .05, .09	1.0, 1.0, 1.0
Q-LB	.05, .07, .09	.21, .36, .45	1.0, 1.0, 1.0	Q-LB	.01, .05, .09	.08, .16, .23	1.0, 1.0, 1.0
Q-Runde	.01, .05, .09	.55, .71, .78	1.0, 1.0, 1.0	Q-Runde	.01, .05, .09	.08, .16, .23	1.0, 1.0, 1.0
T = 1000							
TT-Fix	.02, .07, .13	.02, .06, .10	.78, .86, .90	TT-Fix	.01, .06, .10	.01, .04, .09	.78, .87, .92
TT-MCD	.01, .06, .11	.00, .02, .06	.81, .89, .92	TT-MCD	.01, .06, .10	.00, .04, .08	1.0, 1.0, 1.0
TT-OT	.01, .05, .10	01, .05, .11	1.0, 1.0, 1.0	TT-OT	01, .05, .10	.01, .05, .10	1.0, 1.0, 1.0
MC	.00, .02, .05	.00, .03, .07	.75, .85, .90	MC	.01, .05, .09	.01, .04, .10	1.0, 1.0, 1.0
Q-LB	.04, .06, .08	.26, .41 ,.49	1.0, 1.0, 1.0	Q-LB	.01, .04, .10	.08, .15, .21	1.0, 1.0, 1.0
Q-Runde	.01, .05, .10	.76, .86, .89	1.0, 1.0, 1.0	Q-Runde	.01, .04, .10	.08, .15, .21	1.0, 1.0, 1.0

a. GH = GARCH(1,1).

b. Tail index κ_m of the test equations $m_{i,t}$ under the null (for GARCH this is an upper bound).

c. TT = TTMC test. FIX indicates a pre-chosen tail parameter $\lambda = .05$ in $k_T = [\lambda T/\ln(T)]$. MCD indicates λ is chosen by the Pseudo-MCD method. OT indicates the test is performed by p-value occupation: we reject the null at the α-level if occupation time $\hat{\tau}_T(\alpha) < \alpha$.

d. Values are rejection frequencies at the 1%, 5% and 10% levels. Frequencies under H_1 are adjusted for size distortions under the iid null.

e. The untrimmed CM test.

f. Q-LB = Ljung-Box. Q-Runde = re-scaled Q-LB in Runde (1997). If $\kappa_m > 2$ then Q-Runde = Q-LB. For $\kappa_m < 2$ critical values are computed by simulating 100,000 samples of iid symmetric stable $\{x_t\}_{t=1}^{T}$ with index κ.

Table 4 - Rejection Frequencies : Omitted Variables

Heavy Tails (1.5)				Thin Tails (4.5)		
	H_0 : IID[a]	H_1 : AR			H_0 : IID	H_1 : AR
T = 100						
TT-FIX[b]	.01, .05, .09[c]	.14, .35, .48		TT-FIX	.01, .05, .08	.18, .43, .58
TT-MCD	.01, .05, .11	.17, .36, .53		TT-MCD	.01, .05, .10	.19, .43, .57
TT-OT	.02, .07, .13	.28, .52, .65		TT-OT	.02, .08, .13	.21, .42, .56
MC	.00, .01, .03	.09, .25, .39		MC	.00, .01, .02	.13, .36, .48
Wald	.01, .04, .06	.73, .88, .93		Wald	.01, .04, .08	.65, .83, .90
T = 500						
TT-FIX	.01, .06, .12	.86, .95, .98		TT-FIX	.01, .04, .08	.99, 1.0, 1.0
TT-MCD	.02, .07, .13	.79, .91, .95		TT-MCD	.01, .06, .11	.99, 1.0, 1.0
TT-OT	.02, .06, .13	.99, 1.0, 1.0		TT-OT	.02, .07, .12	.90, .95, .98
MC	.00, .02, .04	.21, .40, .54		MC	.00, .01, .02	86, .95, .97
Wald	.01, .05, .10	1.0, 1.0, 1.0		Wald	.02, .03, .05	1.0, .10, 1.0
T = 1000						
TT-FIX	.02, .08, .14	.99, 1.0, 1.0		TT-FIX	.01, .05, .09	1.0, 1.0, 1.0
TT-MCD	.01, .05, .09	1.0, 1.0, 1.0		TT-MCD	.01, .06, .11	.84, .94, .97
TT-OT	.02, .06, .12	1.0, 1.0, 1.0		TT-OT	.01, .05, .10	1.0, 1.0, 1.0
MC	.01, .04, .08	.36, .54, .64		MC	.00, .01, .03	.97, .99, .99
Wald	.01, .05, .10	.99, .99, 1.0		Wald	.02, .03, .05	1.0, 1.0, 1.0

