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

Jonathan B. Hill[1]
Dept. of Economics University of North Carolina

ABSTRACT

Robust estimation methods are widely available: "QML" robust to breakdown points (Sakata and White 1995, 1998), LAD robust to heavy tails (Ling 2005), GMM robust to heavy tails (Hill and Renault 2010), QML robust to asymmetry and heavy tails (Hill 2011). Robust moment condition tests, however, are not available.

Consistent moment-based tests of regression model functional form require the existence of higher moments to ensure standard asymptotics. We deliver two asymptotic power-one test statistics for heavy tailed time series with a Nonlinear ARX representation, including nonlinear AR-GARCH, and nonlinear ARCH.

Under the null hypothesis the regression errors must have a finite mean, but under the alternative they may have arbitrarily heavy tails. If the errors have an infinite variance then in principle any consistent plug-in is allowed, including those with non-Gaussian limits or a convergence rate less then $\sqrt{n}$.

In one statistic we use an orthogonal test equation transformation, hence irrespective of tail thickness the statistic is robust to a large array of plug-ins. We derive a chi-squared weak limit of the statistic under the null, deliver a consistent bootstrapped p-value for test functionals, and study the finite sample properties.

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

Key words and phrases: conditional moment test; tail trimming; heavy tails.

INTRODUCTION The regression model is

$$y_t = f(x_t, \beta) + \epsilon_t(\beta)$$

where $f : \mathbb{R}^p \times \mathcal{B} \to \mathbb{R}$ for finite $p > 0$ is continuous and differentiable in β, and $x_t \in \mathbb{R}^p$ may contain lags of y_t or other random variables. **Examples**: Nonlinear ARX with Nonlinear GARCHX errors, and Nonlinear ARCHX.

■ We will test

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

■ We allow

$$E\left|y_t\right| < \infty \text{ under either hypothesis to justify a test of } E\left[y_t|x_t\right]$$

$$E\left[\sup_{\beta \in \mathcal{B}} \left|\epsilon_t(\beta)\right|^{\iota}\right] < \infty \text{ under } H_1 : \text{ under mis-specification errors may be arbitrarily heavy tailed.}$$

We develop a consistent Conditional Moment [CM] test against general alternatives for heavy tailed data (cf. Bierens 1990, Lee, White and Granger 1993, de Jong 1996, Corradi and Swanson 2002, Hill 2008a,b).

■ Existing nuisance parameter indexed CM test statistic (Bierens 1982, 1990, Lee, White and Granger 1993)

$$\hat{T}_n\left(\gamma\right) = \frac{1}{\hat{V}_n(\hat{\beta}_n, \gamma)} \left(\sum_{t=1}^{n} \epsilon_t(\hat{\beta}_n) F\left(\gamma'\psi_t\right)\right)^2, \; \psi_t := \psi(x_t) \text{ bounded one-to-one, } \gamma \in \Gamma \subset \mathbb{R}^r, \; \hat{\beta}_n = \beta^0 + O_p(1/n^{1/2}).$$

If $E|\epsilon_t| < \infty$ and H_1 holds then $E[\epsilon_t F(\gamma'\psi_t)] \neq 0$ for all γ on any compact $\Gamma \subset \mathbb{R}^p$, except for countably many γ for any non-polynomial real analytic $F : \mathbb{R} \to \mathbb{R}$ with affine argument $\gamma'\psi_t(x_t)$ (Bierens 1990, Bierens and Ploberger 1997, Stinchcombe and White 1998) : *exponential* (Bierens), *logistic* (Lee, White and Granger).

■ $\hat{T}_n\left(\gamma\right) \xrightarrow{d} \chi^2(1)$ require ϵ_t and $(\partial/\partial\beta)f(x_t, \beta)$ to have finite $4 + \iota^{th}$-moments and $\hat{\beta}_n = \beta^0 + O_p(1/n^{1/2})$.

■ $E|\epsilon_t| < \infty$ ensures $E[\epsilon_t|x_t]$ exists. Under H_1 the error ϵ_t may not be integrable when $E[y_t^2] = \infty$. **Example:** $y_t = x_t + u_t$ for iid u_t, scalar x_t, where $\{u_t, x_t\}$ have finite means and infinite variances. If $f(x_t, \beta) = (x_t + \beta)^2$ then $E|\epsilon_t(\beta)| = \infty$ for any $\beta \in \mathcal{B}$, hence $E[\epsilon_t|x_t]$ is not well defined.

