Heavy-Tail and Plug-In Robust Consistent Conditional Moment Tests of Functional Form

Jonathan B. Hill[1]

Feb. 2012

Abstract We present asymptotic power-one tests of regression model functional form for heavy tailed time series. Under the null hypothesis of correct specification the model errors must have a finite mean, and otherwise only need to have a fractional moment. If the errors have an infinite variance then in principle any consistent plug-in is allowed, depending on the model, including those with non-Gaussian limits and/or a sub-$\sqrt{n}$-convergence rate. One test statistic exploits an orthogonalized test equation that promotes plug-in robustness irrespective of tails. We derive chi-squared weak limits of the statistics, we characterize an empirical process method for smoothing over a trimming parameter, and we study the finite sample properties of the test statistics.

1 Introduction

Consider a regression model

$$y_t = f(x_t, \beta) + \epsilon_t(\beta) \tag{1}$$

where $f : \mathbb{R}^p \times \mathcal{B} \to \mathbb{R}$ is a known response function for finite $p > 0$, continuous and differentiable in $\beta \in \mathcal{B}$ where $\mathcal{B}$ is a compact subset of $\mathbb{R}^q$, and the regressors $x_t \in \mathbb{R}^p$ may contain lags of y_t or other random variables. We are interested in testing whether $f(x_t, \beta)$ is a version of $E[y_t|x_t]$ for unique β^0, without imposing higher moments on y_t, while under mis-specification we only require $E[\sup_{\beta\in\mathcal{B}} |\epsilon_t(\beta)|^{\iota}] < \infty$ and each $E[\sup_{\beta\in\mathcal{B}} |(\partial/\partial\beta_i) f(x_t, \beta)|^{\iota}] < \infty$ for some tiny $\iota > 0$. Heavy tails in macroeconomic, finance, insurance and telecommunication time series are now well documented (Resnick 1987, Embrechts et al 1997, Finkenstadt and Rootzén 2003, Gabaix 2008). Assume $E|y_t| < \infty$ to ensure $E[y_t|x_t]$ exists by the Radon-Nikodym theorem, and consider the hypotheses

$$H_0 : E[y_t|x_t] = f(x_t, \beta^0) \ a.s. \text{ for unique } \beta^0 \in \mathcal{B}, \quad \text{versus} \quad H_1 : \max_{\beta\in\mathcal{B}} P\left(E[y_t|x_t] = f(x_t, \beta)\right) < 1.$$

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

The author thanks an anonymous referee and Co-Editor Xiaohong Chen for constructive remarks.

We develop consistent Conditional Moment [CM] test statistics for general alternatives that are both robust to heavy tails and to a plug-in for β^0. Our focus is Bierens' (1982, 1990) nuisance parameter indexed CM test for the sake of exposition, with neural network foundations in Gallant and White (1989), Hornik et al (1989, 1990), and White (1989a), and extensions to semi- and non-parametric models in Chen and Fan (1999). Let $\{y_t, x_t\}_{t=1}^n$ be the sample with size $n \geq 1$, let $\hat{\beta}_n$ be a consistent estimator of β^0, and define the residual $\epsilon_t(\hat{\beta}_n) := y_t - f(x_t, \hat{\beta}_n)$. The test statistic is

$$\hat{T}_n(\gamma) = \frac{1}{\hat{V}_n(\hat{\beta}_n, \gamma)} \left(\sum_{t=1}^n \epsilon_t(\hat{\beta}_n) F(\gamma'\psi_t) \right)^2 \text{ where } F(\gamma'\psi_t) = \exp\{\gamma'\psi_t\} \text{ and } \psi_t := \psi(x_t), \quad (2)$$

where ψ is a bounded one-to-one Borel function from $\mathbb{R}^p$ to $\mathbb{R}^p$, $\hat{V}_n(\hat{\beta}_n, \gamma)$ estimates $E[(\sum_{t=1}^n \epsilon_t(\hat{\beta}_n)F(\gamma'\psi_t))^2]$, and $\gamma \in \mathbb{R}^p$ is a nuisance parameter.

If $E|\epsilon_t| < \infty$ and $E[\epsilon_t|x_t] \neq 0$ with positive probability then $E[\epsilon_t F(\gamma'\psi_t)] \neq 0$ for all γ on any compact $\Gamma \subset \mathbb{R}^p$ with positive Lebesgue measure, except possibly for γ in a countable subset $S \subset \Gamma$ (Bierens 1990: Lemma 1). This seminal result promotes a consistent test: if ϵ_t and $\sup_{\beta \in \mathcal{B}} |(\partial/\partial\beta_i) f(x_t, \beta)|$ have finite $4 + \iota^{th}$-moments for tiny $\iota > 0$, and the NLLS estimator $\hat{\beta}_n = \beta^0 + O_p(1/n^{1/2})$ then $\hat{T}_n(\gamma) \xrightarrow{d} \chi^2(1)$ under H_0 and $\hat{T}_n(\gamma) \xrightarrow{p} \infty$ under H_1 for all $\gamma \in \Gamma/S$. Such moment and plug-in conditions are practically de rigueur (e.g. Hausman 1978, White 1981, Davidson et al 1983, Newey 1985, White 1987, Bierens 1990, de Jong 1996, Fan and Li 1996, Corradi and Swanson 2002, Hong and Lee 2005).

The property $E[\epsilon_t F(\gamma'\psi_t)] \neq 0$ under H_1 for all but countably many γ carries over to non-polynomial real analytic $F : \mathbb{R} \to \mathbb{R}$, including exponential and trigonometric classes (Lee et al 1993, Bierens and Ploberger 1997, Stinchcombe and White 1998), and compound versions where S may be empty (Hill 2008a,b), and has been discovered elsewhere in the literature on universal approximators (Hornik et al 1989, 1990, Stinchcombe and White 1989, White 1989b, 1990). Stinchcombe and White (1998: Theorem 3.1) show boundedness of ψ ensures $\{F(\gamma'\psi(x_t)) : \gamma \in \Gamma\}$ is weakly dense on the space on which x_t lies, a property exploited to prove F is revealing.[2]

The moment $E|\epsilon_t| < \infty$ is imposed to ensure $E[\epsilon_t|x_t]$ exists under *either* hypothesis, but if $f(x_t, \beta^0)$ is mis-specified then there is no guarantee ϵ_t is integrable when $E[y_t^2] = \infty$ precisely because $f(x_t, \beta^0)$

---

[2] We use the term "revealing" in the sense of "generically totally revealing" in Stinchcombe and White (1998: p. 299). A member h of a function space $\mathcal{H}$ reveals mis-specification $E[y|x] \neq f$ when $E[(y - f)h] \neq 0$. A space $\mathcal{H}$ is generically totally revealing if all but a negligible number of $h \in \mathcal{H}$ have this property. In the index function case $h(x) = F(\gamma'\psi(x))$, where the weight h aligns with γ and the class $\mathcal{H}$ with Γ, this is equivalent to saying all $\gamma \in \Gamma/S$ where S has Lebesgue measure zero.

need not be integrable. Suppose x_t is an integrable scalar with an infinite variance, and $f(x_t, \beta) = (x_t + \beta)^2$. Then $E|\epsilon_t(\beta)| = \infty$ for any $\beta \in \mathcal{B}$, hence $E[\epsilon_t(\beta)|x_t]$ is not well defined for any β. Clearly we only need $E|y_t| < \infty$ to ensure $E[y_t|x_t]$ exists for a test of (1), while heavy tails can lead to empirical size distortions in a variety of test statistics (de Lima 1997, Hill and Aguilar 2011).

In this paper we apply a trimming indicator $\hat{I}_{n,t}(\beta) \in \{0,1\}$ to $\epsilon_t(\beta)$ in order to robustify against heavy tails. Define the weighted and trimmed errors and test statistic

$$\hat{\mathcal{T}}_n(\gamma) = \frac{1}{\hat{S}_n^2(\hat{\beta}_n, \gamma)} \left(\sum_{t=1}^{n} \hat{m}_{n,t}^*(\hat{\beta}_n, \gamma) \right)^2 \quad \text{where} \quad \hat{m}_{n,t}^*(\beta, \gamma) := \epsilon_t(\beta)\hat{I}_{n,t}(\beta) F\left(\gamma'\psi_t\right),$$

where $\hat{S}_n^2(\beta, \gamma)$ is a kernel estimator of $E[(\sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\beta}_n, \gamma))^2]$ defined by

$$\hat{S}_n^2(\beta, \gamma) = \sum_{s,t=1}^{n} \omega\left((s-t)/b_n\right) \left\{\hat{m}_{n,s}^*(\beta, \gamma) - \hat{m}_n^*(\beta, \gamma)\right\} \left\{\hat{m}_{n,t}^*(\beta, \gamma) - \hat{m}_n^*(\beta, \gamma)\right\}$$

with $\hat{m}_n^*(\beta, \gamma) = 1/n \sum_{t=1}^n \hat{m}_{n,t}^*(\beta, \gamma)$, and $\omega(\cdot)$ is a kernel function with bandwidth $b_n \to \infty$ and $b_n/n \to 0$. By exploiting methods in the tail-trimming literature we construct $\hat{I}_{n,t}(\beta)$ in a way that ensures sufficient but *negligible trimming*: $\hat{I}_{n,t}(\beta) = 0$ for asymptotically infinitely many sample extremes of $\epsilon_t(\beta)$ representing a vanishing sample portion. This promotes both Gaussian asymptotics under H_0 *and* a consistent test.

Tail truncation by comparison is not valid when $E[\epsilon_t^2] = \infty$ because sample extremes of ϵ_t are replaced by tail order statistics of ϵ_t that increases with n: too many large values are allowed for Gaussian asymptotics (Csörgo et al 1986). On the other hand, trimming or truncating a *constant* sample portion of $\epsilon_t(\beta)$ leads to bias in general, unless ϵ_t is symmetrically distributed about zero under H_0 and symmetrically trimmed or truncated. In some cases, however, symmetry may be impossible as in a test of ARCH functional form (see Section 4.2).

We assume $F(u)$ is bounded on any compact subset of its support, covering exponential, logistic, and trigonometric weights, but not real analytic functions like $(1 - u)^{-1}$ on $[-1, 1]$. Otherwise we must include $F(\gamma'\psi_t)$ in the trimming indicator $\hat{I}_{n,t}(\beta)$ which sharply complicates proving $\hat{\mathcal{T}}_n(\gamma)$ obtains an asymptotic power of one on Γ/S. A HAC estimator $\hat{S}_n^2(\beta, \gamma)$ is required in general unless ϵ_t is iid under H_0: even if ϵ_t is a martingale difference $\hat{m}_{n,t}(\beta^0, \gamma)$ may not be due to trimming.

In lieu of the test statistic form a unique advantage exists in heavy tailed cases since $1/n \sum_{t=1}^n \hat{m}_{n,t}^*(\beta^0, \gamma)$ is sub-$n^{1/2}$-convergent. Depending on the data generating process, a plug-in $\hat{\beta}_n$ may converge fast enough that it does not impact the limit distribution of $\hat{\mathcal{T}}_n(\gamma)$ under H_0, including estimators with a sub-$n^{1/2}$ rate and/or a non-Gaussian limit. Conversely, if $\hat{\beta}_n \xrightarrow{p} \beta^0$ at a sufficiently slow rate we either assume $\hat{\beta}_n$ is asymptotically linear, or in the spirit of Wooldridge (1990) exploit an orthogonal transformation of $\hat{m}^*_{n,t}(\beta,\gamma)$ that is robust to any $\hat{\beta}_n$ with a minimal convergence rate that may be below $n^{1/2}$ for heavy tailed data. Orthogonal transformations have not been explored in the heavy tail robust inference literature, and they do not require $n^{1/2}$-convergent or asymptotically normal $\hat{\beta}_n$ in heavy tailed cases.

Model (1) covers Nonlinear ARX with random volatility errors of an unknown form, and Nonlinear strong and semi-strong ARCH. Note, however, that we do not test whether $E[y_t|z_{t-1}, z_{t-2}, ...] = f(x_t, \beta^0)$ *a.s.* where $z_t = [y_t, x'_{t+1}]'$ such that the error $\epsilon_t = y_t - f(x_t, \beta^0)$ is a martingale difference under H_0. This rules out testing whether a Nonlinear ARMAX or Nonlinear GARCH model is correctly specified. We can, however, easily extend our main results to allow such tests by mimicking de Jong's (1996: Theorem 2) extension of Bierens' (1990: Lemma 1) main result.

Consistent tests of functional form are widely varied with nonparametric, semiparametric and bootstrap branches. A few contributions not cited above include White (1989a), Chan (1990), Eubank and Spiegelman (1990), Robinson (1991), Yatchew (1992), Härdle and Mammen (1993), Dette (1996), Zheng (1996), Fan and Li (1996, 2000) and Hill (2012). Inherently robust methods include distribution-free tests like indicator or sign-based tests (e.g. Brock et al 1996), the m-out-of-n bootstrap with $m = o(n)$ applied to (2) (Arcones and Gine 1989, Lahiri 1995), and exact small sample tests based on sharp bounds (e.g. Dufour et al 2006, Ibragimov and Müller 2010).

In Section 2 we construct $\hat{I}_{n,t}(\beta)$ and characterize allowed plug-ins. In Section 3 we discuss re-centering after trimming to remove small sample bias that may arise due to trimming. We then construct a p-value occupation time test that allows us to bypass choosing a particular number of extremes to trim and to commit only to a functional form for the sample fractile. Section 4 contains AR and ARCH examples where we present an array of valid plug-ins. In Section 5 we perform a Monte Carlo study and Section 6 contains concluding remarks.

We use the following notation conventions. Let $\Im_t := \sigma(y_\tau, x_{\tau+1} : \tau \leq 1)$, and let M and N be finite integers. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the minimum and maximum eigenvalues of a square matrix $A \in \mathbb{R}^{M \times M}$. The L_p-norm of stochastic $A \in \mathbb{R}^{M \times N}$ is $||A||_p := (\sum_{i=1,j=1}^{M,N} E|A_{i,j}|^p)^{1/p}$, and the spectral norm of $A \in \mathbb{R}^{M \times N}$ is $||A|| = (\lambda_{\max}(A'A))^{1/2}$. For scalar z write $(z)_+ := \max\{0, z\}$, and let $[z]$ be the integer part of z. $K > 0$ is a finite constant and $\iota > 0$ is a tiny constant, the values of which may change from line to line; $L(n)$ is a slowly varying function where $L(n) \to \infty$ as $n \to \infty$, the rate of which may change from line to line.[3] If $\{A_n(\gamma), B_n(\gamma)\}_{n\geq 1}$ are sequences of functions of γ and $\sup_{\gamma\in\Gamma} |A_n(\gamma)/B_n(\gamma)| \to 1$ we write $A_n(\gamma) \sim B_n(\gamma)$ *uniformly on* Γ, and if $\sup_{\gamma\in\Gamma} |A_n(\gamma)/B_n(\gamma)| \stackrel{p}{\to} 1$ we write $A_n(\gamma) \stackrel{p}{\sim} B_n(\gamma)$ *uniformly on* Γ. $\Longrightarrow$ denotes weak convergence on C[Γ], the space of continuous real functions on Γ. The indicator function is $I(a) = 1$ if a is true, and 0 otherwise. A random variable is *symmetric* if its distribution is symmetric about zero.

2 Tail-Weighted Conditional Moment Test

2.1 Tail-Trimmed Equations

Compactly denote the test equation, and the error evaluated at β^0:

$$m_t(\beta,\gamma) := \epsilon_t(\beta)F(\gamma'\psi_t) \text{ and } \epsilon_t = \epsilon_t(\beta^0).$$

By the mean-value-theorem the residuals $\epsilon_t(\hat{\beta}_n)$ reflect the plug-in $\hat{\beta}_n$, the regression error ϵ_t, and the response gradient written variously as

$$g_t(\beta) = [g_{i,t}(\beta)]_{i=1}^q = g(x_t,\beta) := \frac{\partial}{\partial\beta}f(x_t,\beta) \in \mathbb{R}^q.$$

We should therefore trim $\epsilon_t(\beta)$ by setting $\hat{I}_{n,t}(\beta) = 0$ when $\epsilon_t(\beta)$ *or* $g_{i,t}(\beta)$ is an extreme value. This idea is exploited for a class of heavy tail robust M-estimators in Hill (2011b), and similar ideas are developed in Hill and Renault (2010) and Hill and Aguilar (2011).

In the following let $z_t(\beta) \in \{\epsilon_t(\beta), g_{i,t}\beta)\}$, define tail specific observations

$$z_t^{(-)}(\beta) := z_t(\beta)I(z_t(\beta) < 0) \text{ and } z_t^{(+)}(\beta) := z_t(\beta)I(z_t(\beta) \geq 0),$$

and let $z_{(i)}^{(\cdot)}(\beta)$ be the i^{th} sample order statistic of $z_t^{(\cdot)}(\beta)$: $z_{(1)}^{(-)}(\beta) \leq \cdots \leq z_{(n)}^{(-)}(\beta) \leq 0$ and $z_{(1)}^{(+)}(\beta) \geq \cdots \geq z_{(n)}^{(+)}(\beta) \geq 0$. Let $\{k_{j,\epsilon,n} : j = 1,2\}$ and $\{k_{j,i,n} : j = 1,2\}$ be sequences of positive integers taking

[3] Slow variation implies $\lim_{n\to\infty} L(\lambda n)/L(n) = 1$ for any $\lambda > 0$ (e.g. a constant, or $(\ln(n))^a$ for finite $a > 0$: see Resnick 1987). In this paper we always assume $L(n) \to \infty$.

values in $\{1, ..., n\}$, define trimming indicators

$$\hat{I}_{\epsilon,n,t}(\beta) := I\left(\epsilon^{(-)}_{(k_{1,\epsilon,n})}(\beta) \leq \epsilon_t(\beta) \leq \epsilon^{(+)}_{(k_{2,\epsilon,n})}(\beta)\right)$$

$$\hat{I}_{i,n,t}(\beta) := I\left(g^{(-)}_{i,(k_{1,i,n})}(\beta) \leq g_{i,t}(\beta) \leq g^{(+)}_{i,(k_{2,i,n})}(\beta)\right) \text{ and } \hat{I}_{g,n,t}(\beta) := \prod_{i=1}^{q} \hat{I}_{i,n,t}(\beta)$$

$$\hat{I}_{n,t}(\beta) := \hat{I}_{\epsilon,n,t}(\beta) \times \hat{I}_{g,n,t}(\beta),$$

and trimmed test equations

$$\hat{m}^*_{n,t}(\beta,\gamma) := m_t(\beta,\gamma) \times \hat{I}_{n,t}(\beta) = \epsilon_t(\beta) \times \hat{I}_{n,t}(\beta) \times F(\gamma'\psi_t).$$

Thus $\hat{I}_{n,t}(\beta) = 0$ when any $\epsilon_t(\beta)$ or $g_{i,t}(\beta)$ is large. Together with some plug-in $\hat{\beta}_n$ and HAC estimator $\hat{S}^2_n(\hat{\beta}_n,\gamma)$ we obtain our test statistic $\hat{\mathcal{T}}_n(\gamma) = \hat{S}^{-2}_n(\hat{\beta}_n,\gamma)(\sum_{t=1}^n \hat{m}^*_{n,t}(\hat{\beta}_n,\gamma))^2$.

