

Moment Condition Tests for Heavy-Tailed Time Series

Jonathan B. Hill* and Michael Aguilar†
Dept. of Economics
University of North Carolina

September 23, 2010

Abstract

We develop an asymptotically chi-squared test statistic for testing moment conditions $E[m_t(\theta^0)] = 0$ where scalar components of $m_t(\theta^0)$ may have an infinite variance and $m_t(\theta^0)$ may be weakly dependent. In general $E[m_t(\theta^0)]$ need not exist under the alternative. A variety of tests can be heavy-tail robustified by our method, including white noise, GARCH affects, omitted variables, order selection, 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 $\hat{\theta}_T$ may be any consistent estimator like QML, LAD, GMM, and Empirical Likelihood as well as robust estimators like Least Trimmed Squares, Least Absolute Weighted Deviations, and Generalized Method of Tail-Trimmed Moments. Simple rules of thumb for selecting the trimming fractiles are presented, and in many cases when $m_t(\theta^0)$ has infinite variance components the fractiles and/or $\hat{\theta}_T$ can be chosen to ensure $\hat{\theta}_T$ does not influence the test statistic's limit distribution. Thus, in heavy tailed cases $\hat{\theta}_T$ does not need to have a Gaussian limit. We apply our statistic to tests of white noise, omitted variables and volatility spillover and find it obtains correct empirical size, while conventional tests exhibit sharp distortions.

1. INTRODUCTION We propose an asymptotically chi-squared test statistic for testing moment conditions in the presence of heavy tails. Let $m_t : \Theta \rightarrow \mathbb{R}^q$ be parametric estimating equations on compact $\Theta \subset \mathbb{R}^r$, where $q, r \geq 1$. We assume $m_t(\theta)$ is continuous and differentiable for simplicity of exposition. The null hypothesis is

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

with a general alternative

$$H_1 : \text{the null is false.} \quad (2)$$

We allow $E[m_{i,t}^2(\theta^0)] = \infty$ in general, and do not require $E[m_t(\theta)]$ to exist under H_1 for any θ . If $m_t(\theta)$ is integrable uniformly on Θ then the alternative becomes $H_1: E[m_t(\theta^0)]$

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

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

Key words and phrases: moment condition test; model specification test; heavy tails; tail trimming; robust inference; volatility spillover; functional form.

JEL subject classifications. C13, C20, C22.

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

The authors thank Eric Renault for many helpful conversations. All errors, of course, are solely the authors'.

$\neq 0$. In general θ may represent a subset of parameters from a regression model setting, for example when testing for the autoregression order in an AR-GARCH, or $m_t(\theta) = m_t$ may be parameter-free as in a test of white noise on an observable time series. We present a variety of examples in Section 2 and augment them with theory details in Section 4.

Moment condition [MC] tests and tests with MC interpretations exhibit a vast history, covering tests of omitted or instrumental variables, functional form, conditional heteroscedasticity, over-identification, causation, volatility spillover, structural change, order selection and encompassing tests. Notable theory contributions for iid data include Newey (1985), Tauchen (1985), Ramsey (1969), Hausman (1978), White (1981), Hansen (1982), Bierens (1990), Imbens et al (1998) and Kitamura et al (2004) to name a few. Moment equality and inequality tests for time series are developed in de Jong (1996), Floren et al (1998), Corradi and Swanson (2002), Ghysels and Guay (2003), Ghysels and Andreou (2003), and Hill (2008) to name a very few.

Evidence for heavy tails across disciplines is substantial, ranging from financial, macroeconomic, auction, actuarial, meteorological to network telecommunication data. The literature is vast, but consider Campbell and Hentschel (1992), Engle and Ng (1993), Embrechts et al (1997), Resnick (1997), Finkenstadt and Rootzén (2003), and Hill and Shneyerov (2010). In general, by their sample moment form MC tests require higher moments on the equations $m_t(\theta^0)$ to ensure a classic limit distribution, and therefore possibly very severe moment restrictions on an underlying process that may fail for even mildly volatile data.

A simple example is a test of mis-specified ARCH order. Consider an ARCH(1)

$$y_t = h_t(\theta^0)u_t \text{ where } h_t^2(\theta) = \omega + \alpha y_{t-1}^2, u_t \stackrel{iid}{\sim} N(0, 1), \omega > 0 \text{ and } \alpha \in (0, 1),$$

with test equations

$$m_{i,t}(\theta) = (y_t^2 - h_t^2(\theta)) y_{t-i}^2, i = 1, 2, \dots, q.$$

The equations $m_{i,t}(\theta^0)$ exhibit power-law and not exponential tail decay (Cline 1986, Mikosch and Stărică 2000, Hill and Renault 2010a), and have a finite variance only if u_t and y_t have finite 4th and 8th moments respectively. This is highly unrealistic for many financial and macroeconomic time series in lieu of heavy tailed noise and/or GARCH-like feedback (Embrechts et al 1997, Finkenstadt and Rootzén 2003, Hall and Yao 2003, Davis and Mikosch 2009a,b, Linton et al 2010).

In order to conquer the challenge of heavy tails, and arrive at a test statistic that is easily computed and interpreted due to a standard limit distribution, we negligibly trim each equation $m_{i,t}(\theta)$. 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: $k_{j,i,T} \rightarrow \infty$ and $k_{j,i,T}/T \rightarrow 0$ (Leadbetter et al 1983). Define tail specific observations of $m_{i,t}(\theta)$ and their sample order statistics:

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

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

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

trim $m_{i,t}(\theta)$ between its lower $k_{1,i,T}/T^{th}$ and upper $k_{2,i,T}/T^{th}$ sample quantiles:

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

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

and $I(A) = 1$ if A is true, and 0 otherwise.

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

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

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

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

and $\hat{m}_T^*(\theta) := 1/T \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T)$, $k(\cdot)$ is kernel function and $\gamma_T \rightarrow \infty$ is bandwidth.

As long as $m_t(\theta)$ satisfies a mixing condition, the trimming indicators $\hat{I}_{i,T,t}(\theta)$ have good metric entropy properties, and the rate of convergence $\hat{\theta}_T \xrightarrow{p} \theta^0$ is fast enough relative to the rate of convergence of $\sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0)$ that the data generating process of $\hat{\theta}_T$ does not over shadow (1), then \hat{W}_T is asymptotically chi-squared under (1). Further, under the same conditions \hat{W}_T has non-negligible power against a sequence of local alternatives, hence $\hat{W}_T \rightarrow \infty$ under (2) with probability one. This relies closely on the assumption that $\hat{\theta}_T$ is consistent under either hypothesis. See Section 3.

We investigate tests of white noise, omitted variables and volatility spillover in a simulation study in Section 5. The control tests for comparisons are an untrimmed version of \hat{W}_T , the Ljung-Box Q-test of white noise, a Wald test of parametric restrictions, and Hong's (2001) test of volatility spillover. Our simulations serve two purposes. First, they demonstrate heavy tails substantially distort empirical size of non-robust tests (Ljung-Box, Wald, Hong 2001), adding to evidence provided for a variety model specification tests in de Lima (1997) and elsewhere (e.g. Runde 1997). Second, trimming remarkably few large $m_t(\theta)$ leads to sharp empirical size, while still permitting substantial power in many cases, and competitive power in other cases.

If the data generating process of $m_{i,t}(\theta^0)$ is known to be symmetric then trimming is symmetric $k_{1,i,T} = k_{2,i,T}$. Otherwise an asymmetric policy $\{k_{1,i,T}, k_{2,i,T}\}$ should be imposed to ensure identification of the null (1) by the trimmed equation $\hat{m}_{T,t}^*(\theta)$ as $T \rightarrow \infty$. We discuss in Section 3 simple rules of thumb for selecting $k_{j,i,T}$ based on three possible criteria: test statistic convergence rate, rate of identification, and whether the plug-in $\hat{\theta}_T$ influences the limit distribution of \hat{W}_T .

Further, if at least one equation $m_{i,t}(\theta^0)$ has an infinite variance then depending the equation form we can choose $\{k_{1,i,T}, k_{2,i,T}\}$ to slow down $\sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0)$, or choose a comparatively fast plug-in $\hat{\theta}_T \xrightarrow{p} \theta^0$ so that $\hat{\theta}_T$ does not affect \hat{W}_T . Thus, in some cases $\hat{\theta}_T$ does not have to have a Gaussian limit. This is possible because $\sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0)$ is $O(T^{1/2})$ -convergent while super- $T^{1/2}$ -convergent $\hat{\theta}_T$ exist for some heavy tailed time

series, including OLS and LAD with non-Gaussian limits, and HR's (2010a) GMTTM and Hill's (2010a) Least Tail-Trimmed Squares [LTTS] with Gaussian limits.

Our framework is built on the principles of Generalized of Method of Tail-Trimmed Moments [GMTTM] by Hill and Renault (2010a), denoted HR (2010a). A matching theory of robust inference, however, does not exist, in particular inference via tail-trimmed equations that are not necessarily used to estimate θ^0 , and with an arbitrary plug-in $\hat{\theta}_T$ that may not have a Gaussian limit. Valid plug-ins include conventional estimators like GMM, NLLS, QML, LAD, and the Empirical Likelihood and information-theoretic variety like CUE-GMM and Exponential Tilting (Hansen et al 1996, Antoine et al 2007, Kitamura and Stutzer 1997); as well as outlier-robust estimators like Least Trimmed Squares (Ruppert and Carroll 1980, Čížek 2008); and heavy tail-robust estimators like HR's (2010a) GMTTM, Ling's (2005, 2007) Least Absolute Weighted Deviations [LAWD] and Quasi-Maximum Weighted Likelihood [QMWL], Hill's (2010a) LTTS, and R-estimators (Jaeckel 1972, Andrews 2008).

Certainly trimming by a *fixed* quantile of $m_t(\theta)$ simplifies limit theory since $k_{j,i,T}/T \rightarrow (0, 1)$ ensures $\hat{m}_{T,t}(\theta^0)$ has a finite variance even asymptotically. Fixed quantile trimming and truncation are primary tools for outlier robust estimation (e.g. Huber 1977; see Cizek 2008, 2009 and his citation). But there is no guarantee the trimmed equations will identify the null (1) in the sense $1/T \sum_{t=1}^T \hat{m}_{T,t}(\theta^0) \xrightarrow{p} 0$, when the data generating process is nonlinear. Further, we are not claiming the data are contaminated: we trim only to induce a standard limiting distribution for a test statistic. Indeed, only *tail*-trimming robustifies against heavy tails *and* bias in general settings. Bias correction by simulated method of moments requires knowledge of an underlying distribution (Ronchetti and Trojani 2001), and otherwise bias is merely assumed away (Čížek 2009). See HR (2010a). Conversely, the lightest trimming case $k_{j,i,T} \rightarrow k_{j,i}$, a fixed integer, results in too few equations removed to ensure a standard null distribution.

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 Hill (2010a, 2010b) and HR (2010) for detailed literature reviews.

The proposed TTMC statistic is generalistic. If a particular context is entertained then a different statistic form and therefore tail-trimming strategy may be optimal. A robust test of white noise, for example, can easily be couched in terms of (1) and therefore tested by \tilde{W}_T , but it is also conceivable to tail-trim sample covariances for a robust portmanteau statistic. The large variety of possible tests makes entertaining specific trimming strategies cases impossible, and is therefore left by-case for future research.

There are at least four major veins of inference in the presence of heavy tails. First, re-scaled tests obtain non-standard limits like t-ratios, portmanteau statistics, tests of covariance stationarity, unit roots, cointegration and GARCH (Davis and Resnick 1986, Chan and Tran 1989, Loretan and Phillips 1991, Davis et al 1992, Runde 1997, Caner 1998, Hall and Yao 2003). 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. Schmidt and Stadtmüller 2006, Davis and Mikosch 2009b). There are few attempts in the literature to extend such methods to econometric specification contexts, although tail behavior of regression estimators is used to model breakdown points (Jurečková 1981, He et al 1990) and efficiency and majorization (Ibragimov 2007).

The third class is distribution free tests and non-parametric inference, including rank-order tests of unit roots, correlation integral-based tests of dependence, and bootstrapped confidence bands (e.g. Breitung and Gouriéroux 1997, Brock et al 1996, de Lima 1996, Mason and Shao 2001).

The fourth class includes statistics derived from heavy tail robust methods which

therefore have standard limits. Examples are Ling's (2005) and Hill's (2010a) Wald statistics respectively for LAWD and LTTS estimators, and Hill's (2010b) kernel variance estimator for a tail-trimmed sum. In these cases inference and limit theory are developed for a particular estimation problem. 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. Examples are given in the next section.

We use the following notation conventions. If $A_T(\theta)$ is a matrix function of θ we write

$$A_T = A_T(\theta^0).$$

$\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the minimum and maximum eigenvalues of A . The L_p -norm is $\|x\|_p = (\sum_{i,j} E|x_{i,j}|^p)^{1/p}$, and the spectral (matrix) norm is $\|A\| = (\lambda_{\max}(A'A))^{1/2}$. $(z)_+ := \max\{0, z\}$. K denotes a positive finite constant whose value may change from line to line; $\iota > 0$ is a tiny constant; N is a whole number. \xrightarrow{p} and \xrightarrow{d} denote probability and distribution convergence. $[z]$ denotes the integer part of z .

2. EXAMPLES OF MOMENT CONDITION TESTS Evidently any test with a moment condition interpretation can be couched within our tail-trimming framework. We give examples of tests of white noise, omitted variables, functional form, GARCH specification, volatility spillover and over-identification. In Section 4 we complete several examples by showing how the theory developed in Section 3 applies, and verify the major assumptions.

Unless otherwise specified the model is

$$y_t = f(x_t, \theta) + \epsilon_t(\theta), \text{ where } f : \mathbb{R}^p \times \Theta \rightarrow \mathbb{R} \text{ and } \Theta \subset \mathbb{R}^r, \quad (5)$$

where $f(\cdot, \theta)$ is continuous and differentiable, $\epsilon_t = \epsilon_t(\theta^0)$ is an idiosyncratic shock for unique θ^0 , and $x_t \in \mathbb{R}^p$ and $z_t \in \mathbb{R}^s$ are non-redundant regressors, $r, s \geq 1$.

EXAMPLE 1 (White Noise): A test of white noise in the errors is a test of (1) with equations

$$m_t(\theta) = \epsilon_t(\theta) \times [\epsilon_{t-1}(\theta), \dots, \epsilon_{t-q}(\theta)]'.$$

If ϵ_t has an infinite variance then each $m_{i,t}(\theta^0) = \epsilon_t \epsilon_{t-i}$ has an infinite variance, and if there are GARCH affects $E[m_{i,t}^2(\theta^0)] < \infty$ may require ϵ_t to have up to a finite fourth moment.

A test of white noise on zero mean y_t uses $m_t = y_t \times [y_{t-1}, \dots, y_{t-q}]'$, so the test equations may not depend on a parameter θ .

EXAMPLE 2 (Omitted Variables): A simple orthogonality test of omitted variables z_t in an additive form $y_t = f(x_t, \theta^0) + \beta' z_t + \epsilon_t$ checks whether $E[m_t(\theta^0)] = E[\epsilon_t z_t] = 0$. A general test of omitted variables can be treated as in Fan and Li (1996): see Example 4, below.

Consider an AR(1) as a simple example:

$$y_t = \theta^0 y_{t-1} + \epsilon_t, \quad |\theta^0| < 1,$$

with $z_t = [y_{t-2}, \dots, y_{t-p}]'$. The error must satisfy $E[\epsilon_t^2] < \infty$ for a score test of $E[\epsilon_t y_{t-2}] = 0$ based on OLS or GMM. Notice if $\theta^0 \neq 0$ and $E[\epsilon_t^2] = \infty$ then $E[\epsilon_t(\theta) y_{t-2}] = E[\epsilon_t y_{t-2}] + (\theta - \theta^0) E[y_{t-1} y_{t-2}]$ does not exist for any $\theta \neq \theta^0$, even under the null, so the proper alternative is indeed (2).

Now assume the error is IGARCH(1,1) $y_t = \theta^0 y_{t-1} + \epsilon_t$, $|\theta^0| < 1$, where $\epsilon_t = h_t u_t$, $h_t^2 = \omega^0 + \alpha^0 \epsilon_{t-1}^2 + (1 - \alpha^0) h_{t-1}^2$, $\alpha^0 \in (0, 1)$ and $u_t \stackrel{iid}{\sim} N(0, 1)$. Then equations like $m_{i,t}(\theta^0)$

$= \epsilon_t y_{t-i}$ have an infinite variance $E[m_{i,t}^2(\theta^0)] = \infty$ and are not integrable anywhere on Θ/θ^0 since $|E[m_{i,t}(\theta)]| = \infty$ for any $\theta \neq \theta^0$. The GARCH parameters ω and α may be estimated along with θ , as in QML. Thus, implicit in our framework are moment condition tests where θ is a subset of parameters.

EXAMPLE 3 (Neural Test of Neglected Nonlinearity): Consider Lee et al's (1996) version of Bierens' (1990) celebrated consistent test of functional form. Assume y_t is integrable so it has a conditional expectations by the Radon-Nikodym Theorem. By hypothesis $f(x_t, \theta^0)$ is a version of $E[y_t|x_t]$ a.s. for unique $\theta^0 \in \Theta$. Let $\hat{\theta}_T$ be an estimator of θ^0 , and $\hat{\epsilon}_t = y_t - f(x_t, \hat{\theta}_T)$. The test statistic is

$$\hat{R}_T(\gamma) = \left(\frac{1}{T^{1/2}} \sum_{t=1}^T \hat{\epsilon}_t F(\gamma' \psi(x_t)) \right)^2 / \hat{v}_T^2(\gamma) \quad \text{with } F(u) = \frac{1}{1 + \exp\{u\}},$$

where $\hat{v}_T^2(\gamma)$ estimates $E[\{1/T^{1/2} \sum_{t=1}^T \hat{\epsilon}_t F(\gamma' \psi(x_t))\}^2]$ under the null, $\gamma \in \Gamma \subset \mathbb{R}^p$ is a nuisance parameter, and $\psi : \mathbb{R}^r \rightarrow \mathbb{R}^r$ is a bounded one-to-one function.

Define the scalar test equation

$$m_t(\theta, \gamma) := \epsilon_t(\theta) \times F(\gamma' \psi(x_t)).$$

Under the null $E[m_t(\theta^0, \gamma)] = 0$ for every γ by iterated expectations. The statistic $\hat{R}_T(\gamma)$ is grounded on the fact that if ϵ_t is integrable, x_t is finite dimensional, and the null is false $E[\epsilon_t|x_t] \neq 0$ then $F(\gamma' \psi(x_t))$ is "revealing" (Stinchcombe and White 1998: Definition 2.1). That is $E[m_t(\theta^0, \gamma)] = E[\epsilon_t F(\gamma' \psi(x_t))] \neq 0$ for all γ on any compact Γ except for γ in $S \subset \Gamma$ with Lebesgue measure zero (Bierens and Ploberger 1997, Stinchcombe and White 1998). The result carries over to any non-polynomial real analytic function $F : \mathbb{R} \rightarrow \mathbb{R}$ with affine argument $\gamma' \psi(x_t)$ (Stinchcombe and White 1998).

In order for $\hat{R}_T(\gamma) \xrightarrow{d} \chi^2(1)$ under the null, the gradient $(\partial/\partial\theta_i)f(x_t, \theta)$, uniformly on Θ , and the error ϵ_t must have finite $4 + \iota^{th}$ -moments to ensure $\hat{v}_T^2(\gamma)$ is consistent for weakly dependent data (e.g. de Jong 1996, Hill 2008).

EXAMPLE 4 (Hong-White and Fan-Li Tests of Functional Form): Consider the Example 3 framework and define conditional moments (assumed to exist)

$$\mathcal{Y}(x_t) := E[y_t|x_t] \quad \text{and} \quad \mathcal{E}(x_t) := E[\epsilon_t|x_t].$$

Although nuisance parameter indexing within a flexible functional form can consistently reveal mis-specification, Hong and White (1995) note $E[m_t(\theta^0)] = E[\epsilon_t(f(x_t, \theta^0) - \mathcal{Y}(x_t))] = E[(f(x_t, \theta^0) - \mathcal{Y}(x_t))^2] = 0$ if and only if $\mathcal{E}(x_t) = 0$. Thus $f(x_t, \theta^0) - \mathcal{Y}(x_t)$ is revealing and does not depend on a nuisance parameter. They suggest a non-parametric estimator of $\mathcal{Y}(x_t)$ where $f(x_t, \theta^0)$ is known by hypothesis.

A similar test for omitted variables and functional form mis-specification is developed in Fan and Li (1996). They exploit $E[(\epsilon_t \mathcal{E}(x_t))] = E[\mathcal{E}(x_t)]^2 = 0$ if and only if $\mathcal{E}(x_t) = 0$ a.s., and propose a nonparametric estimator of $\mathcal{E}(x_t)$. Both approaches suggest an MC test based on $E[\epsilon_t w_t]$ with w_t identically $f(x_t, \theta^0) - \mathcal{Y}(x_t)$ or $\mathcal{E}(x_t)$.

We do not tackle non-parametric function estimation in this paper: we only allow a parametric plug-in $\hat{\theta}_T$ to keep arguments tight. It seems apparent, however, that the methods developed here have a natural analogue for testing the above moments with nonparametric plug-ins.

EXAMPLE 5 (GARCH Specification): Construction a GARCH model

$$y_t = h_t(\theta^0) \epsilon_t \quad \text{where} \quad h_t^2(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2(\theta), \quad \omega > 0, \quad \alpha, \beta \geq 0,$$

and $\epsilon_t \stackrel{iid}{\sim} (0, 1)$, with parameter set $\theta = [\omega, \alpha, \beta]'$. If the true data generating process is semi-strong GARCH(1,1) then $y_t/h_t(\theta^0)$ is white noise, so Example 1 applies.