SUMMARY AND MAIN RESULTS

We apply a stochastic non-differentiable *asymptotically negligible trimming indicator* $\hat{I}_{n,t}(\beta) \in \{0,1\}$ to $\epsilon_t(\beta)$:

$$\hat{T}_n(\gamma) = \frac{1}{\hat{S}_n^2(\hat{\beta}_n,\gamma)} \left(\sum_{t=1}^{n} \epsilon_t(\hat{\beta}_n) \hat{I}_{n,t}(\hat{\beta}_n) F(\gamma'\psi_t) \right)^2 \quad \text{where} \quad \{0,1\} \ni \hat{I}_{n,t}(\hat{\beta}_n) \overset{a.s.}{\to} 1.$$

$\hat{S}_n^2(\beta,\gamma)$ is a kernel estimator: $\epsilon_t \hat{I}_{n,t} F(\gamma'\psi_t)$ may not be orthogonal due to trimming.

■ The method applies to any moment condition test including white noise, omitted variables, instrumental variables, stochastic breaks, causality and distribution.

■ If $E[\epsilon_t^2] = \infty$ then $\hat{\beta}_n$ may be essentially any consistent plug-in: it may be $o_p(n^{1/2})$, and/or non-Gaussian.

■ Negligibility (*tail*-trimming) $\hat{I}_{n,t}(\hat{\beta}_n) \overset{a.s.}{\to} 1$ ensures

$$H_0 : E\left[\epsilon_t \hat{I}_{n,t} F(\gamma'\psi_t)\right] \to 0 \ \text{ and } \ H_1 : \limsup_{n\to\infty} \left| E\left[\epsilon_t \hat{I}_{n,t} F(\gamma'\psi_t)\right]\right| > 0 \ \ \forall \gamma \in \Gamma/S \ \ (S \text{ has measure zero}).$$

The tail-trimmed moment is asymptotically unbiased, and revealing.

■ *Trimming* $\hat{I}_{n,t}(\hat{\beta}_n) \in \{0,1\}$ and a mixing condition ensure

$$\left\{\hat{T}_n(\gamma) : \gamma \in \Gamma\right\} \Longrightarrow \left\{z(\gamma)^2 : \gamma \in \Gamma\right\}, \text{ chi-squared process on } \mathcal{C}[\Gamma].$$

■ Test functional follows standard asymptotics (bootstrap follows standard theory : Hansen '96, Gonçalves and White '02):

$$\sup_{\gamma\in\Gamma} \left\{\hat{T}_n(\gamma)\right\} \Longrightarrow \sup_{\gamma\in\Gamma} \left\{z(\gamma)^2\right\} \quad \text{and} \quad \int_\Gamma \hat{T}_n(\gamma)\mu(d\gamma) \Longrightarrow \int_\Gamma z(\gamma)^2 \mu(d\gamma).$$

■ ICM test is an easy extension (Bierens 1982, Bierens and Ploberger 1997).

■ Exploits new theory for heavy tail robust inference (Hill 2010, 2011 : CLT, LLN, Hill and Renault 2010: UCLT, ULLN).

TAIL INDICATORS

By the mean-value-theorem $\epsilon_t(\beta)$ contains information from the response gradient

$$g_t(\beta) = [g_{i,t}(\beta)]_{i=1}^{q} := \frac{\partial}{\partial \beta} f(x_t, \beta) \in \mathbb{R}^q \text{ and } g_t = g_t(\beta^0).$$

■ We trim larger $\epsilon_t(\beta)$ and $g_{i,t}(\beta)$: intermediate order sequences $\{k_{j,\epsilon,n} : j = 1,2\}$ and $\{k_{j,i,n} : j = 1,2\}$: $1 \leq k_{1,z,n} + k_{2,z,n} < n$, $k_{j,z,n} \to \infty$ and $k_{z,n} = o(n)$.

■ Denote tail specific observations

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

their *order statistics* $z_{(1)}^{(-)}(\beta) \leq \cdots \leq z_{(n)}^{(-)}(\beta) < 0$ and $z_{(1)}^{(+)}(\beta) \geq \cdots \geq z_{(n)}^{(+)}(\beta) \geq 0$. Trimming *indicators* are

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

■ The compound indicator $\hat{I}_{n,t}(\beta) := \hat{I}_{\epsilon,n,t}(\beta) \times \hat{I}_{g,n,t}(\beta)$ is used in $\hat{T}_n(\gamma) = \dfrac{1}{\hat{S}_n^2(\hat{\beta}_n, \gamma)} \left(\sum_{t=1}^{n} \epsilon_t(\hat{\beta}_n) \hat{I}_{n,t}(\hat{\beta}_n) F(\gamma' \psi_t)\right)^2$.