We determine how many observations of $\epsilon_t(\beta)$ and $g_{i,t}(\beta)$ are extreme values by assuming $\{k_{j,\epsilon,n}\}$ and $\{k_{j,i,n}\}$ are *intermediate order sequences*. If $\{k_{j,z,n}\}$ denotes any one of them, then

$$1 \leq k_{1,z,n} + k_{2,z,n} < n, \quad k_{j,z,n} \to \infty \text{ and } k_{j,z,n}/n \to 0.$$

The fractile $k_{j,z,n}$ represents the number of $m_t(\beta,\gamma)$ trimmed due to a large left- or right-tailed $\epsilon_t(\beta)$ or $g_{i,t}(\beta)$. Since we trim asymptotically infinitely many large values $k_{j,z,n} \to \infty$ we ensure Gaussian asymptotics, while trimming a vanishing sample portion $k_{j,z,n}/n \to 0$ promotes identification of H_0 and H_1.[4] The reader may consult Leadbetter et al (1983: Chapter 2), Hahn et al (1991) and Hill (2011a) for the use of intermediate order statistics in extreme value theory and robust estimation. See Section 3 for details on handling the fractiles $k_{j,z,n}$.

If any z_t is symmetric then symmetric trimming is used:

$$I\left(|z_t(\beta)| \leq z^{(a)}_{(k_{z,n})}(\beta)\right) \text{ where } z^{(a)}_t := |z_t|, \quad k_{z,n} \to \infty \text{ and } k_{z,n}/n \to 0. \tag{3}$$

If a component takes on only one sign then one-sided trimming is appropriate, and if $z_t(\beta)$ has a finite variance then it can be dropped from $\hat{I}_{n,t}(\beta)$. In general tail thickness does not need to be

[4] Consider if ϵ_t is iid and asymmetric under H_0, but symmetrically and non-negligibly trimmed $k_{1,\epsilon,n} = k_{2,\epsilon,n} \sim \lambda n$ where $\lambda \in (0,1)$. Then $\hat{T}_n(\gamma) \xrightarrow{p} \infty$ under H_0 is easily verified. The test statistic reveals mis-specification due entirely to trimming itself.

known because our statistic has the same asymptotic properties for thin or thick tailed data, while unnecessary tail trimming is both irrelevant in theory, and does not appear to affect the test in small samples.

2.2 Plug-In Properties

The plug-in $\hat{\beta}_n$ needs to be consistent for a unique point $\beta^0 \in \mathcal{B}$.[5] In particular, we assume there exists a sequence of positive definite matrices $\{\tilde{V}_n\}$, where $\tilde{V}_n \in \mathbb{R}^{q \times q}$ and $\tilde{V}_{i,i,n} \to \infty$, and

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

As we discuss below, in the presence of heavy tails $\hat{\beta}_n$ need not have $n^{1/2}$-convergent components, and depending on the model may have components with different rates $\tilde{V}_{i,i,n}^{1/2}$ below, at or above $n^{1/2}$.

Precisely how fast convergence $\hat{\beta}_n \xrightarrow{p} \beta^0$ is gauged by exploiting an asymptotic expansion of $\hat{S}_n^{-1}(\hat{\beta}_n, \gamma) \sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\beta}_n, \gamma)$ around β^0. We therefore require the non-random quantile sequences which the order statistics $\epsilon_{(k_{j,\epsilon,n})}^{(\cdot)}(\beta)$ and $g_{i,(k_{j,i,n})}^{(\cdot)}(\beta)$ approach asymptotically. The sequences are positive functions $\{l_{z,n}(\beta), u_{z,n}(\beta)\}$ denoting the lower $k_{1,z,n}/n^{th}$ and upper $k_{2,z,n}/n^{th}$ quantiles of $z_t(\beta)$ in the sense

$$P\left(z_t(\beta) < -l_{z,n}(\beta)\right) = \frac{k_{1,z,n}}{n} \text{ and } P\left(z_t(\beta) > u_{z,n}(\beta)\right) = \frac{k_{2,z,n}}{n}. \tag{4}$$

Distribution smoothness for $\epsilon_t(\beta)$ and $g_{i,t}(\beta)$ ensures $\{l_{z,n}(\beta), u_{z,n}(\beta)\}$ exist for all β and any chosen fractile policy $\{k_{1,z,n}, k_{2,z,n}\}$. See Appendix A for all assumptions. By construction $\{z_{(k_{1,z,n})}^{(-)}(\beta)$, $z_{(k_{2,z,n})}^{(+)}(\beta)\}$ estimate $\{-l_{z,n}(\beta), u_{z,n}(\beta)\}$ and are uniformly consistent, e.g. $\sup_{\beta \in \mathcal{B}} |z_{(k_{2,z,n})}^{(+)}(\beta)/u_{z,n}(\beta) - 1| = O_p(1/k_{1,z,n}^{1/2})$. See Hill (2011b: Lemma C.2).

Now construct indicators and a trimmed test equation used solely for asymptotics: in general write $I_{z,n,t}(\beta) := I(-l_{z,n}(\beta) \leq z_t(\beta) \leq u_{z,n}(\beta))$, and define

$$I_{n,t}(\beta) := I_{\epsilon,n,t}(\beta) \times \prod_{i=1}^{q} I_{i,n,t}(\beta) = I_{\epsilon,n,t}(\beta) \times I_{g,n,t}(\beta) \text{ and } m_{n,t}^*(\beta, \gamma) := m_t(\beta, \gamma) \times I_{n,t}(\beta).$$

[5] Under the alternative β^0 is the unique probability limit of $\hat{\beta}_n$, a "quasi-true" point that optimizes a discrepancy function, for example a likelihood function, method of moments criterion or the Kullback-Leibler Information Criterion. See White (1982) amongst many others.

We also require covariance, Jacobian and scale matrices:

$$S_n^2(\beta,\gamma) := E\left(\sum_{t=1}^n \left\{m_{n,t}^*(\beta,\gamma) - E[m_{n,t}^*(\beta,\gamma)]\right\}\right)^2 \text{ and } J_n(\beta,\gamma) := \frac{\partial}{\partial\beta}E\left[m_{n,t}^*(\beta,\gamma)\right] \in \mathbb{R}^{q\times 1}$$
$$V_n(\beta,\gamma) := n^2 S_n^{-2}(\beta,\gamma) \times J_n(\beta,\gamma)' J_n(\beta,\gamma) \in \mathbb{R}.$$

Now drop β^0 throughout, e.g. $g_t = g_t(\beta^0)$, $m_{n,t}^*(\gamma) = m_{n,t}^*(\beta^0,\gamma)$ and $S_n^2(\gamma) = S_n^2(\beta^0,\gamma)$. We may work with $m_{n,t}^*(\gamma)$ for asymptotic theory purposes since

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{S_n(\gamma)}\sum_{t=1}^n \left\{\hat{m}_{n,t}^*(\gamma) - m_{n,t}^*(\gamma)\right\}\right| = o_p(1),$$

while trimming negligibility and response function smoothness ensure the following expansion:

$$\frac{1}{\hat{S}_n(\hat{\beta}_n,\gamma)}\sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\beta}_n,\gamma) \overset{p}{\sim} \frac{1}{S_n(\gamma)}\sum_{t=1}^n m_{n,t}^*(\gamma) + V_n^{1/2}(\gamma)\left(\hat{\beta}_n - \beta^0\right). \tag{5}$$

See Lemmas B.2 and B.3 in Appendix A. Thus $\hat{\mathcal{T}}_n(\gamma)$ tests H_0 if $\hat{\beta}_n \xrightarrow{p} \beta^0$ fast enough in the sense $\sup_{\gamma\in\Gamma}||V_n(\gamma)\tilde{V}_n^{-1}|| = O(1)$. In the following we detail three plug-in cases denoted P1, P2 and P3.

Case P1 (fast (non)linear plug-ins): In this case $\sup_{\gamma\in\Gamma}||V_n(\gamma)\tilde{V}_n^{-1}|| \to 0$ hence $\hat{\beta}_n$ does not impact $\hat{\mathcal{T}}_n(\gamma)$ asymptotically, which is evidently only possible if ϵ_t and/or $g_{i,t}$ are heavy tailed. If $\{\epsilon_t, g_t\}$ are sufficiently thin tailed then under regularity conditions minimum distance estimators $\hat{\beta}_n$ are $n^{1/2}$-convergent while $V_n(\gamma)/n \to V(\gamma) = S^{-2}(\gamma)J(\gamma)'J(\gamma)$ is finite for each $\gamma \in \Gamma$.[6] In the presence of heavy tails, however, a unique advantage exists since $\sup_{\gamma\in\Gamma}||V_n^{1/2}(\gamma)|| = o(n^{1/2})$ may hold allowing many plug-ins to satisfy $\sup_{\gamma\in\Gamma}||V_n(\gamma)\tilde{V}_n^{-1}|| \to 0$. See Section 4 for examples.

Case P2 (slow linear plug-ins): If $\tilde{V}_n$ is proportional to $V_n(\gamma)$ then $\hat{\beta}_n$ impacts $\hat{\mathcal{T}}_n(\gamma)$ asymptotically. This is the case predominantly encountered in the literature since $\tilde{V}_n/n \to \tilde{V}$ and $V_n(\gamma)/n \to V(\gamma)$ for sufficiently thin tailed $\{\epsilon_t, g_t\}$. At least two solutions exist. First, under the present case $\hat{\beta}_n$ is assumed to be asymptotically linear and normal, covering many minimum discrepancy estimators when $\{\epsilon_t, g_t\}$ are sufficiently thin tailed, or heavy tail robust linear estimators like Quasi-Maximum Tail-Trimmed Likelihood (Hill 2011b). Linearity rules out quantile estimators like LAD and its variants, including Log-LAD for GARCH models with heavy tailed errors (Peng and Yao 2003) and Least Absolute Weighted Deviations for heavy tailed autoregressions (Ling 2005).

[6] The rate of convergence for some minimum discrepancy estimators may be below $n^{1/2}$, even for thin tailed data, in contexts involving weak identification, kernel smoothing and in-fill asymptotics. We implicitly ignored such cases here.

Case P3 ((non)linear plug-ins for orthogonal equations): If $\tilde{V}_n$ is proportional to $V_n(\gamma)$ then our second solution is to exploit Wooldridge's (1990) orthogonal transformation for a new test statistic, ensuring plug-in robustness and allowing nonlinear plug-ins. Other projection techniques are also evidently valid (e.g. Bai 2003).

Define a projection operator $\hat{\mathcal{P}}_{n,t}(\gamma)$ and filtered equations $\hat{m}^{\perp}_{n,t}(\beta,\gamma)$:

$$\hat{\mathcal{P}}_{n,t}(\gamma) = 1 - g_t'(\hat{\beta}_n)\hat{I}_{n,t}(\hat{\beta}_n)\left(\frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\beta}_n)g_t'(\hat{\beta}_n)F(\gamma'\psi_t)\hat{I}_{n,t}(\hat{\beta}_n)\right)^{-1} \times \frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\beta}_n)F(\gamma'\psi_t)\hat{I}_{n,t}(\hat{\beta}_n)$$
$$\hat{m}^{\perp}_{n,t}(\beta,\gamma) = \hat{m}^{*}_{n,t}(\beta,\gamma) \times \hat{\mathcal{P}}_{n,t}(\gamma).$$

The test statistic is now

$$\hat{\mathcal{T}}^{\perp}_n(\gamma) = \frac{1}{\hat{S}^{\perp 2}_n(\hat{\beta}_n,\gamma)}\left(\sum_{t=1}^{n}\hat{m}^{\perp}_{n,t}(\hat{\beta}_n,\gamma)\right)^2,$$

where $\hat{S}^{\perp 2}_n(\beta,\gamma)$ is identically $\hat{S}^2_n(\beta,\gamma)$ computed with $\hat{m}^{\perp}_{n,t}(\beta,\gamma)$.

The asymptotic impact of $\hat{\beta}_n$ is again gauged by using the non-random thresholds $\{l_{z,n}, u_{z,n}\}$ to construct orthogonal equations and their variance and Jacobian:

$$\mathcal{P}_{n,t}(\gamma) := 1 - g_t' I_{n,t}\left(E\left[g_t g_t' F(\gamma'\psi_t) I_{n,t}\right]\right)^{-1} \times E\left[g_t F(\gamma'\psi_t) I_{n,t}\right] \text{ and } m^{\perp}_{n,t}(\beta,\gamma) = m^{*}_{n,t}(\beta,\gamma) \times \mathcal{P}_{n,t}(\gamma)$$
$$S^{\perp 2}_n(\beta,\gamma) := E\left(\sum_{t=1}^{n}\left\{m^{\perp}_{n,t}(\beta,\gamma) - E[m^{\perp}_{n,t}(\beta,\gamma)]\right\}\right)^2 \text{ and } J^{\perp}_n(\beta,\gamma) := \frac{\partial}{\partial\beta}E\left[m^{\perp}_{n,t}(\beta,\gamma)\right] \in \mathbb{R}^{q\times 1}$$
$$V^{\perp}_n(\beta,\gamma) := n^2 S^{\perp -2}_n(\beta,\gamma) \times J^{\perp}_n(\beta,\gamma)' J^{\perp}_n(\beta,\gamma) \in \mathbb{R}.$$

Notice $\mathcal{P}_{n,t}(\gamma)$ is $\sigma(x_t)$-measurable, and uniformly L_1-bounded by Lyapunov's inequality and boundedness of $F(u)$, thus by dominated convergence $E[m^{\perp}_{n,t}(\gamma)] \to 0$ under H_0. By imitating expansion (5) and arguments in Wooldridge (1990), it can easily be shown if $V^{\perp}_n(\gamma)^{1/2}(\hat{\beta}_n - \beta^0) = O_p(1)$ then $\hat{S}^{\perp -1}_n(\hat{\beta}_n,\gamma)\sum_{t=1}^{n}\hat{m}^{\perp}_{n,t}(\hat{\beta}_n,\gamma) \overset{p}{\sim} S^{\perp -1}_n(\gamma)\sum_{t=1}^{n} m^{\perp}_{n,t}(\gamma)$. In general the new statistic $\hat{\mathcal{T}}^{\perp}_n(\gamma)$ is robust to $\hat{\beta}_n$, allowing non-linear estimators, as long as

$$\tilde{V}_n^{1/2}\left(\hat{\beta}_n - \beta^0\right) = O_p(1) \text{ and } \limsup_{n\to\infty}\sup_{\gamma\in\Gamma}\left\|V^{\perp}_n(\gamma)\tilde{V}_n^{-1}\right\| < \infty. \tag{6}$$

2.3 Main Results

Appendix A contains all assumptions concerning the fractiles and non-degeneracy of trimmed moments (F1-F2); identification of the null (I1); the kernel and bandwidth for the HAC estimator (K1);

the plug-in (P1-P3); moments and memory of regression components (R1-R4); and the test weight (W1). We state the main results for both $\hat{\mathcal{T}}_n(\gamma)$ and $\hat{\mathcal{T}}_n^{\perp}(\gamma)$, but for the sake of brevity limit discussions to $\hat{\mathcal{T}}_n(\gamma)$. Throughout Γ is a compact subset of $\mathbb{R}^p$ with positive Lebesgue measure.

Our first result shows tail-trimming does not impact the ability of $F(\gamma'\psi_t)$ to reveal mis-specification.

LEMMA 2.1. *Let $\mu_{n,t}(\gamma)$ denote either $m_{n,t}^*(\gamma)$ or $m_{n,t}^{\perp}(\gamma)$. Under the null $E[\mu_{n,t}(\gamma)] \rightarrow 0$. Further, if test weight property W1 and the alternative H_1 hold then $\liminf_{n\rightarrow\infty} |E[\mu_{n,t}(\gamma)]| > 0$ for all $\gamma \in \Gamma$ except possibly on a set $S \subset \Gamma$ with Lebesgue measure zero.*

Remark: Under H_1 it is possible in small samples for $E[m_{n,t}^*(\gamma)] = 0$ due to excessive trimming, and $|E[m_{n,t}^*(\gamma)]| \rightarrow \infty$ due to heavy tails. The test weight $F(u)$ therefore is still revealing under tail-trimming for *sufficiently large* n.

Next, the test statistics converge to chi-squared processes under H_0 and are consistent. Plug-in cases P1-P3 are discussed in Section 2.2.