Otherwise, a test of GARCH(p, q) against GARCH(r, s) can be constructed as a QML score statistic (see Bollerslev 1986), with equations

$$m_t(\theta) = \left(\frac{y_t^2}{h_t^2(\theta)} - 1 \right) \times h_t^2(\theta) \times \frac{\partial}{\partial \theta} h_t^2(\theta).$$

$$m_t(\theta^0) = (\epsilon_t^2 h_t^2(\theta^0) - h_t^2) \times \frac{\partial}{\partial \theta} h_t^2(\theta).$$

Under the null $h_t^2(\theta^0) = \omega^0$ hence $E[m_{i,t}^2(\theta^0)] < \infty$ if and only if $E[\epsilon_t^4] < \infty$, and $E[m_{i,t}(\theta)]$ does not exist if the GARCH model is mis-specified and $E[\epsilon_t^4] = \infty$.

In lieu of recent efforts to model volatility with heavy tailed errors (Hall and Yao 2003, HR 2010a, Linton et al 2010), a robust GARCH specification test is desired.

EXAMPLE 6 (Volatility Spillover): A rich literature has emerged on testing for market associations and contagion, and stock price/volume relationships during volatile periods (King et al 1994, Brooks 1998, Comte and Lieberman 2000, Hong 2001, Caporale et al 2002, Forbes and Rigobon 2002). Let $\{y_{1,t}, y_{2,t}\}$ be a joint process of interest with GARCH(1,1) coordinates: each $y_{i,t}$ satisfies under the null

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

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. 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. See the simulation study in Section 5 for details.

Define 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 $E[m_{j,t}(\theta^0)] = E[(\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-j}^2 - 1)] = 0$, and $E[m_{j,t}^2(\theta^0)] < \infty$ requires at least $E[\epsilon_{i,t}^4] < \infty$. Under tail-trimming we only need the equations $m_t(\theta^0)$ to be integrable under the null hence $E[\epsilon_{i,t}^2] < \infty$, a substantial improvement over Hong (2001).

EXAMPLE 7 (Over-Identification): By construction \hat{W}_T can be used as Hansen's (1982) test of over-identifying restrictions when $m_t(\theta)$ is used to estimate θ by GMM or GMTTM (Hansen 1982, HR 2010a). The truly interesting aspect of such a test occurs when some equations $m_{i,t}(\theta^0)$ have an infinite variance since we may use a variety of plug-ins that lead to a consistent test: asymptotic power of one against rejection of over-identification. See Section 3, below, for theory details.

3. ROBUST MOMENT CONDITION TESTS Denote by $L_i(\theta), U_i(\theta) \in [0, \infty]$ equation specific support bounds: $-L_i(\theta) \leq m_{i,t}(\theta) \leq U_i(\theta)$ a.s. The potential problem with testing (1) is at least one equation may have an unbounded support and infinite variance.

Equations $m_{i,t}(\theta^0)$ known to have a bounded variance are logically untrimmed for the sake of efficiency. Assume the first $q \in \{1, \dots, q\}$ equations are trimmed:

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

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 than *all* equations are trimmed: $\underline{q} = q$. Trimming equations with finite variance may reduce efficiency (HR 2010a), but the obverse is precisely the crux of this paper: not to trim equations with infinite variance will result in a non-standard or degenerate limit distribution of \hat{W}_T .

In this section we define fractile and threshold sequences, detail plug-in properties, discuss why a HAC estimator is needed in general for the test statistic scale, state the main results and conclude with details on fractile choice.

3.1 Threshold and Fractile Sequences

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

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

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

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

$$\frac{T}{k_{1,i,T}} P(m_{i,t}(\theta) < -l_{i,T}(\theta)) = 1 \text{ and } \frac{T}{k_{2,i,T}} P(m_{i,t}(\theta) > u_{i,T}(\theta)) = 1. \quad (6)$$

Thus, $l_{i,T}(\theta)$ and $u_{i,T}(\theta)$ are identically the equation specific lower $k_{1,i,T}/T^{th} \rightarrow 0$ and upper $k_{2,i,T}/T^{th} \rightarrow 0$ tail quantiles. 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 absolutely continuous marginal distributions. See Appendix A for all assumptions and related discussion.

The TTMC statistic (4) involves $\hat{m}_{T,t}^*(\theta)$ in (3), but asymptotic theory is grounded on deterministic trimming with equations

$$\begin{aligned} m_{T,i,t}^*(\theta) &:= m_{i,t}(\theta) \times I(-l_{i,T}(\theta) \leq m_{i,t}(\theta) \leq u_{i,T}(\theta)) \\ &= m_{i,t}(\theta) \times I_{i,T,t}(\theta) \quad : \quad 1 \leq i \leq \underline{q} \end{aligned} \quad (7)$$

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

In Appendix C we show $\hat{m}_{T,t}^*(\theta^0)$ is sufficiently close to $m_{T,t}^*(\theta^0)$ in the sense

$$S_T^{-1/2} \sum_{t=1}^T \{\hat{m}_{T,t}^*(\theta^0) - m_{T,t}^*(\theta^0)\} = o_p(1),$$

where S_T is the covariance matrix for $\sum_{t=1}^T m_{T,t}^*(\theta^0)$, defined below. Thus, all asymptotic arguments are grounded on $m_{T,t}^*(\theta^0)$, 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:

$$H_0 : E[m_{T,t}^*(\theta^0)] \rightarrow 0. \quad (1')$$

The condition is easily guaranteed by Lebesgue's dominated convergence since tail trimming is negligible and $m_t(\theta^0)$ is integrable by construction under H_0 . Further, if $m_{i,t}(\theta^0)$ exhibit power law tail decay then it is easy to characterize arbitrarily many fractile sequences $\{k_{1,i,T}, k_{2,i,T}\}$ that ensure $E[m_{T,t}^*(\theta^0)] \rightarrow 0$ arbitrarily fast. See Section 3.7, below.

Indeed, (1') is trivial if each $m_{i,t}(\theta^0)$ is symmetrically distributed at zero and trimming is symmetric $l_{i,T}(\theta) = u_{i,T}(\theta)$ since $E[m_{T,t}^*(\theta^0)] = 0$. This is a key distinction between tail and fixed quantile trimming where bias may arise by the latter.

3.2 Plug-In Properties

In simple contexts $m_t(\theta^0) = m_t$ is non-parametric, as in a test of white noise for an observable process $\{y_t\}$. In parametric contexts we assign to the plug-in $\hat{\theta}_T$ a sequence of positive definite scale matrices $\{\tilde{V}_T\}$, $\tilde{V}_T \in \mathbb{R}^{r \times r}$, with diagonal components $\tilde{V}_{i,i,T} \rightarrow \infty$, 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, cf. Akaike 1973, White 1982). As long as $\hat{\theta}_T \xrightarrow{p} \theta^0$ sufficiently fast (i.e. $\tilde{V}_T^{1/2} \rightarrow \infty$ fast enough) 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 $\sum_{t=1}^T \{m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)]\} = O_p(T^{1/2})$. But, as we detail below and in Section 4, in many cases heavy tails introduce a unique advantage for ensuring some estimators $\hat{\theta}_T$ have no impact on \hat{W}_T , allowing $\hat{\theta}_T$ with a non-Gaussian limit.

In general $\hat{\theta}_T$ may not have $T^{1/2}$ -convergent components, and it may have components that converge faster or slower than $T^{1/2}$ (cf. Antoine and Renault 2010, Hill 2010a, HR 2010a). We therefore call $\tilde{V}_T^{1/2}$ the compound rate of convergence and $\|\tilde{V}_T\|^{1/2}$ the maximum rate of convergence.

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 covariance, Jacobian and scale matrices associated with the expansion (cf. Newey and McFadden 1994):

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

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

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

Under regulatory conditions detailed in Appendix A, \hat{S}_T and $\hat{m}_{T,t}^*(\theta)$ obtain an asymptotic expansion

$$\hat{S}_T^{-1}(\hat{\theta}_T) \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) = \left\{ S_T^{-1} \sum_{t=1}^T m_{T,t}^*(\theta^0) + T S_T^{-1/2} J_T(\hat{\theta}_T - \theta^0) \right\} \times (1 + o_p(1)) + o_p(1), \quad (8)$$

where $T S_T^{-1/2} J_T$ satisfies

$$\{T S_T^{-1/2} J_T\} \times V_T^{-1} \times \{T S_T^{-1/2} J_T\}' \rightarrow I_q.$$

See especially the proof Theorem 3.1 in Appendix B, and see Lemmas C.3-C.6 in Appendix C. We ensure \hat{W}_T actually tests (1) by assuming the plug-in rates $\tilde{V}_{i,i,T}^{1/2} \rightarrow \infty$ 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 GMTTM over-identifying restrictions where $m_t(\theta^0)$ are both estimating and test equations provides the intuition. If $\hat{\theta}_T$ is the efficiently weighted GMTTM estimator based on $m_t(\theta^0)$ with trimming fractile $\{k_{1,i,T}, k_{2,i,T}\}$, then $V_T^{1/2}$ is exactly the GMTTM scale: under (1) and regulatory conditions outlined in Appendix A, below, $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{d} N(0, I_r)$. See HR (2010a: Theorem 2.2). 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 $\|V_T \tilde{V}_T^{-1}\| = O(1)$.

In practice this requires knowledge of the asymptotic properties of $\hat{\theta}_T$ and $m_t(\theta^0)$, hence we must have a particular test in mind. Nevertheless, depending on the equation form when a component $m_{i,t}(\theta^0)$ has an infinite variance certain estimators $\hat{\theta}_T$ inherently satisfy $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{p} 0$, or for a chosen $\hat{\theta}_T$ we can ensure $V_T^{1/2}(\hat{\theta}_T - \theta^0) \xrightarrow{p} 0$ by trimming $m_t(\theta)$ enough that $TS_T^{-1/2}J_T$ is slow relative to $\hat{\theta}_T \xrightarrow{p} \theta^0$. See Section 3.7 for fractile choice details. Thus, depending on the test, plug-in and fractiles, $\hat{\theta}_T$ may not influence \tilde{W}_T . In such a case $\hat{\theta}_T$ does not have to have a non-Gaussian limit. 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 we assess the impact of $\hat{\theta}_T$ on \tilde{W}_T by assuming $\hat{\theta}_T$ is grounded on some array of parametric estimating equations

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

that may depend on additional parameters ζ . In general $\tilde{m}_{T,t}(\theta, \zeta)$ is from an estimation problem like GMM, CUE, QML, LAWD or GMTTM. Thus, θ may be a subset of parameters of interest as in Examples 2 and 5. Since ζ does not play any role here, and treatment of it merely deviates from the central theme, without too much loss of generality assume¹

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

In this case we assume $\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0)$ is asymptotically linear in $\sum_{t=1}^T \{\tilde{m}_{T,t}(\theta^0) - E[\tilde{m}_{T,t}(\theta^0)]\}$ which satisfies a Gaussian central limit theorem.

This is important to note: if $\hat{\theta}_T \xrightarrow{p} \theta^0$ relatively fast $\|V_T \tilde{V}_T^{-1}\| \rightarrow 0$ then $\hat{\theta}_T$ need not be asymptotically normal; and otherwise we must assume $\hat{\theta}_T$ is grounded on equations $\tilde{m}_{T,t}(\theta^0)$ that belong to the normal domain of attraction. The latter implies either the data are sufficiently thin tailed that a conventional plug-in like OLS, LAD, QML and GMM has a normal limit, or $\hat{\theta}_T$ is heavy tailed robust, like LAWD, QMWL, LTTS, GMTTM and R-estimators. We show in Section 4 how OLS, LAWD, QML, QMWL, LTTS and GMTTM variously satisfy the required rates of convergence for different test environments, and use those estimators in the simulation study in Section 5. A large variety of estimators can be similarly verified.

We rule out the perverse case $\|V_T \tilde{V}_T^{-1}\| \rightarrow 0$ since that implies some components $\hat{\theta}_{i,T}$, and therefore $\tilde{m}_{T,t}(\theta)$, are so dominant in the limit the test equations $m_{T,t}^*(\theta)$ have no impact on \tilde{W}_T . Nevertheless, in principle this may not be a problem if $\tilde{m}_{T,t}(\theta)$ satisfies

$$E[\tilde{m}_{T,t}(\theta)] \rightarrow 0 \text{ if and only if } \theta = \theta^0.$$

¹This is the same as assuming the "true" value ζ^0 is known. The theory that follows easily allows for a plug-in $\hat{\zeta}_n$ that is consistent $\hat{\zeta}_n \xrightarrow{p} \zeta^0$ sufficiently fast (e.g. no slower than $\hat{\theta}_n \xrightarrow{p} \theta^0$). See plug-in properties P1-P2 in Appendix A.

In this case we would replace $m_{T,t}^*(\theta)$ with $\tilde{m}_{T,t}(\theta)$ and compute the test statistic from $\tilde{m}_{T,t}(\theta)$.

Finally, although we assume $m_t(\theta)$ is continuous and differentiable, we make no such assumptions on $\tilde{m}_{T,t}(\theta)$. Nevertheless, $m_{T,t}(\theta)$ and $\tilde{m}_{T,t}(\theta)$ may have shared elements, so define the total set of unique equations $\mathcal{M}_{T,t}^*(\theta)$:

$$\mathcal{M}_{T,t}^*(\theta) \in \mathbb{R}^s \text{ where } m_{T,t}^*(\theta), \tilde{m}_{T,t}(\theta) \in \mathcal{M}_{T,t}^*(\theta), s \geq \max\{p, q\}. \quad (9)$$

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

3.3 HAC Estimator

A HAC estimator $\hat{S}_T(\theta)$ is not required if the equations $m_t(\theta^0)$ are sufficiently orthogonal that

$$S_T(\theta) = T \times E \left[\left\{ m_{T,s}^*(\theta) - E[m_{T,s}^*(\theta)] \right\} \left\{ m_{T,t}^*(\theta) - E[m_{T,t}^*(\theta)] \right\}' \right] = T \times \Sigma_T(\theta),$$

say. In this case $\hat{S}_T(\hat{\theta}_T) = T \hat{\Sigma}_T(\hat{\theta}_T)$ suffices, where

$$\hat{\Sigma}_T(\hat{\theta}_T) = \frac{1}{T} \sum_{t=1}^T \left\{ \hat{m}_{T,t}^*(\hat{\theta}_T) - \hat{m}_T^*(\hat{\theta}_T) \right\} \left\{ \hat{m}_{T,t}^*(\hat{\theta}_T) - \hat{m}_T^*(\hat{\theta}_T) \right\}'.$$

In general a HAC estimator is preferred even if $\{m_t(\theta^0), \mathfrak{F}_t\}$ is a martingale difference for some non-decreasing sigma-field \mathfrak{F}_t since $m_{T,t}^*(\theta^0)$ may *not* be a martingale difference for each T . Even asymptotically there are two forces: negligible trimming ensures $m_{T,t}^*(\theta^0)$ becomes like the martingale difference $m_t(\theta^0)$, but as T grows there are more cross terms in S_T relating serial dependence in the trimmed equations. Unless more information is provided, in general there is no guarantee $T^{-1}\Sigma_T^{-1}S_T \rightarrow I_q$ fast enough to overcome dependence across the accumulation of observations $T \rightarrow \infty$.

3.4 Main Results

The main results of the paper follow. First, the test statistic is asymptotically chi-squared under the null. See Appendix A for all assumptions concerning distribution properties (D), identification and moment smoothness (I), the HAC kernel (K), and the plug-in (P); and see Appendix B for all proofs.

The simplest case is when $m_t(\theta)$ is non-parametric.

THEOREM 3.1 *Suppose $m_t(\theta) = m_t$, and let D1-D6, I1-I4, and K1 hold. Under the null (1) $\hat{W}_T \xrightarrow{d} \chi^2(q)$ a chi-squared law with q degrees of freedom.*

The general case complicates degrees of freedom.

THEOREM 3.2 *Let D1-D6, I1-I4, K1, and P1 or P2 hold. Under the null (1) $\hat{W}_T \xrightarrow{d} \chi^2(\xi)$ where degrees of freedom ξ depends on the rate of convergence of $\hat{\theta}_T$. In particular, by case if P1 holds such that $\hat{\theta}_T \xrightarrow{p} \theta^0$ fast enough that $\|V_T \tilde{V}_T^{-1}\| \rightarrow 0$ then $\xi = q$ the number of estimating equations; and if P2 holds such that $\tilde{V}_T \sim \mathcal{KV}_T$ 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 and the dimension of θ^0 .*

3.5 Degrees of Freedom

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 plug-in property 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 (e.g. white noise with QMWL; functional form with exactly identified GMTTM), and applies to all cases in our simulation study of Section 5.

COROLLARY 3.3 *Let D1-D6, I1-I4, K1, and P2 hold, and assume all equations are unique ($s = q + p$) and θ^0 is exactly identified ($r = p$). Then under the null (1) $\hat{W}_T \xrightarrow{d} \chi^2(q)$.*

Otherwise, as in (9) decompose $\sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0)$ into components that reveal the impact of $\hat{\theta}_T$. The following uses arguments from the proof of Theorem 3.1 and notation developed in Appendix A. Plug-in case P1 implies $\|V_T \tilde{V}_T^{-1}\| \rightarrow 0$ so $TS_T^{-1/2} J_T(\hat{\theta}_T - \theta^0) \xrightarrow{p} 0$. The plug-in $\hat{\theta}_T$ has no impact on \hat{W}_T asymptotically:

$$\hat{W}_T = \left(\sum_{t=1}^T m_{T,t}^*(\theta^0) \right)' S_T^{-1} \left(\sum_{t=1}^T m_{T,t}^*(\theta^0) \right) (1 + o_p(1)) + o_p(1).$$

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

The more challenging case is plug-in property P2 where $V_T \sim \mathcal{K}\tilde{V}_T$ since we can only say $TS_T^{-1/2} J_T(\hat{\theta}_T - \theta^0) = O_p(1)$. We therefore assume $\hat{\theta}_T$ is asymptotically a linear function in $\tilde{m}_{T,t}(\theta^0)$:

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

where $\tilde{A}_T \in \mathbb{R}^{r \times p}$ satisfies $\tilde{A}_T \tilde{S}_T \tilde{A}_T' \rightarrow I_r$ and \tilde{S}_T is the covariance matrix of $\sum_{t=1}^T \tilde{m}_{T,t}(\theta^0)$. In this case $\hat{S}_T^{-1/2} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T)$ reduces to

$$\begin{aligned} \hat{S}_T^{-1/2}(\hat{\theta}_T) \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) &= S_T^{-1/2} \sum_{t=1}^T \{ m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)] \} \\ &\quad + TS_T^{-1/2} J_T \tilde{V}_T^{-1} \tilde{A}_T \sum_{t=1}^T \{ \tilde{m}_{T,t}(\theta^0) - E[\tilde{m}_{T,t}(\theta^0)] \} (1 + o_p(1)) + o_p(1). \end{aligned}$$

Degrees of freedom are therefore governed by the over-lap of $m_{T,t}^*(\theta^0)$ and $\tilde{m}_{T,t}(\theta^0)$, 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 $\tilde{V}_T = V_T$, $\tilde{A}_T S_T \tilde{A}_T' \rightarrow I_r$ and

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

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. Then $[I_q, T J_T \tilde{V}_T^{-1/2} \tilde{S}_T^{-1/2}]$ 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 ($p = r$).

3.6 Local Alternative

Now consider a class of local alternatives with so-called Pitman drift:

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

We assume $m_t(\theta^0)$ is geometrically β -mixing in Appendix A, so if all $m_{i,t}(\theta^0)$ have finite variances then $S_T \sim T \times S$ for some positive definite matrix $S \in \mathbb{R}^{q \times q}$. In this case $H_{1,L}$ represents a sequence of $T^{1/2}$ -local alternatives.

Otherwise $H_{1,L}$ captures non-degenerate yet $o(T^{1/2})$ -alternatives. This follows from three observations. First, $\|S_T\|/T \rightarrow \infty$ some $m_{i,t}(\theta^0)$ has an infinite variance. Second, S_T has an upper bound by Lemma C.2 in Appendix C:

$$S_T = o\left(T^2 \max\left\{1, \|E[m_{T,t}^*(\theta^0)]\|^2\right\}\right).$$

Third, apply Lebesgue's dominated convergence to deduce $\limsup_{T \geq 1} \|E[m_{T,t}^*(\theta^0)]\| \leq K$ under $H_{1,L}$. Together $H_{1,L}$ captures a sequence of $TS_T^{-1/2}$ -convergent alternatives where $TS_T^{-1/2} \rightarrow \infty$ and $TS_T^{-1/2} = o(T^{1/2})$. This is worth highlighting: the sequence of alternatives converges slower than $T^{1/2}$ when the equations have an infinite variance, and *monotonically slower* with heavier tails since in general $\|S_T\| \rightarrow \infty$ *monotonically faster*.

THEOREM 3.4 *Let D1-D6, I3-I4, K1, P1 or P2 if a plug-in is required, and $H_{1,L}$ hold. Then $\hat{W}_T \xrightarrow{d} \chi_\xi^2(v'v)$ an noncentral chi-squared law with ξ degrees of freedom characterized in Theorems 3.1 and 3.2, and noncentrality parameter $v'v \in [0, \infty)$.*

Remark 1: Theorem 3.4 ensures \hat{W}_T is consistent against arbitrary non-local deviations (2) from the null (1). Under a global alternative where $E[m_t(\theta)]$ may not exist for any θ , or $TS_T^{-1/2} E[m_t(\theta)] \rightarrow \infty$ for any θ , then $P(\hat{W}_T > w) \rightarrow 1$ for all $w > 0$ under H_1 .