■ *Negligible* trimming $k_{j,z,n}/n = o(1)$ ensures *identification* of the null and a *revealing tail-trimmng moment*:

$$H_0 : E\left[\epsilon_t \hat{I}_{n,t} F(\gamma' \psi_t)\right] \to 0 \text{ and } H_1 : \limsup_{n \to \infty} \left| E\left[\epsilon_t \hat{I}_{n,t} F(\gamma' \psi_t)\right] \right| > 0 \;\; \forall \gamma \in \Gamma / S.$$

Trimming $\epsilon_t(\beta)$ by a *fixed* quantile leads to bias unless ϵ_t is symmetrically distributed and trimmed.

■ *Trimming* $k_{j,z,n} \to \infty$ ensures a Gaussian weak limit with samples on $\mathcal{C}[\Gamma]$ *a.s.*:

$$\left\{\hat{T}_n(\gamma) : \gamma \in \Gamma\right\} \Longrightarrow \left\{z(\gamma)^2 : \gamma \in \Gamma\right\}, \text{ chi-squared process on } \mathcal{C}[\Gamma].$$

PLUG-IN ROBUSTNESS : HEAVY TAILS If tails are heavy then $\hat{T}_n(\gamma)$ is *not affected by almost any* $\hat{\beta}_n$.

■ For some positive definite scale matrices $\{\tilde{V}_n\}$, $\tilde{V}_n \in \mathbb{R}^{r \times r}$, and $\tilde{V}_{i,i,n} \to \infty$,

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

In general $\tilde{V}_{i,i,n} \to \infty$ *faster than, equal to, or slower than* $n^{1/2}$ when tails are heavy:

OLS (greater), QML (less), LWAD and Log-LAD (equal: Ling 2005, Peng and Yao 2003), QMWL (equal : Ling 2007), QMTTL (less: Hill 2011), GMTTM (any: Hill and Renault 2010).

■ *Threshold functions* $\{l_{z,n}(\beta), u_{z,n}(\beta)\}$: the lower $k_{1,z,n}/n^{th}$ and upper $k_{2,z,n}/n^{th}$ quantiles:

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

Indicators and a *deterministically trimmed* test equation:

$$I_{n,t}(\beta) := I_{\epsilon,n,t}(\beta) \times \prod_{i=1}^{q} I_{i,n,t}(\beta) = I\left(-l_{\epsilon,n}(\beta) \le \epsilon_t(\beta) \le u_{\epsilon,n}(\beta)\right) \times \prod_{i=1}^{q}\left(-l_{i,n}(\beta) \le g_{i,t}(\beta) \le u_{i,n}(\beta)\right).$$

Long-run *covariance*, *Jacobian* and *scale*:

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

$$V_n(\beta,\gamma) := n^2 S_n^{-2}(\beta,\gamma) \times J_n(\beta,\gamma)' J_n(\beta,\gamma) \in \mathbb{R}^{q\times q}.$$

■ The test equations satisfy

$$\frac{1}{\hat{S}_n(\hat{\beta}_n,\gamma)}\sum_{t=1}^{n}\epsilon_t(\hat{\beta}_n)F(\gamma'\psi_t)\hat{I}_{n,t}\left(\hat{\beta}_n\right) \sim \frac{1}{S_n(\gamma)}\sum_{t=1}^{n}\epsilon_t F(\gamma'\psi_t)I_{n,t}\left(\beta^0\right) + V_n^{-1/2}(\gamma)\left(\hat{\beta}_n - \beta^0\right).$$

Require $\hat{\beta}_n \xrightarrow{p} \beta^0$ fast enough (standard assumption: $\hat{\beta}_n = \beta^0 + O_p(1/n^{1/2})$ in White '81, 82, Newey '85, Bierens '90):

$$\hat{T}_n(\gamma) \text{ tests } H_0 \text{ only if } \sup_{\gamma\in\Gamma}\left\|V_n^{1/2}(\gamma)\times\left(\hat{\beta}_n - \beta^0\right)\right\| = O_p(1) \textit{ hence only if } \sup_{\gamma\in\Gamma}\left\|V_n(\gamma)\tilde{V}_n^{-1}\right\| = O(1).$$

If $E[\epsilon_t^2] < \infty$ and $E||(\partial/\partial\beta)f(x_t,\beta)|_{\beta^0}||^2 < \infty$ then $n^{1/2}(\hat{\beta}_n - \beta^0) = O_p(1)$.