THEOREM 2.2. *Let F1-F2, I1, K1, R1-R4 and W1 hold.*

i. Under H_0 and plug-in cases P1 or P2 there exists a Gaussian process $\{z(\gamma) : \gamma \in \Gamma\}$ on $\mathcal{C}[\Gamma]$ with zero mean, unit variance and covariance function $E[z(\gamma_1)z(\gamma_2)]$ such that $\{\hat{\mathcal{T}}_n(\gamma) : \gamma \in \Gamma\} \Longrightarrow \{z(\gamma)^2 : \gamma \in \Gamma\}$.

ii. Under H_1 and P1 or P2, $\hat{\mathcal{T}}_n(\gamma) \xrightarrow{p} \infty$ $\forall \gamma \in \Gamma/S$ where S has Lebesgue measure zero.

iii. Under plug-in case P3 $\hat{\mathcal{T}}_n^{\perp}(\gamma)$ satisfies cases (i) and (ii).

Remark 1: The literature offers a variety of ways to handle the nuisance parameter γ. Popular choices include randomly selecting $\gamma^* \in \Gamma$ (e.g. Lee et al 1993), or computing a continuous test functional $h(\hat{\mathcal{T}}_n(\gamma))$ like the supremum $\sup_{\gamma\in\Gamma} \hat{\mathcal{T}}_n(\gamma)$ and average $\int_\Gamma \hat{\mathcal{T}}_n(\gamma)\mu(d\gamma)$, where $\mu(\gamma)$ is a continuous measure (Davies 1977, Bierens 1990). In the latter case $h(\hat{\mathcal{T}}_n(\gamma)) \xrightarrow{d} h(z(\gamma)^2) =: h_0$ under H_0 by the mapping theorem.

Hansen's (1996) bootstrapped p-value for non-standard h_0 exploits an iid Gaussian simulator. The method therefore applies only if ϵ_t is a martingale difference under H_0 and the trimmed error $\epsilon_t I_{\epsilon,n,t}$ becomes a martingale difference *sufficiently fast* in the sense $(n/E[m_{n,t}^{*2}(\gamma)])^{1/2} E[\epsilon_t I_{\epsilon,n,t}|\Im_{t-1}] \rightarrow 0$. It therefore suffices for ϵ_t to be iid and symmetric under H_0 and symmetrically trimmed since then $E[\epsilon_t I_{\epsilon,n,t}|\Im_{t-1}] = E[\epsilon_t I_{\epsilon,n,t}] = 0$, or if ϵ_t is asymmetric and $E[\epsilon_t] = 0$ under either hypothesis then ϵ_t can be symmetrically trimmed with re-centering as in Section 3, below. See Hill (2011c: Section C.1), the supplemental appendix to this paper, for details on Hansen's p-value under tail-trimming.

Remark 2: As long as $S_n^2(\gamma) = E[m_{n,t}^*(\gamma)m_{n,t}^*(\gamma)'] \times (1 + o(1))$ then a HAC estimator is not required, including when $\epsilon_t I_{\epsilon,n,t}$ becomes a martingale difference sufficiently fast under H_0 as above. If we do not use a plug-in robust equation then an estimator $\hat{S}_n^2(\hat{\beta}_n, \gamma)$ must control for sampling error associated with $\hat{\beta}_n$. For example, if $\hat{\beta}_n$ is the NLLS estimator then (e.g. Bierens 1990: eq. (14))

$$\hat{S}_n^2(\hat{\beta}_n, \gamma) = \sum_{t=1}^{n} \epsilon_t^2(\hat{\beta}_n)\hat{I}_{n,t}(\hat{\beta}_n) \times \left\{F(\gamma'\psi_t) - \hat{b}_n'\hat{A}_n^{-1}\hat{g}_{n,t}^*(\hat{\beta}_n)\right\}^2, \tag{7}$$

where $\hat{g}_{n,t}^*(\beta) := g_t(\beta)\hat{I}_{g,n,t}(\beta)$, $\hat{b}_n := 1/n\sum_{t=1}^n \hat{g}_{n,t}^*(\hat{\beta}_n)F(\gamma'\psi_t)$ and $\hat{A}_n := 1/n\sum_{t=1}^n \hat{g}_{n,t}^*(\beta)\hat{g}_{n,t}^*(\beta)'$. However, if $S_n^{\perp 2}(\gamma) \sim E[m_{n,t}^{\perp}(\gamma)m_{n,t}^{\perp}(\gamma)']$ then by orthogonality we need only use

$$\hat{S}_n^{\perp 2}(\hat{\beta}_n, \gamma) = \sum_{t=1}^{n} \hat{m}_{n,t}^{\perp}(\hat{\beta}_n, \gamma)\hat{m}_{n,t}^{\perp}(\hat{\beta}_n, \gamma)'. \tag{8}$$

3 Fractile Choice

We must choose how much to trim $k_{j,z,n}$ for each $z_t \in \{\epsilon_t, g_{i,t}\}$ and any given n. We first present a case when symmetric trimming with re-centering is valid even when ϵ_t is asymmetric under H_0. We then discuss an empirical process method that smooths over a class of fractiles.

Symmetric Trimming with Re-Centering If $E[\epsilon_t] = 0$ even under the alternative, and ϵ_t is independent of x_t under H_0, then we may symmetrically trim for simplicity and re-center to eradicate bias that arises due to trimming, and still achieve a consistent test statistic. The test equation is

$$\hat{m}_{n,t}^*(\beta, \gamma) = \left(\epsilon_t(\beta)\hat{I}_{n,t}(\beta) - \frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\beta)\hat{I}_{n,t}(\beta)\right) \times F(\gamma'\psi_t) \tag{9}$$

where $\hat{I}_{n,t}(\beta) = \hat{I}_{\epsilon,n,t}(\beta)\prod_{i=1}^q \hat{I}_{i,n,t}(\beta)$ as before, with symmetric trimming indicators $\hat{I}_{\epsilon,n,t}(\beta) := I(|\epsilon_t(\beta)| \leq \epsilon_{(k_{\epsilon,n})}^{(a)}(\beta))$ and $\hat{I}_{i,n,t}(\beta) := I(|g_{i,t}(\beta)| \leq g_{i,(k_{i,n})}^{(a)}(\beta))$. By independence $m_{n,t}^*(\beta, \gamma) = (\epsilon_t(\beta)I_{n,t}(\beta) - E[\epsilon_t(\beta)I_{n,t}(\beta)]) \times F(\gamma'\psi_t)$ satisfies $E[m_{n,t}^*(\gamma)] = 0$ under H_0 for any $\{k_{\epsilon,n}, k_{i,n}\}$, hence identification I1 is trivially satisfied. Under H_1 the weight $F(u)$ is revealing by Lemma 2.1 since $E[\epsilon_t] = 0$, $F(u)$ is bounded, and trimming is negligible: $\liminf_{n\to\infty} |E[m_{n,t}^*(\gamma)]| = \liminf_{n\to\infty} |E[\epsilon_t I_{n,t} F(\gamma'\psi_t)]| > 0\ \forall\gamma \in \Gamma/S$. A test of linear AR where the errors may be governed by a nonlinear GARCH process, or a test of linear ARCH, provide natural platforms for re-centering. See Section 4 for ARCH.

The moment condition $E[\epsilon_t] = 0$ under either hypothesis rules out some response functions depending on the tails of $\{y_t, x_t\}$. See Section 1 for an example.

P-Value Occupation Time Assume symmetric trimming to reduce notation and define the error moment supremum $\kappa_\epsilon := \arg\sup\{\alpha > 0 : E|\epsilon_t|^\alpha < \infty\}$. Under H_0 any intermediate order sequences $\{k_{\epsilon,n}, k_{i,n}\}$ are valid, but in order for our test to work under H_1 when ϵ_t may be exceptionally heavy tailed $\kappa_\epsilon < 1$, we must impose $k_{\epsilon,n}/n^{2(1-\kappa_\epsilon)/(2-\kappa_\epsilon)} \to \infty$ to ensure sufficient trimming for test consistency (see Assumption F1.b in Appendix A). Thus $k_{\epsilon,n} \sim n/L(n)$ is valid for any slowly varying $L(n) \to \infty$. Consider $k_{\epsilon,n} = k_{i,n} \sim \lambda n/\ln(n)$ where λ is taken from a compact set $\Lambda := [\underline{\lambda}, 1]$ for tiny $\underline{\lambda} > 0$, although any slowly varying $L(n) \to \infty$ may replace $\ln(n)$. The point $\lambda = 0$ is ruled out because the untrimmed $\hat{\mathcal{T}}_n(0)$ is asymptotically non-chi-squared under H_0 when $E[\epsilon_t^2] = \infty$.

We must now commit to some λ. Other than an arbitrary choice, Hill and Aguilar (2011) smooth over a space of feasible $\lambda' s$ by computing p-value occupation time. We construct the occupation time below, and prove its validity for $\hat{\mathcal{T}}_n(\gamma)$ and $\hat{\mathcal{T}}_n^{\perp}(\gamma)$ in Appendix B. The following easily extends to $k_{\epsilon,n} \neq k_{i,n}$, asymmetric trimming, and functionals $h(\hat{\mathcal{T}}_n(\gamma))$ on Γ.

Write $\hat{\mathcal{T}}_n(\gamma, \lambda)$ and $\hat{\mathcal{T}}_n^{\perp}(\gamma, \lambda)$ to reveal dependence on λ, let $p_n(\gamma, \lambda)$ denote the asymptotic p-value $1 - F_\chi(\hat{\mathcal{T}}_n(\gamma, \lambda))$ where F_χ is the $\chi^2(1)$ distribution, and define the α-level occupation time

$$\tau_n(\gamma, \alpha) := \frac{1}{1-\underline{\lambda}} \int_{\underline{\lambda}}^{1} I\left(p_n(\gamma, \lambda) < \alpha\right) d\lambda \in [0,1]\,, \quad \text{where } \alpha \in (0,1).$$

Thus $\tau_n(\gamma, \alpha)$ is the proportion of $\lambda' s$ satisfying $p_n(\gamma, \lambda) < \alpha$ hence rejection of H_0 at level α. Similarly, define the occupation time $\tau_n^{\perp}(\gamma, \alpha)$ for $\hat{\mathcal{T}}_n^{\perp}(\gamma, \lambda)$.

THEOREM 3.1 *Let F1-F2, I1, K1, P1 or P2, R1-R4 and W1 hold. Let $\{u(\lambda) : \lambda \in \Lambda\}$ be a stochastic process that may be different in different places: in each case it has a version that has uniformly continuous sample paths, and $u(\lambda)$ is uniformly distributed on $[0,1]$. Under the null $\tau_n(\gamma, \alpha) \overset{d}{\to} (1-\underline{\lambda})^{-1} \int_{\underline{\lambda}}^{1} I(u(\lambda) < \alpha) d\lambda$ and $\tau_n^{\perp}(\gamma, \alpha) \overset{d}{\to} (1-\underline{\lambda})^{-1} \int_{\underline{\lambda}}^{1} I(u(\lambda) < \alpha) d\lambda$, and under the alternative $\tau_n(\gamma, \alpha) \overset{p}{\to} 1$ and $\tau_n^{\perp}(\gamma, \alpha) \overset{p}{\to} 1$ $\forall \gamma \in \Gamma$ except possibly on subsets with measure zero.*

Remark 1: Since $u(\lambda)$ is a uniform random variable it follows $\lim_{n\to\infty} P(\tau_n(\gamma, \alpha) > \alpha | H_0) < \alpha$. A p-value occupation test therefore rejects H_0 at level α if $\tau_n(\gamma, \alpha) > \alpha$. In practice a discretized version is computed, for example

$$\hat{\tau}_n(\gamma, \alpha) := \frac{1}{n_{\underline{\lambda}}} \sum_{i=1}^{n} I\left(p_n(\gamma, i/n) < \alpha\right) \times I\left(i/n \geq \underline{\lambda}\right) \tag{10}$$

where $n_{\underline{\lambda}} := \sum_{i=1}^{n} I(i/n \geq \underline{\lambda})$ is the number of discretized points in $[\underline{\lambda}, 1]$.

Remark 2: In Section 4 we show $\hat{\beta}_n$ has a larger impact on $\hat{\mathcal{T}}_n(\gamma,\lambda)$ in small samples when the error has an infinite variance $\kappa_\epsilon < 2$, each $g_{i,t}$ has a finite mean $\kappa_i > 1$, and the number of trimmed errors $k_{\epsilon,n}$ is large (see Remark 3 of Lemma 4.1). This translates to the possibility of plug-in sensitivity of $\tau_n(\gamma,\alpha)$ in small samples. We show in our Monte Carlo study of Section 5 that when $\kappa_\epsilon < 2$ and $\kappa_i > 1$ the occupation time $\tau_n(\gamma,\alpha)$ results in size distortions that are eradicated when the plug-in robust $\hat{\tau}_n^\perp(\gamma,\alpha)$ is used.

In Figure 1 we plot sample paths $\{p_n(\gamma,\lambda), p_n^\perp(\gamma,\lambda) : \lambda \in [.01, 1.0]\}$ based on two samples $\{y_t\}_{t=1}^n$ of size $n = 200$: one sample is drawn from an AR(1) process and the other from a Threshold AR(1) process, each with iid Pareto errors ϵ_t and tail index 1.5. See Section 5 for simulation details. We estimate an AR(5) model by OLS, compute $\hat{\mathcal{T}}_n(\gamma,\lambda)$ and $\hat{\mathcal{T}}_n^\perp(\gamma,\lambda)$ with weight $F(\gamma'\psi(x_t)) = \exp\{\gamma'\psi(x_t)\}$, $\psi(x_t) = [1, \arctan(\tilde{x}_t^*)']'$ where $\tilde{x}_t^*$ is centered $\tilde{x}_t = [y_{t-1}, ..., y_{t-5}]'$, and uniformly randomize γ on $[.1, 2]^6$. In this case at the 5% level $\hat{\tau}_n, \hat{\tau}_n^\perp = 0, 0$ for the AR sample hence we fail to reject H_0, and $\hat{\tau}_n, \hat{\tau}_n^\perp = .59, 1.0$ for the SETAR sample hence we reject H_0.

Notice in the AR case $p_n(\gamma,\lambda)$ is smallest for large $\lambda \geq .9$, and $p_n(\gamma,\lambda) < p_n^\perp(\gamma,\lambda)$ for most λ: $p_n(\gamma,\lambda)$ is more likely to lead to a rejection than the plug-in robust $p_n^\perp(\gamma,\lambda)$ and for large λ. Although we only use one AR sample here, in Section 5 we show plug-in sensitivity does indeed lead to over-rejection of H_0.

[Insert Figure 1 about here]

4 Plug-In Choice and Verification of the Assumptions

We first characterize $V_n(\gamma)$ to show how fast $\hat{\beta}_n$ in $\hat{\mathcal{T}}_n(\gamma)$ must be in view of expansion (5). Synonymous derivations carry over to portray $V_n^\perp(\gamma)$. We then verify the assumptions for AR and ARCH models and several plug-in estimators. Define moment suprema $\kappa_\epsilon := \arg\sup\{\alpha > 0 : E|\epsilon_t|^\alpha < \infty\}$ and $\kappa_i := \arg\sup\{\alpha > 0 : E|g_{i,t}|^\alpha < \infty\}$.

LEMMA 4.1. *Let F1-F2, I1, R1-R4 and W1 hold. If* $\kappa_i \leq 1$ *then assume* $P(|g_{i,t}| > g) = d_i g^{\kappa_i}(1 + o(1))$ *for some* $d_i > 0$*. Let* $L(n) \to \infty$ *be slowly varying, and let* $\{\mathfrak{L}_n\}$ *be a sequence of positive constants:* $\liminf_{n\to\infty} \mathfrak{L}_n \geq 1$ *and* $\mathfrak{L}_n = O(\ln(n))$*, and if* ϵ_t *is finite dependent then* $\mathfrak{L}_n = K$*. In this following* $L(n)$ *and* $\mathfrak{L}_n$ *may be different in different places.*

i. Let $\min\{\kappa_i\} > 1$*. If* $\kappa_\epsilon > 2$ *then* $V_n(\gamma) = O(n)$*; if* $\kappa_\epsilon = 2$ *then* $V_n(\gamma) \sim n/L(n)$*; and if* $\kappa_\epsilon < 2$ *then* $V_n(\gamma) \sim Kn(k_{\epsilon,n}/n)^{2/\kappa_\epsilon - 1}/\mathfrak{L}_n$.

ii. Let some $\kappa_i < 1$. *If* $\kappa_\epsilon > 2$ *then* $V_n(\gamma) \sim Kn\max_{i:\kappa_i<1}\{(n/k_{i,n})^{2/\kappa_i-2}\}$; *if* $\kappa_\epsilon = 2$ *then* $V_n(\gamma) \sim Kn\max_{i:\kappa_i<1}\{(n/k_{i,n})^{2/\kappa_i-2}\}/L(n)$; *and if* $\kappa_\epsilon < 2$ *then* $V_n(\gamma) \sim Kn\max_{i:\kappa_i<1}\{(n/k_{i,n})^{2/\kappa_i-2}\} \times (k_{\epsilon,n}/n)^{2/\kappa_\epsilon-1}/\mathfrak{L}_n$.

iii. If $\min\{\kappa_i\} = 1$ *then replace* $\max_{i:\kappa_i<1}\{(n/k_{i,n})^{2/\kappa_i-2}\}$ *in (b) with* $L(n)$.

Remark 1: The term $\mathfrak{L}_n$ arises due to β-mixing dependence *and* heavy tails. Clearly $S_n^2(\gamma) \sim KnE[m_{n,t}^2]$ if ϵ_t is finite dependent or has a finite variance, but otherwise we can only show $S_n^2(\gamma) \sim nE[m_{n,t}^2] \times O(\ln(n))$, cf. Hill (2011b: Lemma B.2).

Remark 2: If $E[\epsilon_t^2] = \infty$ then $V_n(\gamma) = o(n)$ as long as all $\kappa_i > 1$, hence $\hat{\beta}_n$ may be sub-$n^{1/2}$-convergent. This arises, for example, in integrable AR models or ARCH models with square integrable errors as we verify below.