Remark 2: Consistency is not without a price. We must have a consistent plug-in $\hat{\theta}_T \rightarrow \theta^0$ under 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 mis-specified. In the former case we must implicitly correctly specify an encompassing model as in Example 2 of Section 2. In the latter case $\hat{\theta}_T \rightarrow \theta^0$ even if the Example 3 regression model error $E[\epsilon_t|x_t] \neq 0$ with positive probability. Evidently the literature on regression model estimation for heavy tailed data nearly universally 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). HR's (2010a) GMTTM estimator is a notable exceptions since only a mixing condition on the estimating equations is required, implicitly covering regression models with non-martingale difference errors. This topic, however, is well beyond the focus of the present paper.

3.7 Optimal Fractile Selection

Any intermediate order sequences $\{k_{j,i,T}\}$ in principle are valid for trimming. In general choosing between policies $\{k_{1,i,T}, k_{2,i,T}\}$ requires information about the data generating process, and a criterion for defining an "optimal" policy. Even if $m_{i,t}(\theta^0)$ is

symmetrically distributed under the null, where $E[m_{T,t}^*(\theta^0)] = 0$ and any $k_{1,i,T} = k_{2,i,T}$, the question of how fast $k_{i,T} \rightarrow \infty$ still remains.

In the following we discuss rules of thumb for selecting rates $k_{j,i,T} \rightarrow \infty$ and relationships between $k_{1,i,T}$ and $k_{2,i,T}$. We begin with a natural analogue based on GMTTM.

GMTTM Rate: If $m_{T,t}^*(\theta)$ are the tail-trimmed estimating equations for GMTTM, HR (2010a) characterize policies $\{k_{1,i,T}, k_{2,i,T}\}$ that optimize the efficiently weighted GMTTM compound rate of convergence $V_T^{1/2}$ and expedite the rate of identification $E[m_{T,t}^*(\theta^0)] \rightarrow 0$. They focus on equations with Paretian tails if they have an infinite variance, and on AR, ARCH and AR-ARCH models. Although they only consider a small set of models, the rules appear to apply in general: heavy trimming $k_{j,i,T} \sim T/L(T)$ for any slowly varying $L(T) \rightarrow \infty$ when stochastic regressors are not available (e.g. a model of location), or error-regressor feedback exists (e.g. GARCH, ARMA-GARCH); and light trimming $k_{j,i,T} \sim L(T)$ when stochastic regressors exist without feedback with errors (e.g. ARMA with iid errors).

But the test equations $m_t(\theta)$ may not be the same as the estimating equations $\tilde{m}_{T,t}(\theta)$, and the fractiles used to compute the GMTTM $\hat{\theta}_T$ do not have to be the same as those used to trim the test equations. This holds even if $m_t(\theta^0)$ are identical to those used to estimate $\hat{\theta}_T$, as in a test of over-identifying conditions: we may use one set $\{k_{1,i,T}, k_{2,i,T}\}$ to estimating $\hat{\theta}_T$ and another fractile set for the TTMC statistic.

Although the rules of thumb developed in HR (2010a) appear to apply in general, testing moment conditions and using an arbitrary plug-in $\hat{\theta}_T$ provide several unique challenges and advantages inherently neglected in GMTTM. We therefore discuss three criteria for selecting $\{k_{1,i,T}, k_{2,i,T}\}$ that are fundamentally distinct from optimizing the GMTTM rate.

Test Statistic Rate under H_0 : Expansion (9) reveals the accuracy and efficiency of \hat{W}_T depend on the rate of convergence of $\hat{\theta}_T$, the rate of identification $E[m_{T,t}^*(\theta^0)] \rightarrow 0$ for asymmetric equations, and the rate of convergence of $m_T^*(\theta^0) := 1/T \sum_{t=1}^T m_{T,t}^*(\theta^0)$ under the null. We first consider $m_T^*(\theta^0)$ and $E[m_{T,t}^*(\theta^0)]$ separately from the choice of plug-in $\hat{\theta}_T$, then simultaneously with the plug-in choice.

Test Equation Rate: Since the rate of convergence of $m_T^*(\theta^0)$ is $TS_T^{-1/2} = O(T^{1/2})$, in order to optimize the rate the policy should imply $S_T \rightarrow \infty$ slowly hence $k_{j,i,T} \rightarrow \infty$ quickly. The fastest allowed rate is $k_{j,i,T} \sim T/L(T)$ for slowly varying $L(T) \rightarrow \infty$, for example $k_{j,i,T} = [\delta_j T / \ln(T)]$ for $\delta_j > 0$. Similarly, $k_{j,i,T} = [T^{\lambda_j}]$ for large $\lambda_j \in (0, 1)$ augments the rate of convergence of $m_T^*(\theta^0)$. The sluggish rate of convergence of a tail-trimmed mean $TS_T^{-1/2} = O(T^{1/2})$ is well known in the literature for iid (e.g. Hahn et al 1991) and weakly dependent data (Hill 2010b).

Identification Rate: Although policies $[\delta_j T / \ln(T)]$ or $[T^{\lambda_j}]$ optimize or augment the rate of convergence of $m_T^*(\theta^0)$, the nuisance parameters δ_j and λ_j must be chosen. Further, augmenting the rate for $m_T^*(\theta^0)$ does not implicitly expedite the rate of identification $E[m_{T,t}^*(\theta^0)] \rightarrow 0$ under H_0 for asymmetric equations.

Suppose $m_{i,t}(\theta^0)$ is asymmetrically distributed. HR (2010a: Section 4) characterize a relationship between $k_{1,i,T}$ and $k_{2,i,T}$ based on the equation power law tail parameters that ensures identification $E[m_{T,t}^*(\theta^0)] \rightarrow 0$ arbitrarily fast under the null. In Appendix A we impose power law tail decay under condition D1, so assume $m_t = m_t(\theta^0)$ is scalar with an exact Paretian tail on m_t to simplify exposition: for all $m \geq M$ and some $M \geq 1$

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

where $d_i > 0$ and $\min\{\kappa_i\} > 1$. We impose $\min\{\kappa_i\} > 1$ to ensure the equation is integrable under the null. Then any policy $\{k_{1,i,T}, k_{2,i,T}\}$ that satisfies

$$\frac{k_{2,T}^{1-1/\kappa_2}}{k_{1,T}^{1-1/\kappa_1}} = T^{1/\kappa_1-1/\kappa_2} \frac{d_1^{1/\kappa_1}}{d_2^{1/\kappa_2}} \frac{(1-1/\kappa_2)}{(1-1/\kappa_1)} \quad (10)$$

ensures $E[m_{T,t}^*(\theta^0)] \approx 0$ arbitrarily close for each T .

Relation (10) implies the *heavier* tail (e.g. $d_2 > d_1$ and/or $\kappa_2 < \kappa_1$) is *trimmed less* (e.g. $k_{2,T} < k_{1,T}$). The intuition is easily grasped by assuming a heavier right tail $d_2 > d_1$ and/or $\kappa_1 > \kappa_2$ but with symmetric trimming $k_{1,T} = k_{2,T}$. There are a disproportionate number of large positive values produced and therefore trimmed, resulting in $E[m_{T,t}^*] < 0$. The solution is to decrease the number of trimmed right-tail equations $k_{2,T} < k_{1,T}$.

In practice plug-ins for d_j and κ_j can be used to enforce (10). A large variety of estimators $\hat{d}_{j,T}$ and $\hat{\kappa}_{j,T}$ are available in the literature, with only some theory supporting consistency for weakly dependent data in general. Estimators by Hill (1975) and Hall (1982) respectively for κ_j and d_j are consistent for a massive array of dependent, heterogeneous data (Hill 2009a,2010c), but the data sample $\{m_t(\theta^0)\}$ in general requires a plug-in for θ^0 . It is well beyond the scope of the present paper to develop these details, but it is relatively easy to show a plug-in will not affect consistency. See also HR (2010a: Section 4).

Plug-In and Fractile Relation: If the DGP is thin tailed and all available plug-ins are $T^{1/2}$ -convergent, then there is no leverage with which to reduce expansion (9) to

$$\hat{S}_T^{-1} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) = S_T^{-1} \sum_{t=1}^T \{m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)]\} + o_p(1). \quad (11)$$

In heavy tailed cases, however, we can fine tune the choice of test equation fractiles $\{k_{1,i,T}, k_{2,i,T}\}$ to slow down $TS_T^{-1/2} J_T \rightarrow \infty$ relative to the rate $\hat{\theta}_T \rightarrow \theta^0$ to allow (11).

We do not necessarily defend (11) as an objective. Rather, it is certainly a criterion of interest since it removes plug-in influence on the TTMC limit law, and therefore makes asymptotic arguments very simple². Further, it implicitly permits a consistent test of over-identifying restrictions in GMTTM when some $m_{i,t}(\theta^0)$ have an infinite variance. See Hall (2000) for a discussion related to the HAC matrix, and see Newey and McFadden (1994).

How to choose $\{k_{1,i,T}, k_{2,i,T}\}$ and $\hat{\theta}_T$ simultaneously to ensure (11), however, requires a specification for $m_t(\theta^0)$ and knowledge of the data generating process since we must know the diagonal components of \tilde{V}_T and V_T . A variety of examples based on specific tests are presented in the next section. As a brief example, consider an infinite variance stationary AR(1) with Paretian innovations:

$$y_t = \theta^0 y_{t-1} + \epsilon_t, \quad \epsilon_t \text{ is iid and symmetric, } P(|\epsilon_t| > \epsilon) = d\epsilon^{-\kappa} (1 + o(1)), \quad \kappa \in (1, 2),$$

with one test equation, the exactly identified GMM estimating equation:

$$m_t(\theta) = (y_t - \theta^0 y_{t-1}) y_{t-1}.$$

Assume symmetric trimming with one fractile k_T . In this case the scalar $V_T^{1/2}$ is exactly the GMTTM scale. The OLS plug-in $\hat{\theta}_T$ has a rate $\tilde{V}_T^{1/2} \sim KT^{1/\kappa}$ (Davis et al 1992). The

²In the examples of Section 2, Hong's (2001) test of volatility spillover with QML plug-in has this property, but not Bierens' (1990) test of functional form with NLLS plug-in. If some test equation has an infinite variance then our TTMC statistic can always be assured to have this property simply by trimming more equations.

GMTTM rate is $V_T^{1/2} \sim KT^{1/\kappa}$ if the trimming fractile $k_T \sim L(T)$ for slowly varying $L(T) \rightarrow \infty$, and otherwise $V_T/T^{2/\kappa} \rightarrow 0$ for non-slowly varying k_T (e.g. $k_T \sim T^\lambda$, $\lambda \in (0, 1)$). See Lemma 3.1 and Example 5 in HR 2010a). Therefore the OLS plug-in will not influence the test statistic limit if we choose $k_T \sim T^\lambda$ for any $\lambda \in (0, 1)$ since $V_T/\tilde{V}_T \rightarrow 0$. By increasing the equation trimming rate above $L(T)$ to T^λ we literally slow down $TS_T^{-1/2}J_T$ relative to $\hat{\theta}_T \rightarrow \theta^0$, hence (11) applies.

If some other class of test equations is used for testing then the final analysis may change. The point here is a case-by-case study of \tilde{V}_T and how $\{k_{1,i,T}, k_{2,i,T}\}$ relates to V_T can be used to enforce (11) for heavy tailed data.

4. EXAMPLES CONTINUED We now apply Theorem 3.2 under the null to Examples 1, 2, 3 and 6 by fully developing the background theory, and verifying the major assumptions presented in Appendix A. The remaining examples follow similarly.

Since our examples involve both *test equation* trimming, and GMTTM with *estimating equation* trimming and LTTS with least squares *criterion trimming*, we distinguish the possible sets of fractiles for clarity. We use as always $\{k_{1,i,T}, k_{2,i,T}\}$ for the test equations $m_t(\theta)$. We use $\{\tilde{k}_T\}$ to denote fractiles for trimming GMTTM estimating equations $\tilde{m}_{T,t}(\theta)$ and the LTTS criterion. In all cases GMTTM is efficiently weighted.

We make frequent use of the following properties. Under D1-D6 the gradient of the tail-trimmed equation moment J_T is proportion to the mean tail-trimmed gradient (HR 2010a: Lemma C.4):

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

In the special case $m_t(\theta^0) = u_t x_{t-1}$ for iid zero mean $u_t \in \mathbb{R}$ with symmetric distribution, and $x_t \in \mathbb{R}^q$, then symmetrically trimmed $\{m_{T,t}^*(\theta^0), \mathfrak{S}_t\}$ forms an adapted martingale difference array with $\mathfrak{S}_t = \sigma(\{u_\tau, x_\tau\} : \tau \leq t)$ since

$$E [m_{i,T,t}^*(\theta^0) | \mathfrak{S}_{t-1}] = x_{i,t-1} E [u_t I(|u_t x_{i,t-1}| \leq c_{T,i}(\theta^0)) | \mathfrak{S}_{t-1}] = 0. \quad (\text{MDA})$$

EXAMPLES 1 and 2 (White Noise and Omitted Variables): Consider testing for omitted variables in a stationary AR(p) model by testing white noise. The model is

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

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

$$P(|\epsilon_t| > \epsilon) = d\epsilon^{-\kappa} (1 + o(1)) \quad \text{where } d > 0 \text{ and } \kappa \in (1, 2]. \quad (12)$$

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

Define the remaining parameters $\theta := \beta/\beta^{(r)} \in \mathbb{R}^{p-r}$, regressors $w_t := x_t/x_t^{(r)}$, and associated error

$$u_t(\theta) := y_t - \theta' w_t \quad \text{and} \quad u_t = u_t(\theta^0).$$

By convention $u_t = y_t$ if $r = p$. The null hypothesis is

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

The test equations and trimmed version are

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

Notice independence of the errors ϵ_t implies $E[m_{i,t}(\theta^0)] = 0$ under the null requires a finite mean $\kappa > 1$, and $E[m_{i,t}^2(\theta^0)] < \infty$ if and only if $E[\epsilon_t^2] < \infty$.

Impose common symmetric trimming across equations $k_{1,i,T} = k_{2,i,T} = k_T$ since $m_t(\theta^0)$ are identically and symmetrically distributed under the null. The indicators are simply

$$\hat{I}_{i,T,t}(\theta) = I\left(|m_{i,t}(\theta)| \leq m_{i,(k_T)}^{(a)}(\theta)\right) \quad \text{where} \quad m_{i,t}^{(a)}(\theta) := |m_{i,t}(\theta)|.$$

Let $\hat{\beta}_T$ be the OLS, LTTS, LAWD or GMTTM estimator, and define $\hat{\theta}_T = \hat{\beta}_T / \hat{\beta}_T^{(r)}$.

All assumptions detailed in Appendix A hold by nearly trivial arguments. Identification by $m_t(\theta^0)$ I1 holds given error independence and $\kappa > 1$. Under symmetry identification by $m_{T,t}^*(\theta^0)$ I2 is trivial given I1. Non-degeneracy and positive definiteness I3 and moment smoothness I4 both hold given the stationary autoregressive DGP with smoothly distributed iid error.

Error distribution smoothness ensures D1.i. Power-law tail (12) and the functional form of $m_{i,t}(\theta^0)$ imply tail property D1.ii holds by a convolution tail result due to Cline (1986). Equation differentiability D2 holds by construction of $m_t(\theta)$. Since y_t is geometrically β -mixing by a result in Pham and Tran (1985) and ϵ_t is iid, mixing property D3 holds by construction of $m_t(\theta)$. Envelope bound D4 holds since Θ is compact and y_t is integrable. Jacobian rank D5.i holds by construction, and Jacobian smoothness D.5.ii can be shown to hold by the same argument in HR (2010a: Section 5.1).

Metric entropy with L_2 -bracketing D6 follows if we show an L_2 -norm Lipschitz property

$$\left\| I_{i,T,t}(\theta) - I_{i,T,t}(\tilde{\theta}) \right\| \leq K \left\| \theta - \tilde{\theta} \right\|.$$

See Giné and Zinn (1984), Pollard (1984, 2002) and van der Vaart and Wellner (1996). The finite dimensional distributions of $m_t(\theta)$ are absolutely continuous and bounded uniformly on Θ , so we can always assume $l_T(\theta) = u_T(\theta)$ is continuous and differentiable on Θ . Combine this with distribution continuity and density uniform boundedness to deduce the above Lipschitz property.

Finally consider plug-in properties P1 or P2. Let $L(T)$ be a slowly varying function, $L(T) \rightarrow \infty$, that may change from place to place. OLS in general, and LTTS and GMTTM with slowly varying trimming fractiles $\tilde{k}_T \sim L(T)$, all have scale elements $\hat{V}_{i,i,T}^{1/2} \sim KT^{1/\kappa}/L(T)$, $KT^{1/2}L(T)$, or $KT^{1/2}$ respectively if $\kappa < 2$, $\kappa = 2$ and $\kappa > 2$ (Davis et al 1992, Hill 2010a, HR 2010a). Further, LAWD is $T^{1/2}$ -convergent for any $\kappa > 1$ (Ling 2005).

We will show $V_T \sim KT$ if $\kappa > 2$ and $V_T \sim KT(T/k_T)^{2/\kappa-1}$ for any $\kappa \in (1, 2]$. Thus, LAWD does not satisfy either P1 or P2 if $E[\epsilon_t^2] = \infty$, and OLS, LTTS and GMTTM satisfy P1 or P2 if $E[\epsilon_t^2] = \infty$ or $E[\epsilon_t^2] < \infty$.

Assume $p = q = 1$ to simplify notation, the general cases being identical. Apply properties (MDA) and (JAC) to deduce under the null $S_T = T \times E[\epsilon_t^2 \epsilon_{t-1}^2 I_{T,t}(\theta^0)]$ and

$$J_T = E\left[\{\epsilon_t y_{t-2} + \epsilon_{t-1} y_{t-1}\} I_{T,t}(\theta^0)\right] \times (1 + o(1)) = E\left[\epsilon_{t-1}^2 I(|\epsilon_t \epsilon_{t-1}| \leq c_T)\right] \times (1 + o(1)).$$

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

Otherwise consider $\kappa < 2$, where $\kappa = 2$ is similar. Tail (12) implies $c_T = K(T/k_T)^{1/\kappa}$, and by independence and (12) the product convolution $\epsilon_t \epsilon_{t-1}$ has tail (12) with the same index κ (Cline 1986). An application of Karamata's Theorem (Resnick 1987: Theorem 0.6; cf. Problem 4.2.8) therefore implies

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

Similarly, use independence, distribution continuity and Karamata's Theorem to deduce

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

Therefore

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

Therefore LAWD satisfies neither P1 or P2 because it is only $T^{1/2}$ convergent and $V_T^{1/2}/T^{1/2} \rightarrow \infty$. Conversely, if we maximally trim the test equations $k_T \sim T/L(T)$ then $V_T \sim T \times L(T)$, hence OLS, LTTS and GMTTM all satisfy P1: $V_T^{1/2}/\tilde{V}_{i,i,T}^{1/2} \sim T^{-(1/\kappa-1/2)}L(T) \rightarrow 0$. If we trim slightly less $k_T \sim T^\lambda$ for $\lambda \in (0, 1)$ then $V_T^{1/2}/\tilde{V}_{i,i,T}^{1/2} \sim KT^{1/2-1/\kappa+(1-\lambda)(2/\kappa-1)} \rightarrow 0$ if $\lambda > 1/2$ hence again P1 holds.

LEMMA 4.1 (White Noise/Omitted Variables in AR) *The above AR data generating process satisfies I1-I4 and D1-D6. Further, LAWD does not satisfy either P1 or P2 if variance is infinite, and OLS, LTTS, and GMTTM satisfy P1 if $E[\epsilon_t^2] = \infty$ and $k_T \sim T/L(T)$ or $k_T \sim T^\lambda$ for any $\lambda \in (1/2, 1)$.*

EXAMPLE 3 (Neural Test of Neglected Nonlinearity): Recall $y_t = f(x_t, \theta) + \epsilon_t(\theta)$, $f: \mathbb{R}^p \times \Theta \rightarrow \mathbb{R}$, assume y_t is integrable, $\epsilon_t = \epsilon_t(\theta^0)$ has power-law tail (12) if $E[\epsilon_t^2] = \infty$, and $\epsilon_t(\theta)$ has an absolutely continuous distribution with uniformly bounded density on Θ . Assume $f(\cdot, \theta)$ is continuous and twice differentiable, $f(x, \cdot)$ is Borel measurable, and $\{\epsilon_t, x_t\}$ are geometrically β -mixing.

The null states $f(x_t, \theta^0)$ is a version of $E[y_t | x_t]$ for unique interior point θ^0 of compact Θ . Examples include testing for omitted nonlinearity in the conditional mean of an AR(1)-ARCH(1)

$$y_t = \theta^0 y_{t-1} + \epsilon_t, \quad |\theta^0| < 1, \quad (13)$$

$$\epsilon_t = (\omega^0 + \alpha^0 \epsilon_{t-1}^2)^{1/2} u_t, \quad \omega^0 > 0, \quad \alpha^0 \in [0, 1) \text{ and } u_t \stackrel{iid}{\sim} (0, 1)$$

Under mild regulatory conditions $\{y_t, \epsilon_t\}$ are geometrically β -mixing with regularly varying tails (Borkovec and Klüppelberg 2001, Cline 2007).