PLUG-IN ROBUSTNESS : HEAVY TAILS

■ If $E[\epsilon_t^2] = \infty$ or $E||(\partial/\partial\beta)f(x_t,\beta)|_{\beta^0}||^2 = \infty$

$$\frac{1}{n}\sum_{t=1}^{n} m_{n,t}^*(\beta^0,\gamma) \text{ has convergence rate } \frac{n}{S_n(\gamma)} = o(n^{1/2}) \text{ controlled by } k_{j,z,n} \to \infty.$$

The plug-in

$$\hat{\beta}_n \xrightarrow{p} \beta^0 \text{ faster than } \frac{n}{S_n(\gamma)} = n^{1/2} \text{ is trivial, or assured by rapid } k_{j,z,n} \to \infty.$$

■ **EXAMPLE: AR with iid infinite variance error**

$$y_t = \sum_{i=1}^{p} \beta_i^0 y_{t-1} + \epsilon_t, \;\; \epsilon_t \text{ is iid with } E[\epsilon_t] = 0, \text{ stationary, } \epsilon_t \text{ has absolutely continuous distribution}$$

$$P(|\epsilon_t| > \epsilon) = d\epsilon^{-\kappa}(1 + o(1)), \quad \text{where } \kappa \in (1,2) \text{ implies } E\left[\epsilon_t^2\right] = \infty.$$

Test error is

$$\epsilon_t(\beta) = y_t - \sum_{i=1}^{p} \beta_i y_{t-1}$$

OLS, LAD (Davis et al 1992), LTTS (Hill 2011), GMTTM (Hill and Renault 2010):

$$\hat{\beta}_n \xrightarrow{p} \beta^0 \text{ at least as fast as } n^{1/\kappa}/\ln(n) > n^{1/2}.$$

Yet $V_n(\gamma) \sim Kn(k_n/n)^{2/\kappa - 1} = o(n)$ for any $\{k_{\epsilon,n}, k_{i,n}\}$ due to $E|x_t| < \infty$ and

$$\frac{n}{S_n(\gamma)} = o(n^{1/2}) \;\; : \;\; \text{therefore } \sup_{\gamma\in\Gamma}\left\|V_n^{1/2}(\gamma) \times \left(\hat{\beta}_n - \beta^0\right)\right\| \xrightarrow{p} 0.$$

OLS, LAD, LTTS, GMTTM $\hat{\boldsymbol{\beta}}_n$ do not impact $\hat{\boldsymbol{T}}_n(\gamma)$:

$$\hat{T}_n(\gamma) = \frac{1}{\hat{S}_n^2(\hat{\beta}_n,\gamma)}\left(\sum_{t=1}^{n}\epsilon_t(\hat{\beta}_n)\hat{I}_{n,t}(\hat{\beta}_n)F(\gamma'\psi_t)\right)^2 \sim \frac{1}{S_n^2(\gamma)}\left(\sum_{t=1}^{n}\epsilon_t I_{n,t} F(\gamma'\psi_t)\right)^2.$$

This is fundamentally different from "plug-in robustness" via orthogonal transformation (Wooldridge 1990, Bai 2003).

PLUG-IN ROBUSTNESS : HEAVY TAILS

■ **EXAMPLE: Strong-GARCH**

$$y_t = h_t u_t, \text{ where } u_t \overset{iid}{\sim} (0,1), \text{ absolutely continuous distribution}$$

$$P(|u_t| > u) = du^{-\kappa}(1 + o(1)), \quad \text{where } \kappa > 2 \text{ implies } E\left[u_t^4\right] = \infty \text{ if } \kappa \leq 4$$

$$h_t^2 = \omega^0 + \alpha^0 y_{t-1}^2 + \beta^0 h_{t-1}^2, \ \omega^0 > 0, \ \left(\alpha^0, \beta^0\right) \in [0,1).$$

Test error is

$$\epsilon_t(\beta) = u_t^2(\beta) - 1 = \frac{y_t^2}{h_t^2(\beta)} - 1.$$

QML (Hall and Yao 2003), QMWL (Ling 2007), GMTTM (Hill and Renault 2010), QMTTL (Hill 2011):

$$\hat{\beta}_n \xrightarrow{p} \beta^0 \text{ slower than than } n^{1/2} \text{ if } E[\epsilon_t^4] = \infty.$$

Log-LAD (Peng and Yao 2003): $\hat{\beta}_n \xrightarrow{p} \beta^0$ at rate $n^{1/2}$.

QMTTL and GMTTM ($n^{1/2}/L(n)$), Log-LAD ($n^{1/2}$) do not impact; *QML, QMWL* **($n^{1-2/\kappa}/L(n)$)** *are too slow.*

■ Conventional plug-in robstness exists only for $n^{1/2}$-convergent $\hat{\beta}_n$ under standard regularity conditions.

Does not include QMTTL and GMTTM for GARCH with error $E[\epsilon_t^4] = \infty$ (QMTTL, GMTTM: $n^{1/2}/L(n)$).