Remark 3: If $\kappa_\epsilon < 2$ and each $\kappa_i > 1$ then $V_n(\gamma) \sim Kn(k_{\epsilon,n}/n)^{2/\kappa_\epsilon-1}/\mathfrak{L}_n$. Combine this with expansion (5) to deduce a higher error trimming rate $k_{\epsilon,n} \to \infty$ *amplifies* the impact of $\hat{\beta}_n$ on the test statistic $\hat{\mathcal{T}}_n(\gamma)$ in small samples, even when fast plug-in Assumption P1 holds. This suggests the plug-in robust statistic $\hat{\mathcal{T}}_n^{\perp}(\gamma)$ should be used when $k_{\epsilon,n}$ is chosen to be large relative to n. This is supported by experiments in Section 5 where the p-value occupation which smooths over small and large $k_{\epsilon,n}$ performs substantially better when $\hat{\mathcal{T}}_n^{\perp}(\gamma)$ is used.

4.1 Linear AR

Consider a stationary AR(p) $y_t = \beta^{0\prime}x_t + \epsilon_t$ where $x_t = [y_{t-1}, ..., y_{t-p}]'$, ϵ_t is iid and $E[\epsilon_t] = 0$. Assume ϵ_t has an absolutely continuous symmetric distribution with a uniformly bounded density $\sup_{c\in\mathbb{R}} |(\partial/\partial c)P(\epsilon_t \leq c)| < \infty$, and Paretian tail:

$$P(|\epsilon_t| > \epsilon) = d\epsilon^{-\kappa}(1 + o(1)), \; d > 0, \; \kappa > 1. \tag{11}$$

Since y_t is symmetric with a power law tail and the same index κ (Brockwell and Cline 1985), and $g_{i,t} = y_{t-i}$, we use symmetric trimming (3) with common fractiles $k_{\epsilon,n} = k_{y,n}$ denoted k_n. Let $\hat{\beta}_n$ be computed by OLS, LAD, Lease Weighted Absolute Deviations [LWAD] by Ling (2005), Least Tail-Trimmed Squares [LTTS] by Hill (2011b), or Generalized Method of Tail-Trimmed Moments [GMTTM] by Hill and Renault (2010) with estimating equations $[\epsilon_t(\beta)y_{t-i}]_{i=1}^r$ for some $r \geq p$.[7]

[7] Other over-identifying restrictions can easily be included, but the GMTTM rate my differ from what we cite in the proof of Lemma 4.2 if they are not lags of y_t. See Hill and Renault (2010).

LEMMA 4.2. *Assumptions F2, I1, and R1-R4 hold. If* $\kappa < 2$ *then* $V_n(\gamma) \sim Kn(k_n/n)^{2/\kappa-1}$ *and if then* $V_n(\gamma) \sim Kn/L(n)$ *uniformly on* Γ. *Therefore each* $\hat{\beta}_n$ *satisfies P1 and P3 if* $E[\epsilon_t^2] = \infty$ *and P3 if* $E[\epsilon_t^2] < \infty$*; and if* $E[\epsilon_t^2] < \infty$ *then only OLS, LTTS and GMTTM satisfy P2.*

Remarks: The F1 fractile properties are controlled by the analyst. Each plug-in is super-$n^{1/2}$-convergent when $E[\epsilon_t^2] = \infty$, and OLS and LAD have non-Gaussian limits when $E[\epsilon_t^2] = \infty$ (Davis et al 1992, Ling 2005, Hill and Renault 2010, Hill 2011b), while $V_n^{1/2}(\gamma) = o(n^{1/2})$ by Lemma 4.1. Hence each $\hat{\beta}_n$ satisfies fast plug-in P1. However, if ϵ_t has a finite variance then $V_n(\gamma) \sim Kn$ and each $\hat{\beta}_n$ has rate $n^{1/2}$, ruling out LAD and LWAD for the non-orthogonalized $\hat{\mathcal{T}}_n(\gamma)$ since P2 requires estimator linearity (cf. Davis et al 1992).

4.2 Linear ARCH

Now consider a strong-ARCH(p) $y_t = h_t u_t$ where $u_t \overset{iid}{\sim} (0,1)$ and $h_t^2 = \omega^0 + \sum_{i=1}^p \alpha_i^0 y_{t-i}^2 = \beta^{0\prime} x_t$, $\omega^0 > 0$, and $\alpha_i^0 \geq 0$. Assume at least one $\alpha_i^0 > 0$ for brevity, let $\sum_{i=1}^p \alpha_i^0 < 1$, and assume the distribution of u_t is non-degenerate, symmetric, absolutely continuous and bounded $\sup_{c \geq 0} |(\partial/\partial c) P(u_t \leq c)| < \infty$. Let κ_u be the moment supremum $\arg\sup\{\alpha > 0 : E|u_t|^\alpha < \infty\}$. If $\kappa_u \in (2,4]$ then assume u_t has tail (11) with index κ_u.

A test of omitted ARCH nonlinearity can be framed in terms of errors $u_t^2 - 1$ or $y_t^2 - \beta^{0\prime} x_t = (u_t^2 - 1)h_t^2$. Since the former only requires u_t^2 and not y_t^2 to be integrable, consider $\epsilon_t(\beta) := u_t^2(\beta) - 1 := y_t^2/(\beta' x_t) - 1$. In this case $(\partial/\partial\beta)\epsilon_t(\beta)|_{\beta^0} = -u_t^2 x_t/h_t^2$ has tails that depend solely on the iid error u_t since we impose ARCH effects $\alpha_i^0 > 0$: $||x_t/h_t^2|| \leq K$ *a.s.* We therefore do not need to use information from x_t for trimming. The error $\epsilon_t = u_t^2 - 1$ may be asymmetric but we can symmetrically trim with re-centering as in Section 3. The trimmed equation with re-centering assuming ARCH effects is $\hat{m}_{n,t}^*(\beta,\gamma) = \{\epsilon_t \hat{I}_{\epsilon,n,t}(\beta) - 1/n \sum_{t=1}^n \epsilon_t \hat{I}_{\epsilon,n,t}(\beta)\} \times F(\gamma'\psi_t)$ where $\hat{I}_{\epsilon,n,t}(\beta) := I(|\epsilon_t(\beta)| \leq \epsilon_{(k_{\epsilon,n})}^{(a)}(\beta))$.

In the following we consider plug-ins $\hat{\beta}_n$ computed by QML, Log-LAD by Peng and Yao (2003), Quasi-Maximum Tail-Trimmed Likelihood [QMTTL] by Hill (2011b), or GMTTM with QML-type equations $\{u_t^2(\beta) - 1\} z_t(\beta)$ where $z_t(\beta) = [(\beta' x_{t-i})^{-1} x_{t-i}]_{i=0}^r$ for some $r \geq 0$ (Hill and Renault 2010).

LEMMA 4.3. *Assumptions F2, I1 and R1-R4 hold. Further a. GMTTM and QMTTL satisfy P1 if* $\kappa_u \in (2,4]$*, P2 if* $\kappa_u > 4$*, and P3 in general; b. QML satisfies P2 and P3 if* $\kappa_u \geq 4$*, but does not satisfy P1-P3 when* $\kappa_u \in (2,4)$*; c. Log-LAD satisfies P1 if* $E[u_t^4] = \infty$*, it does not satisfy P2 if* $\kappa_u > 4$*, and it satisfies P3 in general.*

Remarks: QML is too slow when the ARCH error has an infinite fourth moment $\kappa_u \in (2,4)$. This arises due both to feedback with the error u_t, and to the F1.b lower bound on the error trimming rate $k_{j,\epsilon,n}/n^{2(1-\kappa_\epsilon)/(2-\kappa_\epsilon)} \rightarrow \infty$ which ensures test consistency when $E|\epsilon_t| = \infty$: the former implies $||\tilde{V}_n|| = Kn^{1-2/\kappa_u} = o(n^{1/2})$ (Hall and Yao 2003), while the latter guarantees $\inf_{\gamma \in \Gamma} ||V_n(\gamma)||/n^{1-2/\kappa_u}$ $\rightarrow \infty$. Each remaining estimator has a Gaussian limit since $\kappa_u > 2$. Log-LAD is not linear so orthogonalization is required when $E[u_t^4] < \infty$.

5 Simulation Study

We now present a small scale simulation study where we test for omitted nonlinearity in three models: linear AR(2) $y_t = .8y_{t-1} - .4y_{t-2} + \epsilon_t$, Self-Exciting Threshold AR(1) [SETAR] $y_t = .8y_{t-1}I(y_{t-1} < 0) - .4y_{t-1}I(y_{t-1} \geq 0) + \epsilon_t$, and Bilinear [BILIN] $y_t = .9y_{t-1}\epsilon_{t-1} + \epsilon_t$. We generate 10,000 samples of size $n \in \{200, 800, 5000\}$ by using a starting value $y_1 = \epsilon_1$, generating $2n$ observations of y_t and retaining the last n. The errors $\{\epsilon_t\}$ are either iid $N(0,1)$, symmetric Pareto $P(\epsilon_t \leq -c) = P(\epsilon_t \geq c)$ $= .5(1+c)^{-\kappa_\epsilon}$ with index $\kappa_\epsilon = 1.5$; or IGARCH(1,1) $\epsilon_t = h_t u_t$ where $h_t^2 = .3 + .4u_{t-1}^2 + .6h_{t-1}^2$ and $u_t \overset{iid}{\sim} N(0,1)$, with starting value $h_1^2 = .3$. The errors ϵ_t therefore have possible moment suprema κ_ϵ $\in \{1.5, 2, \infty\}$. Each process is stationary geometrically ergodic and therefore geometrically β-mixing (Pham and Tran 1985, An and Huang 1996, Meitz and Saikkonen 2008). We estimate an AR(5) model $y_t = \sum_{i=1}^{5} \beta_i^0 y_{t-i} + \epsilon_t$ by OLS for each series, although LTTS and LWAD render essentially identical results (cf. Section 4.1, and Hill 2011b).

5.1 Tail-Trimmed CM Test

Write $x_t := [y_{t-1}, ..., y_{t-p}]'$. Recall from Section 3 $k_{j,\epsilon,n} \sim n/L(n)$ for slowly varying $L(n) \rightarrow \infty$ promotes test consistency when $E|\epsilon_t| = \infty$ under the alternative. Considering ϵ_t and y_{t-i} have the same moment supremum κ_ϵ and are symmetric under H_0, we simply use symmetric trimming with $k_n = [\lambda n/\ln(n)]$ for each ϵ_t and y_{t-i}. We re-center by using $\hat{m}_{n,t}^*(\beta,\gamma)$ defined in (9), and compute the orthogonal equations $\hat{m}_{n,t}^{\perp}(\beta,\gamma)$ with the re-centered $\hat{m}_{n,t}^*(\beta,\gamma)$ and operator $\hat{\mathcal{P}}_{n,t}(\gamma)$ $= 1 - x_t'\hat{I}_{n,t}(\hat{\beta}_n) \times (\sum_{t=1}^n x_t x_t' F(\gamma'\psi_t)\hat{I}_{n,t}(\hat{\beta}_n))^{-1} \times \sum_{t=1}^n x_t F(\gamma'\psi_t)\hat{I}_{n,t}(\hat{\beta}_n)$. We use an exponential weight $F(\gamma'\psi(x_t)) = \exp\{\gamma'\psi(x_t)\}$ and argument $\psi(x_t) = [1, \arctan(x_t^*)']' \in \mathbb{R}^6$ with $x_{i,t}^* = x_{i,t} - 1/n\sum_{t=1}^n x_{i,t}$ (cf. Bierens 1990: Section 5), and then compute $\hat{\mathcal{T}}_n(\gamma)$ and $\hat{\mathcal{T}}_n^{\perp}(\gamma)$. We use scale estimators (7) and (8) with $g_t = x_t$ for the sake of comparison with our choice of additional test statistics discussed below. We randomly draw γ from a uniform distribution on $\Gamma = [.1, 2]^6$ for each sample generated, and fix $\lambda = .025$ or compute p-value occupation times $\hat{\tau}_n(\gamma, \alpha)$ and $\hat{\tau}_n^{\perp}(\gamma, \alpha)$ on $[.01, 1.0]$ a la (10) for nominal levels $\alpha \in \{.01, .05, .10\}$. Notice $\lambda = .025$ implies very few observations are trimmed, e.g. at most 1.5% of a sample of size 800.[8]

5.2 Tests of Functional Form

The remaining tests are based on *untrimmed* versions of $\hat{\mathcal{T}}_n(\gamma)$ and $\hat{\mathcal{T}}_n^{\perp}(\gamma)$ where critical values are obtained from a $\chi^2(1)$ distribution; Hong and White's (1995) non-parametric test, Ramsey's (1969) Regression Error Specification Test [RESET], McLeod and Li's (1983) test, and a test proposed by Tsay (1986). Hong and White's (1995) statistic is $\hat{M}_n = (2 \ln n)^{-1/2}(s_n^{-2} \sum_{t=1}^n \hat{\epsilon}_t \hat{v}_{n,t} - \ln n)$ with components $s_n^2 := 1/n \sum_{t=1}^n \hat{\epsilon}_t^2$ and $\hat{v}_{n,t} := \hat{f}_t - \hat{\beta}_n' x_t$, and nonparametric estimator $\hat{f}_t = \sum_{i=1}^{[\ln(n)]} \phi_i \exp\{\gamma_i' x_t\}$ of $E[y_t|x_t]$, cf. Gallant (1981) and Bierens (1990: Corollary 1). The parameters γ_i are for each sample uniformly randomly selected from Γ, and ϕ is estimated by least squares.[9] If certain regularity conditions hold, including independence of ϵ_t and $E[\epsilon_t^4] < \infty$, then $\hat{M}_n \xrightarrow{d} N(0, 1)$ under H_0, while $\hat{M}_n \to \infty$ in probability under H_1, hence a one-sided test is performed. The RESET test is an F-test on the auxiliary regression $\hat{\epsilon}_t = \phi_0' x_t + \sum_{i=2}^{k_1} \sum_{j=2}^{k_2} \phi_{i,j} x_{t-j}^i + u_t$ where we use $k_1 = k_2 = 3$; the McLeod-Li statistic is $\sum_{t=1}^n (\hat{\epsilon}_t^2 - s_n^2)(\hat{\epsilon}_{t-h}^2 - s_n^2) / \sum_{t=1}^n (\hat{\epsilon}_t^2 - s_n^2)^2$ with lags $h = 3$; and Tsay's test is based on first regressing $\text{vech}[x_t x_t'] = \xi' x_t + u_t$, and then computing $F_n := \sum_{t=1}^n (\hat{\epsilon}_t \hat{u}_t)[\sum_{t=1}^n \hat{u}_t \hat{u}_t']^{-1} \sum_{t=1}^n (\hat{\epsilon}_t \hat{u}_t')$: $F_n \xrightarrow{d} \chi^2(p(p+1)/2)$ under H_0 as long as $E[\epsilon_t^4] < \infty$.

5.3 Simulation Results

See Tables 1-3 for test results, where empirical power is adjusted for size distortions. We only present results for $n \in \{200, 800\}$: see the supplemental appendix Hill (2011c: Section C.4) for $n = 5000$.

Write $\hat{\mathcal{T}}_n$-Fix or $\hat{\mathcal{T}}_n$-OT for tests based on fixed $\lambda = .025$ or occupation time. The results strongly suggest orthogonalization is required if we use occupation time because $\hat{\mathcal{T}}_n$-OT exhibits large size distortions, while $\hat{\mathcal{T}}_n^{\perp}$-OT has fairly sharp size and good power. This follows from the dual impact of sampling error associated with $\hat{\beta}_n$ and the loss of information associated with trimming. Our simulations show this applies in general, irrespective of heavy tails, while Remark 3 of Lemma 4.1 shows when $\kappa_\epsilon = \kappa_i \in (1, 2)$ then a *large amount of trimming k_n amplifies sensitivity* of $\hat{\mathcal{T}}_n$ to $\hat{\beta}_n$ in

[8] If $n = 800$ then $k_n = [.025 \times 800/\ln(800)] = 2$ for each $\{\epsilon_t, y_{t-1}, ..., y_{t-5}\}$. Hence at most $2 \times 6 = 12$ observations are trimmed, which is 1.5% of 800.

[9] See Hong and White (1995: Theorem 3.2) for defense of a slowly varying series length $\ln(n)$.

small samples. Orthogonalization *should* play a stronger role when λ is large, hence $\hat{\mathcal{T}}_n^{\perp}$-OT *should* dominate $\hat{\mathcal{T}}_n$-OT, at least when the variance is infinite.

In heavy tailed cases $\hat{\mathcal{T}}_n$-Fix and $\hat{\mathcal{T}}_n^{\perp}$-OT in general exhibit the highest power, although all tests exhibit low power when the errors are IGARCH and $n \in \{200, 800\}$. It should be noted the Hong-White, RESET, McLeod-Li and Tsay tests are all designed under the assumption ϵ_t is independent under H_0 and $E[\epsilon_t^4] < \infty$, hence IGARCH errors are invalid due both to feedback and heavy tails. If ϵ_t is iid Gaussian then trimming does not affect the power of the CM statistic, although Hong-White, McLeod-Li and Tsay tests exhibit higher power.

The untrimmed CM statistics tend to under-reject H_0 and obtain lower power when the error variance is infinite. RESET and McLeod-Li statistics under-reject when $\kappa_\epsilon < 2$, while RESET performs fairly well for an AR model with IGARCH error, contrary to asymptotic theory. The McLeod-Li statistic radically over-rejects H_0 for AR-IGARCH, merely verifying the statistic was designed for iid normal errors under H_0. Tsay's F-statistic radically over-rejects for iid and GARCH errors with infinite variance: empirical power *and* size are above .60. In these cases heavy tails and/or conditional heteroscedasticity simply appear as nonlinearity (cf. de Lima 1997, Hong and Lee 2005, Hill and Aguilar 2011). Hong and White's (1995) non-parametric test exhibits large, and sometimes massive, size distortions when variance is infinite, even for iid errors.