Assume the test weight $F: \mathbb{R} \rightarrow \mathbb{R}$ is non-polynomial, real analytic, and uniformly bounded on any compact subset of \mathbb{R} , covering exponential, logistic and trigonometric functions (see Stinchcombe and White 1998, cf. Bierens 1990 and Bierens and Ploberger 1997). The argument of $F(\gamma' \psi(x_t))$ is based on any bounded one-to-one $\psi: \mathbb{R}^p \rightarrow \mathbb{R}^p$.

Now define a scalar test equation

$$m_t(\theta, \gamma) := \epsilon_t(\theta)F(\gamma'\psi(x_t))$$

with trimmed version

$$\hat{m}_{T,t}^*(\hat{\theta}_T, \gamma) = \epsilon_t(\hat{\theta}_T)F(\gamma'\psi(x_t))\hat{I}_{T,t}(\hat{\theta}_T).$$

Since $F(\gamma'\psi(x_t))$ is uniformly bounded given boundedness of ψ and Γ , and ϵ_t is symmetrically distributed, the indicator symmetrically trims only according to large values of $\epsilon_t(\theta)$:

$$\hat{I}_{T,t}(\theta) = I\left(|\epsilon_t(\theta)| \leq \epsilon_{(k_T)}^{(a)}(\theta)\right)$$

A tail-trimmed version of Lee et al's (1993) version of Bierens' (1990) test is

$$\hat{W}_T(\gamma) = \left(\frac{1}{\hat{S}_T^{1/2}(\hat{\theta}_T, \gamma)} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T, \gamma)\right)^2$$

where $\hat{S}_T(\hat{\theta}_T, \gamma)$ is the HAC estimator of $E[(\sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T, \gamma))^2]$. Lee et al (1993) argue for randomly selecting γ on compact Γ in a way independent of the sample $\{y_t, x_t\}$.

A power-optimal test, as in Bierens (1990), Andrews and Ploberger (1994) and Hill (2008), uses $\sup_{\gamma \in \Gamma} \hat{W}_T(\gamma)$ or $\int_{\gamma \in \Gamma} \hat{W}_T(\gamma) d\mu(\gamma)$ for some absolutely continuous measure $\mu(\gamma)$ on Γ . This requires weak limit theory for tail-trimmed arrays that is beyond the scope of the present paper. The theory developed here can only cover randomization or arbitrary selection of γ .

Identification I1 holds by error integrability and weight boundedness. I2 automatically follows if the error distribution is symmetric, or asymmetric with Paretian tails and the fractiles are chosen to satisfy (11). I4 follows from smoothness of the response function and error distribution.

Distribution properties D1.i, D2-D4 and D6 all follow since $\epsilon_t(\theta)$ is geometrically β -mixing with a continuous and bounded distribution, and $F(\gamma'\psi(x_t))$ is continuous and bounded with measurable and uniformly bounded derivative.

The remaining properties covering power-laws for heavy tailed equations D1.ii, Jacobian non-degeneracy and smoothness D5, covariance positive definiteness and non-degeneracy I3, and plug-in rate P1 or P2 are all regulatory and depend upon response function and error distribution specifics. We verify the conditions for the AR-ARCH model (13) below.

EXAMPLE 3 (Neural Test for AR-ARCH): Assume the plug-in $\hat{\theta}_T$ is Ling's (2007) QMWL, or HR's (2010a: Sections 3 and 6) GMTTM with QML estimating equations. Assume the iid ARCH innovation u_t is symmetrically distributed with absolutely continuous distribution. If $E[u_t^4] = \infty$ assume u_t has Paretian tail (12) with index $\kappa_u \in (2, 4]$. Both $\{y_t, \epsilon_t\}$ have power-law tail (12) with index κ_ϵ that satisfies $E|\theta^0 + (\alpha^0)^{1/2}u_t|^{\kappa_\epsilon} = 1$ (Cline 2007: Example 3.2). Assume y_t is integrable $\kappa_\epsilon > 1$, effectively assuming θ^0 and α^0 are sufficiently small³.

Since the error ϵ_t is symmetrically distributed under the null the equation $m_t(\theta, \gamma)$ is symmetrically trimmed with fractile k_T that satisfies

$$\frac{T}{k_T} P(|\epsilon_t(\theta)| > c_T) = 1.$$

³Simulation experiments can be used to compute the solution to $\mu(\alpha, \theta, \kappa) := E|\theta + \alpha^{1/2}u_t|^\kappa = 1$. We used 100,000 iid draws $u_t \sim N(0,1)$ to estimate $\mu(\alpha, \theta, \kappa)$ for $\alpha = .5$ and $\theta = .9$, and $\kappa \in \mathcal{K} = \{.01, .02, \dots, 5.0\}$. The solution $\arg \min_{\kappa \in \mathcal{K}} |\hat{\mu}(.5, .9, \kappa) - 1|$ is 1.16. Similarly, $\arg \min_{\kappa \in \mathcal{K}} |\hat{\mu}(.6, .9, \kappa) - 1| = 1.03$ and $\arg \min_{\kappa \in \mathcal{K}} |\hat{\mu}(.35, .97, \kappa) - 1| = 1.05$.

Properties D1.ii, D5 and I3 follow instantly from linearity, error distribution smoothness, error power-law tail (12) and convolution tail theory due to Cline (1986). See also HR (2010a: Section 5).

Plug-in properties P1 or P2 require the scale V_T . Use (MDS) to deduce $S_T = T \times E[m_{T,t}^{*2}(\theta^0, \gamma)]$. Further, $(\partial/\partial\theta)m_t(\theta, \gamma)|_{\theta^0} = -y_t F(\gamma'\psi(y_{t-1}))$ is integrable since y_t is and $F(\gamma'\psi(y_{t-1}))$ is bounded, hence (JAC) implies $J_T \sim -E[y_t F(\gamma'\psi(y_{t-1}))]$. The scale is therefore

$$V_T \sim KT \times (E[\epsilon_t^2 F(\gamma'\psi(x_t))^2 I(|\epsilon_t| \leq c_T)])^{-1}.$$

If $\kappa_\epsilon > 2$ then $E[m_{T,t}^{*2}(\theta^0, \gamma)] \rightarrow K$ hence $V_T \sim KT$. Otherwise, regular variation (12) and boundedness of $F(\gamma'\psi(y_{t-1}))$ ensure by an application of Karamata's Theorem

$$\kappa_\epsilon < 2 : V_T \sim KT (c_T^2 (k_T/T))^{-1} = KT (k_T/T)^{2/\kappa_\epsilon - 1} = o(T)$$

$$\kappa_\epsilon = 2 : V_T \sim T/L(T) = o(T).$$

QMWL is $T^{1/2}$ -convergent under mild additional regulatory conditions (Ling 2007). Therefore P2 holds if ϵ_t has a finite variance, and otherwise P1 applies.

QML is $T^{1/2}$ -convergent if $E[u_t^4] < \infty$, and $T^{1-2/\kappa_u}/L(T)$ -convergent if $E[u_t^4] = \infty$ by Theorem 2.1 of Hall and Yao (2003). Verifying P1 or P2 for QML therefore requires knowledge of κ_ϵ , κ_u and a policy $\{k_T\}$. If $k_T \sim T/L(T)$ then $V_T \sim T/L(T)$ so QML satisfies neither P1 nor P2. If we trim fewer equations with $k_T \sim T^\lambda$ then QML satisfies P1 if $\lambda \in (0, 1 - 2/[\kappa_u(2/\kappa_\epsilon - 1)])$ and P2 if $\lambda = 2/[\kappa_u(2/\kappa_\epsilon - 1)]$. Heavier tailed iid or ARCH errors (respectively $\kappa_u \searrow 2$ or $\kappa_\epsilon \searrow 1$) imply less trimming of the test equations to ensure QML satisfies P1. QML is so slow to converge we must trim few equations since that augments the trimmed equation sample volatility and therefore slows its rate of convergence.

GMTTM based on QML-estimating equations is $T^{1/2}$ -convergent if $E[u_t^4] < \infty$, and maximally $T^{1/2}/L(T)$ -convergent if $E[u_t^4] = \infty$ and the GMTTM fractiles $k_T \sim T/L(T)$. See HR (2010a: Sections 3.3-3.4). Therefore GMTTM satisfies P1 if u_t has an infinite fourth moment, and P2 otherwise.

LEMMA 4.2 (Functional Form for AR-ARCH) *The above AR-ARCH data generating process satisfies I1-I4 and D1-D6. QML satisfies P1 or P2 if sufficiently few observations are trimmed, e.g. $k_T \sim T^\lambda$ with $\lambda \in (0, 1 - 2/[\kappa_u(2/\kappa_\epsilon - 1)])$. QMWL satisfies P1 if ϵ_t has an infinite variance, and P2 otherwise. GMTTM satisfies P1 if u_t has an infinite fourth moment, and P2 otherwise.*

EXAMPLE 6: Volatility Spillover: Hong (2001) uses QML to estimate univariate GARCH(1,1) models

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

Under the null of no volatility spillover $\epsilon_{i,t} \stackrel{iid}{\sim} (0, 1)$, and Hong's alternative is spillover of a CCC-GARCH form:

$$h_{i,t}^2(\theta_i) = \omega_i + \alpha_{i,i} y_{i,t-1}^2 + \beta_{i,i} h_{i,t-1}^2(\theta) + \alpha_{i,j} y_{j,t-1}^2 + \beta_{i,j} h_{j,t-1}^2(\theta).$$

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 (12) with

indices $\kappa_{y_i} > 0$ (Mikosch and Střarica 2000, Basrak et al 2002, Ferdandez and Muriel 2009). Further, similar to Hall and Yao (2003) and HR (2010a), if $E[\epsilon_{i,t}^4] = \infty$ assume $\epsilon_{i,t}$ has tail (12) with index $\kappa_i \in (2, 4]$. Under the stated conditions $\{y_{1,t}, y_{2,t}\}$ are geometrically β -mixing (Boussama 1998, Comte and Leiberman 2003).

Define test equations

$$m_t(\theta) = \left[\left(\frac{y_{1,t}^2}{h_{1,t}^2(\theta)} - 1 \right) \times \left(\frac{y_{2,t-j}^2}{h_{2,t-j}^2(\theta)} - 1 \right) \right]_{i=1}^q. \quad (14)$$

Although Hong (2001) requires $E[\epsilon_{i,t}^8] < \infty$ due to his standardize portmanteau statistic form (see section 5, below), clearly if $E[\epsilon_{i,t}^4] = \infty$ then tail-trimming is appropriate with trimmed equations

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

In general $m_{j,t}(\theta^0)$ has the same tail thickness for each j , so the same fractiles $\{k_{1,T}, k_{2,T}\}$ are used for each equation. Further, $m_{j,t}(\theta^0)$ are skewed right by construction hence asymmetric trimming $k_{1,T} > k_{2,T}$ is imposed. If $\epsilon_{i,t}$ have Paretian tails then the recommendations of Section 3.7 can be followed for selecting $\{k_{1,T}, k_{2,T}\}$ to optimize the rate of identification $E[m_{T,t}^*(\theta^0)] \rightarrow \infty$ under the null.

All conditions are verified as in Examples 1 and 2. The only steps that slightly differ concern V_T . If either $E[\epsilon_{i,t}^4] = \infty$ then $m_{j,t}(\theta^0) = (\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-j}^2 - 1)$ has tail (12) with index $\kappa/2 := \min\{\kappa_1, \kappa_2\}/2 \leq 2$, cf. Breiman (1965) and Cline (1986). Assume $\kappa < 4$, the case $\kappa = 4$ being similar. Define $c_{i,T} = \max\{l_{i,T}(\theta^0), u_{i,T}(\theta^0)\}$ and $k_T = \min\{k_{1,T}, k_{2,T}\}$, and apply Karamata's Theorem to deduce

$$E[m_{T,i,t}^*(\theta^0)] \sim K c_{i,T}^2 P(|(\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-i}^2 - 1)| > c_{i,T}) = K(T/k_T)^{4/\kappa-1},$$

hence by (MDS) $S_{i,i,T} \sim KT(T/k_T)^{4/\kappa-1}$.

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

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

If there are GARCH effects then under the null $J_{j,t}$ is integrable since $\epsilon_{i,t}^2$ is iid and integrable and $\|x_{i,t}\|/h_{i,t}^2 \leq K$ *a.s.* 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 the scale components $V_{i,i,T} \sim KT(k_T/T)^{4/\kappa-1} = o(T)$. QMWL is $T^{1/2}$ -convergent so it satisfies P1 if either $E[\epsilon_{i,t}^4] = \infty$, and otherwise P2. GMTTM with estimating equation 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.

QML is $T^{1/2}$ -convergent if both $E[\epsilon_{i,t}^4] < \infty$, and otherwise $KT^{1-2/\kappa_i}/L(T)$ -convergent (Hall and Yao 2003: Theorem 2.1). Thus, whether and if QML satisfies P1 or P2 depends on error tail thickness κ_i and equation policy $\{k_{1,T}, k_{2,T}\}$. Specifically, $V_{i,i,T} \sim T/(T/k_T)^{4/\kappa-1} < T^{1-2/\kappa_i}/L(T) \sim \tilde{V}_{i,i,T}$ holds when we trim less, exactly as in Lemma 4.2. Consider $k_T \sim T^\lambda$ for any $\lambda \in (0, 1 - 2/\kappa/(4/\kappa - 1)]$.

Notice the substantial improvement of higher moments over Hong (2001): we only need $\epsilon_{i,t} \stackrel{iid}{\sim} (0, 1)$ for test equation integrability under the null (1), while Hong (2001) needs $E[\epsilon_{i,t}^8] < \infty$.

LEMMA 4.3 (Volatility Spillover in GARCH) *The above GARCH data generating process satisfies I1-I4 and D1-D6. Further, QMWL and GMTTM satisfy P1 if either $E[\epsilon_{i,t}^4] = \infty$, and P2 otherwise. QML satisfies P1 or P2 if sufficiently few observations are trimmed, for example $k_T \sim T^\lambda$ for any $\lambda \in (0, 1 - 2/\kappa/(4/\kappa - 1)]$.*

5. SIMULATION STUDY We now use the TTMC test statistic, its untrimmed version, and conventional statistics to perform tests of white-noise, omitted variables, and volatility spillover. The data generating processes are IID, AR(2), and bivariate GARCH(1,1). We simulate 1000 samples of each process for a sample size $T = 1000$.

Let P_κ denote a symmetric Pareto distribution: if $\epsilon_t \sim P_\kappa$ then $P(|\epsilon_t| > \epsilon) = .5(1 + \epsilon)^{-\kappa}$. Random draws from P_κ with $\kappa > 2$ are standardized to ensure $\epsilon_t \stackrel{iid}{\sim} (0, 1)$. The IID and AR models have iid innovations $\epsilon_t \sim P_\kappa$ with index $\kappa \in \{1.5, 2.5\}$.

The bivariate Constant Conditional Correlation GARCH process $\{y_{1,t}, y_{2,t}\}$ has coordinate specification

$$y_{i,t} = \epsilon_{i,t} h_{i,t} \text{ where } \epsilon_{i,t} \stackrel{iid}{\sim} P_{2.5} \text{ or } \epsilon_{i,t} \stackrel{iid}{\sim} N(0, 1)$$

$$h_{i,t}^2 = \omega_i + \alpha_{i,i} y_{i,t-1}^2 + \beta_{i,i} h_{i,t-1}^2 + \alpha_{i,j} y_{j,t-1}^2 + \beta_{i,j} h_{j,t-1}^2, \omega_i > 0, \alpha_{i,j}, \beta_{i,j} > 0.$$

In each model we use one of two possible errors. One error is heavy tailed enough that the conventional statistics used here have non-standard limit distributions under the null. The other error has thin enough tails to promote standard limit distributions. See Table 1 for all model specifications and tail indices.

Each process is stationary geometrically ergodic (Pham and Tran 1985, Boussama 1998, Comte and Lieberman 2003) and therefore geometrically β -mixing (Doukhan 1994). Further, each process has symmetric power-law tails (Hannan and Kanter 1977, Cline 1989, Borkovec and Klüppelberg 2001, Fernández and Muriel 2009). In the IID and AR cases y_t has the same index as ϵ_t . The CCC-GARCH process exhibits power-law tail decay due to Markov-type feedback, and the underlying error when it is Paretian.

TABLE 1 - Data Generating Processes

Model Type	Process Specification for y_t	κ_ϵ^a	κ_y^b
IID	$y_t = \epsilon_t$	$\kappa_\epsilon \in \{1.5, 4.5\}$	$\kappa_y \in \{1.5, 4.5\}$
AR(2)	$y_t = .8 \times y_{t-1} + .4 \times y_{t-2} + \epsilon_t$	$\kappa_\epsilon \in \{1.5, 4.5\}$	$\kappa_y \in \{1.5, 4.5\}$
C-G NULL ^c	$h_{i,t}^2 = .3 + .3\epsilon_{1,t-1}^2 + .6h_{1,t-1}^2$ for both $i = 1, 2$	$\kappa_\epsilon \in \{2.5, \infty\}$	$\kappa_y \in \{1.5, 2.9\}$
C-G ALT1	$h_{1,t}^2 = .3 + .3y_{1,t-1}^2 + .6h_{1,t-1}^2 + .1y_{2,t-1}^2 + .3h_{2,t-1}^2$ $h_{2,t}^2 = .3 + .3y_{2,t-1}^2 + .6h_{2,t-1}^2$	$\kappa_\epsilon \in \{2.5, \infty\}$ $\kappa_\epsilon \in \{2.5, \infty\}$	$1 < \kappa_{y1} \leq \kappa_{y2}$ $\kappa_{y2} \in \{1.5, 2.9\}$
C-G ALT2	$h_{1,t}^2 = .3 + .3y_{1,t-1}^2 + .6h_{1,t-1}^2 + .3y_{2,t-1}^2 + .6h_{2,t-1}^2$ $h_{2,t}^2 = .3 + .3y_{2,t-1}^2 + .6h_{2,t-1}^2$	$\kappa_\epsilon \in \{2.5, \infty\}$ $\kappa_\epsilon \in \{2.5, \infty\}$	$1 < \kappa_{y1} \leq \kappa_{y2}$ $\kappa_{y2} \in \{1.5, 2.9\}$

a. Moment supremum for ϵ_t . If ϵ_t is Paretian this is the tail index.

b. Tail index for IID and AR y_t , and tail indices for CCC-GARCH $(y_{1,t}, y_{2,t})^4, ^5$.

c. C-G = CCC-GARCH. The hypotheses are NULL: no spillover; ALT1: weak spillover from y to x ; ALT2: strong spillover from y to x .

⁴Define a GARCH(1,1) $y_t = h_t \epsilon_t$ with iid ϵ_t and $h_t^2 = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2$. By exploiting a result due to Kesten (1973), Basrak et al (2002: eq. 2.10) show under regulatory conditions $E[(\alpha \epsilon_t^2 + \beta)^{\kappa/2}] = 1$. The conditions are necessarily satisfied by our C-G NULL model for either Paretian or Gaussian

5.1 TTMC Test Equations and Control Tests

We now construct test equations for the TTMC statistic and detail conventional tests for comparisons. In all relevant cases a Bartlett kernel $k(\cdot)$ is used. Summaries of trimming policies, the number of trimmed equations when $T = 1000$, HAC bandwidths and tail indices for $m_{i,t}(\theta^0)$ under the null are presented in Table 2. The test of omitted variables requires a fractile class $k_{i,T} \sim \delta T / \ln(T)$ to ensure OLS and GMTTM apply (Lemma 4.2). We therefore use $k_{i,T} \sim \delta T / \ln(T)$ for all tests.

In all cases estimating and test equations are unique (there are no common components), and all estimators are exactly identified, so TTMC degrees of freedom are q in all cases (Corollary 3.3). Plug-in choices are summarized in Table 3. Throughout $\hat{\theta}$ denotes any plug-in, where $\hat{\theta}_{LS} = \text{OLS}$, $\hat{\theta}_{QW} = \text{QMWL}$, and $\hat{\theta}_{GT} = \text{GMTTM}$. In all cases GMTTM is based on an efficient weight.

White Noise: Let $\{y_t\}$ denote the IID or AR(2) process, or the GARCH component $\{y_{1,t}\}$. We test each y_t for serial correlation with equations

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

Under the null, IID and AR equations $m_{i,t}$ are integrable since y_t is iid with a finite mean, so (1) is valid. Under the null $m_{i,t} = y_t y_{t-i}$ is Paretian with the same index $\kappa_m = \kappa$ as y_t (Cline 1986).

Only the GARCH model C-G NULL with Gaussian shocks has integrable equations $m_{i,t}$ since the errors are independent and $\kappa_y = 2.9$. Thus we only test IID, AR and C-G NULL for white noise. All GARCH models NULL, ALT1 and ALT2 have infinite variance equations.

All equations $m_{i,t}$ are symmetrically distributed under the null, so symmetric trimming is used with one fractile $k_T = \lceil .2T / \ln(T) \rceil$ for IID and AR, and $k_T = \lceil .45T / \ln(T) \rceil$ for GARCH. Simulation evidence not reported here suggests $k_T \sim \delta T / \ln(T)$ with $\delta < .10$ or $\delta > .30$ leads to less sharp empirical size for IID and AR data. Evidently only a few large equation observations from IID and AR data need to be trimmed to ensure the test statistic under the null is approximately chi-squared.

The GARCH model exhibits volatility feedback, hence substantially more trimming is required to ensure sharp empirical size. This simply replicates simulation results in HR

error (Basrak et al 2002: Theorem 3.1). 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 $K \in \{.01, .02, \dots, 10\}$ based on $N = 100,000$ iid random draws ϵ_t from $N(0, 1)$ or $P_{2.5}$. The 1% bands are less than .001 in all cases.