a. The null is no omitted variables in an AR(2). The alternative is the second lag is omitted.
b. FIX indicates pre-chosen tail parameter λ = .05 in $k_T = [\lambda T/\ln(T)]$.
c. Values are rejection frequencies at the 1%, 5% and 10% levels.

Table 5 - Robustness to Lag (White Noise, $T = 1000$)

$q = 1$					
	Heavy Tails (1.5)			Thin Tails (4.5)	
	H_0 : IID	H_1 : AR		H_0 : IID	H_1 : AR
TT-FIX[a]	.01, .06, .10	.93, .99, 1.0		.01, .05, .10	1.0, 1.0, 1.0
TT-MCD	.02, .05, .10	.96, 1.0, 1.0		.01, .05, .10	1.0, 1.0, 1.0
TT-OT	.01, .05, .09	1.0, 1.0, 1.0		.01, .05, .10	1.0, 1.0, 1.0
$q = 5$					
	H_0 : IID	H_1 : AR		H_0 : IID	H_1 : AR
TT-FIX	.02, .07, .13	.78, .86, .90		.01, .06, .10	.78, .87, .92
TT-MCD	.01, .06, .11	.81, .89, .92		.01, .06, .10	1.0, 1.0, 1.0
TT-OT	.01, .05, .10	1.0, 1.0, 1.0		.01, .05, .10	1.0 ,1.0, 1.0
$q = 10$					
	H_0 : IID	H_1 : AR		H_0 : IID	H_1 : AR
TT-FIX	.02, .07, .13	.66, .75, .80		.02, .06, .11	1.0, 1.0, 1.0
TT-MCD	.02, .06, .12	.64, .74, .78		.01, .06, .10	1.0, 1.0, 1.0
TT-OT	.01, .05, .10	.99, .99, 1.0		.01, .05, .10	1.0, 1.0, 1.0

a. FIX indicates pre-chosen tail parameter λ = .05 in $k_T = [\lambda T/\ln(T)]$.

Table 6 - Runde's Q-test[a] of White Noise with Plug-In $\hat{\kappa}_{\tilde{k}_T}$

	H_0 : IID		H_0 : GH		H_1 : AR	
T	Rej. %	$\hat{\kappa}_{\tilde{k}_T}$ (1.5^b)	Rej. %	$\hat{\kappa}_{\tilde{k}_T}$ (1.7)	Rej. %	$\hat{\kappa}_{\tilde{k}_T}$ (1.5)
100	.12, .18, .22[c]	$1.52 \pm .74^d$	.26, .34, .39	$1.75 \pm .88$	.99, 1.0, 1.0	1.78 ± 1.5
500	.09, .15, .18	$1.48 \pm .48$	.48, .56, .62	$1.65 \pm .58$	1.0, 1.0, 1.0	1.69 ± 1.0
1000	.08, .14, .18	$1.46 \pm .38$	.57, .64, .68	$1.64 \pm .45$	1.0, 1.0, 1.0	$1.56 \pm .83$

a. The test statistic is $(T/\ln(T))^{2/\hat{\kappa}^*_{\tilde{k}_T}} \times (T+2)\Sigma_{i=1}^{5}\hat{\rho}(i)^2/(T-i)$ where $\hat{\kappa}^*_{\tilde{k}_T}$ is the Hill (1975) estimator of the tail index κ_y averaged over fractiles $\tilde{k}_T \in \{[aT/\ln(T)], ..., [bT/\ln(T)]\}$. We use $\{a,b\} = \{.1,.25\}$ for IID, $\{a,b\} = \{.1,.3\}$ for GARCH and $\{a,b\} = \{1,2\}$ for AR. The sample size is T = 1000.

b. The tail index κ_y of y_t.

c. Values are rejection frequencies at the 1%, 5% and 10% levels.

d. The average 95% band for $\hat{\kappa}_{\tilde{k}_T}$ based on Hill's (2010b) kernel variance estimator for $E[(\tilde{k}_T^{1/2}(\hat{\kappa}_{\tilde{k}_T} - \kappa))^2]$.