6 Conclusion

We develop tail-trimmed versions of Bierens' (1982, 1990) and Lee, White and Granger's (1993) tests of functional form for heavy tailed time series. The test statistics are robust to heavy tails since trimming ensures standard distribution limits, while negligible trimming ensures the revealing nature of the test weight is not diminished. We may use plug-ins that are sub-$n^{1/2}$-convergent or do not have a Gaussian limit when tails are heavy, depending on the model and error-regressor feedback, and Wooldridge's (1990) orthogonal projection promotes robustness to an even larger set of plug-ins.

A p-value occupation time test allows the analyst to by-pass the need to choose a trimming portion by smoothing over a class of fractiles. A large amount of trimming, however, may have an adverse impact on the test in small samples due to the loss of information coupled with sampling error due to the plug-in. This implies the p-value occupation time may be sensitive to the plug-in in small samples, but when computed with the plug-in robust orthogonal test equation delivers a sharp test in controlled experiments.

Future work may seek to include other trimming techniques like smooth weighting; adaptive methods for selecting the fractiles; and extensions to other classes of tests like Hong and White's (1995) nonparametric test for iid data, and Hong and Lee's (2005) spectral test which accommodates conditional heteroscedasticity of unknown form.

APPENDIX A: Assumptions[10]

Write thresholds and fractiles compactly $c_{z,n}(\cdot) = \max\{l_{z,n}(\cdot), u_{z,n}(\cdot)\}$ and $k_{j,n} = \max\{k_{j,\epsilon,n}, k_{j,1,n}, ..., k_{j,q,n}\}$, define $\sigma_n^2(\beta,\gamma) := E\left[m_{n,t}^{*2}(\beta,\gamma)\right]$ and

$$J_t(\beta,\gamma) := -g_t(\beta) F\left(\gamma'\psi_t\right), \;\; J_{n,t}^*(\beta,\gamma) := J_t(\beta,\gamma) I_{n,t}(\beta), \;\; \hat{J}_{n,t}^*(\beta,\gamma) = J_t(\beta)\hat{I}_{n,t}(\beta)$$
$$J_n^*(\beta,\gamma) := \frac{1}{n}\sum_{t=1}^n J_{n,t}^*(\beta,\gamma), \;\; \hat{J}_n^*(\beta,\gamma) := \frac{1}{n}\sum_{t=1}^n \hat{J}_{n,t}^*(\beta,\gamma).$$

Drop β^0, define $\Im_t = \sigma(x_{\tau+1}, y_\tau : \tau \leq t)$, and let Γ be any compact subset of $\mathbb{R}^p$ with positive Lebesgue measure. Six sets of assumptions are employed. First, the test weight is revealing.

W1 (*weight*).

a. $F : \mathbb{R} \rightarrow \mathbb{R}$ *is Borel measurable, analytic and non-polynomial on some open interval* $R_0 \subseteq \mathbb{R}$ *containing* 0.

b. $\sup_{u\in U}|F(u)| \leq K$ *and* $\inf_{u\in U}|F(u)| > 0$ *on any compact subset* $U \subset S_F$, *with* S_F *the support of* F.

Remark: The W1.b upper bound allows us to exclude $F(\gamma'\psi_t)$ from the trimming indicators which greatly simplifies proving test consistency under trimming, and is mild since it applies to repeatedly cited weights (exponential, logistic, sine, cosine). The lower bound in W1.b helps establish a required stochastic equicontinuity condition for weak convergence when ϵ_t may be heavy tailed, and is easily guaranteed by centering $F(\gamma'\psi_t)$ if necessary.

Second, the plug-in $\hat{\beta}_n$ is consistent. Let $\tilde{m}_{n,t}$ be $\Im_t$-measurable mappings from $\mathcal{B} \subset \mathbb{R}^q$ to $\mathbb{R}^r$, $r \geq q$, and $\{\tilde{V}_n\}$ a sequence of non-random matrices $\tilde{V}_n \in \mathbb{R}^{q\times q}$ where $\tilde{V}_{i,i,n} \rightarrow \infty$. Stack equations

[10] We ignore for notational economy measurability issues that arise when taking a supremum over an index set. Assume all functions in this paper satisfy Pollard's (1984) permissibility criteria, the measure space that governs all random variables is complete, and therefore all majorants are measurable. Probability statements are therefore with respect to outer probability, and expectations over majorants are outer expectations. Cf. Dudley (1978) and Stinchcombe and White (1992).

$\mathcal{M}_{n,t}^{*}(\beta,\gamma) := [m_{n,t}^{*}(\beta,\gamma), \tilde{m}_{n,t}^{\prime}(\beta)]^{\prime} \in \mathbb{R}^{r+1}$, and define the covariances $\tilde{S}_{n}(\beta) := \sum_{s,t=1}^{n} E[\{\tilde{m}_{n,s}(\beta) - E[\tilde{m}_{n,s}(\beta)]\} \times \{\tilde{m}_{n,t}(\beta) - E[\tilde{m}_{n,t}(\beta)]\}^{\prime}]$ and $\mathfrak{S}_{n}^{*}(\beta,\gamma) := \sum_{s,t=1}^{n} E[\{\mathcal{M}_{n,s}^{*}(\beta,\gamma) - E[\mathcal{M}_{n,s}^{*}(\beta,\gamma)]\} \times \{\mathcal{M}_{n,t}^{*}(\beta,\gamma) - E[\mathcal{M}_{n,t}^{*}(\beta,\gamma)]\}^{\prime}]$, hence $[\mathfrak{S}_{i,j,n}^{*}(\beta,\gamma)]_{i=2,j=2}^{r+1,r+1} = \tilde{S}_{n}(\beta)$. We abuse notation since $\mathfrak{S}_{n}^{*}(\beta,\gamma)$ may not exist for some or any β. Let *f.d.d.* denote *finite dimensional distributions.*

P1 (*fast (non)linear plug-ins*). *$\tilde{V}_{n}^{1/2}(\hat{\beta}_{n} - \beta^{0}) = O_{p}(1)$ and $\sup_{\gamma\in\Gamma} ||V_{n}(\gamma)\tilde{V}_{n}^{-1}|| \rightarrow 0$.*

P2 (*slow linear plug-ins*). *$\mathfrak{S}_{n}^{*}(\gamma)$ exists for each n, specifically $\sup_{\gamma\in\Gamma} ||\mathfrak{S}_{n}^{*}(\gamma)|| < \infty$ and $\liminf_{n\rightarrow\infty} \inf_{\gamma\in\Gamma} \lambda_{\min}(\mathfrak{S}_{n}^{*}(\gamma)) > 0$. Further:*

a. $\tilde{V}_{n}^{1/2}(\hat{\beta}_{n} - \beta^{0}) = O_{p}(1)$ and $\tilde{V}_{n} \sim \mathcal{K}(\gamma)V_{n}(\gamma)$, where $\mathcal{K} : \Gamma \rightarrow \mathbb{R}^{q\times q}$ and $\inf_{\gamma\in\Gamma} \lambda_{\min}(\mathcal{K}(\gamma)) > 0$.

b. $\tilde{V}_{n}^{1/2}(\hat{\beta}_{n} - \beta^{0}) = \tilde{A}_{n} \sum_{t=1}^{n} \{\tilde{m}_{n,t} - E[\tilde{m}_{n,t}]\} \times (1 + o_{p}(1)) + o_{p}(1)$ where non-stochastic $\tilde{A}_{n} \in \mathbb{R}^{q\times r}$ has full column rank and $\tilde{A}_{n}\tilde{S}_{n}^{-1}\tilde{A}_{n}^{\prime} \rightarrow I_{q}$.

c. The f.d.d. of $\mathfrak{S}_{n}^{*}(\gamma)^{-1/2} \{\mathcal{M}_{n,t}^{*}(\gamma) - E[\mathcal{M}_{n,t}^{*}(\gamma)]\}$ belong to the same domain of attraction as the f.d.d. of $S_{n}^{-1}(\gamma)\{m_{n,t}^{*}(\gamma) - E[m_{n,t}^{*}(\gamma)]\}$.

P3 (*orthogonal equations and (non)linear plug-ins*). *$\tilde{V}_{n}^{1/2}(\hat{\beta}_{n} - \beta^{0}) = O_{p}(1)$ and $\limsup_{n\rightarrow\infty} \sup_{\gamma\in\Gamma} ||V_{n}^{\perp}(\gamma)\tilde{V}_{n}^{-1}|| < \infty$.*

Remark: $\hat{\beta}_{n}$ effects the limit distribution of $\hat{\mathcal{T}}_{n}(\gamma)$ under P2 hence we assume $\hat{\beta}_{n}$ is linear. P3 is invoked for orthogonalized equations $\hat{m}_{n,t}^{\perp}(\beta,\gamma)$.

Third, identification under trimming.

I1 (*identification by $m_{n,t}^{*}(\gamma)$*). *Under the null $\sup_{\gamma\in\Gamma} |nS_{n}^{-1}(\gamma)E[m_{n,t}^{*}(\gamma)]| \rightarrow 0$.*

Remark: If $m_{t}(\gamma)$ is asymmetric there is no guarantee $E[m_{n,t}^{*}(\gamma)] = 0$, although $E[m_{n,t}^{*}(\gamma)] \rightarrow 0$ under H_{0} by trimming negligibility and dominated convergence. The fractiles $\{k_{j,\epsilon,n}, k_{j,i,n}\}$ must therefore promote I1 for asymptotic normality in view of expansion (5) and mean centering. Since $\sup_{\gamma\in\Gamma}\{S_{n}(\gamma)/n\} = o(1)$ by Lemma B.1, below, I1 implies identification of H_{0} *sufficiently fast*. The property is superfluous if $E[\epsilon_{t}] = 0$ under either hypothesis, ϵ_{t} is independent of x_{t} under H_{0}, and re-centering is used since then $E[m_{n,t}^{*}(\gamma)] = 0$ under H_{0} (see Section 3).

Fourth, the DGP and properties of regression model components.

R1 (*response*). *$f(\cdot,\beta)$ is for each $\beta \in \mathcal{B}$ a Borel measurable function, continuous and differentiable on $\mathcal{B}$ with Borel measurable gradient $g_{t}(\beta) = g(x_{t},\beta) := (\partial/\partial\beta)f(x_{t},\beta)$.*

R2 (*moments*). *$E|y_{t}| < \infty$, and $E(\sup_{\beta\in\mathcal{B}} |f(x_{t},\beta)|^{\iota}) < \infty$ and $E(\sup_{\beta\in\mathcal{B}} |(\partial/\partial\beta_{i})f(x_{t},\beta)|^{\iota}) < \infty$ for each i and some tiny $\iota > 0$.*

R3 (*distribution*).

a. The finite dimensional distributions of $\{y_t, x_t\}$ *are strictly stationary, non-degenerate and absolutely continuous. The density function of* $\epsilon_t(\beta)$ *is uniformly bounded* $\sup_{\beta\in\mathcal{B}}\sup_{a\in\mathbb{R}}\{(\partial/\partial a)P(\epsilon_t(\beta) \leq a)\} < \infty$.

b. Define $\kappa_\epsilon(\beta) := \text{argsup}_{\alpha > 0}\{E|\epsilon_t(\beta)|^\alpha < \infty\} \in (0,\infty]$, *write* $\kappa_\epsilon = \kappa_\epsilon(\beta^0)$, *and let* $\mathcal{B}_{2,\epsilon}$ *denote the set of* β *such that the error variance is infinite* $\kappa_\epsilon(\beta) \leq 2$. *If* $\kappa_\epsilon(\beta) \leq 2$ *then* $P(|\epsilon_t(\beta)| > c) = d(\beta)\epsilon^{-\kappa_\epsilon(\beta)}(1 + o(1))$ *where* $\inf_{\beta\in\mathcal{B}_{2,\epsilon}} d(\beta) > 0$ *and* $\inf_{\beta\in\mathcal{B}_{2,\epsilon}} \kappa_\epsilon(\beta) > 0$, *and* $o(1)$ *is not a function of* β, *hence* $\lim_{c\rightarrow\infty}\sup_{\beta\in\mathcal{B}_{2,\epsilon}} |d(\beta)^{-1}\epsilon^{\kappa_\epsilon(\beta)}P(|\epsilon_t(\beta)| > c) - 1| = 0$.

R4 (*mixing*). $\{y_t, x_t\}$ *are geometrically* β*-mixing:* $\sup_{\mathcal{A}\subset\Im_{t+l}^{+\infty}} E|P(\mathcal{A}|\Im_{-\infty}^t) - P(\mathcal{A})| = o(\rho^l)$ *for* $\rho \in (0,1)$.

Remark 1: Response function smoothness R1 coupled with distribution continuity and boundedness R3.a imply $\sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\beta}_n,\gamma)$ can be asymptotically expanded around β^0, cf. Hill (2011b: Appendices B and C). Power-law tail decay R3.b is mild since it includes weakly dependent processes that satisfy a central limit theorem (Leadbetter et al 1983), and simplifies characterizing tail-trimmed variances in heavy tailed cases by Karamata's Theorem.

Remark 2: The mixing property characterizes nonlinear AR with nonlinear random volatility errors (Pham and Tran 1985, An and Huang 1996, Meitz and Saikkonen 2008).

Fifth, we restrict the fractiles and impose non-degeneracy under trimming. Recall $k_{j,n} = \max\{k_{j,\epsilon,n}, k_{j,1,n}, ..., k_{j,q,n}\}$, the R3.b moment supremum $\kappa_\epsilon > 0$, and $\sigma_n^2(\beta,\gamma) = E[m_{n,t}^{*2}(\beta,\gamma)]$.

F1 (*fractiles*). *a.* $k_{j,\epsilon,n}/\ln(n) \rightarrow \infty$; *b. if* $\kappa_\epsilon \in (0,1)$ *then* $k_{j,\epsilon,n}/n^{2(1-\kappa_\epsilon)/(2-\kappa_\epsilon)} \rightarrow \infty$.

F2 (*non-degenerate trimmed variance*). $\liminf_{n\rightarrow\infty}\inf_{\beta\in\mathcal{B},\gamma\in\Gamma}\{S_n^2(\beta,\gamma)/n\} > 0$ *and* $\sup_{\beta\in\mathcal{B},\gamma\in\Gamma}\{n\sigma_n^2(\beta,\gamma)/S_n^2(\beta,\gamma)\} = O(1)$.

Remark 1: F1.a sets a mild lower bound on $k_{\epsilon,n}$ that is useful for bounding trimmed variances $\sigma_n^2(\beta,\gamma)$ and $S_n^2(\beta,\gamma)$. F1.b sets a harsh lower bound on $k_{\epsilon,n}$ if, under mis-specification, ϵ_t is not integrable: as $\kappa_\epsilon \searrow 0$ we must trim more $k_{\epsilon,n} \nearrow n$ in order to prove a LLN for $m_{n,t}^*(\gamma)$ which is used to prove $\hat{\mathcal{T}}_n(\gamma)$ is consistent. Any $k_{\epsilon,n} \sim n/L(n)$ for slowly varying $L(n) \rightarrow \infty$ satisfies F1.

Remark 2: Distribution non-degeneracy under R3.a coupled with trimming negligibility ensure trimmed moments are not degenerate for sufficiently large n, for example $\liminf_{n\rightarrow\infty}\inf_{\beta\in\mathcal{B},\gamma\in\Gamma}\sigma_n^2(\beta,\gamma) > 0$. The long-run variance $S_n^2(\beta,\gamma)$, however, can in principle be degenerate due to negative dependence, hence F2 is imposed. F2 is standard in the literature on dependent CLT's and exploited here for a CLT for $m_{n,t}^*(\beta,\gamma)$, cf. Dehling et al (1986).

Finally, the kernel $\omega(\cdot)$ and bandwidth b_n.

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

Remark: Assumption K1 includes Bartlett, Parzen, Quadratic Spectral, Tukey-Hanning and other kernels. See de Jong and Davidson (2000) and their references.

APPENDIX B: Proofs of Main Results

We require several preliminary results proved in the supplemental appendix Hill (2011c: Section C.3). Throughout the terms $o_p(1)$, $O_p(1)$, $o(1)$ and $O(1)$, do not depend on β, γ and t. We only state results that concern $\hat{m}_{n,t}^{*}(\beta,\gamma)$ and $m_{n,t}^{*}(\beta,\gamma)$, since companion results extend to $\hat{m}_{n,t}^{\perp}(\beta,\gamma)$ and $m_{n,t}^{\perp}(\beta,\gamma)$. Let F1-F2, K1, R1-R4, and W1.b hold. Recall $\sigma_n^2(\beta,\gamma) = E[m_{n,t}^{*2}(\beta,\gamma)]$.

LEMMA B.1 (variance bounds).

a. $\sigma_n^2(\beta,\gamma) = o\left(n \max\left\{1, (E[m_{n,t}^{*}(\beta,\gamma)])^2\right\}\right)$, $\sup_{\gamma\in\Gamma}\left\{\dfrac{\sigma_n^2(\gamma)}{\max\left\{1,(E[m_{n,t}^{*}(\gamma)])^2\right\}}\right\} = o(n/\ln(n))$;

b. $S_n^2(\gamma) = \mathfrak{L}_n n \sigma_n^2(\gamma) = o(n^2)$ *for some sequence* $\{\mathfrak{L}_n\}$ *that satisfies* $\liminf_{n\rightarrow\infty} \mathfrak{L}_n > 0$, $\mathfrak{L}_n = K$ *if* ϵ_t *is finite dependent or* $E[\epsilon_t^2] < \infty$, *and otherwise* $\mathfrak{L}_n \leq K \ln(n/\min_{j\in\{1,2\}}\{k_{j,\epsilon,n}\}) \leq K\ln(n)$.