⁵Fernández and Muriel (2009) exploit the same result by Kesten (1973) to show bivariate CCC-GARCH have regularly varying tails. A moment condition based on a matrix values stochastic recurrence equation reveals the tail indices, but for CCC-GARCH *with spillover* we have not found a tractable method for simulating the tail indices. We therefore only give the index for the component $y_{2,t}$ without spillover, where $\kappa_{y_1} \leq \kappa_{y_2}$ necessarily follows by the additional feedback. The lower bound $\kappa_{y_1} > 1$ is deduced from simulating 10,000 series $\{y_{1,t}\}_{t=1}^{1000}$ and computing the Hill (1975) two-tailed tail index estimator $\hat{\kappa}_{y_i, \tilde{k}_T} = 1/\tilde{k}_T \sum_{j=1}^{\tilde{k}_T} \ln(y_{i,(i)}^{(a)}/y_{j,(\tilde{k}_T+1)}^{(a)})$, $y_{j,t}^{(a)} := |y_{j,t}|$, with Hill's (2010c) kernel estimator $\hat{v}_{\tilde{k}_T}^2$ of the mean-squared-error $v_{\tilde{k}_T}^2 = E[\tilde{k}_T^{1/2}(\hat{\kappa}_{y_i, \tilde{k}_T} - \kappa_{y_i})^2]$. We use a Bartlett kernel with bandwidth $T^{.225}$. The CCC-GARCH process satisfies the conditions of asymptotic normality of $\hat{\kappa}_{y_i, \tilde{k}_T}$ and consistency of $\hat{v}_{\tilde{k}_T}^2$ since $y_{i,t}$ is geometrically β -mixing with Paretian tail (Boussama 1998, Fernández and Muriel 2009, Hill 2010c): $\tilde{k}_T^{1/2}(\hat{\kappa}_{y_i, \tilde{k}_T} - \kappa) / \hat{v}_{\tilde{k}_T} \xrightarrow{d} N(0, 1)$. The asymptotic 95% confidence bands $\hat{\kappa}_{y_i, \tilde{k}_T} \pm 1.96 \hat{v}_{\tilde{k}_T} / \tilde{k}_T^{1/2}$ are above 1 for over 80% of $\tilde{k}_T \in \{5, \dots, 400\}$, and over 90% of $\tilde{k}_T \in \{5, \dots, 300\}$. Results are available upon request.

(2010a) concerning GMTTM for GARCH models. We find $k_T \sim \delta T / \ln(T)$ with $\delta \in [.4, .5]$ is optimal: too little trimming in small samples results in too many large observations allowed, and too much trimming overwhelms the test statistic's ability to detect white noise by introducing spurious bias in the equations.

Finally, we use a small HAC bandwidth since under the null $m_t(\theta^0)$ is a product of iid symmetrically distribution random variables, hence $m_{T,t}^*$ is a martingale difference with respect to $\mathfrak{F}_t = \sigma(y_\tau : \tau \leq t)$: $E[m_{i,T,t}^* | \mathfrak{F}_{t-1}] = y_{t-i} E[y_t I(|y_t y_{t-i}| \leq c_T) | \mathfrak{F}_{t-1}] = 0$. 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 $\{m_{T,t}^*, \mathfrak{F}_t\}$ is a martingale difference under the null, while conversely a small (large) k_T for IID and AR (GARCH) for any γ_T .

As control tests we compute both 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 .

Omitted Variable: We estimate an AR(2) model $y_t = \sum_{i=1}^2 \beta_i y_{t-i} + \epsilon_t = \beta^0 x_t + \epsilon_t$ for IID and AR data and test for omitted variables. The plugs-in are OLS or GMTTM. See Section 5.2 for all plug-in details. 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} . The equations are therefore

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

A finite mean and independence of the errors $u_t(\theta^0) = \epsilon_t$ under the null ensures (1) is valid. Under the null $m_{i,t}(\theta^0) = \epsilon_t \epsilon_{t-i}$ is Paretian with the same index $\kappa_m = \kappa$ as ϵ_t (Cline 1986).

The equations $m_{i,t}(\theta^0)$ are symmetrically distributed under the null, so symmetric trimming is used. In lieu of Lemma 4.1 we use a trimming policy $k_T = [.015T / \ln(T)]$. Simulation experiments again support the use of a small HAC bandwidth $\gamma_T = [T^{.25}]$ since under the null $m_{i,T,t}^*(\theta^0)$ is a martingale difference with respect to $\mathfrak{F}_t = \sigma(\epsilon_\tau : \tau \leq t)$.

The control tests are an untrimmed version of \hat{W}_T , and Wald statistics \mathcal{W}_T based on OLS or GMTTM:

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

and \hat{V}_T estimates the plug-in scale \tilde{V}_T . Recall for OLS $\hat{V}_T = (x'x)\hat{\sigma}_T^{-2}$ with $\hat{\sigma}_T^2 = 1/T \sum_{t=1}^T (y_t - \hat{\theta}'_{LS} x_t)^2$. In the case of GMTTM define estimating equations $\tilde{m}_t(\theta) = (y_t - \theta' x_t) x_t$, a trimmed version $\hat{m}_{T,t}(\theta) = [\tilde{m}_{i,t}(\theta) \hat{I}_{i,T}(\theta)]_{i=1}^p$ where $\hat{I}_{i,T}(\theta)$ are trimming indicators defined in Section 5.2, below. The sample efficiently weighted GMTTM scale is

$$\hat{V}_T = T^2 \hat{J}'_T \hat{S}_T^{-1} \hat{J}_T$$

where

$$\hat{J}_{i,j,T} = \left[-\frac{1}{T} \sum_{t=1}^T y_{t-i} y_{t-j} \hat{I}_{i,T}(\hat{\theta}_{GT}) \right]_{i,j=1}^p \quad \text{and} \quad \hat{S}_T = \sum_{s,t=1}^T k((s-t)/\gamma_T) \hat{m}_{T,s}(\hat{\theta}_{GT}) \hat{m}_{T,t}(\hat{\theta}_{GT})'.$$

See also HR (2010a: Sections 2 and 3). We use a Bartlett kernel $k(\cdot)$ and the same bandwidth HAC $\gamma_T = [T^{.25}]$ for simplicity.

Volatility Spillover: Similar to Hong (2001), for the bivariate GARCH process $\{y_{1,t}, y_{2,t}\}$ we estimate univariate GARCH(1,1) models

$$y_{i,t} = \epsilon_{i,t} h_{i,t}(\theta^0) \quad \text{and} \quad h_{i,t}^2(\theta) = \omega + \alpha y_{i,t-1}^2 + \beta h_{i,t-1}^2(\theta),$$

and build test equations

$$m_t(\theta) = \left[\left(\frac{y_{1,t}^2}{h_{1,t}^2(\theta)} - 1 \right) \times \left(\frac{y_{2,t-j}^2}{h_{2,t-j}^2(\theta)} - 1 \right) \right]_{i=1}^q, \quad q \in \{1, 5, 10\}.$$

We only use a QMWL plug-in for θ^0 for ease of comparison with Hong's (2001) test detailed below. Under the null $m_{j,t}(\theta^0) = (\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-j}^2 - 1)$, a product of independent mean zero $\epsilon_{i,t}^2 - 1$. We therefore require $\kappa_{\epsilon_i} > 2$ to ensure (1) is valid. Since $\epsilon_{1,t}$ are Gaussian or Paretian, under the null $m_{j,t}(\theta^0)$ has tail index $\kappa_m = \infty$ or $\kappa_m = 2.5/2 = 1.25$ (Cline 1986).

The equations under the null $m_{j,t}(\theta^0) = (\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-j}^2 - 1)$ are skewed right because each $\epsilon_{i,t}^2 - 1$ is skewed right with support $[-1, \infty)$, hence we trim more *left* tail observations on $m_{j,t}(\theta^0)$ then *right*. Experiments not reported here reveal $\{k_{1,T}, k_{2,T}\} = \{[.03T/\ln(T)], [.01T/\ln(T)]\}$ is optimal based on empirical size.

A comparatively large HAC bandwidth is required to absorb the incidental serial association that appears in the residuals $y_{i,t}^2/h_{i,t}^2(\hat{\theta}) - 1$. Evidently this is caused by iterative recursions associated with maximizing the QMWL criterion. We find $\gamma_T = [T^\zeta]$ with $\zeta \in [.30, .40]$ works best, so we use $\gamma_T = [T^{.35}] = 11^6$.

The control tests are an untrimmed version of \hat{W}_T , and Hong's (2001) test with a QMWL plug-in. Hong's theory is designed for QML with a $T^{1/2}$ -rate of convergence, but QML is $o(T^{1/2})$ -convergent if $E[\epsilon_{i,t}^4] = \infty$. His theory and method clearly extend to heavy tail robust and $T^{1/2}$ -convergent QMWL.

Hong (2001: eq. (22)) specifies a centered portmanteau statistic

$$\hat{Q}_T = \frac{T \sum_{i=1}^{T-1} k^2(i/M) \hat{\rho}_{1,2}^2(i) - \sum_{i=1}^{T-1} (1 - i/T) k^2(i/M)}{\left\{ 2 \sum_{i=1}^{T-1} (1 - i/T) (1 - (i+1)/T) k^4(i/M) \right\}^{1/2}}$$

where $k(\cdot)$ is a Bartlett kernel, $M = 20$ and $\hat{\rho}_{1,2}(i)$ is the sample correlation between $\epsilon_{1,t}^2(\hat{\theta}) - 1$ and $\epsilon_{2,t-i}^2(\hat{\theta}) - 1$.

Under the null and Hong's regulatory conditions, including $E[\epsilon_{i,t}^8] < \infty$, the statistic is asymptotically normal $\hat{Q}_T \xrightarrow{d} N(0, 1)$. Conversely, large positive values are indicative of spillover, so a one-sided test is performed. Hong uses $M \in \{10, 20, 30\}$ in his simulations with sample sizes $T \in \{300, 800\}$, and finds little difference in empirical size, but power is sensitive to choice of M for a given design. We simply use the middle of his three values.

⁶This is further supported by separate simulations not displayed here where the true value θ^0 under the null is used, rather than a plug-in. In this case $\gamma_T = [T^{.25}] = 6$ again works exceptionally well.

Table 2: k_T, γ_T, κ_m

Trimming Policy $k_T \sim \delta T / \ln(T)$ and HAC Bandwidth $\gamma_T \sim T^\zeta$				
	White Noise ^a		Omitted Variables	Volatility Spillover
	IID and AR	GARCH	IID and AR	GARCH ^b
k_T	.20 (3 ^c)	.45 (65)	.015 (2)	.03, .01 (4,1)
γ_T	.25 (6 ^d)	.25 (6)	.25 (6)	.35 (22)

Equation Tail Index^e κ_m Under H_0				
	White Noise		Omitted Variables	Volatility Spillover
	IID and AR	GARCH	IID and AR	GARCH
κ_m	1.5, 2.5	2.9	1.5, 2.5	1.25, ∞

- a. TTMC tests of white noise and omitted variables use symmetric trimming $k_T \sim \delta T / \ln(T)$.
b. The TTMC test of volatility spillover uses asymmetric $\{k_{1,T}, k_{1,T}\} \sim \{\delta_1 T / \ln(T), \delta_1 T / \ln(T)\}$.
c. δ and $[\delta 1000 / \ln(1000)]$. d. ζ and $[1000^\zeta]$.
e. In all cases under null the equations $m_{i,t}(\theta^0)$ are a product of independent random variables with tail index identical to either variable. The GARCH model uses Gaussian errors in one case, hence $\kappa_m = \infty$.

5.2 Plug-In

Plug-in choices for all tests are summarized in Table 3. We do not require a plug-in for the test of white-noise on y_t , while the tests of omitted variables and volatility spillover are respectively based on AR and GARCH residuals. We therefore use OLS $\hat{\theta}_{LS}$ and GMTTM $\hat{\theta}_{GT}$ for AR model estimation, and QMWL $\hat{\theta}_{QW}$ for the GARCH model.

In each case test and estimating equations have similar structures, so they either both have infinite variances or both have finite variances. Background theory developed in Section 4 shows the slow convergence plug-in property P2 applies if test equations have finite variances, and rapid convergence P1 applies in the infinite variance case by our choice of test equation fractile type $k_T \sim T^\gamma$.

Complete details follow. Throughout $z_t^{(a)} := |z_t|$.

OLS for AR (omitted variables): Least squares $\hat{\theta}_{LS}$ satisfies P1 or P2 if equations $m_{i,t}(\theta^0)$ have infinite or finite variances, hence if $E[\epsilon_t^2] = \infty$ or $E[\epsilon_t^2] < \infty$. See Lemma 4.1.

GMTTM for AR (omitted variables): GMTTM equations are exactly identified least squares-type

$$\tilde{m}_t(\theta) = (y_t - \theta' x_t) x_t \text{ where } x_t = [y_{t-1}, y_{t-2}]'$$

Each $\tilde{m}_{i,t}(\theta^0) = \epsilon_t x_t$ is mean zero with iid ϵ_t , and symmetrically distributed. We therefore impose symmetric trimming with one common estimating equation fractile $\tilde{k}_T = [T / \ln(T)]$, hence

$$\tilde{m}_{i,T,t}(\theta) = \tilde{m}_{i,t}(\theta) \times I\left(|\tilde{m}_{i,t}(\theta)| \leq \tilde{m}_{i,(\tilde{k}_T+1)}^{(a)}(\theta)\right) = \tilde{m}_{i,t}(\theta) \times \hat{I}_{i,T}(\theta).$$

The efficiently weighted two-step GMTTM criterion is

$$Q_{GT,T}(\theta) := \left(\sum_{t=1}^T \tilde{m}_{i,T,t}(\theta) \right)' \hat{S}_T^{-1} \left(\sum_{at=1}^T \tilde{m}_{i,T,t}(\theta) \right)$$

with first stage HAC $\widehat{S}_T = \sum_{s,t=1}^T k((s-t)/\gamma_T) \widehat{m}_{T,s}(\widehat{\theta}_{NGT}) \widehat{m}_{T,t}(\widehat{\theta}_{NGT})'$ based on the naively weighted GMTTM $\widehat{\theta}_{NGT} = \arg \min_{\theta \in \Theta} \{(\sum_{t=1}^T \widehat{m}_{i,T,t}(\theta))' \times (\sum_{at=1}^T \widehat{m}_{i,T,t}(\theta))\}$, Bartlett $k(\cdot)$ and bandwidth $\gamma_T = [T^{.25}]$.

The fractile class $\widehat{k}_T = [T/\ln(T)]$ optimizes the efficiently weighted GMTTM rate, hence $\widehat{\theta}_{GT}$ is $T^{1/2}$, $T^{1/2}/L(T)$ or $T^{1/\kappa}/L(T)$ convergent if $\kappa_\epsilon > 2$, $\kappa_\epsilon = 2$ or $\kappa_\epsilon < 2$ (HR 2010a: Section 3.1). Similar to the OLS case, $\widehat{\theta}_{GT}$ satisfies P1 or P2 if $m_{i,t}(\theta^0)$ have infinite or finite variances, hence if $E[\epsilon_t^2] = \infty$ or $E[\epsilon_t^2] < \infty$ (Lemma 4.1).

QMWL for GARCH (volatility spillover): Ling's (2007) criterion is QML with a smooth weight $w_{i,t}$:

$$Q_{QW,T}(\theta) := \sum_{t=1}^T w_{i,t} \ln \left(\frac{1}{h_{i,t}(\theta)} \exp \{ -0.5 \epsilon_{i,t}^2 / h_{i,t}^2(\theta) \} \right).$$

Ling (2007) proposes a weight based on Huber's (1977) influence function evaluated at the 5th two-tailed percentile $y_{i,([.05T])}^{(a)}$:

$$w_{i,t} = \left(\max \left\{ 1, \frac{1}{y_{i,([.05T])}^{(a)}} |y_{i,t-1}| I(|y_{i,t-1}| > y_{i,([.05T])}^{(a)}) \right\} \right)^{-4}.$$

Since $\epsilon_{i,t}$ have finite variances the rate of convergence of $\widehat{\theta}_{QW} = \arg \sup_{\theta \in \Theta} \{Q_{QW,T}(\theta)\}$ is $T^{1/2}$. Thus P1 or P2 apply depending on whether test equations $m_{i,t}(\theta^0)$ have an infinite variance, which depends on whether $\epsilon_{i,t}$ have infinite fourth moments: if $E[\epsilon_{i,t}^4] = \infty$ then P1, else P2 applies. See Lemma 4.3.

Table 3: Plug-In $\widehat{\theta}$

	Omitted Var	Vol Spill
	IID and AR	GARCH
TTMC	OLS, GMTTM	QMWL
MC	OLS, GMTTM	QMWL
WALD	OLS, GMTTM	-
Hong	-	QMWL

5.4 Simulation Results

Simulation results are presented in Tables 4 and 5. In the case of infinite variance equations each untrimmed MC test displays empirical size distortions from the nominal sizes of 1%, 5% and 10%. In all cases the null is strongly under-rejected demonstrating a deviation from the chi-squared limit distribution.

The Ljung-Box Q-test of white noise, Wald test of omitted variables and Hong's test of volatility spillover all demonstrate size distortions in the presence of heavy tails. Notice the Q-test for strong-GARCH(1,1) data with index $\kappa_y = 2.9$ substantially over-rejects the null of white noise (see Table 4: top panel, 4th column). Hong's (2001) test of volatility spillover over-rejects the null even when $\epsilon_{i,t} \stackrel{iid}{\sim} N(0,1)$ such that his required regulatory conditions hold (see Table 6: top panel, 5th column).

The TTMC statistic exhibits 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. The TTMC statistic has excellent empirical power for tests of white noise and omitted

variables, while the untrimmed MC statistic has very weak power as a test of omitted variables, and no power as a test of volatility spillover. Indeed, empirical power is *less* than the nominal size, revealing a radical scale defect when tails are heavy. Finally, both Hong's test and the TTMC test for volatility spillover exhibit weak empirical power, although the TTMC dominates for both thin-tailed and heavy-tailed data.

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 $E[m_t(\theta^0)]$. Although $E[m_t(\theta^0)] = 0$ under the null, $E[m_t(\theta)]$ does not have to exist under the alternative for any θ . Under fairly general conditions on the data generating process the resulting test statistic is asymptotically chi-squared, and obtains non-negligible power against a sequence of local alternatives that depends on tail thickness. Hypotheses covered are essentially any testable moment condition, including at least those appearing in tests of omitted variables, functional form, order selection, volatility spillover, white noise, and over-identifying restrictions.

The statistic uses as plug-in a large array of potential estimators $\hat{\theta}_T$ that need not be $T^{1/2}$ -convergent nor asymptotically normal, including conventional M- and MM-estimators, and robust versions based on weighting and trimming like GMM or least squares with tail-trimming. In many cases, depending on heavy tails and the test equation form, plug-ins without Gaussian limits are allowed. Thus, OLS and QML are viable candidates in many cases for very heavy tailed data. This is possible precisely because in the presence of heavy tails super- $T^{1/2}$ -convergent estimators exist, and equation trimming can be exploited to slow down the test equation rate of convergence relative to the plug-in. Both possibilities fail to exist for thin-tailed, stationary data.

We only scratch the surface of possibilities through examples and a simulation study. The TTMC format works very well as a heavy tail robust test of white noise and omitted variables, but exhibits relatively low power as a test of volatility spillover. Although Hong's (2001) test exhibits even lower power, extensions of the methods here may reasonably include a tail-trimmed portmanteau statistic, a task left for future research.

APPENDIX A: Assumptions

Assume $L(T)$ is a slowly varying function, $L(T) \rightarrow \infty$, whose value and rate may change with the context. Write compactly throughout

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

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

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

Asymptotic arguments require the following constructions, some of which are already

defined above. Estimating equation instantaneous and long run covariance matrices are

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

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

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

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

and

$$\mathfrak{S}_T^*(\theta) := \sum_{s,t=1}^T E \left[\{\mathcal{M}_{T,s}^*(\theta) - E[\mathcal{M}_{T,s}^*(\theta)]\} \{\mathcal{M}_{T,t}^*(\theta) - E[\mathcal{M}_{T,t}^*(\theta)]\}' \right]. \quad (15)$$

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

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

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

and a scale matrix is

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

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 ; $S_T^{-1/2}(\theta^0) \sum_{t=1}^T \{m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)]\}$ 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 matrices $\{\tilde{V}_T\}$ on $\mathbb{R}^{r \times r}$, with divergent diagonal components $\tilde{V}_{i,i,T} \rightarrow \infty$.

P1 (fast plug-in convergence). $\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(1)$ and $\|V_T \tilde{V}_T^{-1}\| \rightarrow 0$ 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. $\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = \tilde{A}_T \sum_{t=1}^T \{\tilde{m}_{T,t}(\theta^0) - E[\tilde{m}_{T,t}(\theta^0)]\} \times (1 + o_p(1)) + o_p(1)$ for unique $\theta^0 \in \Theta$ where non-stochastic $\tilde{A}_T \in \mathbb{R}^{r \times p}$ satisfies $\tilde{A}_T \tilde{S}_T^{-1} \tilde{A}_T' \rightarrow I_p$;

c. The limiting finite dimensional distributions for $\mathfrak{S}_T^*(\theta^0)^{-1/2} \{\mathcal{M}_{T,t}^*(\theta^0) - E[\mathcal{M}_{T,t}^*(\theta^0)]\}$ belong to the same class as those for $S_T^{-1} \{m_{T,t}^*(\theta) - E[m_{T,t}^*(\theta)]\}$.