■ If the plug-in is not too "slow":

$$\sup_{\gamma \in \Gamma} \left\| V_n^{1/2}(\gamma) \times \left(\hat{\beta}_n - \beta^0\right) \right\| = O_p(1), \quad \text{but not} \quad \sup_{\gamma \in \Gamma} \left\| V_n^{1/2}(\gamma) \times \left(\hat{\beta}_n - \beta^0\right) \right\| \xrightarrow{p} 0$$

we can assume $\hat{\beta}_n$ is asymptotically linear : **OLS, QML, QMWL, GMTTM, QMTTL**; *not LAD, LWAD, log-LAD, etc.*

PLUG-IN ROBUSTNESS : ORTHOGONAL TRANSFORMATION

■ Projection error operators (e.g. Wooldridge 1990, Bai 2003, Bontemps and Meddahi 2010)

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

$$\mathcal{P}_{n,t}(\gamma) = 1 - g_t' \times \left(E\left[g_t g_t' F(\gamma'\psi_t)I_{n,t}\right]\right)^{-1} \times E\left[g_t F(\gamma'\psi_t)I_{n,t}\right]$$

and equation projections are:

$$\epsilon_t(\hat{\beta}_n)\hat{I}_{n,t}(\hat{\beta}_n)F(\gamma'\psi_t) \times \hat{\mathcal{P}}_{n,t}(\gamma) \quad \text{and} \quad \epsilon_t(\beta)\, I_{n,t}(\beta)F(\gamma'\psi_t) \times \mathcal{P}_{n,t}(\gamma).$$

Then

$$E\left[\epsilon_t(\beta)\, I_{n,t}(\beta)F(\gamma'\psi_t) \times \mathcal{P}_{n,t}(\gamma)\right] \to 0 \quad \textit{if and only if} \quad E\left[\epsilon_t I_{n,t} F(\gamma'\psi_t)\right] \to 0.$$

The test statistic is

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

where $\hat{S}_n^{\perp 2}(\beta,\gamma)$ is identically $\hat{S}_n^2(\beta,\gamma)$ evaluated with $\epsilon_t(\beta)\hat{I}_{n,t}(\beta)F(\gamma'\psi_t) \times \hat{\mathcal{P}}_{n,t}(\gamma)$.

■ $\hat{T}_n^{\perp}(\gamma)$ is robust to any $V_n^{\perp}(\gamma)^{1/2}$-convergent $\hat{\beta}_n$, irrespective of thin (i.e. $V_n^{\perp}(\gamma)^{1/2} \sim Kn^{1/2}$) or heavy tails.

■ Nonlinear $\hat{\beta}_n$ (LAD, LWAD), non-Gaussian (OLS, LAD, EL), and slow (QML, QMWL, QMTTL, GMTTM for GARCH).

■ Together $\hat{T}_n(\gamma)$ and $\hat{T}_n^{\perp}(\gamma)$: heavy tail robust; nonlinear, non-Gaussian and sub-$n^{1/2}$-convergent plug-in robust.

FRACTILE CHOICE

We must choose a fractile policy $\{k_{j,\epsilon,n}, k_{j,i,n}\}$. Consider the class type (optimal for AR and GARCH examples)

$$k_{j,n} = [\lambda_j n/\ln(n)].$$

Write

$$\hat{T}_n(\gamma,\lambda), \quad \hat{I}_{n,t}(\hat{\beta}_n,\lambda) \text{ and } \hat{S}_n^2(\hat{\beta}_n,\gamma,\lambda), \text{ and } I_{n,t}(\lambda) \text{ and } S_n^2(\gamma,\lambda).$$

■ **Wild Bootstrap**

Draw iid $\{z_{r,t}\}_{t=1}^n$, $r = 1, ..., R$,

$$\hat{T}_{n,r}(\gamma,\lambda) := \frac{1}{\hat{S}_n^2(\hat{\beta}_n,\gamma,\lambda)} \left(\sum_{t=1}^n z_{r,t} \times \epsilon_t(\hat{\beta}_n)\hat{I}_{n,t}(\hat{\beta}_n,\lambda) F(\gamma'\psi_t) \right)^2.$$

The bootstrapped p-value under the null is

$$\hat{p}_n^R := \frac{1}{R}\sum_{r=1}^R I\left(\hat{T}_{n,r}(\gamma,\lambda) > \hat{T}_{n,r}(\gamma,\lambda)\right).$$

This bootstrap works only if tail-trimmed $\epsilon_t I_{n,t}(\lambda)$ is asymptotically orthogonal:

$$\frac{1}{S_n^2(\gamma,\lambda)} \times \sum_{t=1}^n E\left[\epsilon_t^2 I_{n,t}(\lambda) F(\gamma'\psi_t)^2\right] \to 1.$$

■ **Block Bootstrap**

Choose any $\{k_{j,\epsilon,n}, k_{j,i,n}\}$, bootstrap distribution of $\hat{T}_n(\gamma,\lambda)$. Cf. Gonçalves and White.