LEMMA B.2 (approximations).

a. $\sup_{\gamma\in\Gamma} |S_n^{-1}(\gamma) \sum_{t=1}^{n}\{\hat{m}_{n,t}^{*}(\gamma) - m_{n,t}^{*}(\gamma)\}| = o_p(1)$.

b. Define $\hat{\mu}_{n,t}^{*}(\beta,\gamma) := \hat{m}_{n,t}^{*}(\beta,\gamma) - \hat{m}_{n}^{*}(\beta,\gamma)$ *and* $\mu_{n,t}^{*}(\beta,\gamma) := m_{n,t}^{*}(\beta,\gamma) - m_{n}^{*}(\beta,\gamma)$. *If additionally P1 or P2 holds* $\sup_{\gamma\in\Gamma} |S_n^{-2}(\gamma) \sum_{s,t=1}^{n} \omega((s-t)/b_n)\{\hat{\mu}_{n,s}^{*}(\hat{\beta}_n,\gamma)\hat{\mu}_{n,t}^{*}(\hat{\beta}_n,\gamma) - \mu_{n,s}^{*}(\gamma)\mu_{n,t}^{*}(\gamma)\}| = o_p(1)$.

LEMMA B.3 (expansion). *Let* $\beta, \tilde{\beta} \in \mathcal{B}$. *For some sequence* $\{\beta_{n,*}\}$ *in* $\mathcal{B}$ *satisfying* $||\beta_{n,*} - \tilde{\beta}|| \leq ||\beta - \tilde{\beta}||$, *and for some tiny* $\iota > 0$ *and arbitrarily large finite* $\delta > 0$ *we have* $\sup_{\gamma\in\Gamma} |\hat{m}_n^{*}(\beta,\gamma) - \hat{m}_n^{*}(\tilde{\beta},\gamma) - \hat{J}_n^{*}(\beta_{n,*},\gamma)'(\beta - \tilde{\beta})| = n^{-\delta} \times ||\beta - \tilde{\beta}||^{1/\iota} \times o_p(1)$.

LEMMA B.4 (Jacobian). *Under P1 or P2* $\sup_{\gamma\in\Gamma} ||J_n^{*}(\hat{\beta}_n,\gamma) - J_n(\gamma)(1 + o_p(1))|| = o_p(1)$.

LEMMA B.5 (HAC). *Under P1 or P2* $\sup_{\gamma\in\Gamma} |\hat{S}_n^2(\hat{\beta}_n,\gamma)/S_n^2(\gamma) - 1| \xrightarrow{p} 0$.

LEMMA B.6 (ULLN). *Let* $\inf_{n\geq N} |E[m_{n,t}^{*}(\gamma)]| > 0$ *for some* $N \in \mathbb{N}$ *and all* $\gamma \in \Gamma/S$ *where* S *has measure zero. Then* $\sup_{\gamma\in\Gamma/S}\{1/n \sum_{t=1}^{n} m_{n,t}^{*}(\gamma)/E\left[m_{n,t}^{*}(\gamma)\right]\} \xrightarrow{p} 1$.

LEMMA B.7 (UCLT). $\{S_n^{-1}(\gamma)\sum_{t=1}^{n}(m_{n,t}^{*}(\gamma) - E[m_{n,t}^{*}(\gamma)]) : \gamma \in \Gamma\} \Longrightarrow \{z(\gamma) : \gamma \in \Gamma\}$, *a scalar* $(0,1)$*-Gaussian process on* $C[\Gamma]$ *with covariance function* $E[z(\gamma_1)z(\gamma_2)]$ *and a.s. bounded* *sample paths. If P2 also holds then* $\{\mathfrak{S}_n^{-1/2}(\gamma)\sum_{t=1}^n\{\mathcal{M}_{n,t}^*(\gamma) - E[\mathcal{M}_{n,t}^*(\gamma)] : \gamma \in \Gamma\} \Longrightarrow \{\mathcal{Z}(\gamma) : \gamma \in \Gamma\}$ *an* $r+1$ *dimensional Gaussian process on* $C[\Gamma]$ *with zero mean, covariance* I_{r+1}*, and covariance function* $E[\mathcal{Z}(\gamma_1)\mathcal{Z}(\gamma_2)']$.

PROOF OF LEMMA 2.1. We only prove the claims for $m_{n,t}^*(\beta,\gamma)$. In view of the $\sigma(x_t)$-measurability of $\mathcal{P}_{n,t}(\gamma)$ and $\sup_{\gamma\in\Gamma} E|\mathcal{P}_{n,t}(\gamma)| < \infty$ the proof extends to $m_{n,t}^{\perp}(\beta,\gamma)$ with few modifications. Under H_0 the claim follows from trimming negligibility and Lebesgue's dominated convergence: $E[m_{n,t}^*(\gamma)] \rightarrow E[m_t(\gamma)] = 0$.

Under the alternative there are two cases: $E|\epsilon_t| < \infty$, or $E|\epsilon_t| = \infty$ such that $E[\epsilon_t|x_t]$ may not exist.

Case 1 ($E|\epsilon_t| < \infty$): Property W1, compactness of Γ and boundedness of ψ imply $F(\gamma'\psi_t)$ is uniformly bounded and revealing: $E[\epsilon_t F(\gamma'\psi_t)] \neq 0$ for all $\gamma \in \Gamma/S$ where S has Lebesgue measure zero. Now invoke boundedness of $F(\gamma'\psi_t)$ with Lebesgue's dominated convergence theorem and negligibility of trimming to deduce $|E[\epsilon_t(1 - I_{n,t}(\beta^0))F(\gamma'\psi_t)]| \rightarrow 0$, hence $E[\epsilon_t I_{n,t}(\beta^0)F(\gamma'\psi_t)] = E\left[\epsilon_t F(\gamma'\psi_t)\right] + o(1) \neq 0$ for all $\gamma \in \Gamma/S$ and all $n \geq N$ for sufficiently large N.

Case 2 ($E|\epsilon_t| = \infty$): Under H_1 since $I_{n,t}(\beta) \rightarrow 1$ *a.s.* and $E|\epsilon_t| = \infty$, by the definition of conditional expectations there exists sufficiently large N such that $\min_{n\geq N} |E[\epsilon_t I_{n,t}(\beta^0)|x_t]| > 0$ with positive probability $\forall n \geq N$. The claim therefore follows by Theorem 1 of Bierens and Ploberger (1997) and Theorem 2.3 of Stinchcombe and White (1998): $\liminf_{n\rightarrow\infty} |E[\epsilon_t I_{n,t}(\beta^0)F(\gamma'\psi_t)]| > 0$ for all $\gamma \in \Gamma/S$. $\mathcal{QED}$.

PROOF OF THEOREM 2.2. Define $M_{n,t}^*(\beta,\gamma) := m_{n,t}^*(\beta,\gamma) - E[m_{n,t}^*(\beta,\gamma)]$ and $\hat{M}_{n,t}^*(\beta,\gamma) := \hat{m}_{n,t}^*(\beta,\gamma) - E[\hat{m}_{n,t}^*(\beta,\gamma)]$. We first state some required properties. Under plug–in properties P1 or P2 $\hat{\beta}_n - \beta^0 = o_p(1)$. Identification I1 imposes under H_0

$$\sup_{\gamma\in\Gamma} \left|S_n^{-1}(\gamma)E[m_{n,t}^*(\gamma)]\right| = o(1/n), \tag{12}$$

which implies the following long-run variance relation uniformly on Γ:

$$E\left(\sum_{t=1}^n M_{n,t}^*(\gamma)\right)^2 = S_n^2(\gamma) - n^2\left(E\left[m_{n,t}^*(\beta,\gamma)\right]\right)^2 = S_n^2(\gamma)\left(1 + o\left(1\right)\right). \tag{13}$$

Uniform expansion Lemma B.3, coupled with Jacobian consistency Lemma B.4 and $\hat{\beta}_n \xrightarrow{p} \beta^0$ imply for any arbitrarily large finite $\delta > 0$,

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \hat{m}^*_{n,t}(\hat{\beta}_n, \gamma) - \hat{m}^*_{n,t}(\gamma) \right\} - J_n(\gamma)' \left(\hat{\beta}_n - \beta^0 \right) (1 + o_p(1)) \right| = o_p\left(n^{-\delta} \right). \tag{14}$$

Finally, by uniform approximation Lemma B.2.a

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{S_n(\gamma)} \sum_{t=1}^{n} \left\{ \hat{m}^*_{n,t}(\gamma) - m^*_{n,t}(\gamma) \right\} \right| = o_p(1), \tag{15}$$

and by Lemma B.5 we have uniform HAC consistency:

$$\sup_{\gamma \in \Gamma} \left| \hat{S}^2_n(\hat{\beta}_n, \gamma)/S^2_n(\gamma) - 1 \right| = o_p(1). \tag{16}$$

Claim i ($\hat{T}_n(\gamma)$: Null H_0): Under fast plug-in case P1 we assume $\sup_{\gamma \in \Gamma} ||V_n(\gamma)\tilde{V}_n^{-1}|| \to 0$, hence

$$\sup_{\gamma \in \Gamma} \left| n S_n^{-1}(\gamma) J_n(\gamma)' \left(\hat{\beta}_n - \beta^0 \right) \right| = o_p(1). \tag{17}$$

Since $\delta > 0$ in (14) may be arbitrarily large, $\liminf_{n \to \infty} \inf_{\gamma \in \Gamma} S_n(\gamma) > 0$ by non-degeneracy F2, and equations (12)-(17) are uniform properties, it follows uniformly on Γ

$$\begin{aligned} \hat{T}_n(\gamma) &\overset{p}{\sim} \left(\frac{1}{S_n(\gamma)} \sum_{t=1}^{n} M^*_{n,t}(\gamma) + \frac{n J_n(\gamma)'}{S_n(\gamma)} \left(\hat{\beta}_n - \beta^0 \right) + o_p\left(\frac{n}{S_n(\gamma)} n^{-\delta} \right) \right)^2 \\ &= \left(\frac{1}{S_n(\gamma)} \sum_{t=1}^{n} M^*_{n,t}(\gamma) + o_p(1) \right)^2 = \mathcal{M}^2_n(\gamma), \end{aligned} \tag{18}$$

say. Now apply variance relation (13), UCLT Lemma B.7 and the mapping theorem to conclude $E[\mathcal{M}^2_n(\gamma)] \to 1$ and $\{\hat{T}_n(\gamma) : \gamma \in \Gamma\} \Longrightarrow \{z^2(\gamma) : \gamma \in \Gamma\}$, where $z(\gamma)$ is $(0,1)$-Gaussian process on $\mathcal{C}[\Gamma]$ with covariance function $E[z(\gamma_1)z(\gamma_2)]$.

Under slow plug-in case P2 a similar argument applies in lieu of plug-in linearity and UCLT Lemma B.7. Since the steps follow conventional arguments we relegate the proof to Hill (2011c: Section C.2).

Claim ii ($\hat{T}_n(\gamma)$: Alternative H_1): Lemma 2.1 ensures $\inf_{n \geq N} \left| E[m^*_{n,t}(\gamma)] \right| > 0$ for some $N \in \mathbb{N}$ and all $\gamma \in \Gamma/S$ where $S \subset \Gamma$ has Lebesgue measure zero. Choose any $\gamma \in \Gamma/S$, assume $n \geq N$

and write

$$\hat{\mathcal{T}}_n(\gamma) = \left(\frac{1}{\hat{S}_n(\hat{\beta}_n,\gamma)}\sum_{t=1}^n \hat{m}^*_{n,t}(\hat{\beta}_n,\gamma)\right)^2 = \frac{n^2\left(E\left[m^*_{n,t}(\gamma)\right]\right)^2}{\hat{S}^2_n(\hat{\beta}_n,\gamma)}\left(\frac{\left|1/n\sum_{t=1}^n \hat{m}^*_{n,t}(\hat{\beta}_n,\gamma)\right|}{\left|E\left[m^*_{n,t}(\gamma)\right]\right|}\right)^2.$$

In lieu of (16) and the Lemma B.1.a,b variance property $n|E[m^*_{n,t}(\gamma)]|/S_n(\gamma) \to \infty$, the proof is complete if we show $\mathcal{M}_n(\hat{\beta}_n,\gamma) := |1/n\sum_{t=1}^n \hat{m}^*_{n,t}(\hat{\beta}_n,\gamma)|/|E[m^*_{n,t}(\gamma)]| \xrightarrow{p} 1$.

By (14), (15) and the triangle inequality $\mathcal{M}_n(\hat{\beta}_n,\gamma)$ is bounded by

$$\frac{1}{\left|E\left[m^*_{n,t}(\gamma)\right]\right|}\left|\frac{1}{n}\sum_{t=1}^n m^*_{n,t}(\gamma)\right| + \frac{1}{\left|E\left[m^*_{n,t}(\gamma)\right]\right|}\left|J_n(\gamma)'\left(\hat{\beta}_n-\beta^0\right)(1+o_p(1))\right| + o_p\left(\frac{S_n(\gamma)}{n\left|E\left[m^*_{n,t}(\gamma)\right]\right|}\right),$$

where $\sup_{\gamma\in\Gamma/S}\{1/n\sum_{t=1}^n m^*_{n,t}(\gamma)/E[m^*_{n,t}(\gamma)]\} \xrightarrow{p} 1$ by Lemma B.6. Further, combine fast or slow plug-in P1 or P2, the construction of $V_n(\gamma)$ and variance relation Lemma B.1.a,b to obtain

$$\frac{\left|J_n(\gamma)'\left(\hat{\beta}_n-\beta^0\right)(1+o_p(1))\right|}{\left|E\left[m^*_{n,t}(\gamma)\right]\right|} \leq \frac{S_n(\gamma)}{n\left|E\left[m^*_{n,t}(\gamma)\right]\right|} nJ_n(\gamma)' S_n^{-1}(\gamma) V_n^{-1/2}(\gamma) \sim K\frac{S_n(\gamma)}{n\left|E\left[m^*_{n,t}(\gamma)\right]\right|} = o(1).$$

Therefore $\mathcal{M}_n(\hat{\beta}_n,\gamma) \xrightarrow{p} 1$.

Claim iii ($\hat{\mathcal{T}}_n^{\perp}(\gamma)$): The argument simply mimics claims (i) and (ii) since under plug-in case P3 it follows $\hat{S}_n^{\perp}(\hat{\beta}_n,\gamma)^{-1}\sum_{t=1}^n \hat{m}^{\perp}_{n,t}(\hat{\beta}_n,\gamma) \overset{p}{\sim} S_n^{\perp}(\gamma)^{-1}\sum_{t=1}^n m^{\perp}_{n,t}(\gamma)$ by construction of the orthogonal equations (Wooldridge 1990), and straightforward generalizations of the supporting lemmas. $\mathcal{QED}$.

The remaining proofs exploit the fact that for each $z_t \in \{\epsilon_t, g_{i,t}\}$ the product $z_tF(\gamma'\psi_t)$ has the same tail decay rate as z_t: by weight boundedness W1.b $P(|z_t\sup_{u\in\mathbb{R}}F(u)| > c) \geq P(|z_tF_t(\gamma)| > c) \geq P(|z_t\inf_{u\in\mathbb{R}}F(u)| > c)$. Further, use $I_{n,t} = I_{\epsilon,n,t}I_{g,n,t}$, dominated convergence and each $I_{z,n,t} \xrightarrow{a.s.} 1$ to deduce $E[|z_tF(\gamma'\psi_t)|^rI_{n,t}] = E[|z_tF(\gamma'\psi_t)|^rI_{z,n,t}] \times (1+o(1))$ for any $r > 0$. Hence higher moments of $z_tF(\gamma'\psi_t)I_{n,t}$ and $z_tI_{z,n,t}$ are equivalent up to a constant scale.

PROOF OF THEOREM 3.1. The claim under H_1 follows from Theorem 2.2. We prove $\tau_n(\alpha) \xrightarrow{d} (1-\underline{\lambda})^{-1}\int_{\underline{\lambda}}^1 I(u(\lambda) < \alpha)d\lambda$ under H_0 for plug-in case P1 since the remaining cases follow similarly. Drop γ and write $\hat{m}^*_{n,t}(\hat{\beta}_n,\lambda)$ and $\hat{S}^2_n(\hat{\beta}_n,\lambda)$ to express dependence on $\lambda \in \Lambda := [\underline{\lambda}, 1]$. Define $\hat{Z}_n(\lambda) := \hat{S}_n^{-1}(\hat{\beta}_n,\lambda)\sum_{t=1}^n \hat{m}^*_{n,t}(\hat{\beta}_n,\lambda)$. We exploit weak convergence on a Polish space[11]: we write $\{\hat{Z}_n(\lambda) : \lambda \in \Lambda\} \Longrightarrow^* \{z(\lambda) : \lambda \in \Lambda\}$ on $l_\infty(\Lambda)$, where $\{z(\lambda) : \lambda \in \Lambda\}$ is a Gaussian process with a version

[11] See Hoffmann-Jørgensen (1991), cf. Dudley (1978).

that has uniformly bounded and uniformly continuous sample paths with respect to $||\cdot||_2$, if $\hat{Z}_n(\lambda)$ converges in *f.d.d.* and tightness applies: $\lim_{\delta\to 0}\limsup_{n\to\infty} P(\sup_{||\lambda-\tilde{\lambda}||\leq\delta}|\hat{Z}_n(\lambda) - \hat{Z}_n(\tilde{\lambda})| > \varepsilon) = 0\ \forall\varepsilon > 0$.

We need only prove $\{\hat{Z}_n(\lambda) : \lambda \in \Lambda\} \Longrightarrow^* \{z(\lambda) : \lambda \in \Lambda\}$ since the claim follows from multiple applications of the mapping theorem. Convergence in *f.d.d.* follows from $\sup_{\lambda\in\Lambda} |\hat{S}_n^{-1}(\hat{\beta}_n, \lambda) \sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\beta}_n, \lambda) - S_n^{-1}(\lambda) \sum_{t=1}^n m_{n,t}^*(\lambda)| \xrightarrow{p} 0$ by (14)-(16) under plug-in case P1, and the proof of UCLT Lemma B.7.