Remark 1: P1 states $\hat{\theta}_T$ is consistent with a compound rate of convergence $\tilde{V}_T^{-1/2}$ faster than GMTTM $V_T^{-1/2}$ in the sense $\|V_T \tilde{V}_T^{-1}\| \rightarrow 0$ for given fractile sequences $\{k_{1,i,T}, k_{2,i,T}\}$. In this case the plug-in estimating equations $\{\tilde{m}_{T,t}(\theta^0)\}$ have no influence on the test statistic, hence we need say nothing further.

Under stationarity $\|V_T\| \sim KT$ if $m_{i,t}^2(\theta^0)$ and $(\partial/\partial\theta)m_t(\theta)|_{\theta^0}$ are integrable. Similarly, a wide range of minimum distance estimators under standard standard regulatory conditions satisfy $\|\tilde{V}_T\| \sim KT$ for sufficiently thin-tailed data. Thus, for stationary data P1 evidently can only occur when a test equation $m_{i,t}(\theta^0)$ has an infinite variance. In heavy tailed cases, however, we can always choose $\{k_{1,i,T}, k_{2,i,T}\}$ to slow down $\|V_T\| = o(\|\tilde{V}_T\|)$, so conventional M-, Method of Moments and Empirical Likelihood estimators may have property P1, including untrimmed NLLS, LAD, QML, GMM, and CUE-GMM, as well as heavy tail robust estimators like QMWL, LAWD and GMTTM.

Consider simple examples of testing error orthogonality in AR(1) and GARCH(1,1) models. The AR model is $y_t = \theta^0 y_{t-1} + \epsilon_t$ with iid ϵ_t and Paretian tail

$$P(|\epsilon_t| > \epsilon) = d\epsilon^{-\kappa} (1 + o(1)), \quad d > 0, \quad \kappa \in (1, 2).$$

Assume the test equation is $m_t(\theta) = (y_t - \theta y_{t-1})y_{t-1}$ under symmetric trimming with fractile k_T . Thus, V_T is identically the exactly identified GMTTM scale. If θ is estimated by least squares then $\tilde{V}_T^{1/2} \sim KT^{1/\kappa}/L(T)$ (Davis et al 1992). The GMTTM scale, however, cannot have a faster rate due to trimming $\|V_T\| \leq K\|\tilde{V}_T\|$, and $\|\tilde{V}_T\|/\|V_T\| \rightarrow \infty$ unless the test equation fractile k_T is slowly varying (e.g. $k_T \sim \ln(T)$). See HR (2010a).

Another example is Ling's (2007) Quasi-Maximum Weighted Likelihood [QMWL] estimator for GARCH(1,1). The estimator is T^{12} -convergent while GMTTM is at best $T^{12}/L(T)$ -convergent when $k_{j,i,T} \sim T/L(T)$, cf. HR (2010a). By comparison the unweighted QML rate of convergence for GARCH is strictly dominated by $T^{12}/L(T)$ when the underlying iid error has an infinite fourth moment (Hall and Yao 2003). Thus, for GARCH models QML satisfies *neither* P1 or P2 when test equation fractiles $k_{j,i,T} \sim T/L(T)$ are chosen. See Section 4 for specific plug-in details for specific models.

Remark 2: P2 imposes proportionality $\tilde{V}_T \sim \mathcal{K}V_T$. In this case since $\hat{\theta}_T \xrightarrow{P} \theta^0$ slowly enough that $\hat{\theta}_T$ effects the test statistic we assume $\hat{\theta}_T$ is asymptotically linear in equations $\tilde{m}_{T,t}(\theta^0)$ in P2.b. Since the test equations $m_t(\theta^0)$ are geometrically β -mixing by D3, below, the tail-trimmed equations $m_{T,t}^*(\theta^0)$ satisfy a Gaussian central limit theorem (Hill 2010b, HR 2010a). Thus, property P2.c ensures $\tilde{m}_{T,t}(\theta^0)$ has the same central limit property, and together P2.b and P2.b imply $\tilde{m}_{T,t}(\theta^0)$ may be estimating equations from conventional and outlier robust M- and MM-estimators under thin tails, and heavy-tailed robust estimators like LAWD and QMWL (Ling 2005, 2007), GMTTM (HR 2010a) and LTTS (Hill 2010a).

Remark 3: The omitted case $\|\tilde{V}_T\|/\|V_T\| \rightarrow 0$ is unsatisfactory since the plug-in equations $\tilde{m}_{T,t}(\theta^0)$ dominate, so a test of (1) cannot be performed. We therefore restrict attention to P1 or P2.

The second set promotes local identification of θ^0 .

I1 (integrability). $m_t(\theta^0)$ is integrable under the null (1).

I2 (identification). Under the null (1) the thresholds $\{l_{i,T}(\theta^0), u_{i,T}(\theta^0)\}$ satisfy a sequence of fixed point bounds: $E[m_{T,t}^*(\theta^0)] = o(\|S_T\|^{1/2}/T)$.

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

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

Remark 1: I1 formally ensures the null (1) makes sense. I2 is required due to the quadratic test statistic form. In general $E[m_{T,t}^*(\theta^0)] \rightarrow 0$ by Lebesgue's dominated

convergence under the null; $E[m_{i,T,t}^*(\theta^0)] = 0$ for any thresholds $l_{i,T}(\theta^0) = u_{i,T}(\theta^0)$ if $m_{i,t}(\theta^0)$ is symmetrically distributed; and $E[m_{i,T,t}^*(\theta^0)] \approx 0$ arbitrarily close for any T by relating the left- and right-tail fractiles $k_{1,i,T}$ and $k_{2,i,T}$. See Section 3.7. In turn $\|S_T\|/T^2 = o(1)$ under the null and mixing D3 by Lemma C.2 in Appendix C.

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 $m_t(\theta)$ and $J_{T,t}^*(\theta)$.

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 $m_{i,t}^2(\theta)$ 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 $\inf_{\theta} \kappa_i(\theta) > 0$, $\kappa_i(\theta^0) > 1$ and $\sup_{\theta} \{d_i^{-1}(\theta)m^{\kappa_i(\theta)}P(|m_{i,t}(\theta)| > m)\} \rightarrow 1$.

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

D3 (mixing). $\mathcal{M}_{T,t}^*(\theta)$ is for each T strictly stationary over $1 \leq t \leq T$, and geometrically β -mixing: $\beta_l := \sup_{A \subset \mathfrak{S}_{t+l}^+} E|P(A|\mathfrak{S}_{-\infty}^t) - P(A)| = o(\rho^l)$ for $\rho \in (0, 1)$, where $\{\mathfrak{S}_t\}$ is some sequence of σ -fields adapted to $\{\mathcal{M}_{T,t}^*(\theta)\}$, and \mathfrak{S}_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_i -bounded $\forall i, j$.

D5 (Jacobian 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 $\{\delta_T\}$, $\delta_T \rightarrow 0$, $\sup_{\|\theta - \theta^0\| \leq \delta_T} \{ \|J_T(\theta)\| / \|J_T\| \} = 1 + o(1)$.

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

Remark 1: D1-D6 are essentially identical to conditions imposed in HR (2010a: D1-D6) for GMTTM. See that source for complete details and examples. Distribution continuity D1 greatly simplifies asymptotics in lieu of the trimming indicators $I_{i,T,t}(\theta)$, cf. Čížek (2008, 2009). Equation differentiability D2 simplifies the discourse and can be removed by borrowing arguments from Pakes and Pollard (1989) and Newey and McFadden (1994).

Remark 2: Mixing D3 and indicator metric entropy property D6⁷ ensure partial sums of $I_{i,T,t}(\theta)$ satisfy a uniform central limit theorem (Pakes and Pollard 1989, Doukhan et al 1995, van der Vaart and Wellner 1994). This is used to prove $\hat{m}_{T,t}^*(\theta)$ uniformly approximates $m_{T,t}^*(\theta)$ sufficiently fast.

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 .

The last concerns kernel properties for the HAC kernel estimator.

⁷The 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 $\mathcal{N}_{[\cdot]}(\varepsilon, \mathcal{B}, \|\cdot\|_2)$ are the number of ε - L_2 -brackets required to cover \mathcal{F} , and metric entropy with L_2 -bracketing is $\mathcal{H}_{[\cdot]}(\varepsilon, \mathcal{B}, \|\cdot\|_2) = \ln(\mathcal{N}_{[\cdot]}(\varepsilon, \mathcal{B}, \|\cdot\|_2))$. See Pollard (1984), van der Vaart and Wellner (1996) and Dudley (1999). Since $\mathcal{H}_{[\cdot]}(\varepsilon, \Theta, \|\cdot\|_2) = O(\ln(\varepsilon))$ clearly $\int_0^1 \mathcal{H}_{[\cdot]}^{1/2}(\varepsilon, \Theta, \|\cdot\|_2) d\varepsilon < \infty$ hence a required stochastic equicontinuity condition for weak convergence of a partial sum of $I_{T,t}(\theta)$ applies (Dudley 1978, Doukhan set al 1995).

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

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

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

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

APPENDIX B: Proofs of Main Results

The following arguments exploit Lemmas C.1-C.8 in Appendix C. Throughout $\{r_T\}$ is a sequence of positive numbers, $r_T \rightarrow 0$ arbitrarily fast, whose rate may change from line to line. For example, we may write $T \times r_T = r_T$. Further, matrix inverses exist under the positive definiteness and rank properties I3 and D5 for large T .

Proof of Theorem 3.1. Let H_0 hold. We prove the claim by case according to plug-in property P1 or P2. Define

$$M_{T,t}^* := m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)] \quad \text{and} \quad \tilde{M}_{T,t}^* := \tilde{m}_{T,t}^*(\theta^0) - E[\tilde{m}_{T,t}^*(\theta^0)]$$

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

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

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

Identification I2 states under the null

$$S_T^{-1/2} E[m_{T,t}^*(\theta^0)] = o(T^{-1}),$$

hence under the null

$$S_T^{-1} \sum_{t=1}^T m_{T,t}^*(\theta^0) = S_T^{-1} \sum_{t=1}^T M_{T,t}^* + o(1). \quad (16)$$

Further, asymptotic expansion Lemma C.4.a coupled with Jacobian consistency Lemma C.5 and $\hat{\theta}_T \xrightarrow{p} \theta^0$ imply for some non-stochastic $r_T \rightarrow 0$ arbitrarily fast

$$\frac{1}{T} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) = \frac{1}{T} \sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0) + J_T(\hat{\theta}_T - \theta^0)(1 + o_p(1)) + o_p(r_T). \quad (17)$$

Finally, by approximation Lemma C.3.a

$$S_T^{-1/2} \sum_{t=1}^T \{\hat{m}_{T,t}^*(\theta^0) - m_{T,t}^*(\theta^0)\} = o_p(1). \quad (18)$$

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

$$TS_T^{-1/2}J_T(\hat{\theta}_T - \theta^0) = \{TS_T^{-1/2}J_T\tilde{V}_T^{-1/2}\}\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = o_p(1) \quad (19)$$

since $\|TS_T^{-1/2}J_T\tilde{V}_T^{-1/2}\| \leq \|V_T\|^{1/2}/\|\tilde{V}_T\|^{1/2} \rightarrow 0$. Further $\hat{S}_T(\hat{\theta}_T) = S_T(1 + o_p(1))$ by HAC consistency Lemma C.6. Since $r_T \rightarrow 0$ in (17) is arbitrarily fast, combine (16)-(19) to obtain

$$\begin{aligned} \hat{W}_T &= T^2 \left(\frac{1}{T} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) \right)' \hat{S}_T^{-1}(\hat{\theta}_T) \left(\frac{1}{T} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) \right) \quad (20) \\ &= T^2 \left(\frac{1}{T} \sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0) + J_T(\hat{\theta}_T - \theta^0)(1 + o_p(1)) + o_p(r_T) \right)' \\ &\quad \times S_T^{-1} \left(\frac{1}{T} \sum_{t=1}^T \hat{m}_{T,t}^*(\theta^0) + J_T(\hat{\theta}_T - \theta^0)(1 + o_p(1)) + o_p(r_T) \right) \\ &\quad \times (1 + o_p(1)) \\ &= \left(S_T^{-1/2} \sum_{t=1}^T M_{T,t}^* + TS_T^{-1/2}J_T(\hat{\theta}_T - \theta^0)(1 + o_p(1)) + o_p(1) \right) \\ &\quad \times \left(S_T^{-1/2} \sum_{t=1}^T M_{T,t}^* + TS_T^{-1/2}J_T(\hat{\theta}_T - \theta^0)(1 + o_p(1)) + o_p(1) \right) \\ &\quad \times (1 + o_p(1)) + o_p(1) \\ &= \left(S_T^{-1/2} \sum_{t=1}^T M_{T,t}^* + o_p(1) \right)' \times \left(S_T^{-1/2} \sum_{t=1}^T M_{T,t}^* + o_p(1) \right) \\ &\quad \times (1 + o_p(1)) + o_p(1) \\ &= \mathcal{Z}'_T \mathcal{Z}_T \times (1 + o_p(1)) + o_p(1), \end{aligned}$$

say. Invoke central limit theorem Lemma C.7 to deduce

$$\mathcal{Z}_T = \left(S_T^{-1/2} \sum_{t=1}^T \{m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)]\} \right)' + o_p(1) \xrightarrow{d} N(0, I_q)$$

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 some non-stochastic sequence $\{\tilde{A}_T\}$, $\tilde{A}_T \in \mathbb{R}^{r \times p}$, satisfies $\tilde{A}_T \tilde{S}_T \tilde{A}'_T \rightarrow I_p$ and

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

where $\tilde{V}_T \sim \mathcal{K}V_T$ for positive definite $\mathcal{K} \in \mathbb{R}^{r \times r}$. Further, HAC consistency Lemma C.6 states

$$\hat{S}_T(\hat{\theta}_T) = \check{S}_T(\hat{\theta}_T) \times (1 + o_p(1)). \quad (21)$$

Substitute for $\hat{\theta}_T - \theta^0$ in (17), and invoke properties (16) and (18) in Case 1 to obtain

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

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 $r_T \rightarrow 0$ arbitrarily fast and $\liminf_{T \geq N} \|S_T\| > 0$ under I3 ensure $o_p(r_T) = o_p(\|S_T\|^{1/2})$. By construction of $V_T^{1/2}$, and $\tilde{V}_T \sim \mathcal{K} V_T$ under P2, there exists positive definite $\mathcal{C} \in \mathbb{R}^{q \times q}$ that satisfies

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

hence

$$E \left(\tilde{\mathcal{B}}_T \sum_{t=1}^T \tilde{M}_{T,t}^* \right) \left(\tilde{\mathcal{B}}_T \sum_{t=1}^T \tilde{M}_{T,t}^* \right)' \sim \mathcal{C} S_T.$$

Recall $\mathcal{M}_{T,t}^*(\theta^0) \in \mathbb{R}^s$ contains all unique equations in $m_{T,t}^*(\theta^0) \in \mathbb{R}^q$ and $\tilde{m}_{T,t}(\theta^0) \in \mathbb{R}^p$, $s \geq \max\{p, q\}$. Let the non-stochastic selection matrix $\mathcal{R}_T \in \mathbb{R}^{q \times s}$ satisfy

$$\mathcal{R}_T \{ \mathcal{M}_{T,t}^*(\theta^0) - E[\mathcal{M}_{T,t}^*(\theta^0)] \} = M_{T,t}^* + \tilde{\mathcal{B}}_T \tilde{M}_{T,t}^*,$$

hence

$$\sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) = \sum_{t=1}^T \mathcal{R}_T \{ \mathcal{M}_{T,t}^*(\theta^0) - E[\mathcal{M}_{T,t}^*(\theta^0)] \} (1 + o_p(1)) + o_p(\|S_T\|^{1/2}). \quad (22)$$

Now define $\mathbb{S}_T := \mathcal{R}_T \mathfrak{S}_T^* \mathcal{R}_T' \in \mathbb{R}^{q \times q}$ where \mathfrak{S}_T^* is the covariance matrix for $\sum_{t=1}^T \mathcal{M}_{T,t}^*(\theta^0)$ defined in (15) in Appendix A. By construction $\|\mathbb{S}_T^{-1} S_T\| = O(1)$, hence by central limit

theorem Lemma C.7

$$\begin{aligned} \mathbb{S}_T^{-1/2} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) &= \mathbb{S}_T^{-1/2} \sum_{t=1}^T \mathcal{R}_T \{ \mathcal{M}_{T,t}^*(\theta^0) - E[\mathcal{M}_{T,t}^*(\theta^0)] \} (1 + o_p(1)) + o_p(1) \\ &\stackrel{d}{\rightarrow} N(0, I_q), \end{aligned} \tag{23}$$

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

Equation (23) implies $E(\mathbb{S}_T^{-1/2} \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T)) \rightarrow 0$ arbitrarily fast by the Helly-Bray theorem, hence

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

But this ensures by HAC consistency (21) and the definition of $\check{S}_T(\hat{\theta}_T)$,

$$\mathbb{S}_T = \check{S}_T(\hat{\theta}_T) \times (1 + o_p(1)) = \hat{S}_T(\hat{\theta}_T) \times (1 + o_p(1)), \tag{24}$$

therefore by (21)-(24) it follows

$$\begin{aligned} \hat{S}_T^{-1/2}(\hat{\theta}_T) \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T) \\ = \mathbb{S}_T^{-1/2} \sum_{t=1}^T \mathcal{R}_T \{ \mathcal{M}_{T,t}^*(\theta^0) - E[\mathcal{M}_{T,t}^*(\theta^0)] \} (1 + o_p(1)) + o_p(1) \stackrel{d}{\rightarrow} N(0, I_q) \end{aligned} \tag{25}$$

Now combine (25) with rank $s - r$ of $\mathcal{R}_T \mathbb{S}_T^{-1} \mathcal{R}'_T$ and invoke the mapping theorem to prove the claim:

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

■

Proof of Theorem 3.2. Notice $H_{1,L}$ does not affect the supporting Lemmas C.1-C.8 since none require null identification I2. The proof of Theorem 3.1 therefore carries over with only minor changes.

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

$$TS_T^{-1/2} E[m_{T,t}^*(\theta^0)] \rightarrow v. \tag{26}$$

Under plug-in rate P1 and $H_{1,L}$, apply (17), (19), (26) and CLT Lemma C.7 to deduce

$$\begin{aligned} S_T^{-1/2} \sum_{t=1}^T m_{T,t}^*(\hat{\theta}_T) &= S_T^{-1/2} \sum_{t=1}^T m_{T,t}^*(\theta^0) + o_p(1) \\ &= S_T^{-1/2} \sum_{t=1}^T \{ m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)] \} + v(1 + o_p(1)) \stackrel{d}{\rightarrow} N(v, I_q), \end{aligned}$$

and so on. ■

APPENDIX C: Supporting Lemmata

In order to prove Theorem 3.1 we 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 $\hat{S}_T(\theta)$. 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. In order to simplify notation assume all equations are trimmed: $\underline{q} = q$.

All of the following come directly from, or after slight adjustments are consequences of, theory developed in HR (2010a,b: Appendices C-E). We present proofs here for the sake of completeness and ease of reference.

First, we bound the maximum threshold $c_T(\theta)$, and relate and bound the instantaneous and long run covariances $\Sigma_T(\theta)$ and $S_T(\theta)$.

LEMMA C.1 (threshold bound) *Under D1 $\sup_{\theta} \{c_T(\theta)/\|\Sigma_T(\theta)\|^{1/2}\} = o(T^{1/2})$ and $\sup_{\theta} \{c_T(\theta)/\|S_T(\theta)\|^{1/2}\} = o(1)$.*

LEMMA C.2 (covariance properties) *Under D3, I3, and $\liminf_{T \geq N} \inf_{\theta} \{\lambda_{\min}(\Sigma_T(\theta))\} > 0$:*

- a. $\limsup_{T \geq N} \sup_{\theta} \|T^{-1}\Sigma_T^{-1}(\theta)S_T(\theta)\| \leq K$;
- b. $\|\Sigma_T(\theta)\| = o(T \times \max\{1, \|E[m_{T,t}^*(\theta)]\|\})$ and $\sup_{\theta} \|\Sigma_T(\theta)\| = o(T \times \max\{1, \sup_{\theta} \|E[m_{T,t}^*(\theta)]\|\})$;
- c. *If additionally I2 holds then $\|S_T\| = o(T^2)$.*

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

LEMMA C.3 (approximations) *Under D1-D4, D6, and P1 or P2:*

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

Recall the kernel function $k_{T,s,t}$, and define $\hat{\mu}_{T,t}^(\theta) := \hat{m}_{T,t}^*(\theta) - \hat{m}_T^*(\theta)$ and $\mu_{T,t}^*(\theta) := m_{T,t}^*(\theta) - m_T^*(\theta)$. If additionally moment smoothness I4 and kernel property K1 hold then:*

- c. $\sup_{\theta \in U^0(\delta_T)} \left\{ \|\hat{m}_T^*(\theta) - m_T^*(\theta)\| / [1 + \|J_T\| \times \|\theta - \theta^0\|] \right\} = o_p(1) \forall \delta > 0$.
- d. $\left\| S_T^{-1} \sum_{s,t=1}^T k_{T,s,t} \left\{ \hat{\mu}_{T,s}^*(\hat{\theta}_T) \hat{\mu}_{T,t}^*(\hat{\theta}_T)' - \mu_{T,s}^*(\theta^0) \mu_{T,t}^*(\theta^0)' \right\} \right\| = o_p(1)$.