FRACTILE CHOICE

■ P-Value Occupation Time

Define a test functional $h_n := h\left(\hat{T}_n(\gamma)\right)$ and its limit under H_0: $h_n \Longrightarrow h^0$.

Define the asymptotic p-value

$$p_n(\lambda) = P\left(h_n > h^0 | H_0\right) \text{ - use } \textit{bootstrapped} \text{ p-value in practice.}$$

Occupation time for some level $\alpha \in (0,1)$ is (here λ is a scalar, but may be a vector)

$$\tau_n(\alpha) := \int_{\lambda \in \Lambda} I\left(p_n(\lambda) \leq \alpha\right) d\lambda \quad \text{and a discretized version } \hat{\tau}_n(\alpha) := \frac{1}{n}\sum_{i=1}^{n} I\left(p_n(i/n) \leq \alpha\right).$$

Under H_0

the occupation time $\sup_{\alpha \in [0,1]} |\hat{\tau}_n(\alpha) - \alpha| \xrightarrow{p} 0$.

■ Arbitrary Symmetric Trimming with Re-Centering

Use $k_n = [\lambda n / \ln(n)]$ with small λ, and trim symmetrically for simplicity:

$$\hat{I}_{\epsilon,n,t}\left(\beta,\lambda\right) := I\left(\left|\epsilon_t\left(\beta\right)\right| \leq \epsilon^{(+)}_{(k_{\epsilon,n})}\left(\beta\right)\right) \quad \text{and} \quad \hat{I}_{i,n,t}\left(\beta,\lambda\right) := I\left(\left|g_{i,t}\left(\beta\right)\right| \leq g^{(+)}_{i,(k_{i,n})}\left(\beta\right)\right).$$

Re-center trimmed test equation to control for bias:

$$\hat{T}_n\left(\gamma\right) = \frac{1}{\hat{S}^2_n(\hat{\beta}_n,\gamma)}\left(\sum_{t=1}^{n}\left\{\epsilon_t(\hat{\beta}_n)\hat{I}_{n,t}(\hat{\beta}_n,\lambda) - \frac{1}{n}\epsilon_t(\hat{\beta}_n)\hat{I}_{n,t}(\hat{\beta}_n,\lambda)\right\} F\left(\gamma'\psi_t\right)\right)^2$$

where $\hat{S}^2_n(\hat{\beta}_n,\gamma)$ reflects re-centering.

Tail trimming is *negligible asymptotically*: re-centering improves test under null, and *does not obscure a revealing moment*:

$$H_1 : \limsup_{n\to\infty} \left|E\left[\left\{\epsilon_t I_{n,t}(\lambda) - E\left[\epsilon_t I_{n,t}(\lambda)\right]\right\} F\left(\gamma'\psi_t\right)\right]\right| > 0 \quad \forall \gamma \in \Gamma/S.$$

SIMULATION STUDY

■ Data Generating Processes

We generate 10,000 samples of size $n \in \{500, 1000\}$ for each process:

TABLE 1 : DGP

AR(2)	$y_t = .8y_{t-1} - .4y_{t-2} + \epsilon_t$
Self Exciting Threshold AR	$y_t = .8y_{t-1}I\left(y_{t-1} < 0\right) - .4y_{t-1}I\left(y_{t-1} \geq 0\right) + \epsilon_t$
Bi-Linear	$y_t = .9y_{t-1}\epsilon_{t-1} + \epsilon_t$
Stochastic Breaks	$y_t = .8y_{t-1}\mathcal{I}_t^* - .4y_{t-1}\left(1 - \mathcal{I}_t^*\right) + \epsilon_t$ if $s_t = 2$

The Stochastic Breaks model requires a uniformly randomized number of breaks $r^* \in \{1, ..., [n/2]\}$, a uniformly randomized set of break periods $T^* := \{t_1^*, ..., t_{r^*}^*\}$, and an indicator $\mathcal{I}_t^* \in \{0, 1\}$ where $\mathcal{I}_1^* = 1$ and all remaining $\mathcal{I}_t^*$ alternate 1 to 0 when $t \in T^*$.

The errors $\{\epsilon_t\}$ are either iid symmetric Pareto

$$P\left(|\epsilon_t| > \epsilon\right) = .5\left(1 + \epsilon\right)^{-\kappa} \quad \text{with} \quad \text{index } \kappa = 1.5;$$

or IGARCH(1,1)

$$\epsilon_t = h_t u_t \quad \text{where} \quad h_t^2 = .3 + .4u_{t-1}^2 + .6h_{t-1}^2 \quad \text{and} \quad u_t \overset{iid}{\sim} N(0, 1);$$

or iid standard normal

$$\epsilon_t \sim N(0, 1).$$

The test errors ϵ_t have moment suprema $\kappa \in \{1.5,\ 2,\ \infty\}$.