Consider tightness and notice by (14)-(17) and plug-in case P1

$$\sup_{\lambda\in\Lambda} \left|\hat{Z}_n(\lambda) - \mathcal{Z}_n(\lambda)\right| \xrightarrow{p} 0 \ \text{ where } \mathcal{Z}_n(\lambda) := \sum_{t=1}^n \frac{1}{S_n(\lambda)} m_t I_{n,t}(\lambda) = \sum_{t=1}^n \mathcal{Z}_{n,t}(\lambda),$$

hence we need only consider $\mathcal{Z}_n(\lambda)$ for tightness. By Lemma B.1.b and $\inf\{\Lambda\} > 0$ it is easy to verify $\inf_{\lambda\in\Lambda} S_n^2(\lambda) = n\sigma_n^2$ for some sequence $\{\sigma_n^2\}$ that satisfies $\liminf_{n\to\infty}\sigma_n^2 > 0$. Therefore

$$\begin{aligned}\left|\sum_{t=1}^n \left\{\mathcal{Z}_{n,t}(\lambda) - \mathcal{Z}_{n,t}(\tilde{\lambda})\right\}\right| &\leq \left|\frac{1}{n^{1/2}\sigma_n}\sum_{t=1}^n m_t \left\{I_{n,t}(\lambda) - I_{n,t}(\tilde{\lambda})\right\}\right| \\ &+ \left|\frac{S_n(\lambda)}{S_n(\tilde{\lambda})} - 1\right| \times \left|\frac{1}{S_n(\lambda)}\sum_{t=1}^n m_t I_{n,t}(\lambda)\right| = \mathcal{A}_{1,n}(\lambda, \tilde{\lambda}) + \mathcal{A}_{2,n}(\lambda, \tilde{\lambda}).\end{aligned}$$

By subadditivity it suffices to prove each $\lim_{\delta\to 0}\limsup_{n\to\infty} P(\sup_{||\lambda-\tilde{\lambda}||\leq\delta}\mathcal{A}_{i,n}(\lambda, \tilde{\lambda}) > \varepsilon) = 0\ \forall\varepsilon > 0$.

Consider $\mathcal{A}_{1,n}(\lambda, \tilde{\lambda})$ and note $I_{n,t}(\lambda)$ can be approximated by a sequence of continuous, differentiable functions (Lighthill 1958, Phillips 1995). Let $\{\mathcal{N}_n\}$ be a sequence of positive numbers to be chosen below, and define a smoothed version of $I_{n,t}(\lambda)$,

$$\mathfrak{I}_{\mathcal{N}_n,n,t}(\lambda) := \int_0^1 I_{n,t}(\varpi)\mathrm{S}\left(\mathcal{N}_n(\varpi - \lambda)\right)\frac{\mathcal{N}_n}{e^{\varpi^2/\mathcal{N}_n^2}}d\varpi = \int_{\lambda-1/\mathcal{N}_n}^{\lambda+1/\mathcal{N}_n} I_{n,t}(\varpi)\left\{\frac{e^{-1/(1-\mathcal{N}_n^2(\varpi-\lambda)^2)}}{\int_{-1}^1 e^{-1/(1-w^2)}dw}\right\}\times\frac{\mathcal{N}_n}{e^{\varpi^2/\mathcal{N}_n^2}}d\varpi,$$

where $\mathrm{S}(u)$ is a so-called "smudge" function used to blot out $I_{n,t}(\varpi)$ when ϖ is outside the interval $(\lambda - 1/\mathcal{N}_n, \lambda + 1/\mathcal{N}_n)$. The term $\{\cdot\}$ after the second equality defines $\mathrm{S}(u)$ on $[-1, 1]$. The random variable $\mathfrak{I}_{\mathcal{N}_n,n,t}(\lambda)$ is $\Im_t$-measurable, uniformly bounded, continuous and differentiable for each $\mathcal{N}_n$, and since $k_n(\lambda) \geq k_n(\tilde{\lambda})$ for $\lambda \geq \tilde{\lambda}$ then $\mathfrak{I}_{\mathcal{N}_n,n,t}(\lambda) \leq \mathfrak{I}_{\mathcal{N}_n,n,t}(\tilde{\lambda})$ *a.s.* Cf. Phillips (1995).

Observe $\mathcal{A}_{1,n}(\lambda,\tilde{\lambda}) = \mathcal{B}_{1,\mathcal{N}_n,n}(\lambda,\tilde{\lambda}) + \mathcal{B}_{2,\mathcal{N}_n,n}(\lambda) + \mathcal{B}_{2,\mathcal{N}_n,n}(\tilde{\lambda})$ where

$$\mathcal{B}_{1,\mathcal{N}_n,n}(\lambda,\tilde{\lambda}) = \sum_{t=1}^{n} \frac{m_t \left\{\mathfrak{I}_{\mathcal{N}_n,n,t}(\lambda) - \mathfrak{I}_{\mathcal{N}_n,n,t}(\tilde{\lambda})\right\}}{n^{1/2}\sigma_n}, \quad \mathcal{B}_{2,\mathcal{N}_n,n}(\lambda) = \sum_{t=1}^{n} \frac{m_t \left\{I_{n,t}(\lambda) - \mathfrak{I}_{\mathcal{N}_n,n,t}(\lambda)\right\}}{n^{1/2}\sigma_n}.$$

Consider $\mathcal{B}_{1,\mathcal{N}_n,n}(\lambda,\tilde{\lambda})$, define $\mathcal{D}_{\mathcal{N}_n,n,t}(\lambda) := (\partial/\partial\lambda)\mathfrak{I}_{\mathcal{N}_n,n,t}(\lambda)$, and let $\{b_n(\lambda,\iota)\}$ for infinitessimal $\iota > 0$ be any sequence of positive numbers that satisfies $P(|m_t| > b_n(\lambda,\iota)) \rightarrow \lambda - \iota \in (0,1)$, hence $\lim_{n\rightarrow\infty}\sup_{\lambda\in\Lambda} b_n(\lambda,\iota) < \infty$. By the mean-value-theorem $\mathfrak{I}_{\mathcal{N}_n,n,t}(\lambda) - \mathfrak{I}_{\mathcal{N}_n,n,t}(\tilde{\lambda}) = \mathcal{D}_{\mathcal{N}_n,n,t}(\lambda_*)(\lambda - \tilde{\lambda})$ for some $\lambda_* \in \Lambda$, $|\lambda - \lambda_*| \leq |\lambda - \tilde{\lambda}|$. But since $\sup_{\lambda\in\Lambda}|I_{n,t}(\lambda) - 1| \overset{a.s.}{\rightarrow} 0$ it must be the case that $\sup_{\lambda\in\Lambda}|\mathcal{D}_{\mathcal{N}nn,n,t}(\lambda)| \rightarrow 0$ $a.s.$ as $n \rightarrow \infty$ for any $\mathcal{N}_n \rightarrow \infty$. Therefore, for N sufficiently large, all $n \geq N$, any $p > 0$ and some $\{b_n(\lambda,\iota)\}$ we have $\sup_{\lambda\in\Lambda} E|m_t\mathcal{D}_{\mathcal{N}_n,n,t}(\lambda)|^p \leq K\sup_{\lambda\in\Lambda} E|m_t I(|m_t| \leq b_n(\lambda,\iota))|^p \leq K\sup_{\lambda\in\Lambda} b_n^p(\lambda,\iota)$ which is bounded on $\mathbb{N}$. This implies $m_t\mathcal{D}_{\mathcal{N}_n,n,t}(\lambda)$ is L_p-bounded for any $p > 2$ uniformly on $\Lambda\times\mathbb{N}$, and geometrically β-mixing under R4. In view of $\liminf_{n\rightarrow\infty}\sigma_n^2 > 0$ we may therefore apply Lemma 3 in Doukhan et al (1995) to obtain $\sup_{\lambda\in\Lambda}|n^{-1/2}\sigma_n^{-1}\sum_{t=1}^{n} m_t\mathcal{D}_{\mathcal{N}_n,n,t}(\lambda)| = O_p(1)$. This suffices to deduce $\lim_{\delta\rightarrow 0}\limsup_{n\rightarrow\infty} P(\sup_{||\lambda-\tilde{\lambda}||\leq\delta}|\mathcal{B}_{1,\mathcal{N}_n,n}(\lambda,\tilde{\lambda})| > \varepsilon)$ is bounded by

$$\lim_{\delta\rightarrow 0}\limsup_{n\rightarrow\infty} P\left(K\sup_{\lambda\in\Lambda}\left|\frac{1}{n^{1/2}\sigma_n}\sum_{t=1}^{n} m_t\mathcal{D}_{\mathcal{N}_n,n,t}(\lambda)\right| \times \delta > \varepsilon\right) = 0.$$

Further, since the rate $\mathcal{N}_n \rightarrow \infty$ is arbitrary, we can always let $\mathcal{N}_n \rightarrow \infty$ so fast that $\limsup_{n\rightarrow\infty} P(\sup_{\lambda\in\Lambda}|\mathcal{B}_{2,\mathcal{N}_n,n}(\lambda)| > \varepsilon) = 0$, cf. Phillips (1995). By subadditivity this proves $\lim_{\delta\rightarrow 0}\limsup_{n\rightarrow\infty} P(\sup_{||\lambda-\tilde{\lambda}||\leq\delta}\mathcal{A}_{1,n}(\lambda,\tilde{\lambda}) > \varepsilon) = 0$ $\forall\varepsilon > 0$.

Now consider $\mathcal{A}_{2,n}(\lambda,\tilde{\lambda})$. By UCLT Lemma B.7 $\sup_{\lambda\in\Lambda}|S_n^{-1}(\lambda)\sum_{t=1}^{n} m_t I_{n,t}(\lambda)| = O_p(1)$ for any compact subset Λ of $(0,1]$. The proof is therefore complete if we show $|S_n(\lambda)/S_n(\tilde{\lambda}) - 1| \leq K|\lambda - \tilde{\lambda}|^{1/2}$. By Lemma B.1.b $S_n^2(\lambda) = \mathfrak{L}_n(\lambda)nE[m_t^2 I_{n,t}(\lambda)]$. Compactness of $\Lambda \subset (0,1]$ ensures $\liminf_{n\rightarrow\infty}\inf_{\lambda\in\Lambda}\mathfrak{L}_n(\lambda) > 0$ and $\sup_{\lambda\in\Lambda}\mathfrak{L}_n(\lambda) = O(\ln(n))$, and by distribution continuity $E[m_t^2 I_{n,t}(\lambda)]$ is differentiable, hence $|S_n(\lambda)/S_n(\tilde{\lambda}) - 1| \leq K(\sup_{\lambda\in\Lambda}\{|G_n(\lambda)|\}/E[m_t^2 I_{n,t}(\lambda)])^{1/2} \times |\lambda - \tilde{\lambda}|^{1/2} =: \mathcal{E}_n|\lambda - \tilde{\lambda}|^{1/2}$ where $G_n(\lambda) := (\partial/\partial\lambda)E[m_t^2 I_{n,t}(\lambda)]$. Since $k_n \sim \lambda n/\ln(n)$ it is easy to verify $\limsup_{n\rightarrow\infty}\sup_{\lambda\in\Lambda}\mathcal{E}_n < \infty$: if $E[m_t^2] < \infty$ then the bound is trivial, and if $E[m_t^2] = \infty$ then use $c_{\epsilon,n} = K(n/k_n)^{1/\kappa} = K(\ln(n))^{1/\kappa}\lambda^{-1/\kappa}$ and Karamata's Theorem (Resnick 1987: theorem 0.6). $\mathcal{QED}$.

PROOF OF LEMMA 4.1. By Lemma B.7 in Hill (2011b) $J_n(\gamma) = -E[g_t F_t(\gamma)I_{n,t}] \times (1 + o(1))$ hence it suffices to bound $(E[g_{i,t}F_t(\gamma)I_{n,t}])^2/S_n^2(\gamma)$. The claim follows from Lemma B.1.b, and the following implication of Karamata's theorem (e.g. Resnick 1987: Theorem .06): if any random variable w_t has tail $P(|w_t| > w) = dw^{-\kappa}(1 + o(1))$, and $w^*_{n,t} := w_t I(|w_t| \leq c_{w,n})$, $P(|w_t| > c_{w,n}) = k_{w,n}/n = o(1)$ and $k_{w,n} \to \infty$, then $E|w^*_{n,t}|^p$ is slowly varying if $p = \kappa$, and $E|w^*_{n,t}|^p \sim K c^p_{w,n}(k_{w,n}/n)$ $= K(n/k_{w,n})^{p/\kappa-1}$ if $p > \kappa$. $\mathcal{QED}$.

PROOF OF LEMMA 4.2. First some preliminaries. Integrability of ϵ_t is assured by $\kappa > 1$, and y_t has tail (11) with the same tail index κ (Brockwell and Cline 1985). Stationarity ensures $\epsilon_t(\beta)$ $= \sum_{i=0}^{\infty} \psi_i(\beta)\epsilon_{t-i}$, where $\sup_{\beta\in\mathcal{B}} |\psi_i(\beta)| \leq K\rho^i$ for $\rho \in (0, 1)$, $\psi_0(\beta^0) = 1$ and $\psi_i(\beta^0) = 0$ $\forall i \geq 1$. Since ϵ_t is iid with tail (11) it is easy to show $\epsilon_t(\beta)$ satisfies uniform power law property R3.b by exploiting convolution tail properties developed in Embrechts and Goldie (1980). Use (4) and (11) to deduce $c_{\epsilon,n} = K\,(n/k_n)^{1/\kappa}$.

F2 follows from the stationary AR data generating process and distribution continuity. I1 holds since $E[m^*_{n,t}(\gamma)] = 0$ by independence, symmetry and symmetric trimming. R1 and R2 hold by construction; (11) and the stated error properties ensure R3; see Pham and Tran (1985) for R4.

Now P1-P3. OLS and LAD are $n^{1/\kappa}$-convergent if $\kappa \in (1, 2]$ (Davis et al 1992); LTTS and GMTTM are $n^{1/\kappa}/L(n)$-convergent if $\kappa \in (1, 2]$ (Hill and Renault 2010, Hill 2011b);[12] and LWAD is $n^{1/2}$-convergent in all cases (Ling 2005). It remains to characterize $V_n(\gamma)$. Each claim follows by application of Lemma 4.1. If $\kappa > 2$ then $V_n(\gamma) \sim Kn$, so OLS, LTTS and GMTTM satisfy P2 (LAD and LWAD are not linear: see Davis et al 1992). If $\kappa \in (1, 2)$ then $V_n(\gamma) \sim Kn\,(k_n/n)^{2/\kappa-1} = o(n)$, while each $\hat{\beta}_n$ satisfies $\tilde{V}^{1/2}_{i,i,n}/n^{1/2} \to \infty$, hence P1 applies for any intermediate order $\{k_n\}$. The case $\kappa = 2$ is similar.

Finally, Lemma 4.1 can be shown to apply to $V^{\perp}_n(\gamma)$ by exploiting the fact that $\epsilon_t g_{i,t} = \epsilon_t y_{t-i}$ have the same tail index as ϵ_t (Embrechts and Goldie 1980). The above arguments therefore extend to $m^{\perp}_{n,t}(\beta, \gamma)$ under P3. $\mathcal{QED}$.

PROOF OF LEMMA 4.3. The ARCH process $\{y_t\}$ is stationary geometrically β-mixing (Carrasco and Chen 2002). In lieu of re-centering after trimming and error independence, all conditions except P1-P3 hold by the arguments used to prove Lemma 4.2.

Consider P1-P3. Note $\epsilon_t = u^2_t - 1$ is iid, it has tail index $\kappa_u/2 \in (1, 2]$ if $E[u^4_t] = \infty$, and $(\partial/\partial\beta)\epsilon_t(\beta)|_{\beta^0} = -u^2_t x_t/h^2_t$ is integrable. Further $S^2_n(\gamma) = nE[m^{*2}_{n,t}(\gamma)]$ by independence and re-

[12] LTTS and GMTTM require trimming fractiles for estimation: GMTTM requires fractiles $\tilde{k}_{i,n}$ for each estimating equation $\tilde{m}_{i,n,t}$, and LTTS requires fractiles $\tilde{k}_{\epsilon,n}$ and $\tilde{k}_{y,n}$ for ϵ_t and y_{t-i}. The given rates of convergence apply if for GMTTM $\tilde{k}_{i,n} \sim \lambda \ln(n)$ (Hill and Renault 2010), and for LTTS $\tilde{k}_{\epsilon,n} \sim \lambda n/\ln(n)$ and $\tilde{k}_{y,n} \sim \lambda \ln(n)$ (Hill 2011b), where $\lambda > 0$ is chosen by the analyst and may be different in different places.

centering. Thus $V_n(\gamma) \sim Kn$ if $E[u_t^4] < \infty$, and otherwise apply Lemma 4.1 to deduce $V_n(\gamma) \sim Kn\,(k_n/n)^{4/\kappa_u - 1}$ if $\kappa_u < 4$ and $V_n(\gamma) \sim n/L(n)$ if $\kappa_u = 4$.

GMTTM with QML-type equations and QMTTL have a scale $||\tilde{V}_n|| \sim n/L(n)$ if $E[u_t^4] = \infty$, hence P1, otherwise $||\tilde{V}_n|| \sim Kn$ hence P2 (Hill and Renault 2010, Hill 2011b). Log-LAD is $n^{1/2}$-convergent if $E[u_t^2] < \infty$, hence P1 if $\kappa_u \leq 4$, and if $\kappa_u > 4$ then it does not satisfy P2 since it is not linear. QML is $n^{1/2}$-convergent if $E[u_t^4] < \infty$ hence P2, and if $E[u_t^4] = \infty$ then the rate is $n^{1-2/\kappa_u}/L(n)$ when $\kappa_u \in (2, 4]$ (Hall and Yao 2003: Theorem 2.1). But if $\kappa_u < 4$ then $n(k_n/n)^{4/\kappa_u - 1} = k_n^{4/\kappa_u - 1} n^{2-4/\kappa_u} > n^{2-4/\kappa_u}/L(n)$ for any slowly varying $L(n) \to \infty$ and intermediate order $\{k_n\}$ hence QML does not satisfy P1 or P2. Synonymous arguments extend to $m_{n,t}^{\perp}(\gamma)$ under P3 by exploiting Lemma 4.1. $\mathcal{QED}$.