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

LEMMA C.4 (expansions) *Under D1-D6:*

- a. $m_T^*(\theta) = m_T^*(\tilde{\theta}) + J_T^*(\theta_*)(\theta - \tilde{\theta}) + r_T \times o_p(1)$ 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 $\iota > 0$.
- b. $E[m_{T,t}^*(\theta)] - E[m_{T,t}^*(\tilde{\theta})] = J_T(\tilde{\theta})(\theta - \tilde{\theta}) + o(\|J_T(\tilde{\theta})\| \times \|\theta - \tilde{\theta}\|)$ for any $\theta, \tilde{\theta} \in \Theta$.

The sample Jacobian of the trimmed equations is consistent.

LEMMA C.5 (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 C.6 (HAC estimator) Under D1-D6, K1, I3, and P1 or P2 $\hat{S}_T(\hat{\theta}_T) = S_T(1 + o_p(1))$ and $\hat{S}_T(\hat{\theta}_T) = \check{S}_T(\hat{\theta}_T)(1 + o_p(1))$.

The test equations satisfy a Gaussian central limit theorem.

LEMMA C.7 (CLT) Under D1 and D3 $r' S_T^{-1/2}(\theta^0) \sum_{t=1}^T \{m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)]\} \xrightarrow{d} N(0, 1)$ for any conformable $r'r = 1$. If P2 also holds then $r' \mathfrak{S}_T^{-1/2}(\theta^0) \sum_{t=1}^T \{\mathcal{M}_{T,t}^*(\theta^0) - E[\mathcal{M}_{T,t}^*(\theta^0)]\} \xrightarrow{d} N(0, 1)$.

Define

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

Finally, $m_T^*\theta$ satisfies a stochastic differentiability property.

LEMMA C.8 (Stochastic Differentiability) Under D1-D6 and I3 for any $\delta \geq 0$

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

Proof of Lemma C.1. The bound $\sup_{\theta} \{c_T(\theta) / \|\Sigma_T(\theta)\|^{1/2}\} = o(T^{1/2})$ is Lemma C.1 of HR (2010a): under power law tail decay D1.ii

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

The second bound $\sup_{\theta} \{c_T(\theta) / \|S_T(\theta)\|^{1/2}\} = o(1)$ follows from covariance relation Lemma C.2.a. ■

Proof of Lemma C.2. Define $z_{T,t}(\theta, r) := r'T^{-1/2} \Sigma_T^{-1/2} (m_{T,t}^*(\theta) - E[m_{T,t}^*(\theta)])$ for any conformable $r'r = 1$, where $\Sigma_T^{-1/2}$ exists by I3 for sufficiently large T .

Claim (a): By β -mixing D3 variance bound Lemma E.1 in HR (2010b) applies: $E[(\sum_{t=1}^T z_{T,t}^2(\theta, r))] \leq K \sum_{t=1}^T E[z_{T,t}^2(\theta, r)] = K$. An identical argument reveals $\sup_{\theta} E[(\sum_{t=1}^T z_{T,t}^2(\theta, r))] \leq K \sup_{\theta} \sum_{t=1}^T E[z_{T,t}^2(\theta, r)] = K$, hence $\sup_{\theta} \|\Sigma_T^{-1}(\theta) S_T(\theta)\| \leq K$.

Claim (b): If $\|\Sigma_T(\theta)\| < \infty$ the claim is trivial, so assume at least one $E[m_{i,t}^2(\theta)]$

$= \infty$, and assume without loss of generality $m_{i,t}(\theta)$ is symmetrically trimmed with two-tailed thresholds $c_{i,T}(\theta)$ and fractiles $k_{i,T}$: $(T/k_{i,T})P(|m_{i,t}(\theta)| > c_{i,T}(\theta)) = 1$. Power-law tail D1.ii implies $c_{i,T}(\theta) = d(\theta)^{1/\kappa_i(\theta)}(T/k_{i,T})^{1/\kappa_i(\theta)}$ for some $\kappa_i(\theta) \in (0, 2]$. Coupled with properties of trimmed variances for regularly varying tails if $\kappa_i(\theta) \in (1, 2)$ then⁸

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

It is easy to show $(T/k_{i,T})^{2/\kappa_i(\theta)-1} = o(T)$ for all $\kappa_i(\theta) \geq 1$. Similarly if $\kappa_i(\theta) = 2$ then $E[(m_{i,T,t}^*(\theta))^2] \sim L(T) \rightarrow \infty$ a slowly varying function which is trivially $o(T)$. Now invoke the Cauchy-Schwartz inequality to deduce $\Sigma_T(\theta) = o(T) = o(T \times \max\{1, \|E[m_{i,T,t}^*(\theta)]\|\})$. If $\kappa_i(\theta) < 1$ then $|E[m_{i,T,t}^*(\theta)]| \sim c_{i,T}(\theta) (k_{i,T}/T) = K (T/k_{i,T})^{1/\inf_{\theta \in \Theta_{2,t}} \kappa_i(\theta)-1}$, hence

$$\frac{E \left[(m_{i,T,t}^*(\theta))^2 \right]}{\left| E \left[m_{i,T,t}^*(\theta) \right] \right|^2} \sim K (T/k_{i,T}) = o(T).$$

The uniform case is identical in lieu of uniform power law tail property D1.ii.

Claim (c): Under I2 claims (a) and (b) together imply the claim. ■

Proof of Lemma C.3. Assume θ and $m_t(\theta)$ are scalars and $m_t(\theta)$ is symmetrically trimmed for notational convenience, and write $\bar{I}_{T,t}(\theta) := 1 - I_{T,t}(\theta)$. Assume θ and $m_t(\theta)$ are scalars and $m_t(\theta)$ is symmetrically trimmed for notational convenience, and write $\bar{I}_{T,t}(\theta) := 1 - I_{T,t}(\theta)$.

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

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

By construction $\|m_t \{\hat{I}_{T,t} - I_{T,t}\}\| \leq 2 \|m_{(k_T)}^{(a)} - c_T\|$, where $m_{(k_T)}^{(a)}/c_T = 1 + O_p(k_T^{-1/2})$ follows under D1-D4 and D6 by Lemma D.2.1 of HR (2010a). Now use threshold bound Lemma C.1 and covariance relation Lemma C.2.a to deduce

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

Next, by construction and the triangle inequality

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

which is $O_p(k_T^{1/2})$ by the threshold construction (6) and an application of HR's (2010a: Lemma D.4) uniform indicator law. Therefore $\sum_{t=1}^T \{\hat{m}_{T,t}^* - m_{T,t}^*\} = o_p(\|S_T\|^{1/2} (T/k_T)^{1/2} k_T^{1/2}) = o_p(\|S_T\|^{1/2} T^{1/2})$.

Claim (b): Define

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

⁸The trimmed moment property follows directly from Karamata's Theorem. See Resnick (1987: Theorem 0.6; Exercise .0.4.2.8).

and repeat the above argument to reach

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

Uniform indicator law Lemma D.4 in HR (2010a) and threshold construction (4) imply the right-hand-side is $O_p(M_T^* k_T^{1/2}/T)$.

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

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

and $\sup_{\theta} |m_{i,(k_T)}^{(a)}(\theta)/c_{i,T}(\theta) - 1| = O_p(k_{i,T}^{-1/2})$ by Lemma D.2.1 of HR (2010a), use threshold bound Lemma C.1, and covariance bound Lemma C.2.b to deduce

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

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

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

Under the null $\|E[m_{T,t}^*(\theta^0)]\| \rightarrow 0$, while under the alternative use moment smoothness I4 to deduce $\sup_{\theta \in U^0(\delta)} \|E[m_{T,t}^*(\theta)]\| > \|E[m_{T,t}^*(\theta^0)]\|$. Under either hypothesis, therefore,

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

Claim (d): We will prove the simpler result

$$\left\| S_T^{-1} \sum_{s,t=1}^T k_{T,s,t} \left\{ \hat{m}_{T,s}^*(\hat{\theta}_T) \hat{m}_{T,t}^*(\hat{\theta}_T) - m_{T,s}^*(\theta^0) m_{T,t}^*(\theta^0) \right\} \right\| = o_p(1).$$

A proof of the claim

$$\left\| S_T^{-1} \sum_{s,t=1}^T k_{T,s,t} \left\{ \hat{\mu}_{T,s}^*(\hat{\theta}_T) \hat{\mu}_{T,t}^*(\hat{\theta}_T)' - \mu_{T,s}^*(\theta^0) \mu_{T,t}^*(\theta^0)' \right\} \right\| = o_p(1)$$

is similar, but with tedious added steps to handle stochastic centering $\hat{\mu}_{T,t}^*(\hat{\theta}_T) = \hat{m}_{T,t}^*(\hat{\theta}_T) - 1/T \sum_{t=1}^T \hat{m}_{T,t}^*(\hat{\theta}_T)$.

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

Step 1: Observe

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

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

$$\begin{aligned} \eta_\delta(x) &:= \frac{1}{(2\delta^2\pi)^{1/2}} \exp\{-x^2\delta^{-2}/2\} \quad \text{and} \quad \eta_{\delta,T,j} := \eta_\delta(j/\gamma_T) \\ \mathcal{A}_{1,T,\delta} &:= \sum_{t=-T+1}^{2T} \left(\frac{1}{\gamma_T^{1/2}} \sum_{l=1-t}^{T-t} k(l/\gamma_T) S_T^{-1/2} m_{t+l} (\hat{I}_{T,t+l}^* - I_{T,t+l}^*) I(0 \leq l \leq \lceil \gamma_T/\delta \rceil) \right) \\ & \quad \times \left(\frac{1}{\gamma_T^{1/2}} \sum_{j=1-t}^{T-t} \eta_{\delta,T,j} S_T^{-1/2} m_{T,t+j}^* I(0 \leq j \leq \lceil \gamma_T/\delta \rceil) \right) \times (1 + o_p(1)). \end{aligned}$$

By CLT Lemma C.7

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

Similarly, approximation Lemma C.3.a coupled with CLT Lemma C.7 and the Helly-Bray theorem imply

$$\left\| \frac{1}{T^{1/2}} S_T^{-1/2} \sum_{t=1}^T m_{T,t}^* (\hat{I}_{T,t}^* - I_{T,t}^*) \right\|_2 = O(1).$$

Now imitate Davidson and de Jong's (2000: Lemmas A.2-A.3) arguments to deduce⁹

$$\lim_{\delta \rightarrow 0} \limsup_{T \rightarrow \infty} \|\mathcal{A}_{1,T} - \mathcal{A}_{1,T,\delta} \times (1 + o_p(1))\|_1 = 0. \quad (27)$$

Next, consider the components of $\mathcal{A}_{1,T,\delta}$. It is straightforward to generalize approximation Lemma C.3.a to a weighted version with $k(t/\gamma_T)$ under K1. Specifically, define $N_T(\delta) := \min\{T, \lceil \gamma_T/\delta \rceil + 1\}$ and note by construction and variance non-degeneracy I3

$$\limsup_{T \geq N} \frac{N_T(\delta)}{\gamma_T} \leq K \quad \text{and} \quad \sup_{\gamma \in \Gamma} \left\{ S_{N_T(\delta)}/N_T^{1/2}(\delta) \right\} \left\{ S_T/T^{1/2} \right\}^{-1} = O(1).$$

⁹Define $X_{T,t} := S_T^{-1/2} m_{T,t}^*$. Davidson and de Jong (2000: p. 414) invoke $E(\sum_{t=1}^T X_{T,t})^2 = O(1)$ under their Lemma A.1, which holds by a mixingale property and McLeish's (1975: Theorem 1.6) maximal inequality. Their proofs reveal $E(\sum_{t=1}^T X_{T,t})^2 = O(1)$ and kernel property K1 need only hold.

Now use stationarity to deduce for any δ

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

Similarly, by a straightforward generalization of CLT Lemma C.7 for any δ

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

Therefore

$$\lim_{\delta \rightarrow 0} \limsup_{T \rightarrow \infty} \| \mathcal{A}_{1,T,\delta} \|_1 = 0. \quad (28)$$

Combine (27) and (28) to conclude $\mathcal{A}_{1,T} = o_p(1)$.

Step 2: Note

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

We will bound the first term, the second is similar. Use the Taylor expansion argument

in the proof of expansion Lemma C.4.a to deduce for some $\|\theta_{T,*} - \theta^0\| \leq \|\hat{\theta}_T - \theta^0\|$

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

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

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

where $\varpi(\xi)$ is defined under K1. Lemma C.3.a and Lemma C.7 render $\mathcal{D}_T(\xi) = O_p(1)$. Further, Jacobian consistency Lemma C.5 with $\|\theta_{T,*} - \theta^0\| \leq \|\hat{\theta}_T - \theta^0\| = O_p(T^{-1/2} \|S_T\|^{1/2} \times \|J_T\|^{-1})$ under P1 or P2, and K1 properties $\sum_{s,t=1}^T |k_{T,s,t}| = o(T^2)$, $\max_{1 \leq s \leq T} \sum_{t=1}^T |k_{T,s,t}| = o(T)$ and $\gamma_T = o(T)$ imply $\mathcal{C}_T(\xi) = o_p(1)$. Therefore $\int_{-\infty}^{\infty} \mathcal{C}_T(\xi) \mathcal{D}_T(\xi) |\varpi(\xi)| d\xi = o_p(1)$ by dominated convergence and K1. Similar arguments extend to the remaining terms. ■

Proof of Lemma C.4.

Claim (a): Assume θ and $m_t(\theta)$ are scalars and $m_t(\theta)$ is symmetrically trimmed to simplify notation.

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

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

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

$$\begin{aligned}
m_T^*(\theta) - m_T^*(\tilde{\theta}) & = J_T^*(\theta_{T,\delta}) \times (\theta - \tilde{\theta}) + \frac{1}{T} \sum_{t=1}^T m_t(\theta) \times \left\{ I_{T,t}(\theta) - I_{T,t}(\tilde{\theta}) \right\} \quad (29) \\
& \quad + \frac{1}{T} \sum_{t=1}^T J_t(\theta_{T,\delta}) \times \left\{ I_{T,t}(\theta) - I_{T,t}(\theta_{T,\delta}) \right\} \times (\theta - \tilde{\theta}).
\end{aligned}$$

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

Consider the second term in (29) and use $I_{T,t}(\theta) - I_{T,t}(\tilde{\theta}) \in \{-1, 0, 1\}$ to bound

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

The threshold construction (6), $I_{T,t}(\theta) \in \{0, 1\}$ and triangle inequality imply for any $p > 0$

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

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

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

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

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

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

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

Repeat the argument for the third term in (29) by invoking envelope bound D4 for $J_t(\theta)$.

Claim (b): Apply Jacobian existence D5.i and the definition of a derivative. ■

Proof of Lemma C.5. Recall $J_T = J_T(\theta^0) = (\partial/\partial\theta)E[m_{T,t}^*(\theta)]|_{\theta^0}$ and write $\hat{m}_T^*(\theta) = 1/T \sum_{t=1}^T \hat{m}_{T,t}^*(\theta)$.

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

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

Minkowski's inequality implies for arbitrary θ

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

Apply asymptotic expansion Lemma C.4.a to deduce for some $\tilde{\theta}_{T,*} \in \{\tilde{\theta}_T - e_i \varepsilon_T, \tilde{\theta}_T + e_i \varepsilon_T\}$

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

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

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

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

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

$$\begin{aligned} & \left\| \left\{ \hat{m}_T(\tilde{\theta}_T + a\varepsilon_T) - \hat{m}_T(\theta^0) \right\} - \left\{ E \left[m_{T,t}^*(\tilde{\theta}_T + a\varepsilon_T) \right] - E \left[m_{T,t}^*(\theta^0) \right] \right\} \right\| \\ & \leq K \left\{ 1 + \|J_T\| \times \left\| \tilde{\theta}_T + a\varepsilon_T - \theta^0 \right\| \right\} \times o_p(1) \times (\mathcal{J}_T(\delta_1, \delta_2) + o_p(1)) \\ & \leq K \left\{ 1 + \|J_T\| \times \left\| \tilde{\theta}_T - \theta^0 \right\| + \|J_T\| \times \|a\varepsilon_T\| \right\} \times (\mathcal{J}_T(\delta_1, \delta_2) + o_p(1)) \\ & = o_p(\varepsilon_T \|J_T\|) + O_p(\varepsilon_T \|J_T\| \times \mathcal{J}_T(\delta_1, \delta_2)). \end{aligned}$$

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

$$\begin{aligned} & \left\| \frac{E \left[m_{T,t}^*(\tilde{\theta}_T + a\varepsilon_T) \right] - E \left[m_{T,t}^*(\theta^0) \right]}{\varepsilon_T} - aJ_T \right\| \\ & = \left\| J_T \varepsilon_T^{-1} \left(\tilde{\theta}_T + a\varepsilon_T - \theta^0 \right) - aJ_T + o_p \left(\|J_T\| \varepsilon_T^{-1} \left(\tilde{\theta}_T + \varepsilon_T - \theta^0 \right) \right) \right\| \\ & = \left\| J_T \varepsilon_T^{-1} \left(\tilde{\theta}_T - \theta^0 \right) \right\| + o_p(\|J_T\|) = o_p(\|J_T\|). \end{aligned}$$

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

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

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

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

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

Define

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

The required limit follows from expansion Lemma C.4.a, and HR's (2010a: Lemma D.3) uniform law of large numbers restricted to $U^0(\delta_1, \delta_2)$. For each $\theta \in U^0(\delta)$ we can always find a sequence $\{\theta_{T,\delta}\} \in U^0(\delta_1, \delta_2)$, $\theta_{T,\delta} \neq \theta^0$ for each finite $T \geq N$, such that

$$\begin{aligned} \frac{E[m_{T,t}^*(\theta_{T,\delta})] - E[m_{T,t}^*(\theta^0)]}{\|\theta_{T,\delta} - \theta^0\|} &= \frac{m_T^*(\theta_{T,\delta}) - m_T^*(\theta^0)}{\|\theta_{T,\delta} - \theta^0\|} + o_p(1) \times \frac{\mathbf{m}_T(\delta_1, \delta_2)}{\|\theta_{T,\delta} - \theta^0\|} \\ &= J_T^*(\theta) \times \frac{(\theta_{T,\delta} - \theta^0)}{\|\theta_{T,\delta} - \theta^0\|} \times (1 + o_p(1)) + o_p(1) \times \frac{\mathbf{m}_T(\delta_1, \delta_2)}{\|\theta_{T,\delta} - \theta^0\|}, \end{aligned}$$

where each $o_p(1)$ term does not depend on θ . Moreover, by moment expansion Lemma C.4.b

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

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

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

Therefore $\mathcal{J}_T(\delta_{1,T}, \delta_{2,T}) \xrightarrow{p} 0$ if $\mathbf{m}_T(\delta_{1,T}, \delta_{2,T})/[\delta_{2,T} \|J_T\|] = O(1)$. By the definition of a derivative, the construction $U^0(\delta_{1,T}, \delta_{2,T}) \subseteq U^0(0, \delta_{2,T}) = U^0(\delta_{2,T})$ and moment smoothness I4

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

Now invoke Jacobian smoothness D5.ii to conclude

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

■

Proof of Lemma C.6. We will only prove $\hat{S}_T(\hat{\theta}_T) = S_T(1 + o_p(1))$, the remaining claim being similar. Define $\hat{\mu}_{T,t}^*(\theta) := \hat{m}_{T,t}^*(\theta) - \hat{m}_T^*(\theta)$, $\mu_{T,t}^*(\theta) := m_{T,t}^*(\theta) - m_T^*(\theta)$ and

$$\begin{aligned} \mathcal{A}_{1,T} &:= S_T^{-1} \sum_{s,t=1}^T k_{T,s,t} \left\{ \hat{\mu}_{T,s}^*(\hat{\theta}_T) \hat{\mu}_{T,t}^*(\hat{\theta}_T) - \mu_{T,s}^*(\theta^0) \mu_{T,t}^*(\theta^0) \right\} \\ \mathcal{A}_{2,T} &:= S_T^{-1} \sum_{s,t=1}^T k_{T,s,t} \mu_{T,s}^*(\theta^0) \mu_{T,t}^*(\theta^0) - I_q. \end{aligned}$$

By the triangle inequality we must show each $\mathcal{A}_{i,T}(\gamma) \xrightarrow{p} 0$. Uniform cross-product approximation Lemma C.3.d states $\mathcal{A}_{1,T} \xrightarrow{p} 0$.

Next, we apply Theorem 2.1 of Davidson and de Jong (2000), denoted DJ, to prove $\mathcal{A}_{2,T} \xrightarrow{p} 0$. It suffices to verify their Assumptions 1-3. DJ's Assumption 1 holds by K1.

Their Assumptions 2 and 3 impose Near Epoch Dependence and relate the property to bandwidth γ_T . Both conditions are only used to promote partial sum variance bounds for a standardized process by invoking McLeish's (1975: Theorem 1.6) maximal inequality.

Define $\mathcal{Z}_{T,t} := S_T^{-1} m_{T,t}^*(\theta^0)$. Under geometric β -mixing D3 $\{m_{T,t}^*(\theta^0), \mathfrak{F}_t\}$ forms a geometric L_2 -mixingale with constants $e_{T,t}$ (cf. McLeish 1975: Theorem 2.1). Therefore

$\{\mathcal{Z}_{T,t}, \mathfrak{F}_t\}$ forms a geometric L_2 -mixingale with constants $\mathcal{E}_{T,t} := T^{-1/2} \Sigma_T^{-1/2} e_{T,t}$. By Lemma C.2.a $\|\Sigma_T^{-1} S_T\| \leq K$ hence $E(\sum_{t=1}^T \mathcal{Z}_{T,t})^2 \leq K$ without any reference to McLeish (1975) or therefore a NED supposition. A careful inspection of DJ's proof of their Theorem 2.1 reveals $E(\sum_{t=1}^T \mathcal{Z}_{T,t})^2 \leq K$ suffices in place of their Assumption 2.