SIMULATION STUDY

■ Tail-Trimmed CM Test

Let $x_t = [y_{t-1}, ..., y_{t-p}]$. We compute $\hat{T}_n(\gamma)$ and $\hat{T}_n^{\perp}(\gamma)$ using an exponential weight $F(\gamma'\psi(x_t)) = \exp\{\gamma'\psi(x_t)\}$ with argument $\psi(x_t) = [1, \arctan(x_t^*)']'$ where $x_{i,t}^* = x_{i,t} - 1/n\sum_{t=1}^{n} x_{i,t}$ (cf. Bierens 1990: Section 5).

Either *no HAC*, or *HAC* estimator with a Bartlett kernel and bandwidth $b_n = [n^b]$, $b \in \{.5, .6, .7\}$.

Either *non-transformed* equations, or *orthogolized* equations.

Fractiles $k_n = [.025 \times n/\ln(n)]$ for ϵ_t and x_t . We *symmetrically trim* and *re-center*.

Either γ is a *uniform draw* from $\Gamma = [.1, 2]^{p+1}$, or *sup-test* with bootstrapped p-value a la Hansen (1996) (not shown here).

■ Tests of Functional Form

Supremum and randomized versions of *untrimmed* $\hat{T}_n(\gamma)$ and $\hat{T}_n^{\perp}(\gamma)$: Bierens (1990), Lee, White and Granger (1993).

Hong and White's (1995) non-parametric test statistic

$$\hat{M}_n = \frac{1}{(2\ln n)^{1/2}}\left(\frac{1}{s_n^2}\sum_{t=1}^{n}\hat{\epsilon}_t\hat{v}_{n,t} - \ln n\right).$$

The components are $s_n^2 := 1/n\sum_{t=1}^{n}\hat{\epsilon}_t^2$ and $\hat{v}_{n,t} := \hat{f}_t - \hat{\beta}_n' x_t$ with nonparametric estimator $\hat{f}_t = \sum_{i=1}^{[\ln(n)]}\phi_i \exp\{\gamma_i' x_t\}$ of $E[y_t|x_t]$. The parameters γ_i are for each sample uniformly randomly selected from Γ, and ϕ is estimated by least squares. The choice of $\hat{f}_t$ is supported by Gallant's (1981) Flexible Fourier Form, and is known to be consistent by Corollary 1 of Bierens (1990).

Ramsey's (1969) RESET; the McLeod and Li (1983) statistic; and Tsay's (1986) F-statistic.

SIMULATION STUDY

Non-robust statistic show empirical size distortions.

Tsay's F-test and Hong and White's test are strongly distorted under heavy tails. McLeod-Li lacks power if IGARCH error.

Orthogonalizing the test equation is irrelevant when tails are heavy: it adds noise, size is slightly distorted, and power is reduced with GARCH errors.

Rejection Frequencies : Linear AR

	iid ϵ_t ($\kappa = 1.5$)			GARCH ϵ_t ($\kappa = 2$)			iid ϵ_t ($\kappa = \infty$)		
n	200	800	5000	200	800	5000	200	800	5000
TT-O,H	.00,.02,.04	.00,.01,.05	.01,.03,.08	.00,.03,.07	.00,.02,.07	.01,.04,.08	.01,.04,.11	.01,.05,.11	.01,.06,.12
TT-O	.00,.01,.04	.00,.02,.06	.01,.03,.09	.00,.01,.05	.00,.03,.09	.00,.03,.09	.00,.02,.05	.00,.03,.07	.01,.04,.09
TT-H	.00,.03,.07	.00,.02,.07	.01,.05,.10	.01,.03,.08	.01,.02,.06	.03,.07,.14	.01,.04,.09	.01,.05,.09	.01,.04,.08
TT	.00,.02,.07	.01,.04,.08	.01,.05,.10	.00,.01,.04	.00,.02,.05	.01,.06,.10	.00,.02,.06	.00,.02,.04	.00,.02,.05
CM-O,H	.00,.03,.08	.00,.03,.07	.00,.02,.06	.01,.05,.13	.00,.03,.08	.00,.02,.06	.02,.07,.14	.01,.06,.13	.01,.06,.11
CM-O	.00,.01,.05	.00,.03,.07	.00,.02,.07	.00,.04,.11	.00,.04,.10	.00,.03,.08	.00,.04,.09	.00,.03,.09	.00,.04,.09
CM-H	.00,.01,.05	.00,.02,.06	..01,.02,.07	.00,.02,.06	.00,.02,.05	.01,.03,.09	.01,.04,.09	.01,.04,.08	.01,.04,.06
CM	.00,.01,.04	.01,.02,.08	.00,.04,.07	.00,.00,.02	.00,.01,.03	.00,.02,.06	.00,.01,.03	.00,.02,.04	.00,.02,.04
HW	.17,.22,.25	.21,.24,.27	.20,.22,.24	.06,.15,.24	.80,.87,.89	.99,.99,.99	.00,.02,.05	.02,.05,.07	.01,.05,.09
RESET	.00,.00,.02	.00,.01,.02	.00,.02,.02	.00,.03,.09	.01,.05,.11	.01,.05,.10	.00,.03,.08	.01,.05,.10	.01,.05,.10
McL-Li	.02,.03,.03	.01,.02,.02	.01,.01,.02	.58,.70,.78	1.0,1.0,1.0	1.0,1.0,1.0	.01,.04,.07	.02,.05,.09	.01,.05,.10
Tsay	.98,.99,1.0	1.0,1.0,1.0	1.0,1.0,1.0	.37,.47,.51	.72,.77,.80	.97,.98,1.0	.01,.05,.10	.01,.05,.10	.01,.05,.10