References

[1] An H.Z., Huang F.C. (1996) The Geometrical Ergodicity of Nonlinear Autoregressive Models, Stat. Sin. 6, 943-956.

[2] Arcones M., Giné E.(1989) The Bootstrap of the Mean with Arbitrary Bootstrap Sample Size, Ann .I H. P. 25, 457-481.

[3] Bai J. (2003) Testing Parametric Conditional Distributions of Dynamic Models, Rev. Econ. Stat. 85, 531-549.

[4] Bierens H.J. (1982) Consistent Model Specification Tests, J. Econometric 20, 105-13.

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

[6] Bierens H.J., Ploberger W. (1997) Asymptotic Theory of Integrated Conditional Moment Tests, Econometrica 65, 1129-1151.

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

[8] Brockwell P.J., Cline D.B.H. (1985) Linear Prediction of ARMA Processes with Infinite Variance, Stoch. Proc. Appl. 19, 281-296.

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

[10] Chan K.S. (1990) Testing for Threshold Autoregression, Ann. Stat. 18, 1886-1894.

[11] Chen X. and Fan Y. (1999) Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series, J. Econometrics 91, 373-401.

[12] Corradi V., Swanson N.R. (2002) A Consistent Test for Nonlinear Out-of-Sample Predictive Accuracy, J. Econometrics 110, 353-381.

[13] Csörgo S., Horváth L., Mason D. (1986) What Portion of the Sample Makes a Partial Sum Asymptotically Stable or Normal? Prob. Theory Rel. Fields 72, 1-16.

[14] Davidson R., MacKinnon J., White H. (1983) Tests for Model Specification in the Presence of Alternative Hypotheses: Some Further Results, J. Econometrics 21, 53-70.

[15] Davies R.B. (1977) Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative, Biometrika 64, 247-254.

[16] Davis R.A., Knight K., Liu J. (1992) M-Estimation for Autoregressions with Infinite Variance, Stoch. Proc. Appl. 40, 145-180.

[17] de Jong R.M. (1996) The Bierens Test under Data Dependence, J. Econometrics 72, 1-32.

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

[19] de Lima P.J.F. (1997) On the Robustness of Nonlinearity Tests to Moment Condition Failure, J. Econometrics 76, 251-280.

[20] Dehling, H., M. Denker, W. Phillip (1986) Central Limit Theorems for Mixing Sequences of Random Variables under Minimal Conditions, Ann. Prob. 14, 1359-1370.

[21] Dette H. (1999) A Consistent Test for the Functional Form of a Regression Based on a Difference of Variance Estimators, Ann. Stat. 27, 1012-1040.

[22] Dufour J.M., Farhat A., Hallin M. (2006) Distribution-Free Bounds for Serial Correlation Coefficients in Heteroscedastic Symmetric Time Series, J. Econometrics 130, 123-142.

[23] Doukhan P., Massart, P., Rio E. (1995) Invariance Principles for Absolutely Regular Empirical Processes, Ann. I. H. P. 31, 393-427.

[24] Dudley R. M. (1978) Central Limit Theorem for Empirical Processes. Ann. Prob. 6, 899-929.

[25] Embrechts P., Goldie C.M. (1980) On Closure and Factorization Properties of Subexponential Distributions, J. Aus. Math. Soc. A, 29, 243-256.

[26] Embrechts P., Klüppelberg C., Mikosch T. (1997) Modelling Extremal Events for Insurance and Finance. Springer-Verlag: Frankfurt.

[27] Eubank R., Spiegelman S. (1990) Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques, J. Amer. Stat. Assoc. 85, 387-392.

[28] Fan Y., Li Q. (1996) Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms, Econometrica 64, 865-890.

[29] Fan Y., Li Q. (2000) Consistent Model Specification Tests: Kernel-Based Tests Versus Bierens' ICM Tests, Econometric Theory 16, 1016-1041, 2000.

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

Parameter Space, Econometric Theory 26, 965-993.

[31] Gabaix, X. (2008) Power Laws, in The New Palgrave Dictionary of Economics, 2nd Edition, S. N. Durlauf and L. E. Blume (eds.), MacMillan.

[32] Gallant A.R. (1981) Unbiased Determination of Production Technologies. J. Econometrics, 20, 285-323.

[33] Gallant A.R., White H. (1989) There Exists a Neural Network That Does Not Make Avoidable Mistakes, Proceedings of the Second Annual IEEE Conference on Neural Net., I:657-664.

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

[35] Hahn M.G., Weiner D.C., Mason D.M. (1991) Sums, Trimmed Sums and Extremes, Birkhäuser: Berlin.

[36] Hansen B.E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis, Econometrica 64, 413-430.

[37] Härdle W., Mammen E. (1993) Comparing Nonparametric Versus Parametric Regression Fits, Ann. Stat. 21, 1926–1947.

[38] Hausman J.A. (1978) Specification Testing in Econometrics, Econometrica 46, 1251-1271.

[39] Hill J.B. (2008a) Consistent and Non-Degenerate Model Specification Tests Against Smooth Transition and Neural Network Alternatives, Ann. D'Econ. Statist. 90, 145-179.

[40] Hill J.B. (2008b) Consistent GMM Residuals-Based Tests of Functional Form, Econometric Rev.: forthcoming.

[41] Hill J.B. (2011a) Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics, Econometric Theory 27, 844–884.

[42] Hill J.B. (2011b) Robust M-Estimation for Heavy Tailed Nonlinear AR-GARCH, Working Paper, University of North Carolina - Chapel Hill.

[43] Hill J.B. (2011c) Supplemental Appendix for Heavy-Tail and Plug-In Robust Consistent Conditional Moments Tests of Functional Form, www.unc.edu/∼jbhill/ap_cm_trim.pdf.

[44] Hill J.B. (2012) Stochastically Weighted Average Conditional Moment Tests of Functional Form: Stud. Nonlin. Dyn. Econometrics 16: forthcoming.

[45] Hill, J.B., Aguilar M. (2011) Moment Condition Tests for Heavy Tailed Time Series, J. Econometrics: Annals Issue on Extreme Value Theory: forthcoming.

[46] Hill J.B., Renault E. (2010) Generalized Method of Moments with Tail Trimming, submitted; Dept. of Economics, University of North Carolina - Chapel Hill.

[47] Hoffmann-Jørgensen J. (1991) Convergence of Stochastic Processes on Polish Spaces, Various Publication Series Vol. 39, Matematisk Institute, Aarhus University.

[48] Hong Y., White H. (1995) Consistent Specification Testing Via Nonparametric Series Regression, Econometrica 63, 1133-1159.

[49] Hong Y., Lee Y-J. (2005) Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form, Rev. Econ. Stud. 72, 499-541.

[50] Hornik K., Stinchcombe M., White H. (1989) Multilayer Feedforward Networks are Universal Approximators, Neural Net. 2, 359-366.

[51] Hornik K., Stinchcombe M., White H. (1990) Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks, Neural Net., 3, 551-560.

[52] Ibragimov R., Müller U.K. (2010) t-Statistic based Correlation and Heterogeneity Robust Inference, J. Bus. Econ. Stat. 28, 453-468.

[53] Lahiri S.N. (1995) On the Asymptotic Beahiour of the Moving Block Bootstrap for Normalized Sums of Heavy-Tailed Random Variables, Ann. Stat. 23, 1331-1349.

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

[55] Lee T., White H., Granger C.W.J. (1993) Testing for Neglected Nonlinearity in Time-Series Models: A Comparison of Neural Network Methods and Alternative Tests, J. Econometrics 56, 269-290.

[56] Lighthill, M.J. (1958). Introduction to Fourier Analysis and Generalised Functions. Cambridge Univ. Press: Cambridge.

[57] Ling S. (2005) Self-Weighted LAD Estimation for Infinite Variance Autoregressive Models, J. R. Stat. Soc. B 67, 381–393.

[58] McLeod A.I., Li W.K. (1983) Diagnostic Checking ARMA Time Series Models Using Squared Residual Autocorrelations, J. Time Ser. Anal. 4, 269-273.

[59] Meitz M., Saikkonen P. (2008) Stability of Nonlinear AR-GARCH Models, J. Time Ser. Anal. 29, 453-475.

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

[61] Peng L., Yao Q. (2003) Least Absolute Deviation Estimation for ARCH and GARCH Models, Biometrika 90, 967–975.

[62] Pham T., Tran L. (1985) Some Mixing Properties of Time Series Models. Stoch. Proc. Appl. 19, 297-303.

[63] Phillips, P.C.B. (1995). Robust Nonstationary Regression, Econometric Theory 11, 912-951.

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

[65] Ramsey J.B. (1969) Tests for Specification Errors in Classical Linear Least-Squares Regression, J. R. Stat. Soc. B 31, 350-371.

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

[67] Stinchcombe M., White H. (1989) Universal Approximation Using Feedforward Networks with Non-Sigmoid Hidden Layer Activation Functions, Proceedings of the International Joint Conference on Neural Net., I, 612-617.

[68] Stinchcombe M.B., White H. (1992) Some Measurability Results for Extrema of Random Functions Over Random Sets, Rev. Economic Stud., 59, 495-514.

[69] Stinchcombe M.B., White H. (1998) Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative, Econometric Theory 14, 295-325.

[70] Tsay R. (1986) Nonlinearity Tests for Time Series, Biometrika 73, 461-466.

[69] White H. (1981) Consequences and Detection of Misspecified Nonlinear Regression Models, J. Amer. Stat. Assoc. 76, 419-433.

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

[72] White H. (1987) Specification Testing in Dynamic Models, in Truman Bewley, ed., Advances in Econometrics. Cambridge University Press: New York.

[73] White H. (1989a) A Consistent Model Selection Procedure Based on m-Testing, in C.W.J. Granger, ed., Modelling Economic Series: Readings in Econometric Methodology, p. 369-403. Oxford University Press: Oxford.

[74] White H. (1989b) Some Asymptotic Results for Learning in Single Hidden Layer Feedforward Network

Models, J. Amer. Stat. Assoc., 84, 1003-1013.

[75] White H. (1990) Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings, Neural Net., 3, 535-549.

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

[77] Yatchew A.J. (1992) Nonparametric Regression Tests Based on Least Squares, Econometric Theory 8, 435-451.

[78] Zheng J.X. (1996) A Consistent Test of Functional Form via Nonparametric Estimation Techniques, J. Econometrics 75, 263-289.

Table 1 - Empirical Size (Linear AR)

	iid ϵ_t $(\kappa = 1.5)^a$		GARCH ϵ_t $(\kappa = 2)$		iid ϵ_t $(\kappa = \infty)$	
n	200	800	200	800	200	800
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
TT-Orth-Fix[b,c]	.00, .01, .04[d]	.00, .02, .06	.00, .01, .05	.00, .03, .09	.00, .02, .05	.00, .03, .07
TT-Fix	.00, .02, .07	.01, .04, .08	.00, .01, .04	.00, .02, .05	.00, .02, .06	.00, .02, .04
TT-Orth-OT	.01, .04, .06	.02, .06, .12	.01, .03, .04	.01, .02, .04	.01, .03, .09	.02, .07, .12
TT-OT	.23, .41, .52	.31, .46, .54	.04, .17, .27	.15, .31, .40	.04, .12, .20	.06, .14, .23
CM-Orth[e]	.00, .01, .05	.00, .03, .07	.00, .04, .11	.00, .04, .10	.00, .04, .09	.00, .03, .09
CM	.00, .01, .04	.01, .02, .08	.00, .00, .02	.00, .01, .03	.00, .01, .03	.00, .02, .04
HW[f]	.17, .22, .25	.21, .24, .27	.06, .15, .24	.80, .87, .89	.00, .02, .05	.02, .05, .07
RESET[g]	.00, .00, .02	.00, .01, .02	.00, .03, .09	.01, .05, .11	.00, .03, .08	.01, .05, .10
McLeod-Li[g]	.02, .03, .03	.01, .02, .02	.58, .70, .78	1.0, 1.0, 1.0	.01, .04, .07	.02, .05, .09
Tsay[g]	.98, .99, 1.0	1.0, 1.0, 1.0	.37, .47, .51	.72, .77, .80	.01, .05, .10	.01, .05, .10

a. Moment supremum of the test error ϵ_t: $\kappa = \sup\{\alpha : E|\epsilon_t|^\alpha < \infty\}$

b. TT = Tail-Trimmed CM test with randomized nuisance parameter γ. Fix = fixed trimming parameter λ.

c. Orth = orthogonal equation transformation. OT = occupation time test over set of λ.

d. Rejection frequencies at 1%, 5% and 10% nominal levels.

e. Untrimmed randomized and sup-CM tests.

f. Hong and White's (1996) nonparametric test.

g. Ramsey's RESET test with 3 lags; McLeod and Li's test with 3 lags; Tsay's F-test.

Table 2 - Empirical Power[a] (Self-Exciting Threshold AR)

	iid ϵ_t ($\kappa = 1.5$)		GARCH ϵ_t ($\kappa = 2$)		iid ϵ_t ($\kappa = \infty$)	
n	200	800	200	800	200	800
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
TT-Orth-Fix	.02, .12, .22	.08, .26, .38	.01, .06, .11	.01, .05, .11	.02, .05, .11	.02, .08, .15
TT-Fix	.12, .18, .24	.21, .32, .39	.01, .05, .11	.03, .12, .23	.02, .07, .12	.09, .27, .43
TT-Orth-OT	.19, .35, .46	.65, .83, .93	.08, .17, .25	.12, .24, .39	.06, .13, .24	.16, .28, .42
TT-OT	.28, .30, .32	.38, .35, .37	.11, .13, .21	.24, .25, .39	.04, .13, .22	.39, .52, .57
CM-Orth	.05, .21, .33	.11, .27, .42	.02, .07, .13	.01, .05, .09	.01, .04, .08	.02, .06, .10
CM	.04, .11, .18	.12, .23, .29	.01, .05, .10	.02, .12, .24	.01, .07, .14	.08, .30, .44
HW	.06, .10, .16	.17, .15, .30	.04, .05, .09	.04, .07, .12	.02, .06, .11	.16, .29, .40
RESET	.03, .14, .28	.08, .28, .45	.02, .12, .24	.15, .38, .53	.20, .54, .73	1.0, 1.0, 1.0
McLeod-Li	.29, .45, .55	.71, .76, .83	.00, .00, .07	.01, .05, .10	.07, .19, .27	.51, .69, .79
Tsay	.02, .02, .02	.00, .00, .00	.13, .17, .19	.15, .17, .21	.45, .65, .70	1.0, 1.0, 1.0

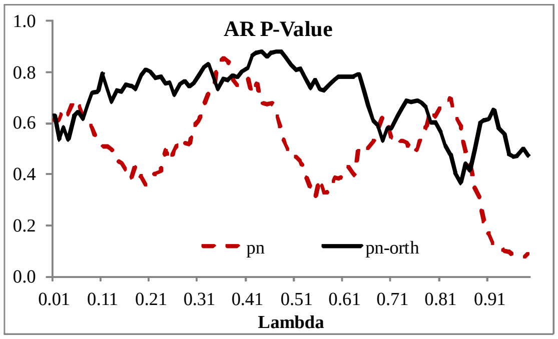
a. The rejection frequencies are adjusted for size distortions based on Table 1.

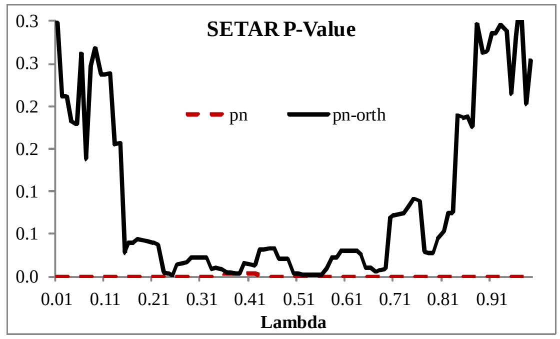
Table 3 - Empirical Power[a] (Bilinear AR)

	iid ϵ_t ($\kappa = 1.5$)		GARCH ϵ_t ($\kappa = 2$)		iid ϵ_t ($\kappa = \infty$)	
n	200	800	200	800	200	800
	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%	1%, 5%, 10%
TT-Orth-Fix	.04, .16, .26	.22, .39, .49	.01, .07, .11	.02, .08, .14	.01, .05, .09	.01, .05, .11
TT-Fix	.04, .13, .21	.13, .31, .42	.02, .09, .17	.02, .09, .18	.01, .04, .07	.01, .04, .10
TT-Orth-OT	.08, .14, .18	.36, .38, .37	.02, .05, .10	.04, .06, .09	.01, .05, .05	.00, .01, .02
TT-OT	.57, .61, .61	.68, .58, .60	.27, .30, .36	.57, .52, .55	.01, .07, .13	.06, .12, .16
CM-Orth	.03, .14, .24	.21, .40, .51	.02, .11, .21	.03, .13, .26	.01, .04, .11	.01, .05, .11
CM	.02, .08, .16	.08, .30, .41	.01, .05, .10	.01, .05, .11	.01, .04, .09	.01, .04, .09
HW	.00, .00, .00	.00, .00, .00	.02, .07, .07	.00, .00, .00	.24, .38, .47	.87, .92, .97
RESET	.02, .07, .14	.01, .06, .14	.03, .07, .11	.01, .05, .09	.02, .06, .12	.03, .17, .28
McLeod-Li	.19, .26, .33	.35, .43, .51	.00, .02, .04	.00, .00, .00	.86, .93, .98	.99, 1.0, 1.0
Tsay	.03, .06, .10	.01, .05, .10	.36, .36, .37	.20, .20, .23	.76, .84, .88	.91, .95, .96

a. The rejection frequencies are adjusted for size distortions based on Table 1.

Figure 1: P-Value Functions $p_n(\lambda)$ and $p_n^{\perp}(\lambda)$





Note: pn = $p_n(\lambda)$, and pn-orth = $p_n^{\perp}(\lambda)$.