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

Proof of Lemma C.7. Define

$$z_T(\lambda) = \sum_{t=1}^T z_{T,t}(\lambda) = \lambda' S_T^{-1/2} \sum_{t=1}^T \{m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)]\}$$

for conformable $\lambda' \lambda = 1$. Since $|z_{T,t}(\lambda)| \leq c_{\epsilon,T} \|S_T\|^{-1/2}$, under mixing and tail decay properties D1 and D3 $\{z_{T,t}(\lambda)\}$ satisfies the conditions of central limit theorem Lemma D.7 in Hill and Renault (2010). This suffices to prove convergence in finite dimensional distributions $z_T(\lambda) \xrightarrow{d} N(0, 1)$, hence $S_T^{-1/2} \sum_{t=1}^T \{m_{T,t}^*(\theta^0) - E[m_{T,t}^*(\theta^0)]\} \xrightarrow{d} N(0, I_q)$ by the Crámer-Wold theorem. See also Theorems 3.2 and 5.1 of Hill (2009b).

Under slow plug-in convergence P2.c and the above argument, the finite dimensional distributions of $\lambda' \mathfrak{S}_T^{-1/2}(\theta^0) \{\mathcal{M}_{T,t}^*(\theta^0) - E[\mathcal{M}_{T,t}^*(\theta^0)]\}$ are asymptotically normal. The claim now follows by the Crámer-Wold theorem. ■

Proof of Lemma C.8. Apply Minkowski's inequality and the Lemma C.3.c uniform approximation to obtain

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

Equation and moment expansions Lemma C.4.a,b imply the last line is bounded by $\sup_{\theta \in U^0(\delta)} \{ \|J_T^*(\theta) - J_T\| / \|J_T\| \} + o_p(1)$. ■

REFERENCES

- [1] Akaike, H. (1973). Information Theory and an Extension of the Likelihood Principle. In: B.N. Petrov and F. Csáki (eds.), Proceedings of the Second International Symposium on Information Theory. Akademiai Kiado: Budapest.
- [2] Andrews, B. (2008). Rank-Based Estimation for Autoregressive Moving Average Time Series Models, *Journal of Time Series Analysis* 29, 51-73.
- [3] Andrews, D.W.K. and W. Ploberger (1994). Optimal Tests when a Nuisance Parameter is

- Present Only under the Alternative, *Econometrica* 82, 1383-1414.
- [4] Antoine, B. and E. Renault (2010). Efficient Minimum Distance Estimation with Multiple Rates of Convergence, *Journal of Econometrics*: forthcoming.
- [5] Antione, B., H. Bonnal and E. Renault (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood, *Journal of Econometrics* 138, 461-487.
- [6] Basrak, B., R.A. Davis, and T. Mikosch (2002). Regular Variation of GARCH Processes, *Stochastic Processes and Their Applications* 12, 908-920.
- [7] Bierens, H. J. (1990). A Consistent Conditional Moment Test of Functional Form, *Econometrica* 58, 1443-1458.
- [8] Bierens, H. J. and W. Ploberger (1997). Asymptotic Theory of Integrated Conditional Moment Tests, *Econometrica* 65, 1129-1151.
- [9] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* 31, 307-327.
- [10] Borkovec, M. and C. Klüppelberg (2001). The Tail of the Stationary Distribution of an Autoregressive Process with ARCH(1) Errors, *Annals of Applied Probability* 11, 1220-124..
- [11] Boussama, F. (1998). Ergodicity, Mixing and Estimation in GARCH Models, Unpublished Ph.D. Dissertation, University of Paris 7.
- [12] Breiman, L. (1965). On Some Limit Theorems Similar to the Arc-Sine Law, *Theory of Probability and its Applications* 10, 351-360.
- [13] Breitung, J. and C. Gouriéroux (1997). Rank Based Tests for Unit Roots, *Journal of Econometrics* 81, 7-27.
- [14] Brooks, C. (1998). Predicting Stock Index Volatility: Can Market Volume Help? *Journal of Forecasting* 17, 59-80.
- [15] Brock, W.A., W.D. Dechert, J.A. Scheinkman, B. LeBaron (1996). A Test for Independence Based on the Correlation Dimension, *Econometric Reviews* 15, 197-235.
- [16] Campbell, J. and L. Hentschel (1992). No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns, *Journal of Financial Economics* 31, 281-318.
- [17] Caporale, M., P. Nikitas, and N. Spagnolo (2002). Testing for Causality-in-Variance: An Application to the East Asian Markets, *International Journal of Finance & Economics* 7, 235-245.
- [18] Caner, M. (1998). Tests for Cointegration with Infinite Variance Errors, *Journal of Econometrics* 86, 155-175.
- [19] Chan, N.H. and L.T. Tran (1989). On the First-Order Autoregressive Process with Infinite Variance, *Econometric Theory* 5, 354-362.
- [20] Čížek, P. (2008). General Trimmed Estimation: Robust Approach to Nonlinear and Limited Dependent Variable Models, *Econometric Theory* 24, 1500-1529.
- [21] Čížek, P. (2009). Generalized Method of Trimmed Moments, Working Paper, CentER, University of Tilburg.
- [22] Cline, D.B.H. (1986). Convolution Tails, Product Tails and Domains of Attraction, *Probability Theory and Related Fields* 72, 529-557.
- [23] Cline, D.B.H. (2007). Regular Variation of Order 1 Nonlinear AR-ARCH Models, *Stochastic Processes and their Applications* 117, 840-861.
- [24] Comte, F., and O. Lieberman (2000). Second-Order Noncausality in Multivariate GARCH Processes, *Journal of Time Series Analysis* 21, 535-557.
- [25] Comte, F., and O. Lieberman (2003). Asymptotic Theory for Multivariate GARCH Processes, *Journal of Multivariate Analysis* 84, 61-84.
- [26] Corradi, V. and N. Swanson (2002). A Consistent Test for Nonlinear Out of Sample Predictive Accuracy, *Journal of Econometrics* 110, 353-381.
- [27] Davis, R.A., K. Knight and J. Liu (1992). M-Estimation for Autoregressions with Infinite Variance, *Stochastic Processes and Their Applications* 40, 145-180.
- [28] Davis, R.A. and T. Mikosch (2009a). Extreme Value Theory for GARCH Processes. In:

- T.G. Andersen, R.A. Davis, J.-P. Kreiss, and T. Mikosch (eds.), Handbook of Financial Time Series, 187-200. Springer: New York.
- [29] Davis, R.A. and T. Mikosch (2009b). The Extremogram: A Correlogram for Extreme Events, *Bernoulli* 15, 977–1009.
- [30] Davis, R.A. and S.I. Resnick (1986). Limit Theory for the Sample Covariance and Correlation Function of Moving Averages, *Annals of Statistics* 14, 533-558.
- [31] Davidson, J. and R. de Jong (2000). Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices, *Econometrica* 68, 407-423.
- [32] de Jong, R. (1996). The Bierens Test under Data Dependence, *Journal of Econometrics* 72, 1-32.
- [33] de Lima, P. (1996). Nuisance Parameter Free Properties of Correlation Integral Based Statistics, *Econometric Reviews* 15, 237-259.
- [34] de Lima, P. (1997). On Robustness of Nonlinearity Tests to Moment Condition Failure, *Journal of Econometrics* 76, 251-280.
- [35] Doukhan, P. (1994). Mixing: Properties and Examples, *Lecture Notes in Statistics* 85. Springer: Berlin.
- [36] Doukhan, P., P. Massart, and E. Rio (1995). Invariance Principles for Absolutely Regular Empirical Processes, *Annales de l'Institut Henri Poincaré* 31, 393-427.
- [37] Dudley, R. M. (1978). Central Limit Theorem for Empirical Processes. *Annals of Probability* 6, 899-929.
- [38] Dudley, R. M. (1999). *Continuous Central Limit Theorems*. Cambridge University Press: Cambridge .
- [39] Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag: Berlin.
- [40] Engle, R. and V. Ng (1993). Measuring and Testing the Impact of News On Volatility, *Journal of Finance* 48, 1749-1778.
- [41] Fan, Y. and Q. Li (1996). Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms, *Econometrica* 64, 865-890.
- [42] Fernández, B. and N. Muriel (2009). Regular Variation and Related Results for the Multivariate GARCH(p,q) Model with Constant Conditional Correlations, *Journal of Multivariate Analysis* 100, 1538-1550.
- [43] Finkenstadt, B., and H. Rootzén (2003). *Extreme Values in Finance, Telecommunications and the Environment*. Chapman and Hall: New York.
- [44] Floren, J-P., E. Renault and N. Touzi (1998). Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes, *Econometric Theory* 14, 744-769.
- [45] Ghysels, E. and A. Andreou (2003). Tests for Breaks in the Conditional Co-Movements of Asset Returns, *Statistica Sinica* 13, 1045-1074.
- [46] Ghysels, E. and A. Guay (2003). Structural Change Tests for Simulated Method of Moments, *Journal of Econometrics* 115, 91-123.
- [47] Giné, E. and J. Zinn (1984). Some Limit Theorems for Empirical Processes, *Annals of Probability* 12, 929-989.
- [48] Hahn, M., J. Kuelbs and D. C. Weiner (1991). Sums, Trimmed Sums, and Extremes (Progress in Probability), Birkhäuser.
- [49] Hall, A. (2000). Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test, *Econometrica* 68, 1517-1527.
- [50] Hall, P. (1982). On Some Estimates of an Exponent of Regular Variation, *Journal of the Royal Statistical Society Series B* 44, 37-42.
- [51] Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH Models with Heavy-Tailed Errors, *Econometrica* 71, 285-317.
- [52] Hansen, L.P. (1982). Large Sample Properties of Method of Moments Estimators, *Econometrica* 50, 1029-1054.

- [53] Hansen, L.P., J. Heaton, and A. Yaron (1996). Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics* 14, 262-280.
- [54] Hausman, J.A. (1978). Specification Testing in Econometrics, *Econometrica* 46, 1251-1271.
- [55] He, X., J. Jurečková, R. Koenker and S. Portnoy (1990). Tail Behavior of Regression Estimators and their Breakdown Points, *Econometrica* 58, 1195-1214.
- [56] Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution, *Annals of Mathematical Statistics* 3, 1163-1174.
- [57] 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.
- [58] Hill, J.B. (2009a). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimator, *Journal of Statistical Planning and Inference* 139, 2091-2110.
- [59] Hill, J.B. (2010a). Least Tail-Trimmed Squares for Infinite Variance Autoregressions, Dept. of Economics, University of North Carolina - Chapel Hill: submitted.
- [60] Hill, J.B. (2010b). Central Limit Theory for Kernel-Self Normalized Tail-Trimmed Sums of Dependent, Heterogeneous Data, Dept. of Economics, University of North Carolina - Chapel Hill: submitted.
- [61] Hill, J.B. (2010c). On Tail Index Estimation for Dependent, Heterogeneous Data, *Econometric Theory* 26, 1398-1436.
- [62] Hill, J.B. and E. Renault (2010a). Generalized Method of Moments with Tail Trimming, Dept. of Economics, University of North Carolina - Chapel Hill: submitted.
- [63] Hill, J.B. and E. Renault (2010b). Appendix E for "Generalized Method of Moments with Tail Trimming", Dept. of Economics, University of North Carolina - Chapel Hill; www.unc.edu/~jbhill/tech_append_gmttm.pdf.
- [64] Hill, J.B. and A. Shneyerov (2010). Are There Common Values in First-Price Auctions? A Tail-Index Nonparametric Test, invited revision for *Journal of Econometrics*.
- [65] Hong, Y. (2001). A Test for Volatility Spillover with Application to Exchange Rates, *Journal of Econometrics* 103, 183-224.
- [66] Hong, Y. and H. White (1995). Consistent Specification Testing Via Nonparametric Series Regression, *Econometrica* 63, 1133-1159.
- [67] Huber, P. J. (1977). *Robust Statistical Procedures*. Philadelphia: Society for Industrial and Applied Mathematics.
- [68] Ibragomiv, R. (2007). Efficiency of Linear Estimators under Heavy-Tailedness: Convolutions of α -Symmetric Distributions, *Econometric Theory* 23, 501-517.
- [69] Imbens, G., R.H. Spady, and P. Johnson (1998). Information Theoretic Approaches to Inference in Moment Condition Models, *Econometrica* 66, 333-357.
- [70] Jaeckel, L.A. (1972). Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals, *Annals of Mathematical Statistics* 43, 1449-1458.
- [71] Jurečková, J. (1981). Tail Behavior of Location Estimators, *Annals of Statistics*, 9, 578-585.
- [72] Kesten, H. (1973) Random Difference Equations and Renewal Theory for Products of Random Matrices, *Acta Mathematica* 131, 207-248.
- [73] King, M., E. Sentana and S. Wadhvani (1994). Volatility and Links between National Stock Markets, *Econometrica* 62, 901-933.
- [74] Kitamura, Y. and M. Stutzer (1997). An Information-Theoretic Alternative to Generalized Method of Moments Estimation, *Econometrica* 65, 861-874.
- [75] Kitamura, Y, T. Gautam, and A. Hyungtaik (2004). Empirical Likelihood-Based Inference in Conditional Moment Restriction Models, *Econometrica* 72, 1667-1714.
- [76] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag: New York.
- [77] Lee T., H. White H., C.W.J. Granger (1993). Testing for Neglected Nonlinearity in Time-Series Models: A Comparison of Neural Network Methods and Alternative Tests, *Journal of*

Econometrics 56, 269-290.

- [78] Ling, S. (2005). Self-Weighted LAD Estimation for Infinite Variance Autoregressive Models, *Journal of the Royal Statistical Society Series B* 67, 381–393.
- [79] Ling, S. (2007). Self-Weighted and Local Quasi-Maximum Likelihood Estimators for ARMA-GARCH/IGARCH Models, *Journal of Econometrics* 150, 849-873.
- [80] 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*: forthcoming.
- [81] Loretan, M. and P.C.B. Phillips (1991). The Durbin Watson Ratio under Infinite Variance Errors, *Journal of Econometrics* 47, 85-114.
- [82] Mason, D.M. and Q-M. Shao (2001). Bootstrapping the Student t-Statistic, *Annals of Probability* 29, 1435-1450.
- [83] Mikosch, T. and C. Stărică (2000). Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process, *Annals of Statistics* 28, 1427-1451.
- [84] Newey, W.K. (1985). Maximum Likelihood Specification Testing and Conditional Moment Tests, *Econometrica* 53, 1047-1070.
- [85] Newey, W.K., and D.L. McFadden (1994). Large Sample Estimation and Hypothesis Testing, in: R. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4. Amsterdam: North-Holland.
- [86] Newey, W.K and D. West (1987). A Simple. Positive Semi-definite. Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.
- [87] Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators, *Econometrica* 57, 1027-1057.
- [88] Pham, T. and L. Tran (1985). Some Mixing Properties of Time Series Models. *Stochastic Processes and Their Applications* 19, 297-303.
- [89] Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlang New York.
- [90] Pollard, D. (2002). Maximal Inequalities via Bracketing with Adaptive Truncation, *Annales de l'Institut Henri Poincaré* 38, 1039-1052.
- [91] Ramsey, J.B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression, *Journal of the Royal Statistical Society Series B* 31, 350-371.
- [92] Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.
- [93] Runde, R. (1997). The Asymptotic Null Distribution of the Box-Pierce Q-Statistic for Random Variables with Infinite Variance: An Application of German Stock Returns, *Journal of Econometrics* 78, 205-216.
- [94] Ruppert, D. and J. Carroll (1980). Trimmed Least Squares Estimation in the Linear Model, *Journal of the American Statistical Association* 75, 828-838.
- [95] Schmidt R. and U. Stadtmüller (2006). Non-Parametric Estimation of Tail Dependence, *Scandinavian Journal of Statistics* 33, 307-335.
- [96] Stinchcombe, M.B. and H. White (1998). Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative, *Econometric Theory* 14, 295-325.
- [97] Tauchen, G. (1985). Diagnostic Testing and Evaluation of Maximum Likelihood Models, *Journal of Econometrics* 30, 415-443 .
- [98] van der Vaart, A.W., and J.A. Wellner (1996). *Weak Convergence and Empirical Processes*, Springer-Verlang: New York.
- [99] White H. (1981). Consequences and Detection of Misspecified Nonlinear Regression Models, *Journal of the American Statistical Association* 76, 419-433.
- [99] White H. (1982). Maximum Likelihood Estimation of Misspecified Models, *Econometrica* 50, pp. 1-25.

Table 4 - White Noise, Volatility Spillover

T = 1000, Heavy-Tailed Data ^a			
White Noise ^b			
	IID ($\epsilon_t \sim P_{1.5}$) ^c	G(1,1) ($\epsilon_t \sim N_{0,1}$)	AR ($\epsilon_t \sim P_{1.5}$)
TTMC	.01,.04,.09 ^d	.01,.06,.10	.85,.93,.95
MC ^e	.00,.02,.05	.00,.03,.07	.75,.85,.90
Q-test ^f	.04,.06,.08	.26,.41,.49	1.0,1.0,1.0
Volatility Spillover ^g (QMWL)			
	C-G H ₀ ($\epsilon_t \sim P_{2.5}$)	C-G H ₁ ¹ ($\epsilon_t \sim P_{2.5}$)	C-G H ₁ ² ($\epsilon_t \sim P_{2.5}$)
TTMC	.01,.05,.09	.18,.30,.34	.30,.43,.55
MC	.00,.00,.04	.00,.01,.02	.00,.01,.02
Hong ^h	.01,.01,.01	.22,.24,.24	.25,.25,.27

- a. All test equations have an infinite variance and finite mean. In particular, tails are heavy enough that the Ljung-Box Q-statistic and Hong's centered Q-statistic have non-normal or degenerate limits under the null.
- b. Test of white noise for IID, AR and G(1,1) = GARCH(1,1): $m_{i,t} = y_t y_{t-i}$, $i = 1, \dots, 5$.
- c. P_κ denotes a Pareto law with index κ .
- d. Values are rejection frequencies at the 1%, 5% and 10% levels.
- e. The untrimmed version of \hat{W}_T .
- f. Ljung-Box Q-test with 5 lags for white noise; Wald test for omitted variables.
- g. Test of volatility spillover for C-G = CCC-GARCH: $m_{i,t}(\theta) = (y_{1,t}^2/h_{1,t}^2(\theta) - 1)(y_{2,t-i}^2/h_{2,t-i}^2(\theta) - 1)$, with lags $i = 1, \dots, 5$. The hypotheses are H₀: no spillover and H₁: strong spillover from y to x. The plug-in is QMWL.
- h. Hong's (2001) centered portmanteau test of volatility spillover with QMWL plug-in.

Table 5 - Omitted Variables

T = 1000, Heavy-Tailed Data ^a					
Omitted Variables ^b (OLS)			Omitted Variables (GMTTM)		
	IID ($\epsilon_t \sim P_{1.5}$)	AR ($\epsilon_t \sim P_{1.5}$)		IID ($\epsilon_t \sim P_{1.5}$)	AR ($\epsilon_t \sim P_{1.5}$)
TTMC	.01,.05,.10	.84,.94,.97	TTMC	.01,.05,.09	.86,.93,.96
MC	.00,.01,.03	.35,.54,.66	MC	.00,.00,.02	.33,.52,.62
Wald	.02,.03,.05	1.0,1.0,1.0	Wald	.01, .04, .09	.89,.96,.99

- a. All test equations have an infinite variance and finite mean, and tails are heavy enough that the Wald statistic has a non-normal or degenerate limit under the null.
- b. Test of omitted variables for IID and AR: $m_{i,t}(\theta) = (y_t - \theta'x_t)(y_{t-i} - \theta'x_{t-i})$, $i = 1, \dots, 5$. The plug-in is either OLS or GMTTM.

Table 6- White Noise, Volatility Spillover, Omitted Variables

T = 1000, Thin-Tailed Data ^a						
White Noise ^b ($\epsilon_t \sim P_{4.5}$)			Volatility Spillover (QMWL $\epsilon_t \sim N_{0,1}$) ^c			
	IID ($\epsilon_t \sim P_{4.5}$)	AR		C-G H ₀	C-G H ₁ ¹	C-G H ₁ ²
TTMC	.01,.06,.10	1.0,1.0,.10	TTMC	.01,.06,.10	.17,.46,.57	.24,.51,.68
MC	.01,.05,.09	1.0,1.0,.10	MC	.01,.03,.07	.06,.14,.26	.11,.25,.40
Q-test	.01,.04,.10	1.0,1.0,.10	Hong ^h	.03,.09,.12	.24,.24,.30	.41,.42,.43
Omitted Variables (OLS, $\epsilon_t \sim P_{2.5}$)			Omitted Variables (GMTTM, $\epsilon_t \sim P_{1.5}$)			
	IID	AR		IID	AR	
TTMC	.01,.05,.09	1.0,1.0,1.0	TTMC	.01,.05,.10	1.0,1.0,1.0	
MC	.01,.04,.08	.98,.99,1.0	MC	.00,.03,.07	.96,.99,1.0	
Wald	.01,.05,.10	.99,.99,1.0	Wald	.01, .05, .11	1.0,1.0,1.0	

- a. All test equations have a finite variance, and all processes are thin-tailed enough that the Lung-Box, Wald and Hong statistics have standard limits under the null.
- b. Since all CCC-GARCH models studied here have infinite variance equations for the test of white noise, we only test IID and AR data for white noise.
- c. $N_{0,1}$ denotes a standard normal law.