Table 7 - Bootstrap White Noise Test[a]

	Heavy Tails				Thin Tails		
	H_0 : IID	H_0 : GH	H_1 : AR		H_0 : IID	H_0 : GH	H_1 : AR
κ_m under H_0	1.5	1.7	1.5		4.5	4.2	4.5
T = 100							
TTMC	.01, .05, .09	.01, .04, .07	.92, .97, .98		.01, .05, .09	.01, .06, .10	.99, .99, 1.0
MC	.01, .04, .08	.01, .03, .06	.92, .96, .98		.02, .06, .10	.01, .06, .10	.99, .99, 1.0
Q-LB	.00, .01, .03	.00, .01, .04	1.0, 1.0, 1.0		.00, .01, .06	.01, .04, .09	1.0, 1.0, 1.0
T = 500							
TTMC	.01, .04, .08	.01, .04, .07	1.0, 1.0, 1.0		.01, .06, .11	.01, .04, .10	1.0, 1.0, 1.0
MC	.00, .03, .06	.00, .01, .02	1.0, 1.0, 1.0		.02, .06, .10	.00, .03, .08	1.0, 1.0, 1.0
Q-LB	.00, .00, .03	.00, .00, .04	1.0, 1.0, 1.0		.01, .03, .08	.00, .03, .08	1.0, 1.0, 1.0
T = 1000							
TTMC	.01, .05, .10	.01, .04, .09	1.0, 1.0, 1.0		.01, .04, .10	.01, .06, .11	1.0, 1.0, 1.0
MC	.00, .03, .07	.00, .01, .02	1.0, 1.0, 1.0		.03, .05, .08	.01, .05, .09	1.0, 1.0, 1.0
Q-LB	.00, .01, .03	.00, .00, .03	1.0, 1.0, 1.0		.01, .04, .09	.00, .03, .09	1.0, 1.0, 1.0

a. The wild bootstrap is used to compute p-values, with 10,000 bootstrap samples for each sample $\{y_t\}$.

Table 8 - Robustness to Asymmetry $(T = 1000, q = 5)$

	$\kappa_1, \kappa_2 = (1.5, 1.5)^a$			$\kappa_1, \kappa_2 = (1.5, 4.5)$			$\kappa_1, \kappa_2 = (4.25, 8.75)$	
White Noise								
	H_0 : IID	H_1 : AR		H_0 : IID	H_1 : AR		H_0 : IID	H_1 : AR
TT-FIX[b]	.01, .04, .09	.83, .91, .93		.01, .07, .13	.80, .86, .91		.01, .05, .10	1.0, 1.0, 1.0
TT-MCD	.01, .05, .11	.85, .93, .96		.01, .06, .12	.81, .85, .92		.01, .05, .11	.99, .99, .99
TT-OT	.01, .05, .10	1.0, 1.0, 1.0		.01, .06, .11	1.0, 1.0, 1.0		.01, .05, .10	1.0, 1.0, 1.0
Omitted Variables								
	H_0 : IID	H_1 : AR		H_0 : IID	H_1 : AR		H_0 : IID	H_1 : AR
TT-FIX	.01, .06, .12	.98, .98, .98		.01, .06, .11	.97, .98, .99		.01, .05, .10	1.0, 1.0, 1.0
TT-MCD	.01, .05, .09	1.0, 1.0, 1.0		.01, .05, .11	.99, .99, .99		.01, .05, .11	.87, .97, .99
TT-OT	.02, .06, .12	1.0, 1.0, 1.0		.01, .06, .11	.99, .99, .99		.01, .05, .10	1.0, 1.0, 1.0

a. Left and right tail indices. b. FIX indicates pre-chosen tail parameter $\lambda = .05$ in $k_T = [\lambda T/\ln(T)]$.