TT = TTCM test. O = orthogonalized; H = HAC is used. CM = untrimmed test.

SIMULATION STUDY

The tail-trimmed test exhibits competitive power even without adjustment for size distortions (not shown).

If variance is infinite Hong and White's test and Tsay's F-test offer almost no power.

Empircal Power (Size Adjusted) : Self-Exciting Threshold AR

	iid ϵ_t $(\kappa = 1.5)$			GARCH ϵ_t $(\kappa = 2)$			iid ϵ_t $(\kappa = \infty)$		
n	200	800	5000	200	800	5000	200	800	5000
TT-O,H	.02,.10,.20	.07,.32,.51	.37,.62,.73	.02,.07,.14	.02,.09,.18	.02,.14,.25	.02,.09,.15	.04,.15,.24	.30,.46,.56
TT-O	.02,.12,.22	.08,.26,.38	.44,.66,.75	.01,.06,.11	.01,.05,.11	.02,.09,.20	.02,.05,.11	.02,.08,.15	.17,.35,.48
TT-H	.11,.24,.33	.28,.43,.50	.55,.65,.71	.01,.07,.12	.06,.21,.34	.44,.65,.72	.05,.17,.28	.30,.50,.60	.81,.91,.95
TT	.12,.18,.24	.21,.32,.39	.46,.57,.64	.01,.05,.11	.03,.12,.23	.35,.58,.71	.02,.07,.12	.09,.27,.43	.72,.87,.95
CM-O,H	.03,.11,.22	.07,.22,.39	.15,.33,.48	.02,.09,.14	.03,.08,.14	.02,.07,.13	.03,.08,.13	.01,.06,.11	.10,.25,.39
CM-O	.05,.21,.33	.11,.27,.42	.18,.41,.54	.02,.07,.13	.01,.05,.09	.01,.04,.10	.01,.04,.08	.02,.06,.10	.06,.21,.33
CM-H	.05,.17,.25	.19,.32,.40	.37,.51,.59	.02,.05,.10	.06,.20,.35	.39,.63,.74	.05,.17,.28	.30,.51,.64	.83,.91,.97
CM	.04,.11,.18	.12,.23,.29	.27,.39,.49	.01,.05,.10	.02,.12,.24	.30,.54,.71	.01,.07,.14	.08,.30,.44	.73,.87,.95
HW	.06,.10,.16	.17,.15,.30	.46,.51,.72	.04,.05,.09	.04,.07,.12	.02,.06,.11	.02,.06,.11	.16,.29,.40	.94,.99,.97
RESET	.03,.14,.28	.08,.28,.45	.24,.44,.63	.02,.12,.24	.15,.38,.53	.31,.58,.70	.20,.54,.73	1.0,1.0,1.0	1.0,1.0,1.0
McL-Li	.29,.45,.55	.71,.76,.83	.86,.93,1.0	.00,.00,.07	.01,.05,.10	.01,.05,.10	.07,.19,.27	.51,.69,.79	1.0,1.0,1.0
Tsay	.02,.02,.02	.00,.00,.00	.00,.00,.00	.13,.17,.19	.15,.17,.21	.04,.07,.10	.45,.65,.70	1.0,1.0,1.0	1.0,1.0,1.0

TT = TTCM test. O = orthogonalized; H = HAC is used. CM = untrimmed test.